16

V-Uniform Ergodicity

In this chapter we introduce the culminating form of the geometric ergodicity theorem, and show that such convergence can be viewed as geometric convergence of an operator norm; simultaneously, we show that the classical concept of uniform (or strong) ergodicity, where the convergence in (13.4) is bounded independently of the starting point, becomes a special case of this operator norm convergence.

We also take up a number of other consequences of the geometric ergodicity properties proven in Chapter 15, and give a range of examples of this behavior. For a number of models, including random walk, time series and state-space models of many kinds, these examples have been held back to this point precisely because the strong form of ergodicity we now make available is met as the norm, rather than as the exception. This is apparent in many of the calculations where we verified the ergodic drift conditions (V2) or (V3): often we showed in these verifications that the stronger form (V4) actually held, so that unwittingly we had proved V-uniform or geometric ergodicity when we merely looked for conditions for ergodicity.

To formalize V-uniform ergodicity, let P_1 and P_2 be Markov transition functions, and for a positive function $\infty > V \geq 1$, define the *V-norm distance* between P_1 and P_2 as

$$|\!|\!| P_1 - P_2 |\!|\!|_V := \sup_{x \in \mathsf{X}} \frac{\| P_1(x,\,\cdot\,) - P_2(x,\,\cdot\,)\|_V}{V(x)} \tag{16.1}$$

We will usually consider the distance $|\!|\!| P^k - \pi |\!|\!|_V$, which strictly speaking is not defined by (16.1), since π is a probability measure, not a kernel. However, if we consider the probability π as a kernel by making the definition

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

then $|\!|\!| P^k - \pi |\!|\!|_V$ is well-defined.

V-uniform ergodicity

An ergodic chain $\boldsymbol{\Phi}$ is called *V-uniformly ergodic* if

$$|\!|\!| P^n - \pi |\!|\!|_V \to 0, \qquad n \to \infty. \tag{16.2}$$

We develop three main consequences of Theorem 15.0.1 in this chapter.

Firstly, we interpret (15.4) in terms of convergence in the operator norm $\|P^k - \pi\|_V$ when V satisfies (15.3), and consider in particular the uniformity of bounds on the geometric convergence in terms of such solutions of (V4). Showing that the choice of V in the term V-uniformly ergodic is not coincidental, we prove

Theorem 16.0.1 *Suppose that $\boldsymbol{\Phi}$ is ψ-irreducible and aperiodic. Then the following are equivalent for any $V \geq 1$:*

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

(ii) *There exists $r > 1$ and $R < \infty$ such that for all $n \in \mathbb{Z}_+$*

$$\|P^n - \pi\|_V \leq Rr^{-n}. \tag{16.3}$$

(iii) *There exists some $n > 0$ such that $\|P^i - \pi\|_V < \infty$ for $i \leq n$ and*

$$\|P^n - \pi\|_V < 1. \tag{16.4}$$

(iv) *The drift condition (V4) holds for some petite set C and some V_0, where V_0 is equivalent to V in the sense that for some $c \geq 1$,*

$$c^{-1}V \leq V_0 \leq cV. \tag{16.5}$$

PROOF That (i), (ii) and (iii) are equivalent follows from Proposition 16.1.3. The fact that (ii) follows from (iv) is proven in Theorem 16.1.2, and the converse, that (ii) implies (iv), is Theorem 16.1.4. □

Secondly, we show that V-uniform ergodicity implies that the chain is *strongly mixing*. In fact, it is shown in Theorem 16.1.5 that for a V-uniformly ergodic chain, there exists R and $\rho < 1$ such that for any g^2, $h^2 \leq V$ and $k, n \in \mathbb{Z}_+$,

$$|\mathsf{E}_x[g(\Phi_k)h(\Phi_{n+k})] - \mathsf{E}_x[g(\Phi_k)]\mathsf{E}_x[h(\Phi_{n+k})]| \leq R\rho^n[1 + \rho^k V(x)].$$

Finally in this chapter, using the form (16.3), we connect concepts of geometric ergodicity with one of the oldest, and strongest, forms of convergence in the study of Markov chains, namely uniform ergodicity (sometimes called strong ergodicity).

