Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 3, 640–652
DOI: 10.1214/16-BJPS327

Boosting, downsizing and optimality of test functions of Markov chains

Thomas R. Boucher

Department of Mathematics, Texas A&M University-Commerce

Abstract. Test functions play an important role in Markov chain theory. Stability of a Markov chain can be demonstrated by constructing a test function of the chain that satisfies a stochastic drift criterion. The test function defines a class of functions of the process for which limit laws hold, yields bounds on the convergence of the Markov chain transition probabilities to the stationary distribution, and provides information concerning the mixing properties of the chain. Under certain conditions, these results can be improved by using a new test function derived from a known test function of a Markov chain.

1 Introduction

Let $\{X_t\}$ be a ψ-irreducible, aperiodic Markov chain on a general state space $\mathcal{X}$. Assume $\mathcal{X}$ is sufficiently tractable, such as a Polish space. Denote the Borel sets on $\mathcal{X}$ by $\mathcal{B}(\mathcal{X})$, the transition kernel of $\{X_t\}$ by P_x and the n-step transition kernel of $\{X_t\}$ by P_x^n. Where the argument x is not relevant it will be suppressed, that is, P and P^n. For a measure μ on $\mathcal{B}(\mathcal{X})$, kernel Q on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$, and function f on $\mathcal{X}$ denote $\mu f := \int_{\mathcal{X}} f(x)\mu(dx)$ and $Q_x f := \int_{\mathcal{X}} f(y)Q(x, dy)$. Expectations conditioned upon an initial $X_0 = x$ or with respect to an initial probability distribution π will be denoted $E_x[\cdot]$ and $E_\pi[\cdot]$, respectively.

A Markov chain is called *ergodic* if a stationary distribution π exists and P^n converges to π in the total variation norm $\|\cdot\|_{\mathrm{TV}}$ for each initial x. If the convergence is uniform in x, the chain is called uniformly ergodic. If the convergence is $O(\rho^n)$ for $0 < \rho < 1$, then the chain is said to be geometrically ergodic. Among the various generalizations of uniform ergodicity is V-*uniform* (or V-*geometric*) ergodicity (see (16.2) in Meyn and Tweedie (1993)), which is equivalent to there existing a *test function* $V : \mathcal{X} \to [1, +\infty)$ so that P^n converges to π in the V-norm $|||\cdot|||_V$ at a uniform geometric rate ρ; i.e., for all n

$$|||P^n - \pi|||_V := \sup_{x\in\mathcal{X}} \sup_{|g|\le V} \frac{|P_x^n g - \pi g|}{V(x)} \le R\rho^n, \qquad R < \infty, \rho < 1. \tag{1}$$

In this paper, we will also suppose the test functions V are *norm-like* meaning that $V(x) \to \infty$ as $\|x\| \to \infty$. This condition focuses on the behavior of the Markov

Key words and phrases. Markov chain, convergence, ergodicity, mixing, test function.
Received September 2015; accepted June 2016.

chain when the process is large in magnitude, which is often of interest in applications.

Stochastic drift criteria involving test functions and *petite sets* are often used to demonstrate the various forms of ergodicity of Markov chains.

Definition 1. A set $A \in \mathcal{B}(\mathcal{X})$ is called *petite* if $\sum_{n\geq 0} a(n) P_x^n(C) \geq \nu(C)$ holds for all $x \in A$, $C \in \mathcal{B}(\mathcal{X})$, with ν a non-trivial measure on $\mathcal{B}(\mathcal{X})$ which may depend upon $\{a(n)\}$, a probability distribution on the nonnegative integers.

The following lemma gives a sufficient condition for V-uniform ergodicity which will be used repeatedly in this paper.

Lemma 1. *Assume $\{X_t\}$ is a ψ-irreducible, aperiodic general state Markov chain. Suppose $V \geq 1$ is bounded on petite sets. If for some integer $k \geq 1$ and petite set A*

$$\sup_{x\in A^C} \frac{P^k V}{V} < 1, \qquad \sup_{x\in A} P^k V < \infty, \qquad \sup_x PV/V < \infty,$$

then $\{X_t\}$ is V-uniformly ergodic.

Proof. Let I_A be the indicator function of the set A. It follows from the assumptions that there exists $K < \infty$ and $0 < \beta < 1$ with $P^k V - V \leq -\beta V + K I_A$. Thus (V4) in Meyn and Tweedie (1993) holds for V and the petite set A, and $\{X_{tk}\}$ is V-uniformly ergodic by Theorem 16.0.1 in Meyn and Tweedie (1993). If $k > 1$ then $\sup_x PV/V < \infty$ implies $|||P|||_V := \sup_x \sup_{|g|\leq V} |Pg|/V < \infty$ and, by Lemma 1 in Boucher and Cline (2007), $\{X_t\}$ is V-uniformly ergodic. $\square$

