Simultaneous drift conditions for Adaptive Markov Chain Monte Carlo algorithms

Yan Bai*

January 2009

Abstract

In the paper, we mainly study the ergodic property of adaptive MCMC algorithms. Suppose that the diminishing adaptation condition and simultaneous polynomial ergodicity hold. We find that either when the number of drift conditions is greater than or equal to two, or when the number of drift conditions having some specific form is one, the adaptive MCMC algorithm is ergodic. For adaptive MCMC algorithm with Markovian adaptation, the algorithm satisfying simultaneous polynomial ergocidity is ergodic without those restrictions. We also discuss some recent results related to this topic, and show that given some condition, the containment condition is necessary for the ergodicity of adaptive algorithm .

1 Introduction

Markov chain Monte Carlo (MCMC) algorithms are widely used for approximately sampling from complicated probability distributions. However, there are some limitations for this method, because it is often necessary to tune the scaling and other parameters before the algorithm will converge efficiently.

Adaptive MCMC methods using regeneration times and other complicated constructions have been proposed by Gilks *et al.* [8], Brockwell and Kadane [6], and elsewhere. More recently, Adaptive MCMC methods has been used to improve convergence via a self-studying procedure. In this direction, a key step was made by Harrio *et al.* [9], who proposed an adaptive Metropolis algorithm attempting to optimize the proposal distribution, and proved that a particular version of this algorithm correctly converges weakly to the target distribution. Andrieu and Robert [2] observed that the algorithm of Haario *et al.* [9] can be viewed as a version of the Robbins-Monro stochastic

*Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, CA. yanbai@utstat.toronto.edu

control algorithm (Robbins and Monro, [12]). The results of Haario *et al.* were then generalized by Atchadé and Rosenthal [4] and Andrieu and Moulines [1], proving convergence of more general adaptive MCMC algorithms.

Roberts and Rosenthal [13] (RR) use a coupling method for adaptive MCMC algorithm to show that its ergodicity is implied by the containment condition and the diminishing adaptation condition, based on the basic assumptions: each transition kernel P_γ on the state space $\mathcal{X}$ and the adaptive parameter space $\mathcal{Y}$ ($\gamma \in \mathcal{Y}$), is ϕ-irreducible, aperiodic and admits a nontrivial unique finite invariant probability measure π. The diminishing adaptation condition is relatively easy to check, and commonly the adaptive strategy is designed artificially. However, the containment condition is really abstract and hard to check. In their paper, they prove that simultaneous strongly aperiodic geometric ergodicity ensures the containment condition.

Based on [13], Yang [16], and Atchadé and Fort [3] (AF) respectively tackle the open problem. Both results are very interesting. Yang assume that the adaptive parameter space is compact under some metric, and connects it with the regeneration decomposition to find the uniform bound of the distance $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|$ for all γ. Once this condition given, the distance $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|$ can be uniformly bounded by the test function. The boundedness of the test function sequence $V(X_n)$ can ensure the containment condition.

Under some situations, to directly check the containment condition is quite hard. AF use similar coupling method as that in [13] to prove an attractive and better result when adaptive MCMC algorithm is restricted to Markovian adaptation. They assume that uniformly strongly aperiodicity, simultaneously drift condition in weakest form, and uniform convergence on any sublevel set of the test function $V(\cdot)$. The idea is that once the chain comes into some "big" sublevel set, the coupling method could be applied conditioned on starting from this set. Indeed, their result implies that the stochastic process $V(X_n)$ is bounded in probability.

In Section 2 we give some necessary notations. In Section 3 we present Yang's, and AF's conditions (respectively (Y1)-(Y4) and (M1)-(M3)) and results (respectively Theorem 3.3 and Theorem 3.4), and give some comments for both results. In Section 4 we provide a necessary condition of the ergodicity conditioned on the condition (A3). In Section 5 we present our conditions and main result (Theorem 5.3).

2 Terminology

Let π be a fixed target probability distribution on the state space $\mathcal{X}$ with σ-algebra $\mathcal{F}$.

Consider the family $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ with the state space $\mathcal{X}$ and the adaptive parameter space $\mathcal{Y}$

where each Markovian transition kernel P_γ is time-homogeneous, ϕ_γ-irreducible and aperiodic with stationary measure $\pi(\cdot)$. Define the filtration $\mathcal{F}_n = \sigma(X_k, \Gamma_k : 0 \le k \le n)$. The *adaptive Markov chain Monte Carlo algorithm* (adaptive MCMC) $\{(X_n, \Gamma_n)\}$ is a chain which at each time n, a random $\mathcal{F}_n$-measurable transition kernel Γ_n is selected basing on the history information with the property:

$$P(X_{n+1} \in A \mid \mathcal{F}_n) = P(X_{n+1} \in A \mid X_n, \Gamma_n) = P_{\Gamma_n}(X_{n+1} \in A \mid X_n). \tag{1}$$

Thus,

$$\begin{aligned} & P\left(X_{n+1} \in dx_{n+1}, \Gamma_{n+1} \in d\gamma_{n+1} \mid \mathcal{F}_n\right) \\ = \; & P(X_{n+1} \in dx_{n+1} \mid X_n, \Gamma_n) P(\Gamma_{n+1} \in d\gamma_{n+1} \mid X_{n+1} = x_{n+1}, \mathcal{F}_n). \end{aligned}$$

So, if Γ_{n+1} is $\sigma(X_n, \Gamma_n)$-measurable, the adaptive MCMC algorithm is called *Markovian Adaptation*, i.e. the pair process $\{(X_n, \Gamma_n)\}$ is a time-inhomogeneous Markov Chain.

Define $P\left(f(X) \in \cdot \mid X_0 = x_0, \Gamma_0 = \gamma_0\right) := P_{(x_0,\gamma_0)}\left(f(X) \in \cdot\right)$ for some measurable function f. Denote the corresponding expectation by $E_{(x_0,\gamma_0)}\left[f(X)\right]$. Define the first return time and the ith return time to the set C from the time n respectively: $\tau_{n,C} = \tau_{n,C}(1) := \min\{k \ge 1 : X_{n+k} \in C\}$ and $\tau_{n,C}(i) := \min\{k > \tau_{n,C}(i-1) : X_{n+k} \in C\}$ for $n \ge 0$ and $i > 1$.

Say the adaptive algorithm $\{X_n\}$ with adaptive scheme $\{\Gamma_n\}$ is *ergodic* if for any initial point $(x_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y}$, $\lim_{n\to\infty} \left\| P_{(x_0,\gamma_0)}(X_n \in \cdot) - \pi(\cdot) \right\| = 0$, where $\|\cdot\|$ is the total variation norm distance, i.e. $\|\mu(\cdot) - \nu(\cdot)\| = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$.

The diminishing adaptation condition: $\lim_{n\to\infty} D_n = 0$ in probability, where

$$D_n := \sup_{x \in \mathcal{X}} \left\| P_{\Gamma_{n+1}}(x, \cdot) - P_{\Gamma_n}(x, \cdot) \right\|, \tag{2}$$

is $\mathcal{F}_{n+1}$-measurable random variable. The condition means that the change of adaptive kernel turns to be zero. It is relatively easy to check, because the adaptive scheme is artificially designed.

