Computable bounds for geometric convergence rates of Markov chains *

Sean P. Meyn[†] R.L. Tweedie[‡]

September 27, 1995

Abstract

Recent results for geometrically ergodic Markov chains show that there exist constants $R < \infty, \rho < 1$ such that

$$\sup_{|f| \leq V} \left| \int P^n(x, dy) f(y) - \int \pi(dy) f(y) \right| \leq RV(x)\rho^n$$

where π is the invariant probability measure and V is any solution of the drift inequalities

$$\int P(x, dy)V(y) \leq \lambda V(x) + b\mathbb{1}_C(x)$$

which are known to guarantee geometric convergence for $\lambda < 1, b < \infty$ and a suitable small set C.

In this paper we identify for the first time computable bounds on R and ρ in terms of λ, b and the minorizing constants which guarantee the smallness of C. In the simplest case where C is an atom $\boldsymbol{\alpha}$ with $P(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geq \delta$ we can choose any $\rho > \vartheta$ where

$$[1 - \vartheta]^{-1} = \frac{1}{(1-\lambda)^2} \Big[1 - \lambda + b + b^2 + \zeta_\alpha (b(1-\lambda) + b^2) \Big]$$

and

$$\zeta_\alpha \leq \Big(\frac{34 - 8\delta^2}{\delta^3} \Big) \Big(\frac{b}{1-\lambda} \Big)^2,$$

and we can then choose $R \leq \rho/[\rho - \vartheta]$. The bounds for general small sets C are similar but more complex. We apply these to simple queueing models and Markov chain Monte Carlo algorithms, although in the latter the bounds are clearly too large for practical application in the case considered.

KEYWORDS: Foster's criterion, irreducible Markov chains, Lyapunov functions, ergodicity, geometric ergodicity, queues, MCMC, Hastings and Metropolis algorithms

RUNNING HEAD: Bounds for Convergence of Markov Chains

---

*Work supported in part by NSF initiation grant # ECS 8910088 and NSF grant DMS-9205687

[†]Postal address: University of Illinois, Coordinated Science Laboratory, 1101 W. Springfield Ave., Urbana, IL 61801, U.S.A.

[‡]Postal Address: Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

1 Bounds for Geometrically Ergodic Chains

Let $\Phi = \{\Phi_n : n \in \mathbb{Z}_+\}$ be a discrete–time Markov chain on a general state space X, endowed with a countably generated σ–field $\mathcal{B}(\mathsf{X})$. Our notation will in general follow that of [16]. We denote by $P^n(x,A),\ n \in \mathbb{Z}_+,\ x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X})$ the (stationary) transition probabilities of the chain, i.e.

$$P^n(x,A) = \mathsf{P}_x\{\Phi_n \in A\}$$

where P_x and E_x denote respectively the probability law and expectation of the chain under the initial condition $\Phi_0 = x$. When the initial distribution is a general probability λ on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ the corresponding quantities are denoted by $\mathsf{P}_\lambda, \mathsf{E}_\lambda$. For any non-negative function f, we write Pf and $P^n f$ respectively for the functions $\int P(x,dy)f(y)$ and $\int P^n(x,dy)f(y)$, and for any (signed) measure μ we write $\mu(f)$ for $\int \mu(dy)f(y)$.

We assume throughout that the chain is ψ-irreducible and aperiodic [16, Chapters 4, 5], and we write $\mathcal{B}^+(\mathsf{X}) = \{A \in \mathcal{B}(\mathsf{X}) : \psi(A) > 0\}$.

In this paper we consider chains which are *geometrically ergodic*: that is, when there exists a unique invariant probability measure π and constants $R_x < \infty$, $\rho < 1$ such that

$$\|P^n(x,\,\cdot\,) - \pi\| \le R_x \rho^n \tag{1}$$

for all $x \in \mathsf{X}, n \in \mathbb{Z}_+$, where for a signed measure μ, we use $\|\mu\| := \sup_{|f|\le 1} |\mu(f)|$ to denote the total variation norm.

Our specific goal is to find computable bounds on the constants R_x and ρ.

Recall from [16, Chapter 5] that for a ψ-irreducible chain every set $A \in \mathcal{B}^+(\mathsf{X})$ contains a *small set* C: that is, a set for which there exists some probability measure ν, some $\delta > 0$ and some m with

$$P^m(x,A) \ge \delta\nu(A), \qquad x \in C,\ A \in \mathcal{B}(\mathsf{X}). \tag{2}$$

Our expressions bounding R_x and ρ will be in terms of the quantities δ and m in (2) together with constants $\lambda < 1, b < \infty$ and a function $V \ge 1$ satisfying the "drift inequality"

$$PV \le \lambda V + b\mathbb{1}_C \tag{3}$$

where C is small. That the existence of solutions to this inequality is equivalent to geometric ergodicity was first shown by Popov [21] in the countable space case and in [19, 35] in the general case, and verifying geometric ergodicity is most often done through such bounds (see Chapters 15 and 16 of [16], or examples such as [4, 5, 3]).

To set up our results we need the concept of a V–norm for the kernels $(P^n - \pi)$.

For a positive function $V \ge 1$, first generalise the idea of total variation by defining the *V-total variation* [16, Chapter 14] of a signed measure μ on $\mathcal{B}(\mathsf{X})$ as

$$\|\mu\|_V := \sup_{|f|\le V} |\mu(f)|.$$

Conditions for the convergence of $\|P^n(x,\,\cdot\,) - \pi\|_V$ for appropriate V are given in [16, Chapter 14]: essentially all that is required to move to this convergence from total variation convergence is that $\pi(V) < \infty$.

Next, for any two kernels P_1 and P_2 on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ define the *V-norm* as

$$|\!|\!| P_1 - P_2 |\!|\!|_V := \sup_{x\in\mathsf{X}} \frac{\|P_1(x,\,\cdot\,) - P_2(x,\,\cdot\,)\|_V}{V(x)}. \tag{4}$$

In studying geometric ergodicity, we will consider the distance $|\!|\!| P^k - \Pi |\!|\!|_V$, where Π is the invariant kernel

$$\Pi(x, A) := \pi(A), \qquad A \in \mathcal{B}(\mathsf{X}),\ x \in \mathsf{X}.$$

The key concept is that of V-uniform ergodicity, introduced in [31, 10] for countable spaces and [15] for general chains. A Markov chain $\boldsymbol{\Phi}$ is called *V-uniformly ergodic* if there exists an invariant probability measure π such that

$$|\!|\!| P^n - \Pi |\!|\!|_V \to 0, \qquad n \to \infty. \tag{5}$$

We then have [10, 15] that V-uniform ergodicity is, for the correct class of functions V, actually equivalent to geometric ergodicity as studied in [20, 18]. Indeed we have from Theorem 16.0.1 of [16]

Theorem 1.1 *For a ψ-irreducible aperiodic chain $\boldsymbol{\Phi}$, the following are equivalent for any $V \geq 1$:*

(i) *$\boldsymbol{\Phi}$ is V-uniformly ergodic.*

(ii) *There exists $\rho < 1$ and $R < \infty$ and an invariant probability kernel Π such that for all $n \in \mathbb{Z}_+$*

$$|\!|\!| P^n - \Pi |\!|\!|_V \leq R\rho^n. \tag{6}$$

(iii) *For some one small set C (and then every small set in $\mathcal{B}^+(\mathsf{X})$) there exists a function $V_C \geq 1$ and constants $\lambda_C < 1, b_C < \infty$ such that the drift condition*

$$PV_C \leq \lambda_C V_C + b_C \mathbb{1}_C \tag{7}$$

holds, and V_C is equivalent to V in the sense that for some $c \geq 1$,

$$c^{-1} V \leq V_C \leq cV. \tag{8}$$

This allows us to place V-uniform ergodicity in an operator theoretic context: for let L_V^∞ denote the vector space of all functions $f\colon \mathsf{X} \to \mathbb{C}$ satisfying

$$|f|_V := \sup_{x\in\mathsf{X}} \frac{|f(x)|}{V(x)} < \infty.$$

If P_1 and P_2 are two transition kernels and if $|\!|\!| P_1 - P_2 |\!|\!|_V$ is finite, then $P_1 - P_2$ is a bounded operator from L_V^∞ to itself, and $|\!|\!| P_1 - P_2 |\!|\!|_V$ is its operator norm.

It then follows from a standard operator norm convergence approach [16], and the fact that $P^n - \Pi = (P - \Pi)^n$, that the convergence must be geometrically fast so that (i) must imply (ii) in Theorem 1.1.

Identifying upper bounds on the rate of convergence ρ and the constant R in (6) for chains on finite state spaces has been a well-studied problem for many years [28], and in special cases the optimal rates of convergence can be found explicitly: for a recent example see [11].

On countable spaces, this problem was studied soon after the original work on geometric ergodicity [36], although no generally applicable results seem available. On these and on more general spaces such as $\mathbb{R}^n$, the problem has attracted considerable attention recently, largely as a result of the use of Markov chain Monte Carlo (MCMC) techniques. In this context one key to stopping rules for simulations is knowledge of the accuracy of n-step approximations, and this motivates the calculation of computable bounds in (6).

However, existing results have been sparse. For example, even for specific situations on continuous spaces in the MCMC area, only the simple facts of convergence in total variation, or the existence of some unknown rate $\rho < 1$ for which convergence is geometrically fast, have been asserted by most authors (see e.g. [23, 8, 33, 3]). Some special cases have been studied through spectral theory of operators [27], but again bounds on the actual value of the "second eigenvalue" have not proved easy to establish in any closed form. In our concluding section we give one such application, although regrettably the actual bounds found are too large for practical purposes as yet.

To our knowledge, the results here are the first explicit general bounds on the rate of convergence, even for countable space chains, although there are several recent approaches which give related results.

Firstly, under the assumption that the whole space is small, so that there is an m, a ν and a δ such that for *all* x

$$P^m(x, \cdot\,) \geq \delta\nu(\,\cdot\,),$$

then (see Theorem 16.0.2 of [16]) there is an elegant coupling argument to show that we have

$$\|P^n(x, \cdot\,) - \pi\| \leq (1-\delta)^{n/m}, \qquad x \in \mathsf{X} : \tag{9}$$

that is, we can choose $R = 1$ and $\rho = (1-\delta)^{1/m}$. This is exploited in, for example, [24, 33, 7, 13]. However, the requirement that X be small is extremely restrictive and is not satisfied for most models of a truly "infinite" nature.

Secondly, and closely related in spirit to our results, in [12] it has been shown since we developed this paper that if one has stochastic monotocicity properties on the chain, then it is possible to get very much tighter bounds which are in very many circumstances exact, using only the same minimal information which we use here. In fact the rate of convergence can be bounded by λ where this is the contraction factor in (3), provided C is a single point (an "atom") at the "bottom" of such a stochastically monotone chain. This can be shown to be best possible under weak (and natural) extra conditions.

Thirdly, in [25] Rosenthal has recently extended the method of argument used in proving (9) for uniformly ergodic chains to find bounds on chains such as those we consider here. His method assumes slightly greater structure than does ours, but the corresponding bounds may well be considerably tighter in many cases as a consequence; no systematic comparison has as yet been undertaken, although the links between the two methods are now understood [26].

Finally, Spieksma [32] and Baxendale [1] have both considered the structure of Kendall's Theorem (Theorem 15.1.1 of [17]) which lies at the heart of the analytic approach to convergence rates. As in [12], Spieksma shows that for a special class of models on countable spaces (which include many single-server queueing models) the rate of convergence can be bounded by λ in (3) when C is a single point: this relies crucially on her Assumption A, which is unfortunately not always satisfied. Baxendale's approach makes no such assumption, and as in this paper, his results give general computable bounds which are complex in nature but could be compared with ours in specific circumstances.