Example 1. Petruccelli and Woolford (1984) investigated the stability of the threshold autoregressive process of order 1 (TAR(1))

$$X_t = \phi_1 X_{t-1} I_{X_{t-1}<0} + \phi_2 X_{t-1} I_{X_{t-1}\geq 0} + \varepsilon_t,$$

where ϕ_1, ϕ_2 are real values and the ε_t are a sequence of i.i.d. random variables with mean 0. Assume their conditions $\phi_1 < 1$, $\phi_2 < 1$, and $\phi_1\phi_2 < 1$. Suppose the ε_t have a continuous density λ that is positive everywhere with $E|\varepsilon_t| < \infty$; then the process is ψ-irreducible and aperiodic, with ψ being Lebesgue measure (Petruccelli and Woolford (1984)). Let $V(x) = 1 + |x|$. Since $\limsup_{|x|\to\infty} P^2 V/V = \max(\phi_1^2, \phi_2^2, \phi_1\phi_2) < 1$ and $\sup_x PV/V < 1 + E|\varepsilon_t|/(1 + |x|) < \infty$ there exists $M < \infty$ so that the assumptions of Lemma 1 are satisfied with $k = 2$, the petite set $A = \{x : |x| \leq M\}$, and $V(x) = 1 + |x|$, implying that $\{X_t\}$ is V-uniformly ergodic.

The test function V plays an important role in applications for V-uniformly ergodic Markov chains. It follows from (1) that the test function plays a role in bounding the distance between the transition probabilities and the stationary distribution. In particular, $V \equiv 1$ is equivalent to uniform ergodicity, while for chains that are not uniformly ergodic the bound on the distance from P^n to π depends upon the initial x through $V(x)$. If $X_0 \sim \mu$ then the upper bound on the distance depends upon μV. This latter point is pertinent, for example, in checking the convergence of Markov chain Monte Carlo algorithms.

Meyn and Tweedie have shown (Meyn and Tweedie (1993, Theorem 17.0.1)) that central limit theorems hold for functions g with $g^2 \leq V$ and that laws of large numbers hold for functions g with $|g| \leq V$. Thus, the test function V bounds the collection of functions of the chain for which limit laws can easily be proven to hold.

V-uniformly ergodic chains satisfy a generalized form of α-mixing termed *V-geometric mixing* (Meyn and Tweedie (1993, Theorem 16.1.5)), where for all positive integers n

$$\sup_{\substack{k\in\mathbb{Z} \\ g^2,h^2\leq V}} \left|E_x\left[g(X_k)h(X_{n+k})\right] - E_x\left[g(X_k)\right]E_x\left[h(X_{n+k})\right]\right| \leq RV(x)\rho^n,$$

with $R < \infty$, $\rho < 1$. The test function V therefore bounds the class of functions for which the mixing occurs and helps determine the distance in the mixing. Note $V \equiv 1$ gives ordinary α-mixing. When $\pi V < \infty$, then if the chain is initialized at its stationary distribution, the mixing occurs at a uniform geometric rate

$$\sup_{\substack{k\in\mathbb{Z} \\ g^2,h^2\leq V}} \left|E_\pi\left[g(X_k)h(X_{n+k})\right] - E_\pi\left[g(X_k)\right]E_\pi\left[h(X_{n+k})\right]\right| \leq R[\pi V]\rho^n,$$

in which the test function continues to play a role in bounding the distance.

Despite the important role played by the test function V in applications, the choice of a test function V is often made as a matter of convenience, the objective being to demonstrate stability of the chain by showing V satisfies a drift condition for ergodicity such as that in Lemma 1. This paper explores the construction of new test functions from known test functions, with the purpose of improving performance in applications to bounding the rate of convergence to the stationary distribution, or bounding the class of functions for which asymptotic results hold.

In each of these applications the test function V is involved in a trade-off between bounds on convergence or mixing and the size of the functions which converge or mix. Thus, "improved performance" of a test function can mean, alternatively, limit laws or mixing applying to the largest possible collection of functions without regard to speed of convergence, or the mixing and convergence rates are the fastest possible through limiting the collection of functions to which these rates apply. Thus, we may be interested in the "largest" test function in the former case or the "smallest" in the latter case. We will call procedures that take us in these directions *boosting* and *downsizing*, respectively.

While a lower bound on downsizing always exists since $V \equiv 1$ corresponds to uniform ergodicity, a bound on boosting is not always guaranteed to exist. Two simple examples will serve to illustrate this. Consider an irreducible, aperiodic finite state chain and any function V with $1 \le V \le N$, $N < \infty$, which is bounded on the states. Then $P_x V < \infty$. The chain is uniformly (and geometrically) ergodic, thus there exist $R < \infty$ and $0 < \rho < 1$ so that

$$\sup_x \frac{|P_x^n V - \pi V|}{V} \le N \sup_x \| P_x^n - \pi \|_{\mathrm{TV}} \le RN\rho^n \le R'\rho^n,$$

implying (1) is satisfied and the chain is V-uniformly ergodic. Thus, there is no boosting limit. Now consider the Markov chain $\{X_t\}$ where the X_t are i.i.d. from a probability distribution with $E|X_t|^r = \infty$ for some $r > 0$. Clearly, (1) is meaningless for $V(x) = 1 + |x|^r$ and any $V' \ge V$, and can hold only for $V' \le V$ which would be the bound on boosting.

The paper is organized as follows. Section 2 defines *optimal* test functions. In Sections 3 and 4, boosting and downsizing of test functions in the direction of optimality are discussed. In Section 5, the paper ends with a brief discussion. Simple examples are interspersed throughout in order to illustrate the points made.