> **Uniform ergodicity**
>
> A chain $\boldsymbol{\Phi}$ is called *uniformly ergodic* if it is V-uniformly ergodic in the special case where $V \equiv 1$; that is, if
>
> $$\sup_{x \in \mathsf{X}} \|P^n(x, \cdot) - \pi\| \to 0, \qquad n \to \infty. \tag{16.6}$$

There are a large number of stability properties all of which hold uniformly over the whole space when the chain is uniformly ergodic.

Theorem 16.0.2 *For any Markov chain $\boldsymbol{\Phi}$ the following are equivalent:*

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

(ii) *There exists $r > 1$ and $R < \infty$ such that for all x*

$$\|P^n(x,\,\cdot\,) - \pi\| \le Rr^{-n}; \tag{16.7}$$

that is, the convergence in (16.6) takes place at a uniform geometric rate.

(iii) *For some $n \in \mathbb{Z}_+$,*

$$\sup_{x\in\mathsf{X}} \|P^n(x,\,\cdot\,) - \pi(\,\cdot\,)\| < 1. \tag{16.8}$$

(iv) *The chain is aperiodic and Doeblin's Condition holds: that is, there is a probability measure ϕ on $\mathcal{B}(\mathsf{X})$ and $\varepsilon < 1$, $\delta > 0$, $m \in \mathbb{Z}_+$ such that whenever $\phi(A) > \varepsilon$*

$$\inf_{x\in\mathsf{X}} P^m(x, A) > \delta. \tag{16.9}$$

(v) *The state space X is ν_m-small for some m.*

(vi) *The chain is aperiodic and there is a petite set C with*

$$\sup_{x\in\mathsf{X}} \mathsf{E}_x[\tau_C] < \infty$$

in which case for every set $A \in \mathcal{B}^+(\mathsf{X})$, $\sup_{x\in\mathsf{X}} \mathsf{E}_x[\tau_A] < \infty$.

(vii) *The chain is aperiodic and there is a petite set C and a $\kappa > 1$ with*

$$\sup_{x\in\mathsf{X}} \mathsf{E}_x[\kappa^{\tau_C}] < \infty,$$

in which case for every $A \in \mathcal{B}^+(\mathsf{X})$ we have for some $\kappa_A > 1$,

$$\sup_{x\in\mathsf{X}} \mathsf{E}_x[\kappa_A^{\tau_A}] < \infty.$$

(viii) *The chain is aperiodic and there is a bounded solution $V \ge 1$ to*

$$\Delta V(x) \le -\beta V(x) + b\mathbb{1}_C(x), \quad x \in \mathsf{X} \tag{16.10}$$

for some $\beta > 0$, $b < \infty$, and some petite set C.

Under (v), we have in particular that for any x,

$$\|P^n(x,\,\cdot\,) - \pi\| \le \rho^{n/m} \tag{16.11}$$

where $\rho = 1 - \nu_m(\mathsf{X})$.

PROOF This cycle of results is proved in Theorem 16.2.1-Theorem 16.2.4. □

Thus we see that uniform convergence can be embedded as a special case of V-geometric ergodicity, with V bounded; and by identifying the minorization that makes the whole space small we can explicitly bound the rate of convergence.

Clearly then, from these results geometric ergodicity is even richer, and the identification of test functions for geometric ergodicity even more valuable than the last chapter indicated. This leads us to devote attention to providing a method of moving from ergodicity with a test function V to e^{sV}-geometric convergence, which in practice appears to be a natural tool for strengthening ergodicity to its geometric counterpart.

Throughout this chapter, we provide examples of geometric or uniform convergence for a variety of models. These should be seen as templates for the use of the verification techniques we have given in the theorems of the past several chapters.

16.1 Operator norm convergence

16.1.1 The operator norm $|\!|\!| \cdot |\!|\!|_V$

We first verify that $|\!|\!| \cdot |\!|\!|_V$ is indeed an operator norm.