It is clear that, other than the results in [12] and [32], both of which require special structure, none of these bounds are likely to be tight, but the methods of proof indicate areas in which more explicit knowledge of specific models may be used to sharpen the results.

In the next section we give the main results, and the proofs and related results then follow. We consider specific numerical values which the bounds produce in a queueing and a uniformly ergodic context, and conclude with a more detailed application, to a specific MCMC context, which is related to rather more detailed studies in [13]. This illustrates the approach needed to apply these methods in practice.

2 Computable Bounds

We first consider the case where (3) holds for a small set $\boldsymbol{\alpha}$ which is an "atom": that is, $P(x,\,\cdot\,) \equiv P(\boldsymbol{\alpha},\,\cdot\,)$ for all $x \in \boldsymbol{\alpha}$. More general cases will later be reduced to this using the Nummelin splitting technique [16, Chapter 5].

Our central result is

Theorem 2.1 *Suppose that for some atom $\boldsymbol{\alpha} \in \mathcal{B}(\mathsf{X})$ we have $\lambda < 1, b < \infty$ and a function $V \geq 1$ such that*

$$PV \leq \lambda V + b\mathbb{1}_{\alpha}. \tag{10}$$

Let $\vartheta = 1 - M_\alpha^{-1}$, where

$$M_\alpha = \frac{1}{(1-\lambda)^2}\Big[1 - \lambda + b + b^2 + \zeta_\alpha(b(1-\lambda) + b^2)\Big] \tag{11}$$

and

$$\zeta_\alpha = \sup_{|z|\leq 1} \Big|\sum_{n=0}^{\infty}[P^n(\boldsymbol{\alpha},\boldsymbol{\alpha}) - P^{n-1}(\boldsymbol{\alpha},\boldsymbol{\alpha})]z^n\Big|. \tag{12}$$

Then $\boldsymbol{\Phi}$ is V-uniformly ergodic, and for any $\rho > \vartheta$,

$$\|P^n - \Pi\|_V \leq \frac{\rho}{\rho - \vartheta}\rho^n, \qquad n \in \mathbb{Z}_+. \tag{13}$$

This theorem, which is proved in Section 3 and Section 4, gives an explicit bound for the constant and the convergence rate, provided we can also find a bound for ζ_α. Although in specific cases ζ_α can be estimated precisely (see for example Section 9), which is why we give the bound M_α in the form of (11), we have found no such results in the renewal theory literature: in Section 5 we therefore prove the general result

Theorem 2.2 *Suppose that (10) holds for an atom* $\boldsymbol{\alpha} \in \mathcal{B}(\mathsf{X})$*, and also that the atom is strongly aperiodic: that is, for some* δ

$$P(\boldsymbol{\alpha}, \boldsymbol{\alpha}) > \delta. \tag{14}$$

Then

$$\zeta_\alpha \leq \frac{32 - 8\delta^2}{\delta^3}\Big(\frac{b}{1-\lambda}\Big)^2. \tag{15}$$

It is instructive to consider the possible performance of these bounds. Under the plausible configuration of smaller values of $\delta, 1-\lambda$ and larger values of b, the leading term is likely to be of the order

$$32[b/(1-\lambda)]^4\delta^{-3}.$$

This is clearly not close to the value if $1/(1-\lambda)$ which is known for stochastically monotone chains [12] and indicates that the bounds cannot be expected to be tight.

The conversion of these bounds to the general strongly aperiodic case is given in Section 6. The situation there is more complex, and the result we have is

Theorem 2.3 *Suppose that* $C \in \mathcal{B}(\mathsf{X})$ *satisfies*

$$P(x, \cdot\,) \geq \delta\nu(\,\cdot\,), \qquad x \in C \tag{16}$$

for some $\delta > 0$ *and probability measure* ν *concentrated on* C*; and that there is drift to* C *in the sense that for some* $\lambda_C < 1$*, some* $b_C < \infty$ *and a function* $V \geq 1$*,*

$$PV \leq \lambda_C V + b_C \mathbb{1}_C \tag{17}$$

where C, V *also satisfy*

$$V(x) \leq \mathrm{v}_C < \infty, \qquad x \in C. \tag{18}$$

Then $\boldsymbol{\Phi}$ *is* V*-uniformly ergodic, and*

$$|||P^n - \Pi|||_V \leq (1+\gamma_C)\frac{\rho}{\rho - \vartheta}\rho^n, \qquad n \in \mathbb{Z}_+. \tag{19}$$

for any $\rho > \vartheta = 1 - M_C^{-1}$*, for*

$$M_C = \frac{1}{(1-\check{\lambda})^2}\Big[1 - \check{\lambda} + \check{b} + \check{b}^2 + \overline{\zeta}_C(\check{b}(1-\check{\lambda}) + \check{b}^2)\Big]. \tag{20}$$

defined either in terms of the constants

$$\begin{aligned}\gamma_C &= \delta^{-2}[4b_C + 2\delta\lambda_C \mathrm{v}_C];\\ \check{\lambda} &= [\lambda_C + \gamma_C]/[1+\gamma_C] < 1;\\ \check{b} &= \mathrm{v}_C + \gamma_C < \infty;\end{aligned} \tag{21}$$

and the bound

$$\overline{\zeta}_C \leq \frac{4-\delta^2}{\delta^5}\Big(\frac{b_C}{1-\lambda_C}\Big)^2; \tag{22}$$

or, in the case where

$$\eta := \inf_{x\in C} P(x,C) - \delta > 0 \tag{23}$$

in terms of the constants

$$\begin{aligned} b^*_C &= [1-\delta]^{-1}[b_C + \delta(\lambda_C \mathrm{v}_C - \nu(V))]; \\ \gamma_C &= [\delta\eta]^{-1}[1-\delta]b^*_C; \qquad b^*_\alpha = \nu(V) - \lambda_C; \\ \check\lambda &= [\lambda_C + \gamma_C]/[1+\gamma_C] < 1; \qquad \check b = b^*_\alpha + \gamma_C < \infty; \end{aligned} \tag{24}$$

and the bound

$$\overline{\zeta}_C = \frac{1-\eta^2}{2\delta^4\eta}\Big(\frac{b_C}{1-\lambda_C}\Big)^2. \tag{25}$$

These expressions give computable bounds on the rates of convergence for strongly aperiodic geometrically ergodic chains.

Because of their generality, the bounds in Theorem 2.3 are of necessity far from tight: indeed, in the form using (21) and (22), we can consider again the probable leading terms and find that we are likely to produce a value of M_C of the order of

$$\overline{\zeta}_C\Big[\frac{\check b}{(1-\check\lambda)}\Big]^2 \approx \frac{4}{\delta^5}\Big[\frac{\check b}{(1-\check\lambda)}\Big]^2\Big[\frac{b_C}{1-\lambda_C}\Big]^2. \tag{26}$$

Since $\check b/(1-\check\lambda)$ is itself of order $\delta^{-4}16b_C^2/(1-\lambda_C)$ this gives a lower bound on M_C of order at least

$$2^{12}b_C^6/[\delta^{13}(1-\lambda_C)^4]. \tag{27}$$

In the case where we can use (24) and (25), then we have similarly that $\gamma_C \approx b_C/\delta\eta$, so $\check b/(1-\check\lambda) \approx b_C^2/[\delta^2\eta^2(1-\lambda_C)]$. Moreover, $\overline{\zeta}_C \approx b_C^2/[2\delta^4\eta(1-\lambda_C)^2]$. Thus M_C is at least

$$b_C^6/[2\delta^8\eta^5(1-\lambda_C)^4]. \tag{28}$$

which we might find to be considerably smaller than (27) in some circumstances.

There is thus clearly a premium on making δ and η as large as possible, and this is somewhat independent of the choice of V. Even so we always have $\delta+\eta \le 1$, and so the intrinsic capacity of this calculation seems doomed to be never better than $M_C \ge 2^{13} \approx 10^4$, regardless of the value of λ_C or b_C.

This shows that in general one will want to use more structure to get explicit bounds. In particular it will pay handsomely to get far better bounds than (22) or (25) for $\overline{\zeta}_C$, as we do in some of the examples below; and in particular we find that it is certainly worth attempting to make the small set C as small as possible in order to maximise the value of δ that can be chosen, provided that this can be done without making λ_C too close to unity or b_C too large. This tradeoff is illustrated in detail in our last section.

In Section 7, we extend the results from strongly aperiodic chains to general aperiodic chains. In this case the bounds become somewhat less explicit unless there is an atom in the space, although this is an important special case since it means that for chains on countable spaces, or for chains such as queueing systems with identified regeneration points, we do have a complete solution to the problem in principle. The result we prove is

Theorem 2.4 *Suppose again that (17) holds, and that there exists an atom α such that for some $N \geq 1$ and $\delta_C > 0$*

$$\sum_{j=1}^{N} P^j(x,\alpha) \geq \delta_C, \qquad x \in C. \tag{29}$$

Define the constants

$$\begin{aligned} \delta_N &= \delta_C/N^2; \\ b_k &= b_C(1+\delta_N^{-1})^k, \qquad k=0,\dots,N; \\ \lambda_k &= 1-(1-\lambda_C)/\textstyle\prod_{i=0}^{k-1}(1+b_i/\delta_N), \qquad k=0,\dots,N. \end{aligned} \tag{30}$$

Then there exists a function V_N with

$$V \leq V_N \leq V + b_N/\delta_N$$

such that

$$PV_N \leq \lambda_N V_N + b_N \mathbb{1}_\alpha. \tag{31}$$

Thus Theorem 2.1 holds using λ_N, b_N and with V_N in place of V; so that in terms of V we have

$$|||P^n - \Pi|||_V \leq [1 + b_N/\delta_N]\frac{\rho}{\vartheta - \rho}\rho^n, \qquad n \in \mathbb{Z}_+. \tag{32}$$

for $\rho > \vartheta$ where ϑ is defined as in Theorem 2.1 using λ_N, b_N.

This result is not quite explicit. It still involves ζ_α for the atom in (29) and the theorem contains no assumptions that will bound this: we need either special pleading, or extra conditions such as the strong aperiodicity in Theorem 2.2, for a completely computable bound.

3 Bounding the convergence rate for a bounded operator

Theorem 2.1 is a consequence of two sets of observations, the first using the spectral theory of operators and the second using probabilistic bounds most of which are inherent in Chapters 14–16 of [16].

As we saw in Theorem 1.1, (10) implies V-uniform convergence of the operator $\overline{P} := P - \Pi$ at *some* geometric rate. It therefore follows that the norm of the inverse $(I - z\overline{P})^{-1}$ is bounded for $|z|$ in some region containing the unit circle, and so at least for $|z|$ on the circle itself. To use this fact we will generalize a result of [22], which enables us to move explicitly from a bound M on the unit circle to a bound in a larger circle, as given in (13). This is the operator-theoretic observation. The probabilistic observations then come in generating the bound on the unit circle, which will give the form of M_α in (11).