2 Optimal test functions

Define *optimal* test functions to be those which provide optimal results uniformly over $\mathcal{X}$ in the applications mentioned in Section 1. In each of these applications, the test function V was involved in a trade-off between bounds on convergence or mixing and the size of the functions which converge or mix. Thus, when discussing optimal test functions, optimality of a test function is taken to mean, alternatively, limit laws applying to the largest possible collection of functions, or the mixing and convergence implied by the test function are the fastest possible.

The choice of a test function is often made as a matter of convenience. Rather than out of a concern for optimality, the objective is to demonstrate stability of the chain by showing V satisfies a drift condition for ergodicity such as that in (1). For a V-uniformly ergodic chain, there is always a nonempty collection of test functions $\mathcal{V} := \{V : \{X_t\}$ is V-uniformly ergodic$\}$. The chosen V may be but one of many elements of this collection and not necessarily optimal in a specific application. For two functions V_1 and V_2, write $V_1 \le V_2$ if $V_1(x) \le V_2(x)$ for all x in $\mathcal{X}$. Call a test function $V_* \in \mathcal{V}$ a minimal element of $\mathcal{V}$ if V_* is such that for any $V \in \mathcal{V}$ with $V \le V_*$ it holds that $V = V_*$. A test function $V^* \in \mathcal{V}$ will be named a maximal element of $\mathcal{V}$ if V^* is such that for any $V \in \mathcal{V}$ with $V \ge V^*$ it holds that $V = V^*$.

Proposition 1. *Suppose $\{X_t\}$ is V-uniformly ergodic. Then there exists a minimal element $V_* \in \mathcal{V}$.*

Proof. By assumption $\mathcal{V}$ is not empty. Recall $V_1 \le V_2$ if $V_1(x) \le V_2(x)$ for all x in $\mathcal{X}$. The relation $\le$ is transitive and antisymmetric. Thus, $\le$ is a partial order and $(\mathcal{V}, \le)$ is a nonempty partially ordered set. The existence of a maximal linearly ordered subset is guaranteed by the Hausdorff Maximal Principle. Since $V \ge 1$, this provides a lower bound on the maximal linearly ordered subset and thus all linearly ordered subsets. By Zorn's lemma $\mathcal{V}$ has a minimal element V_*. □

Since $V_* \in \mathcal{V}$ it follows that $\{X_t\}$ is V_*-uniformly ergodic. Notice that V_* need not equal the lower bound of $V \equiv 1$, otherwise this would imply all V-uniformly ergodic chains are uniformly ergodic, which is not the case. Also, V_* is minimal in the sense of uniform bounds in x; for specific x there may be better choices of V. With these limitations, and while Proposition 1 is not a constructive proof for V_*, it does suggest there is value in exploring conditions where a known test function can be improved upon in the direction of this minimal test function.

The existence of the mimimal test function V_* follows from the weak conditions of Proposition 1 since the bound of $V \equiv 1$ exists trivially. With stronger conditions, a statement can be made about a maximal element V^*; of course, any V with $PV = \infty$ would not work as a test function.

Proposition 2. *Suppose $\{X_t\}$ is a V-uniformly ergodic Markov chain on a compact state space $\mathcal{X}$. Assume $V_\perp(x) := \inf_{V\in\mathcal{V}} V(x)$ and $V^\top(x) := \sup_{V\in\mathcal{V}} V(x)$ are measurable and $PV^\top < \infty$. Then given $\varepsilon > 0$, there exist sets $A_\perp$, $A^\top$ each with probability greater than $1 - 2\varepsilon$ so that*

(i) *there is a $V \in \mathcal{V}$ such that $0 \le V - V_\perp < \varepsilon$ on $A_\perp$,*
(ii) *there is a $V \in \mathcal{V}$ such that $0 \le V^\top - V < \varepsilon$ on $A^\top$.*

Proof. Assume $V_\perp$, $V^\top$ are not in $\mathcal{V}$, else the conclusion is trivial with $V = V_\perp$, $A_\perp = \mathcal{X}$, and $V = V_\top$, $A_\top = \mathcal{X}$, respectively. Choose $V_1 \in \mathcal{V}$, $V_2 \in \mathcal{V}$. Then by Meyn and Tweedie (1993, Theorem 16.0.1 and Lemma 15.2.8) there exist $0 < \lambda_1 < 1$, $0 < \lambda_2 < 1$, $b_1 < \infty$, $b_2 < \infty$ so that $PV_i \le \lambda_i V_i + b_i$ for $i = 1, 2$. Let $(V_1 \wedge V_2)(x) = \min[V_1(x), V_2(x)]$ and $(V_1 \vee V_2)(x) = \max[V_1(x), V_2(x)]$. Then $\mathcal{V}$ is a lattice since $P(V_1 \wedge V_2) \le (\lambda_1 \vee \lambda_2)(V_1 \wedge V_2) + (b_1 \vee b_2)$ and $P(V_1 \vee V_2) \le (\lambda_1 \vee \lambda_2)(V_1 \vee V_2) + (b_1 \vee b_2)$ implies both $V_1 \wedge V_2$ and $V_1 \vee V_2$ are in $\mathcal{V}$.