For a non-negative function V, the process $V(X_n)$ *is bounded in probability* if given any $(x_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y}$,

$$\forall \delta > 0, \exists N > 0,\ M > 0 \text{ such that for } n > N, P_{(x_0,\gamma_0)}(V(X_n) > M) < \delta.$$

The "ϵ-convergence time"

$$M_\epsilon(x, \gamma) := \inf_n \left\{ n \in \mathbb{N}^+ : \left\| P_\gamma^n(x, \cdot) - \pi(\cdot) \right\| < \epsilon \right\} : \mathcal{X} \times \mathcal{Y} \to \mathbb{N}^+. \tag{3}$$

The containment condition is that for any $\epsilon > 0$, the stochastic process $\{M_\epsilon(X_n, \Gamma_n)\}_n$ is bounded in probability given any starting point given any $(x_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y}$, i.e. given any $(x_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y}$, $\forall \epsilon > 0$, $\forall \delta > 0$, $\exists K > 0$, such that $P_{(x_0,\gamma_0)}(M_\epsilon(X_n, \Gamma_n) > K) < \delta$ for all n. Indeed, the containment

containment condition means that the sequence $M_\epsilon(X_n, \Gamma_n)$ is tightness. However, since $M_\epsilon(x, \gamma)$ is integer-value, so the tightness is equivalent to bounded in probability. From intuition, the condition means that the time that transition kernels get close to the target within ϵ is bounded. Call the adaptive parameter process Γ_n *is bounded in probability* if $\forall \epsilon > 0, \exists N > 0$, some compact set $B \subset \mathcal{Y}$, such that for $n > N, P(\Gamma_n \in B^c) < \epsilon$.

3 Simultaneous Drift Conditions

Before studying the simultaneous drift conditions, we present RR's result about adaptive MCMC algorithms.

Theorem 3.1 (Roberts and Rosenthal [13])**.** *Any adaptive MCMC algorithm satisfying the containment condition and the diminishing adaptation, is ergodic to the stationary measure π.*

Remark 3.1. In [13], RR give one condition: simultaneously strongly aperiodically geometrically ergodic which can ensure the containment condition. In the definition, the drift conditions have the same form: $P_\gamma V(x) \leq \lambda V(x) + b\mathbf{1}_C(x)$ for all $\gamma \in \mathcal{Y}$. However, if Γ_n is bounded in probability, the containment condition is implied by that in each compact subset B of $\mathcal{Y}$, the drift conditions have the same form: $P_\gamma V_B(x) \leq \lambda_B V_B(x) + b_B \mathbf{1}_C(x)$. More generally, see the following corollary.

Corollary 3.2. *Suppose that the parameter space $\mathcal{Y}$ is a metric space, and the adaptive parameter $\{\Gamma_n\}$ is bounded in probability; for any compact set $K \subset \mathcal{Y}$, for any $\epsilon > 0$, the local ϵ-convergence time $\widetilde{M}_\epsilon(X_n) := \inf_m \left\{m \in \mathbb{N}^+ : \sup_{\gamma \in K} \left\|P_\gamma^m(X_n, \cdot) - \pi(\cdot)\right\| < \epsilon\right\}$ is bounded in probability; and the diminishing adaptation. Then the adaptive MCMC algorithm (X_n, Γ_n) is ergodic.*

The proof is trivial and omitted.

RR propose one open problem in [13]. Yang [16] gives the following conditions to tackle the problem:

- Y1: There exist a constant $\delta > 0$, and set $C \in \mathcal{F}$, and a probability measure $\nu_\gamma(\cdot)$ for $\gamma \in \mathcal{Y}$, such that $P_\gamma(x, \cdot) \geq \delta \mathbf{1}_C(x) \nu_\gamma(\cdot)$ for $\gamma \in \mathcal{Y}$;
- Y2: all kernels uniformly satisfy the weakest drift condition: $P_\gamma V \leq V - 1 + b\mathbf{1}_C$, where $V : \mathcal{X} \to [1, \infty)$ and $\pi(V) < \infty$;
- Y3: $\mathcal{Y}$ is compact under the metric $d(\gamma_1, \gamma_2) = \sup_{x \in \mathcal{X}} \|P_{\gamma_1}(x, \cdot) - P_{\gamma_2}(x, \cdot)\|$;
- Y4: the stochastic process $\{V(X_n)\}_n$ is bounded in probability.

Theorem 3.3 (Yang [16]). *Suppose the diminishing adaptation condition holds. The conditions (Y1)-(Y4) ensure the ergodicity of adaptive MCMC algorithms.*

Remark 3.2.

1. In Yang's proof, both (Y1) and (Y2) can ensure that each transition kernel is ergodic to π. Both (Y3) and (Y4) imply that the total variation distance between P_γ and π converges to zero uniformly on $\mathcal{Y}$.

2. Consider the condition $\pi(V) < \infty$ in (Y2). For each P_γ, since the chain $X_n^{(\gamma)}$ from the transition kernel P_γ is recurrent, for any $A \subset \mathcal{X}$ with $\pi(A) > 0$, $\pi(V) = \int_A \pi(dy) E_\gamma \left[\sum_{i=0}^{\tau_A - 1} V(X_i^{(\gamma)}) | X_0^{(\gamma)} = y\right]$. If there exists a small set $C_1 \subset \mathcal{X}$ with $\sup_{x\in C_1} E_\gamma \left[\sum_{i=0}^{\tau_{C_1}-1} V(X_i^{(\gamma)}) | X_0^{(\gamma)} = x\right] < \infty$, define $U_\gamma(x) = E_\gamma \left[\sum_{i=0}^{\sigma_A} V(X_i^{(\gamma)}) | X_0^{(\gamma)} = x\right]$. Hence, $P_\gamma U_\gamma - U_\gamma \leq -V(x) + b_1 \mathbf{1}_{C_1}$ where $b_1 = \sup_{x\in C_1} E_\gamma \left[\sum_{i=0}^{\tau_A-1} V(X_i^{(\gamma)}) | X_0^{(\gamma)} = x\right]$. It is well known that U_γ is the minimal pointwise test function. If there exists an upper bound test function $U \geq U_\gamma$ for any $\gamma \in \mathcal{Y}$, then there exists simultaneous drift condition. E.g. if $\sup_{\gamma\in\mathcal{Y}} \sup_{x\in C_1} E_\gamma \left[\sum_{i=0}^{\tau_A-1} V(X_i^{(\gamma)}) | X_0^{(\gamma)} = x\right] < \infty$, then $U = \sup_{\gamma\in\mathcal{Y}} U_\gamma$. Under this situation, (Ys) is a special case of the condition (As) in Section 5. We show that the condition (Y3) is unnecessary (See Theorem 5.3, Remark 5.2 and Remark 5.11). (Y1) is also too strong and can be replaced by (A1) (See Theorem 5.3).

AF [3] also give the following conditions to study the ergodicity of adaptive MCMC with Markovian adaptation:

M1: there exists a probability measure $\nu(\cdot)$, a constant $\delta > 0$, and set $C \in \mathcal{F}$ such that $P_\gamma(x,\cdot) \geq \delta \mathbf{1}_C(x)\nu(\cdot)$ for $\gamma \in \mathcal{Y}$;

M2: there exists a measurable function $V : \mathcal{X} \to [1,\infty)$ and a positive constant $b > 0$ such that for any $\gamma \in \mathcal{Y}$, $(P_\gamma V)(x) - V(x) \leq -1 + b\mathbf{1}_C(x)$;

M3: for any sublevel set $\mathcal{D}_l = \{x \in \mathcal{X} : V(x) \leq l\}$ of V, $\lim_{n\to\infty} \sup_{\mathcal{D}_l \times \mathcal{Y}} \left\| P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|_{TV} = 0$.

Theorem 3.4 (Atchadé and Fort [3]). *Suppose the diminishing adaptation condition holds. The conditions (M1)-(M3) imply the ergodicity of adaptive MCMC algorithm with Markovian adaptation.*

Remark 3.3.

1. The condition (M1) can be replaced with the condition (Y1) in the theorem. The reason is that for large enough $M > 0$, on the any set $D_M = \{x : V(x) \leq M\} \supseteq C$, the simultaneous drift conditions $(P_\gamma V)(x) - V(x) \leq -1 + M\mathbf{1}_{D_M}(x)$, also holds. Only the expectation of the hitting time

of the sublevel set D_M is used in the proof (see details in [3]).