Let $D(r) \subset \mathbb{C}$ denote the open disk centered at the origin of radius r; when $r = 1$ we set $D = D(r)$. We extend a remark given on p. 416 of [22] in

Theorem 3.1 *Suppose that $\mathcal{A}\colon L_V^\infty \to L_V^\infty$ is a bounded operator, that the inverse $(I - z\mathcal{A})^{-1}$ exists for each z in the closed unit disk $\overline{D}$, and that for some finite M*

$$|\!|\!|(I - z\mathcal{A})^{-1}|\!|\!|_V \le M, \qquad z \in \overline{D}.$$

Then $(I - z\mathcal{A})^{-1}$ exists in the larger open disk $D(\frac{M}{M-1})$, and for any $1 < r < \frac{M}{M-1}$,

(i) $|\!|\!|(I - z\mathcal{A})^{-1}|\!|\!|_V \le \dfrac{M}{r - (r-1)M}, \qquad |z| = r;$

(ii) $|\!|\!|\mathcal{A}^n|\!|\!|_V \le \dfrac{M}{r - (r-1)M} r^{-n}, \qquad n \in \mathbb{Z}_+.$

Proof Suppose that $M = \sup_{z \in \overline{D}} |\!|\!|(I - z\mathcal{A})^{-1}|\!|\!|_V$. The function $(I - z\mathcal{A})^{-1}V(x)$ of z is analytic in the open unit disk for any fixed $x \in \mathsf{X}$. By the maximum modulus principle, we must have

$$\sup_{|z|=1} |(I - z\mathcal{A})^{-1}V(x)| \ge |(I - z\mathcal{A})^{-1}V(x)|\big|_{z=0} = V(x),$$

which immediately implies that $M \ge 1$.

For any z we have

$$(I - z\mathcal{A}) = z(\frac{|z|}{z}I - \mathcal{A})[I - (I\frac{|z|}{z} - \mathcal{A})^{-1}(\frac{|z|}{z} - \frac{1}{z})]. \tag{33}$$

The first factor $(\frac{|z|}{z}I - \mathcal{A})$ is invertible by assumption, and the second is invertible if

$$|\!|\!|(I\frac{|z|}{z} - \mathcal{A})^{-1}(\frac{|z|}{z} - \frac{1}{z})|\!|\!|_V < 1.$$

By the conditions of the theorem, this holds if $M(|z| - 1)/|z| < 1$, or equivalently, if $|z| < M/(M-1)$. Hence we have established that $(I - z\mathcal{A})^{-1}$ exists for z in this range.

Now we have from (33), whenever $|z| = r < M/(M-1)$,

$$\begin{aligned}
|\!|\!|(I - z\mathcal{A})^{-1}|\!|\!|_V &\le \frac{1}{r}M\frac{1}{1 - |\!|\!|(I\frac{|z|}{z} - \mathcal{A})^{-1}(\frac{|z|}{z} - \frac{1}{z})|\!|\!|_V} \\
&\le \frac{1}{r}M\frac{1}{1 - (1 - \frac{1}{r})M} \\
&= M\frac{1}{r - (r-1)M},
\end{aligned}$$

which is the desired bound.

As also observed by Spieksma [32] for countable chains, (ii) now follows from (i) and Cauchy's inequality. For if we have for some M_r

$$|\!|\!|(I - z\mathcal{A})^{-1}|\!|\!|_V \le M_r$$

when $|z| = r$, then averaging over the circle of radius r in the complex plane we obtain for any $|f| \le V$,

$$\frac{1}{2\pi}\int_0^{2\pi}(I - re^{-i\theta}\mathcal{A})^{-1} f e^{in\theta}\, d\theta \leq M_r V$$

and since the integral is precisely $r^n \mathcal{A}^n f$,

$$|\mathcal{A}^n f(x)| = \leq r^{-n} M_r V(x)$$

which proves (ii). □

4 Bounding $(I - z\overline{P})^{-1}$ on $|z| \leq 1$

To apply Theorem 3.1 to the operator $\overline{P} = P - \Pi$ we now obtain an upper bound M_α on the norm of the inverse $(I - z\overline{P})^{-1}$ for $|z| \leq 1$ when the drift condition (10) holds.

This bound follows from the Regenerative Decomposition Theorem 13.2.5 of [16]. We have in convolution notation

$$P^n(x, f) = {}_\alpha P^n(x, f) + a_x * u * t_f(n) \tag{34}$$

where for $n \geq 1$, writing τ_α for the first return time to $\boldsymbol{\alpha}$,

$$\begin{aligned} {}_\alpha P^n(x, f) &= \mathsf{E}_x[f(\Phi_n)\mathbb{1}(\tau_\alpha \geq n)] \\ a_x(n) &= \mathsf{P}_x(\tau_\alpha = n) \\ u(n) &= \mathsf{P}_x(\Phi_n = \boldsymbol{\alpha}) \\ t_f(n) &= \mathsf{E}_\alpha[f(\Phi_n)\mathbb{1}(\tau_\alpha \geq n)] \end{aligned}$$

and for convenience we set $t_f(0) = 0$. For any $f \in L^1(\pi)$ define $\overline{f} = f - \pi(f)$.

We first bound the V-norm $|||(I - z\overline{P})^{-1}|||_V$ in terms of these quantities, and then move to bound the quantities themselves using (10).

Proposition 4.1 *If an atom $\boldsymbol{\alpha} \in \mathsf{X}$ exists then for $|z| \leq 1$,*

$$|||(I - z\overline{P})^{-1}|||_V \leq 1 + \sup_{\substack{|f| \leq V \\ x \in \mathsf{X}}} \frac{1}{V(x)} \left\{ \mathsf{E}_x\Big[\sum_{n=1}^{\tau_\alpha} |\overline{f}(\Phi_n)|\Big] + \zeta_\alpha \mathsf{E}_\alpha\Big[\sum_{n=1}^{\tau_\alpha} n|\overline{f}(\Phi_n)|\Big] \right\} \tag{35}$$

where $\zeta_\alpha = \sup_{z \in \overline{D}} |\sum (u(n) - u(n-1)) z^n|$ as in (12).

Proof From the first entrance last exit decomposition (34) we have for $|s| < 1$, $|f| \leq V$,

$$\begin{aligned} (I - z\overline{P})^{-1} f(x) &= \sum_{n=0}^{\infty} z^n \overline{P}^n f(x) \\ &= f(x) + \sum_{n=1}^{\infty} z^n P^n \overline{f}(x) \\ &= f(x) + P(I - z\mathbb{1}_{\alpha^c} P)^{-1}\overline{f}(x) \\ &\quad + A_x(z)\Big((1 - z)U(z)\Big)\frac{1}{1 - z} P(I - z\mathbb{1}_{\alpha^c} P)^{-1}\overline{f}(\boldsymbol{\alpha}) \end{aligned} \tag{36}$$

where

$$U(z) = \sum_{n=1}^{\infty} z^n u(n) \qquad A_x(z) = \sum_{n=1}^{\infty} z^n \mathsf{P}_x(\tau_\alpha = n).$$

We have that $|A_x(z)| \le 1$ for any $z \in \overline{D}$, $x \in \mathsf{X}$. The term $P(I - z\mathbb{1}_{\alpha^c}P)^{-1}\overline{f}\,(x)$ is less than or equal to $\mathsf{E}_x\Big[\sum_{n=1}^{\tau_\alpha} |\overline{f}(\Phi_n)|\Big]$ for $|z| \le 1$, and $|(1-z)U(z)| \le \zeta_\alpha$ for any $|z| \le 1$. Hence the proposition will be established if we can obtain appropriate bounds on

$$\xi(z) := \frac{1}{1-z} P(I - z\mathbb{1}_{\alpha^c}P)^{-1}\overline{f}\,(\alpha)$$

Writing $\beta_n = \mathsf{E}_\alpha[\overline{f}(\Phi_n)\mathbb{1}(n \le \tau_\alpha)] = {}_\alpha P^n(\alpha, \overline{f})$ we have that $\xi(z) = \sum_{n=1}^{\infty} \beta_n \frac{z^n}{1-z}$. Since, from Theorem 10.0.1 of [16], $\sum_{n=1}^{\infty} \beta_n = \pi(\alpha)^{-1}\pi(\overline{f}) = 0$, it follows that for $|z| < 1$,

$$|\xi(z)| = |\sum_{n=1}^{\infty} \beta_n \frac{z^n - 1}{z - 1}| \le \sum_{n=1}^{\infty} |\beta_n| n = \sum_{n=1}^{\infty} \mathsf{E}_\alpha[|\overline{f}(\Phi_n)|\mathbb{1}(n \le \tau_\alpha)]n$$

This combined with (36) proves the proposition. □

The most difficult term to handle in (35) is the final one. The following result, which is a generalization of Corollary 3.1 of [6], enables us to control the second expression in this term.

Proposition 4.2 *For any positive function h and any set $A \in \mathcal{B}^+(\mathsf{X})$ we have*

$$\int_A \mathsf{E}_x\Big[\sum_{n=1}^{\tau_A} n h(\Phi_n)\Big]\,\pi(dx) = \int_{\mathsf{X}} \mathsf{E}_x\Big[\sum_{n=1}^{\tau_A} h(\Phi_n)\Big]\,\pi(dx)$$

Proof By Fubini's Theorem first, and then using the Markov property and the fact that $\{\tau_A \ge k\} \in \mathcal{F}^{\Phi}_{k-1}$ we have

$$\begin{aligned}
\int_A \mathsf{E}_x\Big[\sum_{n=1}^{\tau_A} n h(\Phi_n)\Big]\,\pi(dx) &= \int_A \mathsf{E}_x\Big[\sum_{n=1}^{\tau_A}\sum_{k=1}^{n} h(\Phi_n)\Big]\,\pi(dx) \\
&= \int_A \mathsf{E}_x\Big[\sum_{k=1}^{\tau_A}\sum_{n=k}^{\tau_A} h(\Phi_n)\Big]\,\pi(dx) \\
&= \int_A \mathsf{E}_x\Big[\sum_{k=1}^{\tau_A} \mathsf{E}_{\Phi_{k-1}}\Big[\sum_{n=1}^{\tau_A} h(\Phi_n)\Big]\Big]\,\pi(dx).
\end{aligned}$$

By Theorem 10.0.1 of [16] the right hand side is equal to

$$\int_{\mathsf{X}} \mathsf{E}_x\Big[\sum_{n=1}^{\tau_A} h(\Phi_n)\Big]\,\pi(dx)$$

which proves the proposition. □

We are now able to use (10) to relate (35) to the required bound M_α in (11).

Proposition 4.3 *If (10) holds then the following bounds are satisfied for all $x \in \mathsf{X}$:*

(i) $\dfrac{\pi(V)}{\pi(\boldsymbol{\alpha})} \le \dfrac{b}{1-\lambda}$;

(ii) $\mathsf{E}_x\Big[\sum_{n=1}^{\tau_\alpha} V(\Phi_n)\Big] \le \dfrac{\lambda}{1-\lambda} V(x) + \dfrac{b}{1-\lambda}\mathbb{1}_\alpha(x)$;

(iii) $\mathsf{E}_\alpha\Big[\sum_{n=1}^{\tau_\alpha} nV(\Phi_n)\Big] \le \dfrac{\lambda}{1-\lambda}\dfrac{\pi(V)}{\pi(\boldsymbol{\alpha})} + \dfrac{b}{1-\lambda}$;

(iv) $\mathsf{E}_x[\tau_\alpha] \le \dfrac{1}{\log(1/\lambda)}\Big\{\log V(x) + \dfrac{b}{\lambda}\mathbb{1}_\alpha(x)\Big\}$;

(v) $\mathsf{E}_\alpha[\sum_{n=1}^{\tau_\alpha} n] \le \dfrac{1}{\log(1/\lambda)}\Big\{\dfrac{\pi(\log(V))}{\pi(\boldsymbol{\alpha})} + \dfrac{b}{\lambda}\Big\}$.

PROOF **(i)** This follows directly from Theorem 14.3.7 of [16] with $f = (1-\lambda)V$ and $s = b\mathbb{1}_\alpha$.