By definition of $V_\perp$, given $\varepsilon > 0$ for each $x \in \mathcal{X}$ there exists $V_x \in \mathcal{V}$ with $V_x(x) \le V_\perp(x) + \varepsilon/9$. Since $V_\perp$ and each V_x are measurable and $\mathcal{X}$ compact, there exist a continuous function $h_\perp$, for each x a continuous function h_x, and measurable sets A, and for each x a measurable set B_x, all being of probability greater than $1 - \varepsilon/3$, with $|V_\perp I_A - h_\perp I_A| < \varepsilon/3$ and $|V_x I_{B_x} - h_x I_{B_x}| < \varepsilon/3$.

Since each h_x and $h_\perp$ are continuous, for each $x \in \mathcal{X}$ there exists an open set O_x with $|h_x(x) - h_x(y)| < \varepsilon/9$ and $|h_\perp(x) - h_\perp(y)| < \varepsilon/9$ for all $y \in O_x$. Thus $|h_x(y) - h_\perp(y)| < \varepsilon/3$ for all $y \in O_x$. Since $\mathcal{X}$ is compact a finite collection $O_{x_1}, \ldots, O_{x_n}$ of these cover $\mathcal{X}$. Define $V := V_{x_1} \wedge \cdots \wedge V_{x_n}$ and $B_i := B_{x_i}$, for

$i = 1, \ldots, n$. Then $V \in \mathcal{V}$ since $\mathcal{V}$ is a lattice, and for any $x \in [\bigcup_{i=1}^n B_i] \cap A$ there is a x_i so that $x \in O_{x_i}$, implying

$$\begin{aligned} 0 &\le V - V_\perp \\ &\le [V_{x_i} - V_{x_i} I_{B_i}] + [V_{x_i} I_{B_i} - h_{x_i} I_{B_i}] + [h_{x_i} I_{B_i} - h_{x_i}] \\ &\quad + [h_{x_i} - h_\perp] + [h_\perp - h_\perp I_A] + [h_\perp I_A - V_\perp I_A] + [V_\perp I_A - V_\perp] < \varepsilon. \end{aligned}$$

Since $([\bigcup_{i=1}^n B_i] \cap A) \supseteq (B_1 \cap A)$ and the measure of $B_1 \cap A$ is greater than $1 - 2\varepsilon$, the conclusion (i) follows with $A_\perp = [\bigcup_{i=1}^n B_i] \cap A$. The conclusion (ii) follows by the dual argument applied to the collection $-\mathcal{V}$. □

Notice that $V_\perp$, $V_\top$ are not guaranteed to be in $\mathcal{V}$ while V_*, V^* are by definition. Thus, when $\mathcal{X}$ is compact and $\mathcal{V}$ consists of appropriately smooth functions, we can come arbitrarily close to $V_\perp$ and thus V_*, and likewise $V_\top$ and V^*, except on a set of arbitrarily small, though still positive, measure.

Through analogy with continuous functions, one might expect $V_\perp = V_* \equiv 1$ and to attain the lower bound uniformly over all of compact $\mathcal{X}$ under suitable continuity assumptions on the chain. The simplest cases occur where P has certain continuity properties and the state space is a "manageable size" relative to P.

Definition 2. A process $\{X_t\}$ is said to be a *T-chain* if there exists a substochastic transition kernel T and a probability distribution $\{a(n)\}$ on the nonnegative integers satisfying $\sum_{n \ge 0} a(n) P^n(x, \cdot) \ge T(x, \cdot)$ for $x \in \mathcal{X}$, where $T(x, \cdot) > 0$ for all x and $T(\cdot, A)$ is a lower semicontinuous function for any $A \in \mathcal{B}(\mathcal{X})$.

The importance of petite sets in establishing V-uniform ergodicity of a Markov chain is given in Lemma 1, for instance, and is discussed thoroughly in Meyn and Tweedie (1993). The continuity properties of T-chains ensure that petite sets contain "reasonable" sets such as compact sets, sublevel sets of V, sets mapped in a finite time to a compact set or sublevel set of V, and sets mapped in a finite time to a compact set or sublevel set of V with a probability arbitrarily close to one.

A common minorisation condition for uniform ergodicity is due to Doeblin, which can be stated as there exists an integer n, $\varepsilon > 0$, and a probability measure μ on $\mathcal{X}$ so that $P^n(x, A) \ge \varepsilon \mu(A)$, for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$ (see for example, pg. 397ff. in Meyn and Tweedie (1993)). This is equivalent to petiteness of $\mathcal{X}$ for ψ-irreducible, aperiodic T-chains (Meyn and Tweedie (1993, Theorem 5.5.7)). Thus, for these chains the downsizing limit $V \equiv 1$ is attained.