2. Since

$$\left|P_{(x_0,\gamma_0)}(V(X_n) > M) - \pi(D_M^c)\right| \le \left\|P_{(x_0,\gamma_0)}(X_n \in \cdot) - \pi(\cdot)\right\|.$$

M can be taken extremely large such that $\pi(D_M^c) < \epsilon$. (M1-M3) and the diminishing adaptation condition imply that R.H.S. of the above equation converges to zero. So, $V(X_n)$ is bounded in probability.

3. In Section 4 we show that under some condition, the containment condition is a necessary condition of ergodicity of adaptive MCMC provided that (M3). From another view, AF's proof does apply the coupling method to check the containment condition by using the diminishing condition and simultaneous drift conditions.

4 The necessary condition of the ergodicity

In this section, we study the necessary condition of the ergodicity of adaptive algorithm. One example in [5] (Example 3.1) is given to show that only the diminishing adaptation condition can not ensure the ergodicity. In that example, the containment condition is not satisfied. There is another example to show that the containment condition is also not necessary in [15]. In the following theorem, we prove that under some additional condition similar to (M3), the containment condition is necessary for the ergodicity of adaptive algorithms.

Theorem 4.1 (The necessity of the containment condition)**.** *Suppose that there exists an increasing sequence of sets $\mathcal{D}_k \uparrow \mathcal{X}$ on the state space $\mathcal{X}$, such that for any $k > 0$,*

$$\lim_{n\to\infty} \sup_{\mathcal{D}_k\times\mathcal{Y}} \left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\| = 0. \tag{4}$$

If the adaptive MCMC algorithm is ergodic then the containment condition holds.

Proof: Fix $\epsilon > 0$. For any $\delta > 0$, taking $K > 0$ such that $\pi(\mathcal{D}_K^c) < \delta/2$. For the set $\mathcal{D}_K$, there exists M such that

$$\sup_{\mathcal{D}_K\times\mathcal{Y}} \left\|P_\gamma^M(x,\cdot) - \pi(\cdot)\right\| < \epsilon.$$

Hence, for any $(x_0,\gamma_0) \in \mathcal{X}\times\mathcal{Y}$, by the ergodicity of the adaptive MCMC $\{X_n\}_n$, there exists some $N > 0$ such that $n > N$,

$$\left|P_{(x_0,\gamma_0)}(X_n \in \mathcal{D}_K^c) - \pi(\mathcal{D}_K^c)\right| < \delta/2.$$

So, for $(X_n,\Gamma_n) \in (\mathcal{D}_K,\mathcal{Y})$,

$$[X_n \in \mathcal{D}_K] = [(X_n,\Gamma_n) \in \mathcal{D}_K\times\mathcal{Y}] \subset [M_\epsilon(X_n,\Gamma_n) \le M].$$

Hence,

$$\begin{aligned}
& P_{(x_0,\gamma_0)}\left(M_\epsilon(X_n,\Gamma_n) > M\right) \\
\leq\; & P_{(x_0,\gamma_0)}\left((X_n,\Gamma_n) \in (\mathcal{D}_K \times \mathcal{Y})^c\right) \\
=\; & P_{(x_0,\gamma_0)}\left(X_n \in \mathcal{D}_K^c\right) \\
\leq\; & \left|P_{(x_0,\gamma_0)}\left(X_n \in \mathcal{D}_K^c\right) - \pi(\mathcal{D}_K^c)\right| + \pi(\mathcal{D}_K^c) < \delta.
\end{aligned}$$

Therefore, the containment condition holds. □

Corollary 4.2. *Suppose that the parameter space $\mathcal{Y}$ is a metric space, and the adaptive scheme Γ_n is bounded in probability. Suppose that there exists an increasing sequence of sets $(\mathcal{D}_k, \mathcal{Y}_k) \uparrow \mathcal{X} \times \mathcal{Y}$ such that any $k > 0$,*

$$\lim_{n\to\infty} \sup_{\mathcal{D}_k \times \mathcal{Y}_k} \left\|P_\gamma^n\left(x,\cdot\right) - \pi(\cdot)\right\| = 0.$$

If the adaptive MCMC algorithm is ergodic then the containment condition holds.

Proof: Using the same technique in Theorem 4.1, for large enough $M > 0$,

$$\begin{aligned}
& P_{(x_0,\gamma_0)}\left(M_\epsilon(X_n,\Gamma_n) > M\right) \\
\leq\; & P_{(x_0,\gamma_0)}\left((X_n,\Gamma_n) \in (D_k \times \mathcal{Y}_k)^c\right) \\
\leq\; & P_{(x_0,\gamma_0)}\left(X_n \in D_k^c\right) + P_{(x_0,\gamma_0)}\left(\Gamma_n \in \mathcal{Y}_k^c\right) \\
\leq\; & \left|P_{(x_0,\gamma_0)}\left(X_n \in \mathcal{D}_K^c\right) - \pi(\mathcal{D}_K^c)\right| + \pi(\mathcal{D}_K^c) + P_{(x_0,\gamma_0)}\left(\Gamma_n \in \mathcal{Y}_k^c\right).
\end{aligned}$$

Since Γ_n is bounded in probability, the result holds. □

Here, we use one example to explain the importance of the additional conditions in Theorem 4.1 and Corollary 4.2. This example is from Theorem 5.2 in [15].

Example 4.3. *The target distribution is Unif$(0,1)$. The Metropolis proposal is $Q_k(x,\cdot) \sim$ Unif$(x - \frac{k}{2}, x + \frac{k}{2})$. The adaptation is that $\Gamma_n = k\mathbf{1}\left(\sum_{j=1}^{k-1} r_j \leq n < \sum_{j=1}^{k} r_j\right)$ where $r_j = \inf_m \left\{\left\|P_j^m(x,\cdot) - \pi(\cdot)\right\| \leq \frac{1}{j}\right\}$.*

In Theorem 5.2 of [15], the algorithm is proved to be ergodic and the containment condition does not satisfy. Another required to be concerned is that Γ_n is not bounded in probability, because $\Gamma_n \to \infty$ a.s. Moreover, for any subset D of $(0,1)$, thel equation (4) does not satisfy.

5 Simultaneous Polynomial Ergodicity

Although the ergodicity of adaptive MCMC algorithms, to some degree, is solved in [16] and [3], there are still some properties unknown about simultaneous polynomial ergodicity. In the section, we find that the conditions (Y4) and (M3) are implied for the adaptive MCMC with simultaneous polynomial ergodicity. Before studying it, let us recall the result about a quantitative bound for time-homogeneous Markov chain with polynomial convergence rate by Fort and Moulines [7] (FM).

Theorem 5.1 (Fort and Moulines [7])**.** *Suppose that the time-homogeneous transition kernel P satisfies the following conditions:*

- *P is π-irreducible for an invariant probability measure π;*
- *There exist some sets $C \in \mathcal{B}(\mathcal{X})$ and $D \in \mathcal{B}(\mathcal{X})$, $C \subset D$, $\pi(C) > 0$ and an integer $m \geq 1$, such that for any $(x, x') \in \triangle := C \times D \cup D \times C$, $A \in \mathcal{B}(\mathcal{X})$,*

$$P^m(x, A) \wedge P^m(x', A) \geq \rho_{x,x'}(A) \tag{5}$$

where for some kernel $\rho_{x,x'}(dy)$ from $\triangle$ to $\mathcal{X}$, and $\epsilon^- := \inf_{(x,x')\in\triangle} \rho_{x,x'}(\mathcal{X}) > 0$.