(ii) From an obvious extension of the Comparison Theorem (Theorem 14.2.2) of [16] with $f = (1-\lambda)V$, $s = b\mathbb{1}_\alpha$, and $\tau = \tau_\alpha$, we have the bound

$$(1-\lambda)\mathsf{E}_x\Big[\sum_{n=0}^{\tau_\alpha - 1} V(\Phi_n)\Big] \le V(x) - \mathsf{E}_x[V(\Phi_{\tau_\alpha})] + b\mathbb{1}_\alpha(x)$$

Subtracting $(1-\lambda)[V(x) - \mathsf{E}_x[V(\Phi_{\tau_\alpha})]$ from each side gives

$$(1-\lambda)\mathsf{E}_x\Big[\sum_{n=1}^{\tau_\alpha} V(\Phi_n)\Big] \le \lambda V(x) - \lambda\mathsf{E}_x[V(\Phi_{\tau_\alpha})] + b\mathbb{1}_\alpha(x),$$

and this gives (ii).

(iii) This follows from (ii) and Proposition 4.2.

(iv) By Jensen's inequality and (10),

$$P\log V \le \log V + \log(\lambda) + \frac{b}{\lambda}\mathbb{1}_\alpha.$$

Hence the result again follows from the Comparison Theorem of [16].

(v) As in the proof of (iii), this follows by combining (iv) and Proposition 4.2. □

By putting together these estimates we now have

Proof of Theorem 2.1: From Proposition 4.1 – Proposition 4.3, we have for $|z| \le 1$,

$$\begin{aligned} \|(I - z\overline{P})^{-1}\|_V &\le 1 + \sup_{x\in\mathsf{X}} \frac{1}{V(x)}\Big\{\Big(\frac{\lambda}{1-\lambda}V(x) + \frac{b}{1-\lambda}\mathbb{1}_\alpha(x)\Big)(1+\pi(V)) \\ &\qquad + \zeta_\alpha\Big(\frac{\lambda}{1-\lambda}\frac{\pi(V)}{\pi(\boldsymbol{\alpha})} + \frac{b}{1-\lambda}\Big)(1+\pi(V))\Big\} \\ &\le 1 + \Big(\frac{\lambda}{1-\lambda} + \frac{b}{1-\lambda}\Big)\Big(1 + \frac{b}{1-\lambda}\Big) \end{aligned}$$

$$
\begin{aligned}
&+\zeta_\alpha\Big(\frac{\lambda}{1-\lambda}\frac{b}{1-\lambda}+\frac{b}{1-\lambda}\Big)\Big(1+\frac{b}{1-\lambda}\Big)\\
=\;& \frac{1}{1-\lambda}+\frac{b}{1-\lambda}+\frac{b\lambda}{(1-\lambda)^2}+\frac{b^2}{(1-\lambda)^2}\\
&+\zeta_\alpha\frac{b}{1-\lambda}\Big(1+\frac{b}{1-\lambda}\Big)\\
=\;& \frac{1}{(1-\lambda)^2}\Big[1-\lambda+b+b^2+\zeta_\alpha(b(1-\lambda)+b^2)\Big].
\end{aligned}
$$

The result then follows directly from Theorem 3.1 on letting $\mathcal{A}=\overline{P}$. □

5 Bounding the renewal variation and ζ_α

The bounds in the previous section are in general relatively tight, as can be seen from their derivation. The only term which is not explicitly controlled here is ζ_α, and in the this section we consider bounds on ζ_α in terms of the known quantities for a model satisfying (10).

To do this we introduce the renewal variation $\mathrm{Var}\,(u)$ [16, Chapter 13], defined by

$$\mathrm{Var}\,(u)=\sum_{n=0}^{\infty}|u(n)-u(n-1)|, \tag{37}$$

where $u(-1)=0$ by convention. It is immediate that $\mathrm{Var}\,(u)$ is a bound for ζ_α:

$$\zeta_\alpha=\sup_{|z|\le 1}|\sum_n z^n(u(n)-u(n-1))|\le\sum_{n=0}^{\infty}|u(n)-u(n-1)|.$$

We will use $\mathrm{Var}\,(u)$ as the bound for ζ_α in what follows. In the case of stochastically monotone chains with $\boldsymbol{\alpha}$ at the "bottom" of the space as in [12] this is exact since $u(n)\downarrow\pi(\boldsymbol{\alpha})$. We note that in Assumption A of Spieksma [32] it is in effect assumed that the quantity $|\sum_n z^n(u(n)-u(n-1))|$ is bounded on a suitably large disk $D(r)\subset\mathbb{C}$ and thus the need to find a bound on ζ_α is avoided.

To bound $\mathrm{Var}\,(u)$ we will use a variation of the coupling technique in [16, Chapter 13]. For convenience let τ denote the time of first return to $\boldsymbol{\alpha}$, and let the corresponding return time distribution be given by

$$p(n)=\mathsf{P}_\alpha(\tau=n). \tag{38}$$

As is usual in coupling arguments, we will consider two independent renewal sequences with the properties of u: that is, consider $S_1(n)$ and $S_2(n)$, where the initial variables $S_1(0),S_2(0)$ may have arbitrary distributions, and where the increments $Y_i(n)=S_i(n+1)-S_i(n)$ are independent and each have the distribution

$$\mathsf{P}(Y_i(n)=m)=p(m)$$

for each copy $i = 1, 2$ and each $n > 0$.

Next suppose the first component commences with zero delay (i.e. $S_1(0) = 0$) and the second is a delayed sequence with deterministic delay of one time unit (i.e. $S_2(0) = 1$). Let T_{01} denote the coupling time of these two sequences: that is, the first time that both sequences renew simultaneously. Then we have as in Proposition 13.4.1 of [16] that

$$\mathrm{Var}\,(u) \le \mathsf{E}[T_{01}]. \tag{39}$$

We now recast this bound in terms of the forward recurrence time chains for these renewal sequences. For $i = 1, 2$, let

$$V_i^+(n) := \inf(S_i(m) - n : S_i(m) > n), \qquad n \ge 0$$

denote these chains, and construct the bivariate chain $V^*(n) = \{V_1^+(n), V_2^+(n)\}$. If we put

$$\tau_{1,1} := \inf(n : V^*(n) = (1,1))$$

then in the specific case of an initial delay $V_1^+(0) = 1$ and $V_2^+(0) = 2$,

$$T_{01} = \tau_{1,1}$$

so that from (39),

$$\mathrm{Var}\,(u) \le \mathsf{E}_{1,2}[\tau_{1,1}]. \tag{40}$$

(Note that in (13.70) of [16] the less accurate bound $\mathsf{E}[T_{01}] \le \mathsf{E}_{1,2}[\tau_{1,1}] + 1$ is used.) The key calculation is the following, which expresses $\mathsf{E}_{1,2}[\tau_{1,1}]$ in terms of $\mathsf{E}_{1,1}[\tau_{1,1}]$. We have by symmetry

$$\begin{aligned}
\mathsf{E}_{1,1}[\tau_{1,1}] &= \textstyle\sum_{k=-\infty}^{\infty} \mathsf{E}_{1,1}[\tau_{1,1}\mathbb{1}\{Y_1(1) = Y_2(1) + k\}] \\
&= \mathsf{E}_{1,1}[\tau_{1,1}\mathbb{1}\{Y_1(1) = Y_2(1)\}] + 2\textstyle\sum_{k=1}^{\infty} \mathsf{E}_{1,1}[\tau_{1,1}\mathbb{1}\{Y_1(1) = Y_2(1) + k\}] \\
&= \textstyle\sum_{n=1}^{\infty} np(n)^2 + 2\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} p(n)p(n+k)\mathsf{E}_{1,1+k}[n + \tau_{1,1}] \\
&= \textstyle\sum_{n=1}^{\infty} np(n)^2 + 2\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} np(n)p(n+k) \\
&+ \; 2\textstyle\sum_{n=1}^{\infty} p(n)p(n+1)\mathsf{E}_{1,2}[\tau_{1,1}] + 2\sum_{k=2}^{\infty}\sum_{n=1}^{\infty} p(n)p(n+k)\mathsf{E}_{1,1+k}[\tau_{1,1}]
\end{aligned} \tag{41}$$

so that in particular, using only the fact that $\mathsf{E}_{1,1+k}[\tau_{1,1}] \ge k$ in the last term, we have

$$\begin{aligned}
\mathsf{E}_{1,1}[\tau_{1,1}] \;\ge\; & \textstyle\sum_{n=1}^{\infty} np(n)^2 + 2\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} np(n)p(n+k) \\
& +2\textstyle\sum_{k=2}^{\infty}\sum_{n=1}^{\infty} kp(n)p(n+k) + 2\sum_{n=1}^{\infty} p(n)p(n+1)\mathsf{E}_{1,2}[\tau_{1,1}].
\end{aligned} \tag{42}$$

Thus if we can control the terms $p(n)$ we will be able to bound $\mathsf{E}_{1,2}[\tau_{1,1}]$ in terms of $\mathsf{E}_{1,1}[\tau_{1,1}]$. We can then use the fact that if π^* denotes the invariant measure for the bivariate forward recurrence time chain $V^*(n)$, then from Kac's Theorem (Theorem 10.2.2 of [16])

$$\pi^*(1,1) = \Big[\mathsf{E}_{1,1}[\tau_{1,1}]\Big]^{-1}. \tag{43}$$

But since the bivariate chain consists of independent copies of the forward recurrence time chain, we further have

$$\pi^*(1,1) = [\pi(\boldsymbol{\alpha})]^2. \tag{44}$$

Thus we will be able to bound $\mathrm{Var}\,(u)$ in terms of $\pi(\boldsymbol{\alpha})$.

In the examples below we show various model-dependent ways to do this: here we will develop a bound that holds for arbitrary strongly aperiodic renewal sequences, although this introduces some undesirable inequalities that can be avoided by direct use of (42) if we are able to assume knowledge of, say, higher values of $p(k)$, and in particular of $p(2)$ as well as $p(1)$.

With the structure above, we proceed to the

Proof of Theorem 2.2: Suppose only that (10) holds and that $P(\boldsymbol{\alpha},\boldsymbol{\alpha}) \geq \delta$. Then (15) will hold if we can show

$$\mathrm{Var}\,(u) \leq \frac{32-8\delta^2}{\delta^3}\Big(\frac{b}{1-\lambda}\Big)^2 \tag{45}$$

In order to bound $\mathrm{Var}\,(u)$ in this way, using only the minorization bound on $P(\boldsymbol{\alpha},\boldsymbol{\alpha})$, we need a further and somewhat artificial construction on the original chain. Let us "split" the atom $\boldsymbol{\alpha}$ (in the simplest possible way) into two equal parts $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ by tossing a fair coin each time $\boldsymbol{\alpha}$ is reached, and by putting, for all $A \subseteq \boldsymbol{\alpha}^c$,

$$P(\boldsymbol{\alpha}_i, A) = P(\boldsymbol{\alpha}, A), \qquad i = 0, 1.$$

Rather than using a formal split chain we will assume for convenience of notation in this proof that this is the structure of the original chain, so that in particular

$$\pi(\boldsymbol{\alpha}_i) = \pi(\boldsymbol{\alpha})/2, \qquad i = 0, 1 \tag{46}$$

and

$$P(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j) = P(\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_i) = \delta/2, \qquad i,j = 0,1. \tag{47}$$

Now focusing on $\boldsymbol{\alpha}_0$, let us consider the renewal sequence $u^*(n)$ given by

$$u^*(n) = P^n(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_0) \tag{48}$$

corresponding to returns to $\boldsymbol{\alpha}_0$, with corresponding renewal times

$$p^*(m) = \mathsf{P}_{\alpha_0}(\tau_{\alpha_0} = m), \qquad m \in \mathbb{Z}_+ : \tag{49}$$

this is related to our original renewal sequence by

$$u(n) = P^n(\boldsymbol{\alpha},\boldsymbol{\alpha}) = 2P^n(\boldsymbol{\alpha},\boldsymbol{\alpha}_0) = 2u^*(n), \qquad n \geq 1,$$

and so we have

$$\mathrm{Var}\,(u) \;=\; 2\mathrm{Var}\,(u^*). \tag{50}$$

Our reason for this somewhat tortuous construction is simple: (47) enables us to assert for $n > 1$ that

$$p^*(n) \geq P(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)[P(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1)]^{n-2}P(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_0) = [\delta/2]^n \tag{51}$$

and this control of $p^*(n)$ allows us to control $\mathrm{Var}\,(u^*)$, using (42).