Example 2. Roberts and Rosenthal (2004) analyze the behavior of the Metropolis-Hastings algorithm on $\mathbb{R}$. Supposing the stationary distribution π has density π_u which is finite everywhere and the proposal density $q(x, y)$ is positive and continuous, they show that the chain is π-irreducible and aperiodic. Further, if π_u is continuous and $\mathcal{X}$ is a compact subset of $\mathbb{R}$ then these assumptions imply the existence of $\varepsilon > 0$ with $q(x, \cdot) \geq \varepsilon$ and $0 < k < \infty$ with $q(y, x)/[\pi_u(x)q(x, y)] \geq 1/k$ so

$$P(x, A) = \int_A P(x, dy) \geq \int_A \varepsilon \min\left\{1, \frac{\pi_u(dy)}{k}\right\} > 0,$$

for any set A in $\mathcal{B}(\mathcal{X})$, satisfying the minorisation condition for uniform ergodicity. In fact, the chain is a π-irreducible and aperiodic T-chain in a strong sense since $P(\cdot, A)$ has a constant and therefore uniformly continuous component. The state-space $\mathcal{X}$ is petite; thus, the lower bound $V \equiv 1$ is attained and the chain is uniformly ergodic. On the other hand, if f is such that $Q_x f := \int_{\mathcal{X}} f(y)q(x, y)\, dy = \infty$ then $P_x f = \infty$ since we can bound $\pi_u(y)/[\pi_u(x)q(x, y)] \geq \delta > 0$ and so $P_x f \geq \delta Q_x f$, and there is a maximal element. For example, if $\mathcal{X} = \mathbb{R}$ and $q(x, y)$ is the Cauchy density centered at x, then $Q_x f = \infty$ for $f(x) = x$.

3 Boosting test functions

For chains which are V-uniformly ergodic, to demonstrate limit laws or mixing for a larger collection of functions it is necessary to prove V-uniform ergodicity using a test function V nearer to V^*. One way to accomplish this is through the technique of *exponential boosting* (cf. Borovkov and Hordijk (2004), Cline and Pu (2001), Meyn and Tweedie (1993)) where, under appropriate conditions, the test function V is boosted to an exponential $V' := e^{V^s}$ or $V' := e^{sV}$, $s > 0$. As an example.

Proposition 3. *Assume $\{X_t\}$ satisfies the assumptions of Lemma 1 for a function V. If there exists $q > 0$ so that $\sup_x P_x^k e^{qV}/e^{qV} < \infty$ for some $k \geq 1$ and $\sup_{x\in A} P_x^k e^{qV} < \infty$ for the petite set A of Lemma 1, then there exists $0 < s < q$ so that $\{X_t\}$ is V'-uniformly ergodic with $V'(x) = e^{sV(x)}$.*

Proof. Note from the assumptions that there exists $\beta > 0$ with

$$\sup_{x\in A^C} P_x^k V - V \leq -\beta.$$

Since $(y^s - 1)/s \to \ln(y)$ as $s \to 0$, then given $\varepsilon > 0$ it holds that $e^{sV(X_k)}/e^{sV(x)} \leq 1 + (1+\varepsilon)s[V(X_k) - V(x)]$ for small s, in particular $s < q$. Taking expectations on both sides yields

$$\sup_{x\in A^C} \frac{P_x^k V'}{V'} \leq 1 + (1+\varepsilon)s(-\beta) < 1.$$

By assumption V' is bounded on A and the conclusion follows from Lemma 1. ☐

The drift conditions for V-uniform ergodicity in Lemma 1 and Proposition 3 reflect the fact that it can be more profitable to work with the k-step chain $\{X_{tk}\}$ rather than the one-step chain $\{X_t\}$. It would be helpful to combine the advantages of the continuity properties of T-chains with the advantages of working with the k-step chain. This motivates results for cases where it is known the one-step chain $\{X_t\}$ is a T-chain, yet it is easier to analyze stability of the k-step chain $\{X_{tk}\}$, or vice versa. The following result shows that "T-chain-ness" of the two are almost equivalent. The difference is that more is required of $\{X_t\}$ to imply $\{X_{tk}\}$ is a T-chain; it is required that $\{X_t\}$ be weak Feller, which is stronger than a T-chain, requiring that the chain maps bounded continuous functions to bounded continuous functions and that the support of the irreducibility measure has a non-empty interior, in addition to being irreducible and aperiodic.

Proposition 4. *Consider a Markov chain $\{X_t\}$.*

(i) *If $\{X_{tk}\}$ is a T-chain for some integer k, then $\{X_t\}$ is a T-chain.*
(ii) *If $\{X_t\}$ is weak Feller, aperiodic, and ψ-irreducible for some measure ψ whose support has a non-empty interior, then $\{X_{tk}\}$ is a ψ-irreducible, aperiodic T-chain for all integers $k \geq 1$.*

Proof. To prove (i), since $\{X_{tk}\}$ is a T-chain, there exist a probability distribution $\{a(n)\}$ and a continuous component T with $\sum_{n\geq 0} P^{nk}(x, A)a(n) \geq T(x, A)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$. Define $a'(j) = a(n)$, $j = nk$, $n = 1, 2, 3, \ldots$, and $a'(j) = 0$, else. Then $\sum_{n\geq 0} P^{n}(x, A)a'(n) \geq T(x, A)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$, implying that $\{X_t\}$ is a T-chain.