- *Let $q \geq 1$. There exist some measurable functions $V_k : \mathcal{X} \to \mathbb{R}^+ \setminus \{0\}$ for $k \in \{0, 1, \ldots, q\}$, and for $k \in \{0, 1, \ldots, q-1\}$, for some constants $0 < a_k < 1$, $b_k < \infty$, and $c_k > 0$ such that*

$$\begin{gathered} PV_{k+1}(x) \leq V_{k+1}(x) - V_k(x) + b_k \mathbf{1}_C(x), \inf_{x\in\mathcal{X}} V_k(x) \geq c_k > 0, \\ V_k(x) - b_k \geq a_k V_k(x), x \in D^c, \\ \sup_D V_q < \infty. \end{gathered} \tag{6}$$

- *$\pi(V_q^\beta) < \infty$ for some $\beta \in (0, 1]$.*

Then, for any $x \in \mathcal{X}$, $n \geq m$,

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq \min_{1\leq l\leq q} B_l^{(\beta)}(x, n), \tag{7}$$

with

$$B_l^{(\beta)}(x, n) = \frac{\epsilon^+ \left\langle (I - A_m^{(\beta)})^{-1} \delta_x \otimes \pi(W^\beta), e_l \right\rangle}{S(l, n+1-m)^\beta + \sum_{j\geq n+1-m} (1-\epsilon^+)^{j-(n-m)} (S(l, j+1)^\beta - S(l, j)^\beta)},$$

where $< \cdot, \cdot >$ denotes the inner product in $\mathbb{R}^{q+1}$, $\{e_l\}$, $0 \leq l \leq q$ is the canonical basis on $\mathbb{R}^{q+1}$, I is the identity matrix;

$$\delta_x \otimes \pi(W^\beta) := \int \delta_x(dy)\pi(dy')W^\beta(y, y')$$

where $W^{\beta}(x,x') := \left(W_0^{\beta}(x,x'), \cdots, W_q^{\beta}(x,x')\right)^T$ *with* $W_0(x,x') := 1$ *and*

$$W_l(x,x') = \mathbf{1}_{\triangle}(x,x') + \mathbf{1}_{\triangle^c}(x,x') \left(\prod_{k=0}^{l-1} a_k\right)^{-1} (m(V_0))^{-1} \left(V_l(x) + V_l(x')\right) \text{ for } 1 \le l \le q$$

where $m(V_0) := \inf_{(x,x') \in \triangle^c} \{V_0(x) + V_0(x')\}$;

$$S(0,k) := 1 \text{ and } S(i,k) := \sum_{j=1}^{k} S(i-1,j), i \ge 1;$$

$$A_m^{(\beta)} := \begin{pmatrix} A_m^{(\beta)}(0) & 0 & \cdots & 0 & 0 \\ A_m^{(\beta)}(1) & A_m^{(\beta)}(0) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_m^{(\beta)}(q-1) & A_m^{(\beta)}(q-2) & \cdots & A_m^{(\beta)}(0) & 0 \\ A_m^{(\beta)}(q) & A_m^{(\beta)}(q-1) & \cdots & A_m^{(\beta)}(1) & A_m^{(\beta)}(0) \end{pmatrix},$$

where $A_m^{(\beta)}(l) := \sup_{(x,x') \in \triangle} \sum_{i=0}^{l} S(i,m)^{\beta} \left(1 - \rho_{x,x'}(\mathcal{X})\right) \int R_{x,x'}(x,dy) R_{x,x'}(x',dy') W_{l-i}^{\beta}(y,y')$, *where the residual kernel*

$$R_{x,x'}(u,dy) := \left(1 - \rho_{x,x'}(\mathcal{X})\right)^{-1} \left(P_{\gamma}^{m}(u,dy) - \rho_{x,x'}(dy)\right);$$

and $\epsilon^+ := \sup_{(x,x') \in \triangle} \rho_{x,x'}(\mathcal{X})$.

Remark 5.1. In the $B_l^{(\beta)}(x,n)$, ϵ^+ depends on the set $\triangle$ and the measure $\rho_{x,x'}$; the matrix $(I - A_m^{(\beta)})^{-1}$ depends on the set $\triangle$, the transition kernel P, $\rho_{x,x'}$ and the test functions V_k; $\delta_x \otimes \pi(W^{\beta})$ depends on the set $\triangle$ and the test functions V_k.

Consider the special case of the theorem: $\rho_{x,x'}(dy) = \delta\nu(dy)$ where ν is a probability measure with $\nu(C) > 0$, and $\triangle := C \times C$.

1. $\epsilon^+ = \epsilon^- = (1-\delta)$.

2. $I - A_m^{(\beta)}$ is a lower triangle matrix so $(I - A_m^{(\beta)})^{-1} = \left(b_{ij}^{(\beta)}\right)_{i,j=1,\ldots,q+1}$ is also a lower triangle matrix, and fixing $k \ge 0$ all $b_{i,i-k}^{(\beta)}$ are equal. $b_{ii}^{(\beta)} = \frac{1}{1 - A_m^{(\beta)}(0)}$. For $i > j$, $b_{ij}^{(\beta)}$ is the polynomial combination of $A_m^{(\beta)}(0), \cdots, A_m^{(\beta)}(i+1)$ divided by $(1 - A_m^{(\beta)}(0))^i$. By some algebras, we can obtain that $b_{21}^{(\beta)} = \frac{A_m^{(\beta)}(1)}{(1 - A_m^{(\beta)}(0))^2}$. So, by calculating $B_1^{(\beta)}(x,n)$, we can get the quantitative bound with a simple form. $B_1^{(\beta)}(x,n)$ only involves two test functions $V_0(x)$ and $V_1(x)$.

Remark 5.2. From Equation (6), $V_0(x) \ge b_0/(1-\alpha_0) > b_0$ because $0 < \alpha_0 < 1$. Consider the drift condition: $PV_1 - V_1 \le -V_0 + b_0 \mathbf{1}_C$. Since $\pi P = \pi$, $\pi(V_0) \le b_0 \pi(C) \le b_0$. Hence, the V_0 in the theorem can not be constant.

Remark 5.3. Without the condition $\pi(V^\beta) < \infty$, the bound in Equation (7) can also be obtained. However, the bound is possibly infinity. The subscript l of $B_l^{(\beta)}(x,n)$ and β can explain the polynomial rate. The related rate is $S(l, n+1-m)^\beta = O((n+1-m)^{l\beta})$. It can be observed that $B_l^{(\beta)}(x,n)$ involves test functions $V_0(x), \cdots, V_l(x)$, and $\limsup_n n^{\beta l} B_l^{(\beta)}(x,n) < \infty$. The maximal rate of convergence is equal to $q\beta$.

5.1 Conditions

The following conditions derive from Theorem 5.1, and some changes are added to apply for adaptive MCMC algorithms. Define the adaptive MCMC algorithm $\{X_n\}$ is *simultaneously polynomially ergodic* (S.P.E.) if the conditions (A1)-(A4) are satisfied.

A1: each P_γ is ϕ_γ-irreducible with stationary distribution $\pi(\cdot)$;

Remark 5.4. From MT Proposition 10.1.2, if P_γ is φ-irreducible, then P_γ is π-irreducible and the invariant measure π is a maximal irreducibility measure.

Remark 5.5. By Theorem 5.2.2 in Meyn and Tweedie [11] (MT), if P is ψ-irreducible (ψ is a maximal irreducibility measure), then for any set $A \in \mathcal{F}$ with $\psi(A) > 0$, there exists a ν-small set $C \subset A$ $P^m(x,\cdot) \geq \delta \mathbf{1}_C(x)\nu(\cdot)$, where $\delta > 0$ and ν is a probability measure with $\nu(C) > 0$.