If we now consider $\tau^*_{1,1}$ to be the return time to $(1,1)$ for the process of returns to $\boldsymbol{\alpha}_0$ we have as in (42), using (47) and (51)

$$\begin{aligned}\mathsf{E}_{1,1}[\tau^*_{1,1}] &\geq \textstyle\sum_{n=1}^{\infty} np^*(n)^2 + 2\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} np^*(n)p^*(n+k) \\ &+ \textstyle 2\sum_{k=2}^{\infty}\sum_{n=1}^{\infty} kp^*(n)p(n+k)^* + 2\sum_{n=1}^{\infty} p^*(n)p^*(n+1)\mathsf{E}_{1,2}[\tau^*_{1,1}] \\ &= \tfrac{[\delta/2]^2[1+2[\delta/2]^3]}{[1-(\delta/2)^2]^2} + \tfrac{2[\delta/2]^3}{[1-(\delta/2)^2]}\mathsf{E}_{1,2}[\tau^*_{1,1}].\end{aligned} \tag{52}$$

Thus we have successively from (50), (40), (52), (43), (44) and (46)

$$\begin{aligned}\mathrm{Var}\,(u) &= 2\mathrm{Var}\,(u^*) \\ &\leq 2\mathsf{E}_{1,2}[\tau^*_{1,1}] \\ &\leq \mathsf{E}_{1,1}[\tau^*_{1,1}](8-2\delta^2)/\delta^3 \\ &\leq (4/\pi(\boldsymbol{\alpha})^2)(8-2\delta^2)/\delta^3\end{aligned} \tag{53}$$

where we have ignored the possible increase in accuracy from the first term in the last line of (52). We now need one last step. From Proposition 4.3 (i), and the fact that $V \geq 1$, we have

$$\pi(\boldsymbol{\alpha}) \geq (1-\lambda)/b \tag{54}$$

and so, finally, from (53) and (54) we find (45) as required. □

These bounds involve at least two approximations which are likely to be very poor. The first is the bounding $p^*(n)$ by $[\delta/2]^n$, which ignores a great deal of the probability of the event $\{\tau_{\alpha_0} = n\}$. The second is the bounding of $\mathsf{E}_{1,2}[\tau^*_{1,1}]$ as in (52), which is clearly not tight, although it does perhaps pick up the main part of the bound in (42).

We see in the examples below that the weakness in this estimate of $\mathrm{Var}\,(u)$ warrants every effort to improve this term in our estimate of M_α in practical situations.

6 Bounding strongly aperiodic chains

We have from Theorem 2.1 and Theorem 2.2 a computable bound under the conditions that

(i) there exists an atom with $P(\boldsymbol{\alpha},\boldsymbol{\alpha}) \geq \delta$; and

(ii) there is a known solution to (10) with drift to this same atom.

In this section we find bounds under the much more general condition that for some $\lambda < 1$, some $b < \infty$ and a function $V \geq 1$,

$$PV \leq \lambda_C V + b_C \mathbb{1}_C \tag{55}$$

for some (not necessarily atomic) set C satisfying

$$P(x,\,\cdot\,) \geq \delta\nu(\,\cdot\,), \qquad x \in C \tag{56}$$

for a $\delta > 0$ and a probability measure ν concentrated on C, and

$$\sup_{x\in C} V(x) := \mathrm{v}_C < \infty. \tag{57}$$

Thus through (56) we assume that the chain is strongly aperiodic (see [16, Chapter 5]), and the conditions of Theorem 2.3 all hold.

We first consider a chain containing an atom $\boldsymbol{\alpha}$ and a small set C such that $\boldsymbol{\alpha} \subseteq C$ and for some $\delta > 0$,

$$P(x, \boldsymbol{\alpha}) \geq \delta, \qquad x \in C: \tag{58}$$

this is the special case of (56) with $\boldsymbol{\alpha}$ a singleton and ν its Dirac measure, and will be shown below to be an appropriate route to analyzing general strongly aperiodic chains using the Nummelin splitting.

Let us further look at the bounding constant b_C separately on and off the atom in (55) and define the two constants $b_\alpha^* \leq b_C, b_C^* \leq b_C$ such that

$$PV \leq \lambda_C V + b_C^* \mathbb{1}_{C\cap\alpha^c} + b_\alpha^* \mathbb{1}_\alpha. \tag{59}$$

In practice we may often have $b_C = b_\alpha^* = b_C^*$ but it is worthwhile separating them in what follows.

With this structure we have

Theorem 6.1 *Under (58) and (55), there exists a function V' with $V \leq V' \leq V + b_C^*/\delta$ such that*

$$PV' \leq \lambda V' + b\mathbb{1}_\alpha \tag{60}$$

where

$$\lambda = [\lambda_C + b_C^*/\delta]/[1 + b_C^*/\delta] < 1 \tag{61}$$

and

$$b = b_\alpha^* + b_C^*/\delta < \infty. \tag{62}$$

Thus Theorem 2.1 and Theorem 2.2 hold with this choice of λ, b and with V' in place of V; and so for $\rho > \vartheta$

$$\|P^n - \Pi\|_V \leq [1 + b_C^*/\delta]\frac{\rho}{\vartheta - \rho}\rho^n, \qquad n \in \mathbb{Z}_+. \tag{63}$$

PROOF The construction we used is based on that in [34]. For a B to be chosen later, set

$$\begin{aligned} V'(x) &= V(x), \qquad x \in \boldsymbol{\alpha}; \\ V'(x) &= V(x) + B, \qquad x \in \boldsymbol{\alpha}^c, \end{aligned} \tag{64}$$

so that for $x \in C^c$

$$\begin{aligned} PV' &\leq PV + B \\ &\leq \lambda_C V + B \\ &\leq \lambda V' \end{aligned} \tag{65}$$

provided $\lambda \geq [B+\lambda_C]/[B+1]$. For $x \in \boldsymbol{\alpha}$, we have

$$\begin{aligned} PV' &\leq PV + B \\ &\leq \lambda_C V + b^*_\alpha + B \\ &\leq \lambda V' + b^*_\alpha + B \end{aligned} \tag{66}$$

since $\lambda \geq \lambda_C$. Here we are ignoring a presumably small negative term $(\lambda - \lambda_C)V(\boldsymbol{\alpha})$.

In order to establish an appropriate value for B we calculate for $x \in C \cap \boldsymbol{\alpha}^c$,

$$\begin{aligned} PV'(x) &= P(V+B)(x) - P(x,\boldsymbol{\alpha})B \\ &\leq (\lambda_C V + B) + b^*_C - \delta B \\ &\leq \lambda V' + b^*_C - \delta B \\ &\leq \lambda V' \end{aligned} \tag{67}$$

provided we choose the value $B \geq b^*_C/\delta$. Using the exact values $B = b^*_C/\delta$ and $\lambda = [B+\lambda_C]/[B+1]$, we get the values in (61) and (62) as required.

Since we now have drift to $\boldsymbol{\alpha}$ as in (60), Theorem 2.1 holds, and from (58) with $x = \boldsymbol{\alpha}$ (14) holds and so Theorem 2.2 is also valid. Now in terms of our original function V we have for $\rho > \vartheta$ given by (13)

$$\begin{aligned} \|P^n(x,\,\cdot\,) - \pi\|_V &\leq \|P^n(x,\,\cdot\,) - \pi\|_{V'} \\ &\leq V'(x)\tfrac{\rho}{\vartheta-\rho}\rho^n \\ &\leq V(x)[1 + b^*_C/\delta]\tfrac{\rho}{\vartheta-\rho}\rho^n \end{aligned} \tag{68}$$

so that (63) holds as stated. □

We now move from the atomic to the general strongly aperiodic case using a splitting argument. Suppose that (55) holds but that C satisfies only (56) and (57). If we now write $\beta = \delta/2$, then obviously we still have

$$P(x,\,\cdot\,) \geq \beta\nu(\,\cdot\,), \qquad x \in C \tag{69}$$

and also, since $\nu(C) = 1$,

$$P(x, C) - \beta\nu(C) \geq \beta. \tag{70}$$

Recall now the construction of the Nummelin splitting [16, Chapter 5] based on (69).

We define a "split" chain $\check{\boldsymbol{\Phi}}$ on a state space $\check{\mathsf{X}} = \mathsf{X}_0 \times \mathsf{X}_1$ consisting of two "copies" of X. We write $x_0 \in \mathsf{X}_0$ and $x_1 \in \mathsf{X}_1$ for the two "copies of x" and $A_0 \subseteq \mathsf{X}_0$ and $A_1 \subseteq \mathsf{X}_1$ for the "copies of A" for any set A in $\mathcal{B}(\mathsf{X})$. If μ is any measure on $\mathcal{B}(\mathsf{X})$, then we split μ into two measures on each of X_0 and X_1 by defining the measure

$$\begin{aligned} \mu^*(A_0) &= [1-\beta]\mu(A \cap C) + \mu(A \cap C^c), \\ \mu^*(A_1) &= \beta\mu(A \cap C), \end{aligned} \tag{71}$$

where β and C are the constant and the set in (69). Rather more simply, any function V on X is split by setting $V(x_0) = V(x_1) = V(x)$.

The split chain is then governed by the transition law $\check{P}$ which is defined by

$$\check{P}(x_0, \cdot\,) = P(x, \cdot\,)^*, \qquad x_0 \in \mathsf{X}_0 \backslash C_0; \tag{72}$$

$$\check{P}(x_0, \cdot\,) = [1-\beta]^{-1}[P(x, \cdot\,)^* - \beta\nu^*(\,\cdot\,)], \qquad x_0 \in C_0; \tag{73}$$

$$\check{P}(x_1, \cdot\,) = \nu^*(\,\cdot\,), \qquad x_1 \in \mathsf{X}_1. \tag{74}$$

For $A = A_0 \cup A_1$ we will use the fact that

$$\check{P}(x_0, A) \;=\; [1-\beta]^{-1}[P(x,A) - \beta\nu(A)] \le [1-\beta]^{-1}P(x,A). \tag{75}$$

On C_0^c we have $\check{P}(x_0, \cdot\,) = P(x, \cdot\,)$, and we can essentially ignore $\check{P}(x_1, \cdot\,)$ on C_1^c and shall do so, thus in effect putting $\mathsf{X}_1 = C_1$.

This splitting introduces an atom $\check{\alpha} = C_1 \subset C$ into the split space. Moreover, because we have split using β, we have from (70), (73) and (74) that for $x \in C$

$$\begin{aligned}\check{P}(x_1, \check{\alpha}) &= \beta\nu(C) = \beta; \\ \check{P}(x_0, \check{\alpha}) &= [\beta/(1-\beta)][P(x,C) - \beta\nu(C)] \\ &\ge [\beta^2/(1-\beta)]\end{aligned} \tag{76}$$

which shows that, since $\beta < 1/2$, $\check{\alpha}$ is reached from every point in the split representation $C_0 \cup C_1$ of C with probability at least

$$\check{\delta} := \beta^2/(1-\beta). \tag{77}$$

Note that if we had used a splitting using δ rather than β we might have had $\check{P}(x_0, \check{\alpha}) = 0$ from (56). The choice of $\beta = \delta/2$ can be shown to be close to optimal in this argument, although it is likely that as in the examples below, we can often use δ itself in practice to improve the constants in the bound.