Lemma 16.1.1 *Let L_V^∞ denote the vector space of all functions $f: \mathsf{X} \to \mathbb{R}_+$ satisfying*

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

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.

PROOF The definition of $|\!|\!| \cdot |\!|\!|_V$ may be restated as

$$\begin{aligned}
|\!|\!| P_1 - P_2 |\!|\!|_V &= \sup_{x \in \mathsf{X}} \Big\{ \frac{\sup_{|g| \le V} |P_1(x,g) - P_2(x,g)|}{V(x)} \Big\} \\
&= \sup_{|g| \le V} \sup_{x \in \mathsf{X}} \frac{|P_1(x,g) - P_2(x,g)|}{V(x)} \\
&= \sup_{|g| \le V} |P_1(\,\cdot\,,g) - P_2(\,\cdot\,,g)|_V \\
&= \sup_{|g|_V \le 1} |P_1(\,\cdot\,,g) - P_2(\,\cdot\,,g)|_V
\end{aligned}$$

which is by definition the operator norm of $P_1 - P_2$ viewed as a mapping from L_V^∞ to itself. □

We can put this concept together with the results of the last chapter to show

Theorem 16.1.2 *Suppose that $\boldsymbol{\Phi}$ is ψ-irreducible and aperiodic and (V4) is satisfied with C petite and V everywhere finite. Then for some $r > 1$,*

$$\sum r^n |\!|\!| P^n - \pi |\!|\!|_V < \infty, \tag{16.12}$$

and hence $\boldsymbol{\Phi}$ is V-uniformly ergodic.

PROOF This is largely a restatement of the result in Theorem 15.4.1. From Theorem 15.4.1 for some $R < \infty$, $\rho < 1$,

$$\|P^n(x, \cdot\,) - \pi\|_V \le RV(x)\rho^n, \qquad n \in \mathbb{Z}_+,$$

and the theorem follows from the definition of $|\!|\!| \cdot |\!|\!|_V$. □

Because $|\!|\!| \cdot |\!|\!|_V$ is a norm it is now easy to show that V-uniformly ergodic chains are always geometrically ergodic, and in fact V-geometrically ergodic.

Proposition 16.1.3 *Suppose that π is an invariant probability and that for some n_0,*

$$|\!|\!|P - \pi|\!|\!|_V < \infty \qquad \textit{and} \qquad |\!|\!|P^{n_0} - \pi|\!|\!|_V < 1.$$

Then there exists $r > 1$ such that

$$\sum_{n=1}^{\infty} r^n |\!|\!|P^n - \pi|\!|\!|_V < \infty.$$

PROOF Since $|\!|\!| \cdot |\!|\!|_V$ is an operator norm we have for any $m, n \in \mathbb{Z}_+$, using the invariance of π,

$$|\!|\!|P^{n+m} - \pi|\!|\!|_V = |\!|\!|(P-\pi)^n(P-\pi)^m|\!|\!|_V \le |\!|\!|P^n - \pi|\!|\!|_V |\!|\!|P^m - \pi|\!|\!|_V$$

For arbitrary $n \in \mathbb{Z}_+$ write $n = kn_0 + i$ with $1 \le i \le n_0$. Then since we have $|\!|\!|P^{n_0} - \pi|\!|\!|_V = \gamma < 1$, and $|\!|\!|P - \pi|\!|\!|_V \le M < \infty$ this implies that (choosing $M \ge 1$ with no loss of generality),

$$\begin{aligned} |\!|\!|P^n - \pi|\!|\!|_V &\le |\!|\!|P - \pi|\!|\!|_V^i |\!|\!|P^{n_0} - \pi|\!|\!|_V^k \\ &\le M^i \gamma^k \\ &\le M^{n_0}\gamma^{-1}(\gamma^{1/n_0})^n \end{aligned}$$

which gives the claimed geometric convergence result. □

Next we conclude the proof that V-uniform ergodicity is essentially equivalent to V solving the drift condition (V4).