A2: there is a set $C \subset \mathcal{X}$, some integer $m \in \mathbb{N}$, some constant $\delta > 0$, and some probability measure $\nu_\gamma(\cdot)$ on $\mathcal{X}$ such that:

$$\pi(C) > 0, \text{ and } P_\gamma^m(x,\cdot) \geq \delta \mathbf{1}_C(x)\nu_\gamma(\cdot) \text{ for } \gamma \in \mathcal{Y}; \tag{8}$$

Remark 5.6. In Theorem 5.1, there is one condition (Equation (5)) ensuring the splitting technique. Here we consider the special case of that condition: $\rho_{x,x'}(dy) = \delta\nu_\gamma(dy)$ and $\triangle = C \times C$. Thus, by Remark 5.1, the bound of $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|$ depends on C, m, the minorization constant δ, $\pi(\cdot)$, ν_γ, and test functions $V_l(x)$ so we assume that they are uniform on all the transition kernels.

A3: there is $q \in \mathbb{N}$ and measurable functions: $V_0, V_1, \ldots, V_q : \mathcal{X} \to (0,\infty)$ where $1 \leq V_0 \leq V_1 \leq \cdots \leq V_q$, such that for $k = 0, 1, \ldots, q-1$, there are $0 < \alpha_k < 1$, $b_k < \infty$, and $c_k > 0$ such that:

$$P_\gamma V_{k+1}(x) \leq V_{k+1}(x) - V_k(x) + b_k \mathbf{1}_C(x),\ V_k(x) \geq c_k \text{ for } x \in \mathcal{X} \text{ and } \gamma \in \mathcal{Y}; \tag{9}$$

$$V_k(x) - b_k \geq \alpha_k V_k(x) \text{ for } x \in \mathcal{X}/C; \tag{10}$$

$$\sup_{x \in C} V_q(x) < \infty. \tag{11}$$

Remark 5.7. Obviously, each P_γ is ν_γ-irreducible. $\nu_\gamma(V_l) \leq \mathbf{1}_C(x)\frac{1}{\delta}P_\gamma^m V_l(x) \leq \frac{1}{\delta}\sup_{x\in C} V_l(x) + \frac{mb_{l-1}}{\delta}$.

A4: $\pi(V_q^\beta) < \infty$ for some $\beta \in (0,1]$.

5.2 Main Result

Although the ergodic property of adaptive MCMC algorithms with Markovian adaptation has been basically resolved in [3], there still are some parts unknown about the general adaptive MCMC algorithms. Here, we discuss the properties of adaptive MCMC algorithms with S.P.E..

Lemma 5.2. *Suppose that the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is S.P.E.. If for any $l \in \{1,\dots,q\}$, the stochastic process $V_l(X_n)$ is bounded in probability, then the containment condition is satisfied.*

The proof is in Section 5.4.

Remark 5.8. By Lemma 5.2, the condition (M3) is satisfied provided that (A1)-(A4). Therefore, any adaptive MCMC algorithm with Markovian adaptation and diminishing adaptation satisfying S.P.E. is ergodic by Theorem 3.4.

Remark 5.9. In Lemma 5.2, (A4) will be used to show that the uniform quantitative bound of $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|$ is finite.

Theorem 5.3. *Suppose the an adaptive MCMC algorithm satisfies the diminishing adaptation. Then, the algorithm is ergodic under either of the following cases:*
(i) S.P.E., and the number q of simultaneous drift conditions is strictly greater than two;
(ii) S.P.E., and when the number of simultaneous drift conditions is greater than or equal to two, there exists an increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $V_1(\cdot) \leq f(V_0(\cdot))$;
(iii) (A1), (A2), (A4), and the simultaneous drift condition has the form

$$P_\gamma V(x) - V(x) \leq -cV^\alpha(x) + b\mathbf{1}_C(x) \tag{12}$$

where $cV^\alpha(x) \geq b$ on C^c, for any $\alpha \in (0,1)$;
(iv) (A1), (A2), (A4), and when the simultaneous drift condition has the form

$$P_\gamma V_1(x) - V_1(x) \leq -V_0(x) + b\mathbf{1}_C(x) \tag{13}$$

where $V_0(x) \geq b$ on C^c, and the process $\{V_1(X_n)\}$ is bounded in probability.

The theorem consists of Theorem 5.7, Theorem 5.8, Theorem 5.10, and Lemma 5.2.

Remark 5.10. In the theorem, the containment condition is directly implied by S.P.E..

Remark 5.11. Polynomial ergodicity means that for some rate function $r(n)$, $\lim_{n\to\infty} r(n) \left\| P_\gamma^n(x,\cdot) - \pi(\cdot) \right\| = 0$. Indeed, the above theorem presents that if all the transition kernels converge in polynomial rate $r(n) = O(n^p)$ for $p > 0$.

1. In the case (i), $p > 1$. In the case (ii), $p \geq 1$.

2. In the case (iii), the polynomial rate depends on the value α, and $p = 1/(1-\alpha) - 1$. Jarner and Roberts [10] (JR) prove that for $\alpha \geq 1/2$, this kind of drift condition can imply the nested drift conditions, and the number of drift conditions is equal to $[\frac{1}{1-\alpha}]$ where $[x]$ is the maximal integer not exceeding x. Eg. if $\alpha \geq \frac{2}{3}$, then part (iii) implies part (i); if $\alpha \geq \frac{1}{2}$, and assume that V_0 and V_1 has the relation described in part (ii), then part (iii) implies part (ii).

5.3 Examples

In the following example (Example 3.1 in [5]), we show that under some situations, the simultaneous drift conditions (Y2), (M2) and (A3) do not hold.

Consider the Metropolis-Hastings algorithm on the state space $\mathcal{X} = (0, +\infty)$, and the adaptive parameter space $\mathcal{Y} = \{-1, 1\}$. The target density $\pi(x) \propto \frac{\mathbf{1}(x>0)}{1+x^2}$. Let $\{Z_n\}_n$ be i.i.d. standard normal. the proposal values are given

$$Y_n^{\Gamma_{n-1}} = X_{n-1}^{\Gamma_{n-1}} + Z_n.$$

i.e. if $\Gamma_{n-1} = 1$ then $Y_n = X_{n-1} + Z_n$, while if $\Gamma_{n-1} = -1$ then $Y_n = \frac{1}{1/X_{n-1} + Z_n}$. The adaptation is defined by

$$\Gamma_n = -\Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} < 1/n) + \Gamma_{n-1}\mathbf{1}(X_n^{\Gamma_{n-1}} \geq 1/n),$$

i.e. we change Γ from 1 to -1 when $X < 1/n$, and change Γ from -1 to 1 when $X > n$. otherwise we do not change Γ.

By some algebras, we have that if $\Gamma_{n-1} = 1$,

$$\begin{aligned} P_{\Gamma_{n-1}}(x, A) &= \int_0^\infty \mathbf{1}_A(y) \left(\frac{1+x^2}{1+y^2} \wedge 1 \right) \phi(y-x) dy + \\ &\quad \delta_x(A) \int_0^\infty \left[1 - \frac{1+x^2}{1+y^2} \wedge 1 \right] \phi(y-x) dy; \end{aligned}$$

if $\Gamma_{n-1} = -1$,

$$\begin{aligned} P_{\Gamma_{n-1}}(x, A) &= \int_0^\infty \mathbf{1}_A(y) \left(\frac{1+x^{-2}}{1+y^{-2}} \wedge 1 \right) \phi(y^{-1} - x^{-1})/y^2 dy + \\ &\quad \delta_x(A) \int_0^\infty \left[1 - \frac{1+x^{-2}}{1+y^{-2}} \wedge 1 \right] \phi(y^{-1} - x^{-1})/y^2 dy, \end{aligned}$$

where $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$.