As for (ii), pick an integer k. Since P maps bounded continuous functions to bounded continuous functions, then by induction so does P^k, implying that $\{X_{tk}\}$ is a weak Feller chain. Since $\{X_t\}$ is ψ-irreducible and aperiodic, so is $\{X_{tk}\}$, and since the support of ψ has non-empty interior, we have by Meyn and Tweedie (1993, Theorem 6.2.9) that $\{X_{tk}\}$ is a T-chain. □

Example 1 (cont.). Suppose f is a bounded continuous function and the ε_t have probability density λ, then

$$P_x f = \int f(y)\lambda\big(y - [\phi_1 x I(x < 0) + \phi_2 x I(x \geq 0)]\big)\, dy.$$

Since $\phi_1 x I(x < 0) + \phi_2 x I(x \geq 0)$ and λ are continuous functions, it follows from this and the assumptions that the integrand is continuous and integrable. By dominated convergence then, $P_x f$ is continuous. Clearly $P_x f$ is bounded, implying $\{X_t\}$ is weak Feller and it follows from Proposition 4(ii) that $\{X_{tk}\}$ is a T-chain.

Example 2 (cont.). It was shown that $\{X_t\}$ is a π-irreducible and aperiodic T-chain. The probability of a movement from x to y is

$$\alpha(x, y) = \min\big(1, \pi_u(y)q(y, x)/\pi_u(x)q(x, y)\big).$$

Also, for a bounded continuous function f,

$$P_x f = \int_{\mathcal{X}} f(y)\big[q(x,y)\alpha(x,y)I(x \neq y) + q(x,y)[1-\alpha(x,y)]I(x=y)\big]\,dy,$$

since $q(x,\cdot)$ is continuous. Clearly $P_x f$ is bounded. The integrand is continuous in y apart from $y = x$. Dominated convergence allows us to interchange a limit operation and integration; thus, $P_x f$ is continuous and $\{X_t\}$ is weak Feller. Since $\{X_t\}$ is π-irreducible and aperiodic (recall the irreducibility measure was Lebesgue measure), then $\{X_{tk}\}$ is a T-chain by Proposition 4(ii).

The following is a consequence of results due to Cline and Pu (2001, Theorems 3 and 4), combined with Proposition 4. The key is to find a bound $W(x)$ on $V(X_k)$ that satisfies a uniform integrability condition with respect to $V(x)$, where $\{X_{tk}\}$ is a T-chain. With this in place, exponential boosting is possible. A similar result can be found in Theorem 3 of Borovkov and Hordijk (2004).

Proposition 5. *Assume $\{X_t\}$ is weak Feller, aperiodic and ψ-irreducible for some measure ψ whose support has a non-empty interior. Assume $V \geq 1$ is norm-like. Assume there exists a random variable $W(x)$ and $k < \infty$ so that $V(X_k) \leq W(x)$ whenever $X_0 = x$ and there exists $r > 0$ so that $e^{r[V(X_1)-V(x)]}$ is uniformly integrable.*

(i) *If $\{|W(x) - V(x)| + e^{r[W(x)-V(x)]}\}$ is uniformly integrable and*

$$\limsup_{\|x\|\to\infty} E\big[W(x) - V(x)\big] < 0,$$

then there exist $s > 0$ and $V'(x) = e^{sV(x)}$ so that $\{X_t\}$ is V-uniformly ergodic.

(ii) *If $\{|\log[W(x)/V(x)]| + e^{[W(x)]^r - [V(x)]^r}\}$ is uniformly integrable and*

$$\limsup_{\|x\|\to\infty} E\big(\log[W(x)/V(x)]\big) < 0,$$

then there exist $s > 0$ and $V'(x) = e^{[V(x)]^s}$ so that $\{X_t\}$ is V-uniformly ergodic.

Proof. That $\{X_{tk}\}$ is a ψ-irreducible, aperiodic T-chain follows from Proposition 4. The assumptions in (i) imply $\{X_{tk}\}$ satisfies Theorem 3 in Cline and Pu (2001). The assumptions in (ii) imply $\{X_{tk}\}$ satisfies Theorem 4 in Cline and Pu (2001). The conclusions follow from Lemma 1. □

Example 1 (cont.). If the ε_t also have $Ee^{q|\varepsilon_t|} < \infty$ for some $q > 0$, then the assumptions of Proposition 5(i) are satisfied for $k = 2$, $V(x) = 1 + |x|$ and

$$W(x) = 1 + \max(\phi_1^2, \phi_2^2, \phi_1\phi_2)|x| + |\varepsilon_1| + |\varepsilon_2|.$$

The TAR(1) process is V'-uniformly ergodic with $V'(x) = e^{s[1+|x|]}$ for some $0 < s < q$. In fact, $V''(x) = e^{s|x|}$ would be adequate.

If the ε_t have $Ee^{|\varepsilon_t|^q} < \infty$ for some $q > 0$, then the assumptions of Proposition 5(ii) are satisfied for the same k and $W(x)$. The TAR(1) process is V'-uniformly ergodic with $V'(x) = e^{[1+|x|]^s}$, for some $0 < s < q$.

Remark. The condition $\max(\phi_1, \phi_2, \phi_1\phi_2) < 1$ in Example 1 implies the skeleton

$$x_t := \phi_1 x_{t-1} I(x_{t-1} < 0) + \phi_2 x_{t-1} I(x_{t-1} \geq 0)$$

of $\{X_{tk}\}$ converges to zero at an exponential rate. Combining this with the exponential condition $Ee^{|\varepsilon_t|^q} < \infty$, and considering known results on the equivalence of the stability of $\{X_t\}$ and $\{X_{tk}\}$, it is not surprising that it is possible to exponentially boost the test function V.