Theorem 16.1.4 *Suppose that $\boldsymbol{\Phi}$ is ψ-irreducible, and that for some $V \ge 1$ there exists $r > 1$ and $R < \infty$ such that for all $n \in \mathbb{Z}_+$*

$$|\!|\!|P^n - \pi|\!|\!|_V \le Rr^{-n}. \qquad (16.13)$$

Then the drift condition (V4) holds for some V_0, where V_0 is equivalent to V in the sense that for some $c \ge 1$,

$$c^{-1}V \le V_0 \le cV. \qquad (16.14)$$

PROOF Fix $C \in \mathcal{B}^+(\mathsf{X})$ as any petite set. Then we have from (16.13) the bound

$$P^n(x, C) \ge \pi(C) - R\rho^n V(x)$$

and hence the sublevel sets of V are petite, so V is unbounded off petite sets.

From the bound

$$P^n V \le R\rho^n V + \pi(V) \qquad (16.15)$$

we see that (15.35) holds for the n-skeleton whenever $R\rho^n < 1$. Fix n with $R\rho^n < e^{-1}$, and set

$$V_0(x) := \sum_{i=0}^{n-1} \exp[i/n] P^i V.$$

We have that $V_0 > V$, and from (16.15),

$$V_0 \le e^1 nRV + n\pi(V),$$

which shows that V_0 is equivalent to V in the required sense of (16.14).

From the drift (16.15) which holds for the n-skeleton we have

$$\begin{aligned} PV_0 &= \sum_{i=1}^{n} \exp[i/n - 1/n] P^i V \\ &= \exp[-1/n] \sum_{i=1}^{n-1} \exp[i/n] P^i V + \exp[1-1/n] P^n V \\ &\le \exp[-1/n] \sum_{i=1}^{n-1} \exp[i/n] P^i V + \exp[-1/n] V + \exp[1-1/n]\pi(V) \\ &= \exp[-1/n] V_0 + \exp[1-1/n]\pi(V) \end{aligned}$$

This shows that (15.35) also holds for $\boldsymbol{\Phi}$, and hence by Lemma 15.2.8 the drift condition (V4) holds with this V_0, and some petite set C. □

Thus we have proved the equivalence of (ii) and (iv) in Theorem 16.0.1.

16.1.2 V-geometric mixing and V-uniform ergodicity

In addition to the very strong total variation norm convergence that V-uniformly ergodic chains satisfy by definition, several other ergodic theorems and mixing results may be obtained for these stochastic processes. Much of Chapter 17 will be devoted to proving that the Central Limit Theorem, the Law of the Iterated Logorithm, and an invariance principle holds for V-uniformly ergodic chains. These results are obtained by applying the ergodic theorems developed in this chapter, and by exploiting the V-geometric regularity of these chains. Here we will consider a relatively simple result which is a direct consequence of the operator norm convergence (16.2).

A stochastic process $\mathbf{X}$ taking values in X is called *strong mixing* if there exists a sequence of positive numbers $\{\delta(n) : n \ge 0\}$ tending to zero for which

$$\sup |\mathsf{E}[g(X_k)h(X_{n+k})] - \mathsf{E}[g(X_k)]\mathsf{E}[h(X_{n+k})]| \le \delta(n), \qquad n \in \mathbb{Z}_+,$$

where the supremum is taken over all $k \in \mathbb{Z}_+$, and all g and h such that $|g(x)|$, $|h(x)| \le 1$ for all $x \in \mathsf{X}$.

In the following result we show that V-uniformly ergodic chains satisfy a much stronger property. We will call $\boldsymbol{\Phi}$ *V-geometrically mixing* if there exists $R < \infty$, $\rho < 1$ such that

$$\sup |\mathsf{E}_x[g(\Phi_k)h(\Phi_{n+k})] - \mathsf{E}_x[g(\Phi_k)]\mathsf{E}_x[h(\Phi_{n+k})]| \le RV(x)\rho^n, \qquad n \in \mathbb{Z}_+,$$

where we now extend the supremum to include all $k \in \mathbb{Z}_+$, and all g and h such that $g^2(x)$, $h^2(x) \le V(x)$ for all $x \in \mathsf{X}$.