Obviously, P_γ is π-irreducible and aperiodic. By Theorem 2.2 of Roberts and Tweedie [14], $P_1(x,\cdot) \geq \epsilon\lambda(\cdot)$ for x in any compact set on $\mathbb{R}^+$, where λ is Lebegue measure. Although for P_{-1}, the conditions of that theorem are not satisfied, the minorization condition is still satisfied, because

$$\begin{aligned} & \int_0^\infty \mathbf{1}_A(y)\left(\frac{1+x^{-2}}{1+y^{-2}} \wedge 1\right)\phi(y^{-1}-x^{-1})/y^2 dy \\ = & \int_0^\infty \mathbf{1}_A(y)\left((1+x^{-2}) \wedge (1+y^{-2})\right)\frac{\phi(y^{-1}-x^{-1})}{1+y^2}dy \\ \geq & \int_0^\infty \mathbf{1}_A(y)\frac{\phi(y^{-1}-x^{-1})}{1+y^2}dy \geq \epsilon\mu(A), \end{aligned}$$

where $\mu(\cdot)$ is probability measure by normalizing $\frac{\phi(y^{-1}-x^{-1})}{1+y^2}$, which is absolutely continuous w.r.t. λ. So, Equation (8) is satisfied. The algorithm is not ergodic to the stationary measure $\pi(\cdot)$ (See details in [5]). Hence, S.P.E. is not satisfied.

5.4 Proof of Theorem 5.3

Proof of Lemma 5.2: We use the notation in Theorem 5.1.

Since $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is S.P.E., we know that for each $\gamma \in \mathcal{Y}$, $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\| \leq B_1^{(\beta)}(x,n)$.

By the definition of ϵ^+ and ϵ^-, $\epsilon^+ = \epsilon^- = \delta$.

From Remark 5.1, $\rho_{x,x'}(dy) = \delta\nu_\gamma(dy)$ so that $\rho_{x,x'}(\mathcal{X}) = \delta$. Since the matrix $I - A_m^{(\beta)}$ is a lower triangle matrix, so is $(I - A_m^{(\beta)})^{-1} := (b_{ij}^{(\beta)})_{i,j=0,\cdots,q}$. By the definition of $B_l^{(\beta)}(x,n)$,

$$\begin{aligned} B_l^{(\beta)}(x,n) &= \frac{\epsilon^+\sum_{k=0}^l b_{lk}^{(\beta)}\int\pi(dy)W_k^\beta(x,y)}{S(l,n+1-m)^\beta + \sum_{j\geq n+1-m}(1-\epsilon^+)^{j-(n-m)}(S(l,j+1)^\beta - S(l,j)^\beta)} \\ &\leq \frac{\epsilon^+}{S(l,n+1-m)^\beta}\sum_{k=0}^l b_{lk}^{(\beta)}\int\pi(dy)W_k^\beta(x,y). \end{aligned}$$

Consider the term:

$$\begin{aligned} & \int\pi(dy)W_k^\beta(x,y) \\ = & \mathbf{1}_C(x)\pi(C) + \mathbf{1}_C(x)\int_{C^c}\pi(dy)\left(m(V_0)\prod_{k=0}^{l-1}a_k\right)^{-\beta}(V_k(x)+V_k(y))^\beta + \\ & \mathbf{1}_{C^c}(x)\int\pi(dy)\left(m(V_0)\prod_{i=0}^{l-1}a_i\right)^{-\beta}(V_k(x)+V_k(y))^\beta \\ \leq & \pi(C) + \left(m(V_0)\prod_{i=0}^{l-1}a_i\right)^{-\beta}\left[V_k^\beta(x) + \pi(V_k^\beta)\right], \end{aligned}$$

because $\left(x^{\beta}\right)' \leq 1$ for $x \geq 1$ and $\beta \in (0,1]$ is concave so $(x+1)^{\beta} - x^{\beta} \leq 1$ for $x \geq 1$, then $(V_0(x) + V_0(x'))^{\beta} \leq V_0^{\beta}(x) + V_0^{\beta}(x')$.

By the definition of $m(V_0)$, $m(V_0) \geq 2$.

By induction, we obtain that $b_{10}^{(\beta)} = \frac{A_m^{(\beta)}(1)}{(1-A_m^{(\beta)}(0))^2}$, and $b_{11}^{(\beta)} = \frac{1}{1-A_m^{(\beta)}(0)}$.

Since $0 < W_k(x,y) \leq 1$, $A_m^{(\beta)}(0) \leq 1 - \epsilon^- = 1 - \delta$ so that $\frac{1}{1-A_m^{(\beta)}(0)} \leq \frac{1}{\delta}$.

Consider the term

$$\begin{aligned}
& A_m^{(\beta)}(1) \\
\leq \;& m^{\beta} A_m^{(\beta)}(0) + \sup_{(x,x') \in \triangle} (1-\delta) \int R_{x,x'}(x,dy) R_{x,x'}(x',dy') W_1^{\beta}(y,y') \\
\leq \;& \left(m^{\beta} + 1\right)(1-\delta).
\end{aligned}$$

Hence,

$$b_{10}^{(\beta)} \leq \frac{m^{\beta}+1}{1-\delta}, \text{ and } b_{11}^{(\beta)} \leq \frac{1}{1-\delta}.$$

Thus,

$$\begin{aligned}
& B_1^{(\beta)}(x,n) \\
\leq \;& \frac{\delta}{S(1,n+1-m)^{\beta}} \left(b_{10}^{(\beta)} \int \pi(dy) W_0^{\beta}(x,dy) + b_{11}^{(\beta)} \int \pi(dy) W_1^{\beta}(x,dy) \right) \\
\leq \;& \frac{1}{(1-\delta) S(1,n+1-m)^{\beta}} \left((m^{\beta}+2)\pi(C) + 2^{-\beta} \left[(m^{\beta}+1)(V_0^{\beta} + \pi(V_0^{\beta})) + a_0^{-\beta}(V_1^{\beta} + \pi(V_1^{\beta})) \right] \right).
\end{aligned}$$

Therefore, the boundedness of the process $V_1(X_k)$ implies that the $B_1^{(\beta)}(x,n)$ converges to zero uniformly on $\mathcal{Y}$. The containment condition holds. □

Lemma 5.4 (Dynkin's Formula for adaptive MCMC)**.** *Consider an adaptive MCMC $\{X_n\}$ with adaptive scheme $\{\Gamma_n\}$ (satisfying adaptive MCMC property: Equation (1)). $f(\cdot)$ is a fixed Borel measurable function of X_k. For each $(x,\gamma) \in \mathcal{X} \times \mathcal{Y}$ and $n \in \mathbb{Z}^+$,*

$$E_{(x,\gamma)}\left[f(X_{\tau^n})\right] = E_{(x,\gamma)}\left[f(X_0)\right] + E_{(x,\gamma)}\left[\sum_{i=1}^{\tau^n} \left(E_{\Gamma_{i-1}}[f(X_i)|X_{i-1}] - f(X_{i-1})\right)\right],$$

where τ be any stopping time for X_n, and $\tau^n = \tau \wedge n \wedge \inf\{k \geq 0 : f(X_k) \geq n\}$.