We now consider the drift inequalities for the split chain. We still have trivially that (59) holds for $x \in C_0^c$, from (72). Since $\mathrm{v}_C = \sup_{x\in C} V(x) < \infty$ from (57), and ν is concentrated on C, we have from (74) that for $x_1 \in C_1$

$$\check{P}V(x_1) = \int \nu(dy)V(y) \le \lambda_C V(x_1) + \mathrm{v}_C \tag{78}$$

whilst for $x_0 \in C_0$, from (75)

$$\begin{aligned}\check{P}V(x_0) &\le [1-\beta]^{-1}PV(x_0) \\ &\le \lambda_C V(x_0) + [1-\beta]^{-1}[b_C + \beta\lambda_C \mathrm{v}_C].\end{aligned} \tag{79}$$

Thus writing

$$\check{b}_C^* = [1-\beta]^{-1}[b_C + \beta\lambda_C\mathrm{v}_C], \qquad \check{b}_\alpha^* = \mathrm{v}_C, \tag{80}$$

we have that (59) holds for the split chain with λ_C unchanged and b_C^*, b_α^* replaced by $\check{b}_C, \check{b}_\alpha^*$: that is,

$$\check{P}V \le \lambda_C V + \check{b}_C^* \mathbb{1}_{C\cap\check{\alpha}^c} + \check{b}_\alpha^* \mathbb{1}_{\check{\alpha}}. \tag{81}$$

We can now put these calculations together with Theorem 6.1 to give

Proof of Theorem 2.3: From (81) and (77) we can now use Theorem 6.1 for the split chain to find that for some $V \le V' \le V + \check{b}^*_C/\check{\delta}$

$$\check{P}V \le \check{\lambda}V' + \check{b}\mathbb{1}_{\check{\alpha}} \tag{82}$$

where now we have

$$\check{\lambda} = \frac{\lambda_C + \gamma_C}{1 + \gamma_C}$$

with γ_C given by

$$\gamma_C = \check{b}^*_C/\check{\delta} = \delta^{-2}[4b_C + 2\delta\lambda_C \mathrm{v}_C]$$

from (77) and (80); and from (80) also

$$\check{b} = \check{b}^*_{\alpha} + \check{b}^*_C/\check{\delta} = \mathrm{v}_C + \gamma_C.$$

Thus we have immediately that (20) will hold with $\check{\lambda}, \check{b}$ as in (21).

Hence from (63)

$$|||\check{P}^n - \check{\Pi}|||_V \quad \le \quad [1 + \check{b}/\check{\delta}]\tfrac{\rho}{\vartheta - \rho}\rho^n, \qquad n \in \mathbb{Z}_+ \tag{83}$$

and since

$$|||P^n - \Pi|||_V \quad \le \quad |||\check{P}^n - \check{\Pi}|||_V \tag{84}$$

we have shown that (19) holds with $\rho > \vartheta$ given by (20), provided $\overline{\zeta}_C \ge \mathrm{Var}\,(\check{u})$ where $\mathrm{Var}\,(\check{u})$ is the total variation of the process of renewals to $\check{\boldsymbol{\alpha}}$ in the split chain.

We finally show that (22) provides such a bound for $\mathrm{Var}\,(\check{u})$. If we put $\check{p}(n)$ for the return time distribution for this process, then since

$$\check{P}(x_0, C_0) = P(x, C) - \beta\nu(C) > \beta$$

from (71) and (73), we have (essentially as in (51)) from (76) and (70),

$$\begin{aligned} \check{p}(n) \quad &\ge \quad \int_{C_0} \cdots \int_{C_0} \check{P}(\check{\boldsymbol{\alpha}}, dx_0^1)\check{P}(x_0^1, dx_0^2)\ldots\check{P}(x_0^{n-1}, \check{\boldsymbol{\alpha}}) \\ &\ge \quad (1-\beta)\beta^{n-2}\beta^2/(1-\beta) \\ &= \quad \beta^n. \end{aligned} \tag{85}$$

Now we note that by the splitting construction

$$\check{\pi}(\check{\boldsymbol{\alpha}}) = \delta\pi(C). \tag{86}$$

Moreover, integrating the the original drift equations against π we have

$$(1 - \lambda_C)\pi(V) \le b_C\pi(C)$$

and since $V \ge 1$ we have

$$[\check{\pi}(\check{\boldsymbol{\alpha}})]^{-2} \le [b_C/\delta(1 - \lambda_C)]^2. \tag{87}$$

Emulating (53) we thus have

$$\mathrm{Var}\,(\check{u}) \leq \frac{4-\delta^2}{\delta^3}[\check{\pi}(\check{\alpha})]^{-2} \leq \frac{4-\delta^2}{\delta^3}\Big(\frac{b_C}{\delta(1-\lambda_C)}\Big)^2 \tag{88}$$

which is the required bound; and the theorem is proved for the bounds in (21) and (22).

To prove the bounds using (24) and (25), we refine these calculations when $\eta > 0$ in (23). To show (81) holds with b^*_C, b^*_α defined as in (24), note as in (78) that for $x_1 \in \check{\alpha}$

$$\check{P}V(x_1) = \int \nu(dy)V(y) \leq \lambda_C V(x_1) + b^*_\alpha$$

whilst for $x_0 \in C\backslash\check{\alpha}$, refining (79),

$$\begin{aligned}\check{P}V(x_0) &= [1-\delta]^{-1}[PV(x) - \delta\nu(V)] \\ &\leq [1-\delta]^{-1}[\lambda_C V(x) + b_C - \delta\nu(V)].\end{aligned}$$

As in (76), we also have for each $x \in C\backslash\check{\alpha}$ that

$$\check{P}(x,\check{\alpha}) \geq \check{\delta} := [\delta\eta]/[1-\delta]. \tag{89}$$

From (81) and (89) we can now use Theorem 6.1 to find that for some $V \leq V' \leq V + b^*_C/\check{\delta}$, (82) holds with $\check{b} = b^*_\alpha + b^*_C/\check{\delta}$ and $\check{\lambda} = (\lambda_C + \gamma_C)/(1+\gamma_C)$ with $\gamma_C = b^*_C/\check{\delta}$.

It remains to prove the bound on $\overline{\zeta}_C$. We now have that $\check{p}(n) \geq \delta\eta^{n-1}$ and as in (53) and (88) this gives

$$\overline{\zeta}_C \leq \frac{1-\eta^2}{2\delta^2\eta}[\check{\pi}(\check{\alpha})]^{-2}. \tag{90}$$

Combining (90) and (87) gives the required result. □

7 Bounding general aperiodic chains

In the general aperiodic case we assume only that C in the drift condition (3) satisfies the m-step minorization condition (2). In principle we can reduce this to the strongly aperiodic situation solved above, since the m-skeleton is strongly aperiodic. In developing computable bounds in practice this presents some problems, since the drift condition is typically available for the one-step chain and the strong aperiodicity is for the m-skeleton.

In this section we develop computable bounds in two situations of practical interest and a general solution which requires more information than we have so far assumed.

Our first result indicates how to deal directly with the m-skeleton.

Proposition 7.1 *For a general aperiodic ψ-irreducible chain assume that for some $m \geq 1$, $\delta > 0$, and a probability measure ν concentrated on C,*

$$P^m(x,\,\cdot\,) \geq \delta\nu(\,\cdot\,), \qquad x \in C \tag{91}$$

and that there is a function $V \geq 1$ and constants $\lambda_C < 1, b_C < \infty$ such that the m-step drift

$$P^m V \leq \lambda_C V + b_C \mathbb{1}_C \tag{92}$$

holds. Assume also that for some $d < \infty$

$$PV \leq dV \tag{93}$$

which certainly holds if V is also a solution to (3).

Then defining ϑ and γ_C as in Theorem 2.3, we have for any $\rho > \vartheta$

$$\|P^n - \Pi\|_V \leq (\frac{1+\gamma_C}{\rho - \vartheta})[d + \frac{b_C}{1-\lambda_C}]^m \rho^{n/m} \tag{94}$$

PROOF From Theorem 2.3 applied to the m-skeleton, for any $\rho > \vartheta$

$$\|P^{mk} - \Pi\|_V \leq (\frac{(1+\gamma_C)\rho}{\rho - \vartheta})\rho^k, \qquad k \in \mathbb{Z}_+.$$

Since $\|\cdot\|_V$ is an operator norm, it is sub-multiplicative, and hence for any $n = mk + i \in \mathbb{Z}_+$,

$$\begin{aligned} \|P^n - \Pi\|_V &= \|(P^{mk} - \Pi)(P - \Pi)^i\|_V \\ &\leq \|(P^{mk} - \Pi)\|_V \|(P - \Pi)^i\|_V \\ &\leq (\tfrac{(1+\gamma_C)\rho}{\rho - \vartheta})\rho^k (\|P - \Pi\|_V)^i. \end{aligned}$$

Since $PV \leq dV$

$$\|P - \Pi\|_V \leq \|P\|_V + \|\Pi\|_V \leq d + \frac{b_C}{1-\lambda_C},$$

and hence the result is proved. □

In many situations we will be able to find a solution to the m-step drift equation and this bound is then practicable.

However, in general we are likely to need to consider the case where we have only the one-step drift equation (17) holding, for C satisfying (91), and we now indicate how to control this situation.

As is shown in Theorem 15.3.4 of [16], by iterating we have

$$P^m V \leq \lambda_C^m V + b_C \sum_0^{m-1} P^i \mathbb{1}_C \leq \lambda_C V + m b_C \mathbb{1}_{C(m)}$$

where the set

$$C(m) = \{y : \sum_{i=0}^{m-1} P^i(y, C) \geq [\lambda_C - \lambda_C^m]/b_C\}.$$

Thus we do have an m-step drift equation holding, but the set $C(m)$ may not satisfy (2) for the m-skeleton, and so we cannot use Proposition 7.1 immediately.

As shown in Theorem 15.3.4 of [16], however, $C(m)$ is at least a small set for the m-skeleton. Following the proof of Theorem 5.5.7 of [16], there must therefore exist an integer N° and a $\delta^\circ > 0$ such that

$$P^{N^\circ m}(x, \cdot\,) \geq \delta^\circ \nu(\,\cdot\,), \qquad x \in C(m) \tag{95}$$

where ν is actually the measure concentrated on C in (91).

There is no general prescription that we are aware of for finding N° and δ°: if this situation occurs they will need to be calculated separately. However, if they are known then by splitting the m-skeleton over C and using the same arguments as in (81), we can reduce the situation to one where (17) holds for a set C (now given by $C(m)$ in the split space) satisfying, from (95), the condition that

$$P^{N^\circ}(x, \check{\alpha}) \geq \delta^\circ \delta, x \in C(m). \tag{96}$$

Now we can find explicit bounds in this situation (as is shown in Theorem 2.4, which we have yet to prove), and so we can construct a bound for the rate of convergence of the m-skeleton as in the proof of Theorem 2.3, and then transfer this to the original chain as in Proposition 7.1.

We will not try to identify the outcome of this program explicitly, and we feel it is unlikely to be a task undertaken except in pressing circumstances. We conclude by providing the missing link in the chain, which is of independent interest in many practical situations for countable chains or for chains with a true atom in the space.

Proof of Theorem 2.4: Again we use a construction based on [34], and generalise the calculations in Theorem 6.1, although here we will not try to specify the improvement from differentiating between possible values for b_C over different parts of C such as $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^c$.