Theorem 16.1.5 *If Φ is V-uniformly ergodic then there exists $R < \infty$ and $\rho < 1$ such that for any g^2, $h^2 \le V$ and $k, n \in \mathbb{Z}_+$,*

$$|\mathsf{E}_x[g(\Phi_k)h(\Phi_{n+k})] - \mathsf{E}_x[g(\Phi_k)]\mathsf{E}_x[h(\Phi_{n+k})]| \le R\rho^n[1 + \rho^k V(x)],$$

and hence the chain Φ is V-geometrically mixing.

Proof For any $h^2 \le V$, $g^2 \le V$ let $\overline{h} = h - \pi(h)$, $\overline{g} = g - \pi(g)$. We have by $\sqrt{V}$-uniform ergodicity as in Lemma 15.2.9 that for some $R' < \infty$, $\rho < 1$,

$$\begin{aligned} |\mathsf{E}_x[\overline{h}(\Phi_k)\overline{g}(\Phi_{k+n})]| &= \left|\mathsf{E}_x\left[\overline{h}(\Phi_k)\mathsf{E}_{\Phi_k}\left[\overline{g}(\Phi_n)\right]\right]\right| \\ &\le R'\rho^n \mathsf{E}_x\left[\left|\overline{h}(\Phi_k)\right|\sqrt{V(\Phi_k)}\right]. \end{aligned}$$

Since $|\overline{h}| \le \left(1 + \int V^{\frac{1}{2}} d\pi\right) V^{\frac{1}{2}}$ we can set $R'' = R'\left(1 + \int V^{\frac{1}{2}} d\pi\right)$ and apply (15.35) to obtain the bound

$$\begin{aligned} |\mathsf{E}_x[\overline{h}(\Phi_k)\overline{g}(\Phi_{k+n})]| &\le R''\rho^n \mathsf{E}_x\left[V(\Phi_k)\right] \\ &\le R''\rho^n \left\{\frac{L}{1-\lambda} + \lambda^k V(x)\right\}. \end{aligned}$$

Assuming without loss of generality that $\rho \ge \lambda$, and using the bounds

$$\begin{aligned} |\pi(h) - \mathsf{E}_x[h(\Phi_k)]| &\le R'''\rho^k\sqrt{V(x)} \\ |\pi(g) - \mathsf{E}_x[g(\Phi_{k+n})]| &\le R'''\rho^{k+n}\sqrt{V(x)} \end{aligned}$$

gives the result for some $R < \infty$. □

It follows from Theorem 16.1.5 that if the chain is V-uniformly ergodic then for some $R_1 < \infty$,

$$|\mathsf{E}_x[\overline{h}(\Phi_k)\overline{g}(\Phi_{k+n})]| \le R_1\rho^n[1 + \rho^k V(x)], \qquad k, n \in \mathbb{Z}_+ \tag{16.16}$$

where $\overline{h} = h - \pi(h)$, $\overline{g} = g - \pi(g)$.

By integrating both sides of (16.16) over X, the initial condition x may be replaced with a finite bound for any initial distribution μ with $\mu(V) < \infty$, and a mixing condition will be satisfied for such initial conditions. In the particular case where $\mu = \pi$ we have by stationarity and finiteness of $\pi(V)$ (see Theorem 14.3.7),

$$|\mathsf{E}_\pi[\overline{h}(\Phi_k)\overline{g}(\Phi_{k+n})]| \le R_2\rho^n, \qquad k, n \in \mathbb{Z}_+. \tag{16.17}$$

for some $R_2 < \infty$; and hence the stationary version of the process satisfies a geometric mixing condition under (V4).

16.1.3 V-uniform ergodicity for regenerative models

In order to establish geometric ergodicity for specific models, we will obviously use the drift criterion (V4) to establish the required convergence. We begin by illustrating this for two regenerative models: we give many further examples later in the chapter.