Proof: For each $n \in \mathbb{Z}^+$,

$$\begin{aligned}
f(X_{\tau^n}) &= f(X_0) + \sum_{i=1}^{\tau^n} (f(X_i) - f(X_{i-1})) \\
&= f(X_0) + \sum_{i=1}^{n} \mathbf{1}(\tau^n \geq i)(f(X_i) - f(X_{i-1})).
\end{aligned}$$

Since τ is a stopping time w.r.t. X_n, $\{\tau^n \geq i\} \in \mathcal{F}_{i-1}$. Hence, by Equation (1),

$$\begin{aligned} & E_{(x,\gamma)}\left[f(X_{\tau^n})\right] \\ = \;& E_{(x,\gamma)}\left[f(X_0)\right] + E_{(x,\gamma)}\left[\sum_{i=1}^{n} E_{(x,\gamma)}\left[f(X_i) - f(X_{i-1})|\mathcal{F}_{i-1}\right]\mathbf{1}(\tau^n \geq i)\right] \\ = \;& E_{(x,\gamma)}\left[f(X_0)\right] + E_{(x,\gamma)}\left[\sum_{i=1}^{\tau^n} \left(E_{\Gamma_{i-1}}[f(X_i)|X_{i-1}] - f(X_{i-1})\right)\right]. \end{aligned}$$

□

Lemma 5.5 (Comparison Lemma for adaptive MCMC)**.** *Consider an adaptive MCMC $\{X_n\}$ with adaptive scheme $\{\Gamma_n\}$. Suppose that the non-negative functions V, f, s satisfying the relationship*

$$P_\gamma V(x) - V(x) \leq -f(x) + s(x), \quad x \in \mathcal{X} \text{ and } \gamma \in \mathcal{Y}.$$

Then for each $x \in \mathcal{X}$ and any stopping time τ of X_n we have

$$E_{(x_0,\gamma_0)}\left[\sum_{k=0}^{\tau-1} f(X_{n+k})|X_n,\Gamma_n\right] \leq V(x_0) + E_{(x_0,\gamma_0)}\left[\sum_{k=0}^{\tau-1} s(X_{n+k})|X_n,\Gamma_n\right].$$

Proof: The proof is same as Theorem 14.2.2 in MT [11], and omitted. □

The following proposition shows the relations between the moments of the hitting time and the test function V-modulated moments for adaptive MCMC algorithms with S.P.E., which is derived from the result for Markov chain in JR [10].

Proposition 5.6. *If the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is S.P.E., then*

$$\begin{aligned} \frac{E_{(x_0,\gamma_0)}\left[\tau_{n,C}^k|X_n,\Gamma_n\right]}{kc_{q-k}} & \leq E_{(x_0,\gamma_0)}\left[\sum_{j=0}^{\tau_{n,C}-1}(j+1)^{k-1}V_{q-k}(X_{n+j}) \mid X_n,\Gamma_n\right] \\ & \leq d_{q-k}\left(V_q(X_n) + \sum_{j=1}^{k} b_{q-j}\mathbf{1}_C(X_n)\right), \end{aligned} \tag{14}$$

for $k = 1, \ldots, q$, and some constant d_{q-k}.

Proof: Since $V_k \geq c_k$,

$$\sum_{i=0}^{\tau_{n,C}-1} (i+1)^{k-1} \geq \int_0^{\tau_{n,C}} x^{k-1}dx = k^{-1}\tau_{n,C}^k.$$

So, the first inequality holds.

By Lemmas 5.4 and 5.5 and S.P.E., we obtain

$$\begin{aligned} & E_{(x_0,\gamma_0)}\left[V_k(X_{\tau_{n,C}}) \mid X_n,\Gamma_n\right] \\ \leq \;& V_k(X_n) - E_{(x_0,\gamma_0)}\left[\sum_{i=0}^{\tau_{n,C}-1} V_{k-1}(X_{n+i}) \mid X_n,\Gamma_n\right] + b_{k-1}\mathbf{1}_C(X_n). \end{aligned} \tag{15}$$

Hence,

$$\begin{aligned} E_{(x_0,\gamma_0)}\left[\tau_{n,C} \mid X_n, \Gamma_n\right]/c_{k-1} \leq & \quad E_{(x_0,\gamma_0)}\left[\sum_{i=0}^{\tau_{n,C}-1} V_{k-1}(X_{n+i}) \mid X_n, \Gamma_n\right] \\ \leq & \quad V_k(X_n) + b_{k-1}\mathbf{1}_C(X_n). \end{aligned} \tag{16}$$

From Equation (16) and $V_k(x) \geq c_k > 0$, we have that

$$\sup_{x\in C} E_{(x_0,\gamma_0)}\left[\sum_{i=0}^{\tau_{n,C}-1} V_{k-1}(X_{n+i}) \mid X_n = x\right] \quad < \quad \infty, \tag{17}$$

$$\sup_{x\in C} E_{(x_0,\gamma_0)}\left[\tau_{n,C} \mid X_n = x\right] \quad \leq \quad \left(\sup_{x\in C} V_k(x) + b_{k-1}\right)/c_{k-1} < \infty. \tag{18}$$

Since Equation (18), $\sup_{x\in C}\sum_{i=0}^{\infty} P_{(x_0,\gamma_0)}\left(\tau_{n,C} > i \mid X_n = x\right) < \infty$.
So, $\sup_{x\in C} P_{(x_0,\gamma_0)}\left(\tau_{n,C} > i \mid X_n = x\right) = o\left(i^{-1}\right)$. Hence,

$$\sup_{x\in C} P_{(x_0,\gamma_0)}\left(\tau_{n,C} < \infty \mid X_n = x\right) = 1. \tag{19}$$

Using induction and the technique in [10], the results hold. □

Theorem 5.7. *Suppose that the family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is S.P.E. for $q > 2$. Then, the containment condition holds.*

Proof: For $k = 1, \ldots, q$, take large enough $M > 0$ such that $C \subset \{x : V_{q-k}(x) \leq M\}$,

$$\begin{aligned} & P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M\right) \\ = & \sum_{i=0}^{n} P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M, \tau_{i,C} > n-i, X_i \in C\right) + \\ & P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M, \tau_{0,C} > n, X_0 \notin C\right). \end{aligned}$$

By Proposition 5.6, for $i = 0, \cdots, n$,

$$\begin{aligned} & P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M, \tau_{i,C} > n-i \mid X_i \in C\right) \\ \leq & P_{(x_0,\gamma_0)}\left(\sum_{j=0}^{\tau_{i,C}-1}(j+1)^{k-1}V_{q-k}(X_{i+j}) > (n-i)^{k-1}M + c_{q-k}\sum_{j=0}^{n-i-1}(j+1)^{k-1}, \tau_{i,C} > n-i \mid X_i \in C\right) \\ \leq & P_{(x_0,\gamma_0)}\left(\sum_{j=0}^{\tau_{i,C}-1}(j+1)^{k-1}V_{q-k}(X_{i+j}) > (n-i)^{k-1}M + c_{q-k}\sum_{j=0}^{n-i-1}(j+1)^{k-1} \mid X_i \in C\right) \\ \leq & \frac{\sup_{x\in C} E_{(x_0,\gamma_0)}\left[E_{(x_0,\gamma_0)}\left[\sum_{j=0}^{\tau_{i,C}-1}(j+1)^{k-1}V_{q-k}(X_{i+j}) \mid X_i, \Gamma_i\right] \mid X_i = x\right]}{(n-i)^{k-1}M + c_{q-k}\sum_{j=0}^{n-i-1}(j+1)^{k-1}} P_{(x_0,\gamma_0)}(X_i \in C) \\ \leq & \frac{d_{q-k}\left(\sup_{x\in C} V_q(x) + \sum_{j=1}^{k} b_{q-j}\mathbf{1}_C(x)\right)}{(n-i)^{k-1}M + c_{q-k}\sum_{j=0}^{n-i-1}(j+1)^{k-1}} P_{(x_0,\gamma_0)}(X_i \in C), \end{aligned}$$

and

$$\begin{aligned}
& P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M, \tau_{0,C} > n \mid X_0 \notin C\right) \\
\leq \quad & \frac{d_{q-k}\left(V_q(x_0) + \sum_{j=1}^{k} b_{q-j}\mathbf{1}_C(x_0)\right)}{n^{k-1}M + c_{q-k}\sum_{j=0}^{n-1}(j+1)^{k-1}} P_{(x_0,\gamma_0)}(X_0 \notin C).
\end{aligned}$$

By simple algebra,

$$(n-i)^{k-1}M + c_{q-k}\sum_{j=0}^{n-i-1}(j+1)^{k-1} = O\left((n-i)^{k-1}\left(M + c_{q-k}(n-i)\right)\right).$$

Therefore,

$$\begin{aligned}
& P_{(x_0,\gamma_0)}\left(V_{q-k}(X_n) > M\right) \\
\leq \quad & d_{q-k}\left(\sup_{x\in C\cup\{x_0\}} V_q(x) + \sum_{j=1}^{k} b_{q-j}\right)\cdot\left(\sum_{i=0}^{n} \frac{\left[P_{(x_0,\gamma_0)}(X_i\in C)\right]^2}{(n-i)^{k-1}\left(M+c_{q-k}(n-i)\right)} + \frac{\delta_{C^c}(x_0)}{n^{k-1}\left(M+c_{q-k}n\right)}\right).
\end{aligned} \tag{20}$$

Whenever $q > 2$, k can be equal to 2. The summation of L.H.S. of Equation (20) is finite given M. Hence, taking large enough $M > 0$, the probability will be small enough. So, the sequence $\{V_{q-2}(X_n)\}_n$ is bounded in probability. By Lemma 5.2, the containment condition holds. □

Remark 5.12. In the proof, the parameter β (A4) is not used.