The uniform integrability conditions used in Proposition 5 are implied in Example 1 by the exponential stability conditions on the skeleton and exponential expectation conditions on the errors. When using this "skeleton + noise" approach, exponential stability conditions on both the skeleton and the error distribution are often sufficient, though not always necessary, in order for boosting to be possible.

Example 1 (cont.). To see the lack of necessity, consider the following. Cline and Pu (2001, Example 12) analyze a TAR(1) with additive constants,

$$X_t = (\alpha_1 + \phi_1 X_{t-1}) I(X_{t-1} < 0) + (\alpha_2 + \phi_2 X_{t-1}) I(X_{t-1} \geq 0) + \varepsilon_t,$$

introduced in Chan et al. (1985), where $\phi_1\phi_2 = 1$, $\phi_1 < 0$. Chan et al. proved $\{X_t\}$ is ergodic if and only if $\phi_2\alpha_1 + \alpha_2 > 0$. Cline and Pu showed that if $\phi_2\alpha_1 + \alpha_2 > 0$ and $Ee^{q|\varepsilon_t|} < \infty$ then exponential boosting is possible, though the skeleton is only additively and not exponentially stable.

If $\max(\phi_1, \phi_2, \phi_1\phi_2) > 1$ then the skeleton is not even additively stable, $\{X_t\}$ is transient and boosting is not possible, so that some stability of the skeleton is required. If $\max(\phi_1, \phi_2, \phi_1\phi_2) < 1$, the skeleton is exponentially stable but if the ε_t follow a heavy-tailed distribution such as a t-distribution then the errors are not exponentially stable and exponential boosting is not possible. For this process, exponential stability of the errors is required while exponential stability of the skeleton is not.

4 Downsizing test functions

To gain optimal results concerning convergence rates of probabilities or mixing rates, it is necessary to prove V-uniform ergodicity using a "smaller" test function V as near as possible to V_*. This provides the impetus for the investigation into conditions under which a test function V of a V-uniformly ergodic Markov chain can be *downsized*, in particular *logarithmically* downsized.

Meyn and Tweedie (1993, Proposition 15.2.9) consider $V' = \sqrt{V}$ and show a V-uniformly ergodic Markov chain is also V'-uniformly ergodic. By induction this implies a V-uniformly ergodic chain is also V'-uniformly ergodic for $V' = V^{1/2k}$, for integers $k \geq 1$. This can be generalized further. A different perspective considers functions $V' \geq 1$ increasing as V does, which are dominated by V, and whose drift is dominated by the drift of V. Lastly, we could take $V^{1/n}$ to its limit and consider test functions V' which are logarithmic functions of V. This also provides a nice symmetry with exponential boosting.

Proposition 6. *Suppose $\{X_t\}$ and V satisfy the assumptions of Lemma 1 for an integer k.*

(i) *For a positive integer n let $V' = V^{1/n}$. Then $\{X_t\}$ is V'-uniformly ergodic.*
(ii) *For any other $V' \geq 1$ with $V' \to \infty$ as $V \to \infty$, $V' < V$ and $PV' - V' \leq PV - V$, then $\{X_t\}$ is V'-uniformly ergodic.*
(iii) *Assume $\sup_x P\ln(V)/\ln(V) < \infty$. Let $V' = 1 + s\ln(V)$ for $s > 0$. Then $\{X_t\}$ is V'-uniformly ergodic.*

Proof. To prove (i), note that (Meyn and Tweedie (1993, Lemma 15.2.8)) implies $\{X_t\}$ is V-uniformly ergodic if and only if $V \geq 1$ is unbounded off petite sets and $PV \leq \lambda V + L$ for some $0 < \lambda < 1$ and $L < \infty$. For a fixed integer n and a real number L, there exists $L_* < \infty$ with $L \leq \sum_{k=1}^{n-1} \binom{n}{k} \lambda^{k/n} L_*^{n-k}$. This implies that $(\lambda V + L)^{1/n} \leq \lambda^{1/n} V^{1/n} + L_*$. Also, V' is a strictly concave function of V and so by Jensen's Inequality, $PV' < \lambda^{1/n} V' + L_*$. Clearly $V' \geq 1$ is unbounded off petite sets. Thus, we have $\{X_t\}$ is V'-uniformly ergodic. As for (ii), the assumptions imply there exists $0 < \lambda < 1$ with $\lambda(V - V') + P(V' - V) < 0$, from which it follows that $PV' < \lambda V' + L$. It also follows from the assumptions that V' is unbounded off petite sets; thus, $\{X_t\}$ is V'-uniformly ergodic. Finally, (iii) assumes that there exists a petite set A with $\sup_{x \in A^C} P^k V/V \leq 1 - \beta$, for some $0 < \beta < 1$. Then $\sup_{x \in A^C} P^k V'/V' < 1$ follows from $\sup_{x \in A^C} P^k \ln(V) - \ln(V) \leq \ln(\sup_{x \in A^C} P^k V/V) < 0$. Also, since $\sup_{x \in A} P^k V < \infty$ and $V \geq 1$, this implies $\sup_{x \in A} P^k V' < \infty$. Finally, $\sup_x P\ln(V)/\ln(V) < \infty$ implies $\sup_x PV'/V' < \infty$ when $k > 1$ and $\{X_t\}$ is V'-uniformly ergodic by Lemma 1. □

As an application and final example, consider the following.