We write $C_0 = \boldsymbol{\alpha}$, and $C_k = \{y : P^k(y, \boldsymbol{\alpha}) \geq k\delta_N\}$; and we set

$$\tilde{C}_k = \bigcup_0^k C_j, \qquad \hat{C}_k = \tilde{C}_k \backslash \tilde{C}_{k-1}, \qquad k = 0, \ldots, N.$$

For $x \in C$, from (29) there is some $k \leq n$ such that

$$P^k(x, \boldsymbol{\alpha}) \geq \delta_C / N \geq k\delta_N$$

and so $C \subseteq \cup_0^N C_k = \tilde{C}_N$; thus from (17), we certainly have

$$PV \leq \lambda_C V + b_C \mathbb{1}_{\tilde{C}_N}. \tag{97}$$

We will successively develop functions V_j such that this drift equation holds with $\tilde{C}_N$ replaced by $\tilde{C}_{N-j}$, and with λ_C replaced by λ_j and b_C replaced by b_j as in (30), based on the method of Theorem 6.1. The final iterate of this operation gives the result we seek.

The crucial observation we need is that for any $k = 1, \ldots, N-1$, we have for $x \in C_{k+1}$

$$\begin{aligned} (k+1)\delta_N &\leq P^{k+1}(x, \boldsymbol{\alpha}) \\ &= \textstyle\int_{C_k} P(x, dy) P^k(x, \boldsymbol{\alpha}) + \int_{C_k^c} P(x, dy) P^k(x, \boldsymbol{\alpha}) \\ &\leq P(x, C_k) + k\delta_N \end{aligned} \tag{98}$$

so that for $k = 0, \ldots, N-1$,

$$P(x, C_k) \geq \delta_N, \qquad x \in C_{k+1}. \tag{99}$$

In following the proof of Theorem 6.1, we will use (99) successively to play the role of (58).

Let us define V_1 by

$$\begin{aligned} V_1(x) &= V(x), \qquad x \in \tilde{C}_{N-1}; \\ V_1(x) &= V(x) + b_C/\delta_N, \qquad x \in \tilde{C}^c_{N-1}. \end{aligned}$$

If we define $\lambda_1 = [\lambda_C + b_C/\delta_N]/[1 + b_C/\delta_N]$ and $b_1 = b_C + b_C/\delta_N$ then as in (64)–(67) we find

$$PV_1 \leq \lambda_1 V_1 + b_1 \mathbb{1}_{\tilde{C}_{N-1}}. \tag{100}$$

Rewrite these constants in the form (30):

$$\begin{aligned} b_1 &= b_C(1 + 1/\delta_N) \\ \lambda_1 &= [\lambda_C - 1 + (1 + b_C/\delta_N)]/[1 + b_C/\delta_N]. \end{aligned}$$

If we iterate this construction we find that by induction we get constants λ_k and b_k given by

$$\begin{aligned} b_{k+1} &= b_k(1 + 1/\delta_N) = b_C(1 + 1/\delta_N)^{k+1} \\ \lambda_{k+1} &= [\lambda_k + b_k/\delta_N]/[1 + b_k/\delta_N] \\ &= \Big\{ \Big[\lambda_C - 1 + \textstyle\prod_{i=0}^{k-1}(1 + b_i/\delta_N)\Big] \Big/ \Big[\textstyle\prod_{i=0}^{k-1}(1 + b_i/\delta_N)\Big] \\ &\qquad -1 + (1 + b_k/\delta_N)\Big\} \Big\{1 + b_k/\delta_N\Big\}^{-1} \\ &= \Big[\lambda_C - 1 + \textstyle\prod_{i=0}^{k}(1 + b_i/\delta_N)\Big] \Big/ \textstyle\prod_{i=0}^{k}(1 + b_i/\delta_N) \end{aligned}$$

so that (30) defines the coefficients for

$$PV_k \leq \lambda_k V_k + b_k \mathbb{1}_{\tilde{C}_{N-k}} \tag{101}$$

where the functions V_k are defined successively by

$$\begin{aligned} V_k(x) &= V_{k-1}(x), \qquad x \in \tilde{C}_{N-k}; \\ V_k(x) &= V_{k-1}(x) + b_{k-1}/\delta_N, \qquad x \in \tilde{C}^c_{N-k}. \end{aligned}$$

Thus the result will hold as required for the N^{th} iteration, at which time we will have the function

$$\begin{aligned} V_N(\boldsymbol{\alpha}) &= V(\boldsymbol{\alpha}); \\ V_N(x) &= V(x) + \textstyle\sum_{N-k}^{N-1} b_j, \qquad x \in \hat{C}_k; \\ V_N(x) &= V(x) + \textstyle\sum_{N-k}^{N-1} b_j, \qquad x \in \tilde{C}^c_N \end{aligned} \tag{102}$$

and so (31) holds.

Finally, note that (32) follows exactly as does (63), and the theorem is proved. □

8 Bounding the M/M/1 queue

In this and the next two sections we evaluate the general bounding procedure for three types of chain: for the number of customers in the M/M/1 queue, or the Bernoulli random walk on $\mathbb{Z}_+$; for a general uniformly ergodic chain so that in effect the atom in the space is easy to identify; and for an MCMC example where the chain is truly continuous in nature.

For the first, we have

$$P(x, x-1) = p > 1/2, \qquad P(x, x+1) = q = 1-p, \qquad x \geq 1$$

with the boundary condition $P(0,0) = p, P(0,1) = q$.

For this model we know (see Chapter 15 of [16]) that in (10) with $\boldsymbol{\alpha} = \{0\}$ we can choose $V(x) = (1+\gamma)^x$ for some $\gamma > 0$: and for each choice of γ we get for $x \geq 1$ the value

$$\lambda_\gamma = p(1+\gamma)^{-1} + q(1+\gamma).$$

Choosing γ to minimize this parameter leads to

$$\gamma = \sqrt{p/q} - 1.$$

This then shows that (10) holds for $V(x) = [p/q]^{x/2}$ with the parameters

$$\lambda = 2\sqrt{pq}, \qquad b = p - \sqrt{pq}. \tag{103}$$

In this case we can get a considerably better bound on $\mathrm{Var}\,(u)$ than that given by (15), since we do not need to create the artificial sequence $p^*(n)$ through splitting the state $\{0\}$.

By just choosing the simplest term in (42) and using the facts that $p(1) = p, p(2) = qp$ and $\pi(0) = 1 - q/p$, we have

$$\begin{aligned} 2\mathsf{E}_{1,2}[\tau_{1,1}] &\leq \mathsf{E}_{1,1}[\tau_{1,1}]/p(1)p(2) - 1 \\ &= [\pi(0)]^{-2}/p(1)p(2) - 1 \\ &= [p/(p-q)]^2/p^2 q - 1 \end{aligned}$$

and hence

$$2\mathrm{Var}\,(u) \leq \frac{1}{[(p-q)]^2 q} - 1. \tag{104}$$

As examples let us consider two numerical cases.

Suppose firstly that $p = 2/3$. Then $\lambda = 0.943$ which can be shown to be the best possible rate of convergence (see [12, 32]).

Now we have in this case $b = 0.195$ and $\mathrm{Var}\,(u) \leq 13$ using (104). Substituting in (11) gives us

$$M_0 \approx 88 + 15\mathrm{Var}\,(u) \leq 283$$

so that we get by this method a bound of $\vartheta \leq 0.996$ for the rate of convergence.

Suppose secondly that $p = 0.9$. Then $\lambda = 0.6$ and $b = 0.6$ also; whilst (104) gives $\mathrm{Var}\,(u) \leq 7.3$. This time substituting in (11) gives us

$$M_0 \approx 8.5 + 3.75\mathrm{Var}\,(u) \leq 36$$

leading to a bound of $\vartheta \leq 0.972$ for the rate of convergence.

Suppose that we had actually used the method of splitting $\{0\}$ to bound $\mathrm{Var}\,(u)$ as in (15). In the case $p = 2/3$ we would have found $\mathrm{Var}\,(u) \leq 1116$; in the case $p = 0.9$ we would have found $\mathrm{Var}\,(u) \leq 79$. In either case there is a severe deterioration of the bound for M_0, and the value of more accurate estimates of $\mathrm{Var}\,(u)$ is obvious.

9 Bounding a uniformly ergodic chain

Let us consider next a generic uniformly recurrent chain. Suppose there is some atom in the space such that

$$P(x, \boldsymbol{\alpha}) = 1/2, \qquad x \in \mathsf{X}$$

and let the remainder of the transitions be arbitrary: here the value of 1/2 is for convenience alone. From the coupling bound (9) we know that we can choose $R = 1$ and $\rho = 1/2$ for this example.

In this case, since the renewal sequence corresponds exactly to geometric inter-arrival times with rate 1/2, we have that $u(n) = 1/2$ for all $n \geq 1$, and so the only positive term in $\mathrm{Var}\,(u)$ is the first: thus

$$\mathrm{Var}\,(u) = u(0) - u(1) = 1/2. \tag{105}$$

For this chain, choosing $V(x) = 1 + c$ for $x \in \boldsymbol{\alpha}^c$ and $V(\boldsymbol{\alpha}) = 1$ we see that we can take

$$\lambda = [2 + c]/[2 + 2c], \qquad b = c\lambda$$

in (10). This immediately illustrates the problem we will face in getting a tight bound from (11): as $c \to \infty$ we have $\lambda \to 1/2$, but b will go to infinity thus giving very poor bounds. By substituting in (11) we get from the expressions for λ, b

$$M_{\alpha}(c) = 6c^{-1} + 12 + 6c + c^2 + \mathrm{Var}\,(u)[6 + 5c + c^2] \tag{106}$$

Thus at best, since $c > 0$ we will get a bound of the convergence rate of $\rho > 0.933$. Using $\mathrm{Var}\,(u) = 1/2$ we see that (106) is minimized at approximately $c = 0.75$, and here we get an approximate best value of

$$M_{\alpha}(c) \approx 25.06 + 10.3\mathrm{Var}\,(u) \approx 30.2$$

leading to a bound for $\vartheta \approx 0.967$, and a bound on the constant of $R = 323.3$ at the value $\rho = 0.97$.

Here we have been able to use the true value for $\mathrm{Var}\,(u)$. If we use the approximation methods of Section 5 then inevitably we get less reasonable values. Indeed, since in this special case we can follow through the argument for the atom $\boldsymbol{\alpha}$ itself, rather than splitting $\boldsymbol{\alpha}$ artificially, we get as in (42)

$$\mathsf{E}_{1,1}[\tau_{1,1}] \geq \frac{\delta^2}{(1-\delta^2)^2} + \frac{2\delta^3}{(1-\delta^2)}\mathsf{E}_{1,2}[\tau_{1,1}] = 4/9 + \mathsf{E}_{1,2}[\tau_{1,1}]/3$$

and since now $\pi(\boldsymbol{\alpha}) = 1/2$ we find

$$\mathrm{Var}\,(u) \leq 3/[\pi(\boldsymbol{\alpha})]^2 - 12/9 = 10.67.$$

Using this value in (106) gives us a minimal bound on $M_\alpha(c)$ around 115 at $c = 0.3$, so that now $\vartheta \leq 0.991$.

The effect of splitting the atom at zero and using the bound (15) is again large. We have from this approach

$$\mathrm{Var}\,(u) \leq 258(2+c)^2$$

for all c, so this method yields only $\mathrm{Var}\,(u) \leq 1032$.

This again indicates, rather dramatically, the benefit that flows from attempting in special cases to get a better estimate of $\mathrm{Var}\,(u)$ than that given in (15).

10 Bounding Metropolis algorithms

The examples considered in the previous two sections used results where the rates depended on behaviour at a natural atom. Clearly the bounds in Theorem 2.3 on continuous spaces will be larger, using only the minimal information and the methods we employ. In this section we illustrate this by giving an application to Hastings and Metropolis algorithms, which have recently received considerable attention using general Markov chain theory [29, 33, 3].

We will indicate how the bounds are calculated in practice for such algorithms. We do find, regrettably, that the order of magnitude of the bounds is not of practical value and these techniques cannot be used at this stage to give bounds of value for real applications.