For many models with some degree of spatial homogeneity, the crucial condition leading to geometric convergence involves exponential bounds on the increments of the process. Let us say that the distribution function G of a random variable is in $\mathcal{G}^+(\gamma)$ if G has a Laplace-Stieltjes transform convergent in $[0, \gamma]$: that is, if

$$\int_0^\infty e^{st} G(dt) < \infty, \qquad 0 < s \le \gamma, \tag{16.18}$$

where $\gamma > 0$.

Forward recurrence time chains Consider the forward recurrence time δ-skeleton chain V_δ^+ defined by (RT3), based on increments with spread-out distribution Γ.

Suppose that $\Gamma \in \mathcal{G}^+(\gamma)$. By choosing $V(x) = e^{\gamma x}$ we have immediately that (V4) holds for $x \in C$ with $C = [0, \delta]$, and also

$$[V(x)]^{-1} \int P(x, dy) V(y) = e^{\gamma(x-\delta)}/e^{\gamma x} = e^{-\gamma\delta} < 1, \qquad x > \delta.$$

Thus (V4) also holds on C^c, and we conclude that the chain is $e^{\gamma x}$-uniformly ergodic. Moreover, from Theorem 16.0.1 we also have that

$$\int |P^n(x, dy) e^{\gamma y} - \pi(dy) e^{\gamma y}| < e^{\gamma x} r^{-n},$$

so that the moment-generating functions of the model, and moreover all polynomial moments, converge geometrically quickly to their limits with known bounds on the state-dependent constants.

This is the same result we showed in Section 15.1.4 for the forward recurrence time chain on $\mathbb{Z}_+$; here we have used the drift conditions rather than the direct calculation of hitting times to establish geometric ergodicity.

It is obvious from its construction that for this chain the condition $\Gamma \in \mathcal{G}^+(\gamma)$ is also necessary for geometric ergodicity.

The condition for uniform ergodicity for the forward recurrence time chain is also trivial to establish, from the criterion in Theorem 16.0.2 (vi). We will only have this condition holding if Γ is of bounded range so that $\Gamma[0, c] = 1$ for some finite c; in this case we may take the state space X equal to the compact absorbing set $[0, c]$. The existence of such a compact absorbing subset is typical of many uniformly ergodic chains in practice.

Random walk on $\mathbb{R}_+$ Consider now the random walk on $[0, \infty)$, defined by (RWHL1). Suppose that the model has an increment distribution Γ such that

(a) the mean increment $\beta = \int x\, \Gamma(dx) < 0$;

(b) the distribution Γ is in $\mathcal{G}^+(\gamma)$, for some $\gamma > 0$.

Let us choose $V(x) = \exp(sx)$, where $0 < s < \gamma$ is to be selected. Then we have

$$\begin{aligned} \textstyle\int P(x, dy) \Delta V(y)/V(x) &= \textstyle\int_{-x}^\infty \Gamma(dw)[\exp(sw) - 1] \\ &\qquad + \Gamma(-\infty, -x][\exp(-sx) - 1] \\ &\le \textstyle\int_{-\infty}^\infty \Gamma(dw)[\exp(sw) - 1] \\ &\qquad + \textstyle\int_{-\infty}^{-x} \Gamma(dw)[1 - \exp(sw)]. \end{aligned} \tag{16.19}$$

But now if we let $s \downarrow 0$ then

$$s^{-1} \int_{-\infty}^{\infty} \Gamma(dw)[\exp(sw) - 1] \to \beta < 0.$$

Thus choosing s_0 sufficiently small that $\int_{-\infty}^{\infty} \Gamma(dw)[\exp(s_0 w) - 1] = \xi < 0$, and then choosing c large enough that

$$\Gamma(-\infty, -x] \leq -\xi/2, \qquad x \geq c$$

we have that (V4) holds with $C = [0, c]$. Since C is petite for this chain, the random walk is $\exp(s_0 x)$-uniformly ergodic when (a) and (b) hold.

It is then again a consequence of Theorem 16.0.1 that the moment generating function, and indeed all moments, of the chain converge geometrically quickly.