Remark 5.13. When $q = 2$, from the proof of Theorem 5.7, the sequence $\{V_0(X_n)\}_n$ is bounded in probability which is not helpful to uniformly bound the distance $\left\|P_\gamma^n(x,\cdot) - \pi(\cdot)\right\|$. However, if $V_0(\cdot)$ is a nice function (non-decreasing) of $V_1(\cdot)$, then the sequence $\{V_1(X_n)\}_n$ is bounded in probability. In Theorem 5.10, we discuss this situation for simultaneously single polynomial drift condition.

Theorem 5.8. *Suppose that $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is S.P.E. for $q = 2$. Suppose that there exists a strictly increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that $V_1(x) \leq f(V_0(x))$ for all $x \in \mathcal{X}$. Then, the containment condition is implied.*

Proof: From Equation (20), we have that $\{V_0(X_n)\}_n$ is bounded in probability. Since $V_1(x) \leq f(V_0(x))$,

$$P_{(x_0,\gamma_0)}\left(V_1(X_n) > f(M)\right) \leq P_{(x_0,\gamma_0)}\left(f(V_0(X_n)) > f(M)\right) = P_{(x_0,\gamma_0)}\left(V_0(X_n) > M\right),$$

because $f(\cdot)$ is strictly increasing. By the boundedness of $V_0(X_n)$, for any $\epsilon > 0$, there exists $N > 0$ and some $M > 0$ such that for $n > N$, $P_{(x_0,\gamma_0)}\left(V_1(X_n) > f(M)\right) \leq \epsilon$. Therefore, $\{V_1(X_n)\}_n$ is bounded in probability. By Lemma 5.2, the containment condition is satisfied. □

Consider the single polynomial drift condition, see [10]: $P_\gamma V(x) - V(x) \le -cV^\alpha(x) + b\mathbf{1}_C(x)$ where $0 \le \alpha < 1$. Because the moments of the hitting time to the set C is (see details in [10]), for any $1 \le \xi \le 1/(1-\alpha)$,

$$E_x\left[\sum_{i=0}^{\tau_C - 1}(i+1)^{\xi-1}V(X_i)\right] < V(x) + b\mathbf{1}_C(x).$$

The polynomial rate function $r(n) = n^{\xi-1}$. If $\alpha = 0$, then $r(n)$ is a constant. Under this situation, it is difficult to utilize the technique in Theorem 5.7 to prove $V(X_n)$ is bounded in probability. Thus, we assume $\alpha \in (0,1)$.

Proposition 5.9. *Suppose the family* $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ *is S.P.E. with only one simultaneous drift condition, and it has the form*

$$(P_\gamma V)(x) - V(x) \le -cV^\alpha(x) + b\mathbf{1}_C(x), \tag{21}$$

for any $\alpha \in (0,1)$, *then*

$$E_{(x_0,\gamma_0)}\left[\sum_{i=0}^{\tau_{n,C}-1}(i+1)^{\xi-1}V^{1-\xi(1-\alpha)}(X_{n+i}) \mid X_n, \Gamma_n\right] \le c_\xi(C)(V(X_n)+1). \tag{22}$$

Proof: The proof applies the techniques in Lemma 3.5 and Theorem 3.6 of [10]. From their results, Equation (21) can be transformed into a group of polynomial drift conditions with the same forms in the definition of S.P.E.. □

Theorem 5.10. *Suppose the conditions in Proposition 5.9 are satisfied. Then, the containment condition is implied.*

Proof: Using the same techniques in Theorem 5.7, we have that

$$\begin{aligned} & P_{(x_0,\gamma_0)}\left(V^{1-\xi(1-\alpha)}(X_n) > M\right) \\ \le \; & c_\xi\left(\sup_{x\in C\cup\{x_0\}} V(x) + 1\right)\cdot\left(\sum_{i=0}^n \frac{\left[P_{(x_0,\gamma_0)}(X_i\in C)\right]^2}{(n-i)^{\xi-1}(M+n-i)} + \frac{\delta_{C^c}(x_0)}{n^{\xi-1}(M+n)}\right). \end{aligned} \tag{23}$$

Therefore, for any $\alpha \in (0,1)$, for any $\xi \in (1, 1/(1-\alpha))$, the sequence $\left\{V^{1-\xi(1-\alpha)}(X_n)\right\}_n$ is bounded in probability. Since $1-\xi(1-\alpha) > 0$, the process $\{V(X_n)\}$ is bounded in probability. By Lemma 5.2, the containment condition holds. □

Acknowledgements

Special thanks are due to Professor Jeffrey S. Rosenthal. Without his assistance this work would not have been completed. The author is very grateful to G. Fort for her helpful suggestions.

References

[1] Andrieu, C and Moulines, E. On the ergodicity properties of some adaptive Markov Chain Monte Carlo algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.

[2] Andrieu, C and Robert, C.P. Controlled mcmc for optimal sampling. *Preprint*, 2002.

[3] Atchadé, Y.F. and Fort, G. Limit Theorems for some adaptive MCMC algorithms with subgeometric kernels. *Preprint*, 2008.

[4] Atchadé, Y.F. and Rosenthal, J.S. On Adaptive Markov Chain Monte Carlo Algorithms. *Bernoulli*, 11(5):815–828, 2005.

[5] Bai, Y., Roberts, G.O., and Rosenthal, J.S. On the containment condition of adaptive Markov Chain Monte Carlo algorithms. *Preprint*, 2008.

[6] Brockwell, A.E. and Kadane, J.B. Indentification of regeneration times in mcmc simulation, with application to adaptive schemes. *J. Comp. Graph. Stat*, 14:436–458, 2005.

[7] Fort, G. and Moulines, E. Computable bounds for subgeometrical and geometrical ergodicity. 2000.

[8] Gilks, W.R., and Roberts, G.O. and Sahu, S.K. Adaptive Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 93:1045–1054, 1998.

[9] Haario, H., and Saksman, E., and Tamminen, J. An adaptive metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[10] Jarner, S.F. and Roberts, G.O. Polynomial convergence rates of markov chains. *Ann. Appl. Probab.*, 12(1):224–247, 2002.

[11] Meyn, S.P. and Tweedie, R.L. *Markov Chains and Stochastic Stability.* London: Springer-Verlag, 1993.

[12] Robbins, H. and Monro, S. A stochastic approximation method. *Ann. Math. Stat.*, 22:400–407, 1951.

[13] Roberts, G.O. and Rosenthal, J.S. Coupling and Ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Prob.*, 44:458–475, 2007.

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

[15] Yang, Chao. Ergodicity of Adaptive MCMC and its applications. *Ph.D. thesis*, 2008.

[16] Yang, Chao. Recurrent and Ergodic Properties of Adaptive MCMC. *Preprint*, 2008.