Hastings and Metropolis algorithms ([9, 14]) allow simulation of a probability measure π which is only known up to an unknown constant factor: that is, if densities exist, when only $\pi(x)/\pi(y)$ is known. This is especially relevant when π is the posterior distribution in a Bayesian context: see [2, 30, 29, 23, 33] for a more detailed introduction.

To implement such algorithms on $\mathbb{R}^k$, say, we first consider a "candidate" transition kernel Q with transition probability density $q(x,y)$ which generates potential transitions for a Markov chain. A "candidate transition" generated according to Q is then accepted with probability $\alpha(x,y)$; otherwise the chain does not move. The key is that $\alpha(x,y)$ can be chosen to ensure that π is invariant for the final chain.

We will consider here only the Metropolis algorithm: other examples are studied in detail in [13]. The Metropolis algorithm utilises a symmetric candidate transition Q: that is, one for which $q(x,y) = q(y,x)$. The most common usage of such chains occurs (cf. [33]) if Q is not merely symmetric but satisfies the random walk condition

$$q(x,y) = q(x-y) = q(y-x) \tag{107}$$

for some fixed density q. Thus actual transitions of the Metropolis chain with random walk candidate distribution take place according to a law P with transition density

$$p(x,y) = q(x-y)\alpha(x,y), \qquad y \neq x \tag{108}$$

and with probability of remaining at the same point given by

$$P(x,\{x\}) = \int q(x-y)[1-\alpha(x,y)]\mu^{\text{Leb}}(dy). \tag{109}$$

The acceptance probabilities then take the form

$$\alpha(x,y) = \begin{cases} \min\{\frac{\pi(y)}{\pi(x)}, 1\} & \pi(x)q(x,y) > 0 \\ 1 & \pi(x)q(x,y) = 0 \end{cases} \tag{110}$$

The key observation for this algorithm is that, with this choice of α, the "target" measure π is stationary for P. As an example of such an algorithm for which we can calculate bounds on the rate of convergence, we will consider the case in which π is $N(0,1)$. A natural choice of symmetric candidate distribution is then the centred Normal: that is, $Q(x,\cdot\,)$ is a $N(x,1)$ distribution. In this situation we have

$$\alpha(x,y) = \min(1, e^{-1/2(y^2-x^2)}), \qquad x,y > 0,$$

and we know from Theorem 3.4 of [13] that the chain is geometrically ergodic. We will apply the second set of bounds in Theorem 2.3 to generate bounds on this convergence.

It is intuitively sensible to apply (17) for the symmetric small set $C_\Delta = (-x_\Delta, x_\Delta)$ for some $x_\Delta > 0$ to be chosen.

The size of the bound (28) seems best controlled by controlling x_Δ, which determines δ_C and η. If one were interested only in establishing existence of geometric ergodicity, ν could be chosen to be Lebesgue so that δ_C would be given by $e^{-2x_\Delta^2}$. A considerable improvement is obtained by choosing the Normal minorising distribution

$$\nu(\cdot) = N_{x\in C_\Delta}(0, 1/\sqrt{2})/N_\Delta$$

with N_Δ the normalising constant

$$N_\Delta = [1/\sqrt{2\pi}]\int_{-x_\Delta}^{x_\Delta} e^{-x^2}\,dx \ :$$

this gives $\delta_C = N_\Delta e^{-x_\Delta^2}$. The infimum η occurs at $x = x_\Delta$, so that $\eta = 1/2 - \boldsymbol{\Phi}(-2x_\Delta) - N_\Delta e^{-x_\Delta^2}$.

To choose an optimal value of x_Δ we now need to focus on V and λ. We will take the test function $V(y) = e^{s|y|}$ for some s to be chosen also. Then straightforward calculations show

$$\begin{aligned}
\lambda(x,s) &:= \int P(x,y)V(y)/V(x)dy \\
&= e^{s^2/2}\left(\boldsymbol{\Phi}(-s) - \boldsymbol{\Phi}(-x-s) + e^{-2xs}(\boldsymbol{\Phi}(-x+s) - \boldsymbol{\Phi}(-2x+s))\right) \\
&+ \tfrac{1}{\sqrt{2}}\left(e^{(x^2-6xs+s^2)/4}\boldsymbol{\Phi}((-3x+s)/\sqrt{2}) + e^{(x-s)^2/4}\boldsymbol{\Phi}((-x+s)/\sqrt{2})\right) \\
&+ 1/2 + \boldsymbol{\Phi}(-2x) - \tfrac{1}{\sqrt{2}}e^{x^2/4}(\boldsymbol{\Phi}(-3x/\sqrt{2}) + \boldsymbol{\Phi}(-x/\sqrt{2}))
\end{aligned}$$

Now instead of specifying s and calculating λ_C, it appears more effective to first specify λ_C which controls the size of M_C and determines the remaining constants. Thus we first calculate s by equating $\sup_{x \in C_\Delta^c} \lambda(x, s)$ to the given λ_C, which is easy since the supremum occurs at x_Δ . Since $PV(x) \leq \lambda_c V(x) + (\lambda(x, s) - \lambda_C)V(x)$ for all $x \in C_\Delta$ we then find

$$b_C = \sup_{x \in C_\Delta} e^{s|x|}(\lambda(x, s) - \lambda_c)$$

which has a supremum at zero given by

$$b_C = \sqrt{2}\Phi(s/\sqrt{2})e^{s^2/4} + 1 - 1/\sqrt{2} - \lambda_C;$$

and then we can calculate

$$b_\alpha^* = \left((\sqrt{2}e^{s^2/4})/(N_\Delta)\right)\left(\Phi(\sqrt{2}(x_\Delta - s/2) + \Phi(s/\sqrt{2}) - 1\right).$$

The terms $b_c^*, \gamma_c, \check{\lambda}, \check{b}, \overline{\zeta}, M_C$ then follow as in (24), (25) and (20).

The terms comprising (28) and M_C are given below, for several combinations of x_Δ and λ_C.

x_Δ	λ_C	s	b_c^6	δ_c^{-8}	η^{-5}	$(1-\lambda)^{-4}$	M_C
1.10	0.75	1.9	5.5E2	7.1E5	4.1E2	2.6E2	4.6E13
1.30	0.75	1.4	0.14	2.1E7	1.4E2	2.6E2	1.0E13
1.15	0.95	0.48	6.1E-4	1.5E6	3.0E2	1.6E5	6.5E10
1.30	0.95	0.38	1.6E-4	2.1E7	1.4E2	1.6E5	9.0E10
1.20	0.99	0.31	1.4E-5	3.4E6	2.2E2	1.0E8	2.1E12

As indicated, we find a minimum value of $M_C = 6.5E10$, obtained for $x_\Delta = 1.15$ and $\lambda_c = 0.95$.

Although this bound is not useful for practical purposes, it is unfortunately relatively close to the best that can be expected using these computations. We note that this procedure gives $\delta_C = 0.17$ and $\eta = 0.32$, so that $\delta_c^{-8}\eta^{-5} = 1.4E9$ in our best case; (28) shows that with this constraint our final result cannot be improved greatly.

Acknowledgement

We are grateful to Kerrie Mengersen for much discussion of this problem in the MCMC context, and especially for assistance on the example in Section 10, which is based on ideas in [13]; to Robert Lund for careful reading of the initial manuscript, which led to some improvements in the final bounds calculated, as well as to the approach in [12]; to Jeff Rosenthal and Peter Baxendale for providing access to their recent work [25] and [1]; and to Floske Spieksma for stimulating discussion and considerable effort as we tried (with limited success) to improve the renewal variance bounds in Section 5.

REFERENCES

[1] Peter H. Baxendale. Uniform estimates for geometric ergodicity of recurrent Markov processes. Unpublished report, University of Southern California, 1993.

[2] J.E. Besag and P.J. Green. Spatial statistics and Bayesian computation (with discussion). *J. Roy. Statist. Soc. Ser. B*, 55, 1993.

[3] K.S. Chan. Asymptotic behaviour of the Gibbs sampler. *J. Amer. Statist. Assoc.*, 88:320–326, 1993.

[4] R. Chen and R. S. Tsay. On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.*, 1:613–634, 1991.

[5] R. Chen and R. S. Tsay. Functional-coefficient autoregeressive models. *J. Amer. Statist. Assoc.*, 88:298–308, 1993.

[6] R. Cogburn. A uniform theory for sums of Markov chain transition probabilities. *Ann. Probab.*, 3:191–214, 1975.

[7] P. Diaconis, P.D. Feigin, and R.L. Tweedie. Rates of convergence in successive substitution sampling schemes. 1993. In preparation.

[8] Hani Doss. Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Preprint, Florida State University, 1991.

[9] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[10] A. Hordijk and F.M. Spieksma. On ergodicity and recurrence properties of a Markov chain with an application. *Adv. Appl. Probab.*, 24:343–376, 1992.

[11] Jun Liu. Eigenanalysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance resampling. Research Report R-427, Department of Statistics, Harvard University, 1992.

[12] R. Lund and R.L. Tweedie. Geometric convergence rates for stochastically ordered Markov chains.

[13] K.L. Mengersen and R.L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. 1993. Submitted for publication.

[14] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chemical Physics*, 21:1087–1091, 1953.

[15] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes I: Discrete time chains. *Adv. Appl. Probab.*, 24:542–574, 1992.

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

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

[18] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge, 1984.

[19] E. Nummelin and P. Tuominen. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Applns.*, 12:187–202, 1982.

[20] E. Nummelin and R. L. Tweedie. Geometric ergodicity and R-positivity for general Markov chains. *Ann. Probab.*, 6:404–420, 1978.

[21] N. Popov. Conditions for geometric ergodicity of countable Markov chains. *Soviet Math. Dokl.*, 18:676–679, 1977.

[22] F. Riesz and B. S. Nagy. *Functional Analysis.* Ungar, New York, NY, 1955.

[23] Gareth O. Roberts and Nicholas G.Polson. A note on the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B*, to appear, 1994.

[24] J. S. Rosenthal. *Rates of Convergence for Gibbs Sampler and Other Markov Chains.* PhD thesis, Harvard University, 1992.

[25] J.S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. Technical report 9321, Department of Statistics, University of Toronto, 1993.

[26] J.S. Rosenthal and R.L. Tweedie. Methods of calculating convergence rates for Markov chains. Technical report in preparation, Colorado State University, 1994.

[27] Mark J. Schervish and Bradley P. Carlin. On the convergence of successive substitution sampling. *J. Comp. Graphical Stats.*, 1:111–127, 1993.

[28] E. Seneta. *Non-negative Matrices and Markov Chains.* Springer-Verlag, New York, Second edition, 1980.

[29] A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *Amer. Statist.*, 46:84–88, 1992.

[30] A.F.M. Smith and G.O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B*, 55, 1993.

[31] F. M. Spieksma. *Geometrically Ergodic Markov Chains and the Optimal Control of Queues.* PhD thesis, University of Leiden, 1991.

[32] F.M. Spieksma. Spectral conditions and bounds for the rate of convergence of countable Markov chains. Technical Report, University of Leiden., 1993.

[33] L. Tierney. Exploring posterior distributions using Markov chains. Technical Report 560, University of Minnesota, 1991.

[34] R. L. Tweedie. Hitting times of Markov chains, with application to state-dependent queues. *Bull. Austral. Math. Soc.*, 17:97–107, 1977.

[35] R. L. Tweedie. Criteria for rates of convergence of Markov chains with application to queueing and storage theory. In J. F. C. Kingman and G. E. H. Reuter, editors, *Probability, Statistics and Analysis.* London Mathematics Society Lecture Note Series, Cambridge University Press, Cambridge, 1983.

[36] D. Vere-Jones. On the spectra of some linear operators associated with queueing systems. *Z. Wahrscheinlichkeitstheorie und Verw. Geb.*, 2:12–21, 1963.