Example 3. Mengersen and Tweedie (1996) consider the Metropolis algorithm with geometric target distribution π for a parameter p

$$\pi(j) = (1-p)p^j, \qquad j = 0, 1, 2, \ldots,$$

and symmetric candidate distribution

$$q(i, i-1) = q(i, i+1) = 1/2, \qquad i > 0; \qquad q(0,1) = q(0,0) = 1/2.$$

They use their Theorem 4.1(ii) to find the drift constant $\lambda = 1 - \beta$ in Proposition 1 for the test function $V(j) = \alpha^j$ and with $A := \{0\}$, yielding

$$\lambda = \min_{\alpha > 0} \lambda_\alpha = \sqrt{p} + (1-p)/2, \qquad \text{at } \alpha = \frac{1}{\sqrt{p}} > 1.$$

Since A is an atom, and therefore small and petite, applying Proposition 6(iii) here yields that the Markov chain created by the algorithm is V'-uniformly ergodic with $V'(j) = 1 + s\log(V(j)) = 1 + s\log(\alpha^j) = 1 - sj\log(p)/2$, a linear rather than an exponential test function. This can be verified by direct calculation of the drift for this test function

$$\lambda = \sum_{j\geq 0} p(i,j)V'(j)/V'(i) = 1 + (1-p)\frac{s\log(p)/4}{1 - si\log(p)/2} < 1,$$

since $s > 0$, $i \geq 0$, $0 < p < 1$. Now, while logarithmic-downsizing as in Proposition 6(iii) gives the best *qualitative* bound, it may not be the most useful in calculating *quantitative* bounds, as is the case in this example. For the test function $V(j) = \alpha^j$, the converse results in Mengersen and Tweedie (1996) show that $\lambda = \sqrt{p} + (1-p)/2$ is the exact rate of convergence. To improve upon this result, a different test function can be chosen. Using Proposition 6(i), then the algorithm is V''-uniformly ergodic with $V''(j) = [V(j)]^{1/n} = \alpha^{j/n}$. This yields a drift $\lambda_n = \sqrt{p/n} + (1-p)/2$ which converges to $(1-p)/2$ as $n \to \infty$; the drift of the downsized test function beating the drift rate of the original test function $V(j) = \alpha^j$ by $\sqrt{p}$.

Also note that Proposition 6(iii) implies the chain is likewise V'''-uniformly ergodic with $V'''(x) = 1 + s_2\ln[1 + s_1\ln(1 + |x|)]$. Thus by iteration we can approach $V \equiv 1$ as closely as we like on any compact set, but without actually reaching it unless there are additional continuity assumptions on the chain. The barrier separating uniformly ergodic and non-uniformly ergodic Markov chains cannot be breached.

5 Discussion

This paper concentrates on conditions whereby the test function V can be altered, but the geometric rate of convergence/mixing is retained. It is not always the case that V can be altered and the rate of convergence retained. Meyn and Tweedie (1993) give an example of a uniformly ergodic Markov chain which converges at a geometric rate in the total variation norm and a simple unbounded function f of this chain which converges but loses the geometric rate of convergence. Thus, when boosting is not possible limit laws may still be demonstrated for a larger class of functions in exchange for sacrifice on the rate of convergence. Conversely, if ergodicity were originally demonstrated using this function f as a test function, by downsizing to $f' \equiv 1$, the geometric rate of convergence would be gained. It is also not always possible to either boost or downsize a test function. Proposition 15 in Roberts and Rosenthal (2004) has an example where the test function V cannot be changed.

The emphasis in this paper has been on qualitative rates of convergence. Roberts and Rosenthal (2004) discuss computable rates of convergence, and show the constant bounding the computed rate of convergence is related to the test function V.

Rosenthal also (Rosenthal (2003, Proposition 3)) derives a computable exponential boosting result for uniformly ergodic reversible Markov chains.

References

Borovkov, A. A. and Hordijk, A. (2004). Characterization and sufficient conditions for normed ergodicity of Markov chains. *Adv. in Appl. Probab.* **36**, 227–242. MR2035781

Boucher, T. R. and Cline, D. B. H. (2007). Stability of cyclic threshold autoregressive time series models. *Statist. Sinica* **17**, 43–62. MR2352503

Chan, K. S., Petruccelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22**, 267–279. MR0789351

Cline, D. B. H. and Pu, H. (2001). Stability of nonlinear time series: What does noise have to do with it? In *Selected Proceedings of the Symposium on Inference for Stochastic Processes, Vol. 37*, 151–170. MR2002508

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101–121. MR1389882

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. London: Springer. MR1287609

Petruccelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *J. Appl. Probab.* **21**, 270–286. MR0741130

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1**, 20–71. MR2095565

Rosenthal, J. S. (2003). Asymptotic variance and convergence rates of nearly-periodic MCMC algorithms. *J. Amer. Statist. Assoc.* **98**, 169–177. MR1965683

Department of Mathematics
Texas A&M University-Commerce
Commerce, Texas 75428
USA
E-mail: thomas.boucher@tamuc.edu