Thus we see that the behavior of the Bernoulli walk in Section 15.5 is due, essentially, to the bounded and hence exponential nature of its increment distribution.

We will show in Section 16.3 that one can generalize this result to general chains, giving conditions for geometric ergodicity in terms of exponentially decreasing "tails" of the increment distributions.

16.2 Uniform ergodicity

16.2.1 Equivalent conditions for uniform ergodicity

From the definition (16.6), a Markov chain is uniformly ergodic if $\|P^n - \pi\|_V \to 0$ as $n \to \infty$ when $V \equiv 1$. This simple observation immediately enables us to establish the first three equivalences in Theorem 16.0.2, which relate convergence properties of the chain.

Theorem 16.2.1 *The following are equivalent, without any a priori assumption of ψ-irreducibility or aperiodicity:*

(i) *Φ is uniformly ergodic.*

(ii) *there exists $\rho < 1$ and $R < \infty$ such that for all x*

$$\|P^n(x, \cdot\,) - \pi\| \leq R\rho^n.$$

(iii) *for some $n \in \mathbb{Z}_+$,*

$$\sup_{x \in \mathsf{X}} \|P^n(x, \cdot\,) - \pi(\,\cdot\,)\| < 1.$$

Proof Obviously (i) implies (iii); but from Proposition 16.1.3 we see that (iii) implies (ii), which clearly implies (i) as required. □

Note that uniform ergodicity implies, trivially, that the chain actually is π-irreducible and aperiodic, since for $\pi(A) > 0$ there exists n with $P^n(x, A) \geq \pi(A)/2$ for all x.

We next prove that (v)-(viii) of Theorem 16.0.2 are equivalent to uniform ergodicity.

Theorem 16.2.2 *The following are equivalent for a ψ-irreducible aperiodic chain:*

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

(ii) *the state space X is petite.*

(iii) *there is a petite set C with $\sup_{x \in \mathsf{X}} \mathsf{E}_x[\tau_C] < \infty$, in which case for every $A \in \mathcal{B}^+(\mathsf{X})$ we have $\sup_{x \in \mathsf{X}} \mathsf{E}_x[\tau_A] < \infty$.*

(iv) *there is a petite set C and a $\kappa > 1$ with $\sup_{x \in \mathsf{X}} \mathsf{E}_x[\kappa^{\tau_C}] < \infty$ in which case for every $A \in \mathcal{B}^+(\mathsf{X})$ we have $\sup_{x \in \mathsf{X}} \mathsf{E}_x[\kappa_A^{\tau_A}] < \infty$ for some $\kappa_A > 1$.*

(v) *there is an everywhere bounded solution V to (16.10) for some petite set C.*

PROOF Observe that the drift inequality (11.17) given in (V2) and the drift inequality (16.10) are identical for bounded V. The equivalence of (iii) and (v) is thus a consequence of Theorem 11.3.11, whilst (iv) implies (iii) trivially and Theorem 15.2.6 shows that (v) implies (iv): such connections between boundedness of τ_A and solutions of (16.10) are by now standard.

To see that (i) implies (ii), observe that if (i) holds then $\boldsymbol{\Phi}$ is π-irreducible and hence there exists a small set $A \in \mathcal{B}^+(\mathsf{X})$. Then, by (i) again, for some $n_0 \in \mathbb{Z}_+$, $\inf_{x \in \mathsf{X}} P^{n_0}(x, A) > 0$ which shows that X is small from Theorem 5.2.4.

The implication that (ii) implies (v) is equally simple. Let $V \equiv 1$, $\beta = b = \frac{1}{2}$, and $C = \mathsf{X}$. We then have

$$\Delta V = -\beta V + b \mathbb{1}_C,$$

giving a bounded solution to (16.10) as required.

Finally, when (v) holds, we immediately have uniform geometric ergodicity by Theorem 16.1.2. □

Historically, one of the most significant conditions for ergodicity of Markov chains is *Doeblin's Condition.*

> Doeblin's Condition
>
> Suppose there exists a probability measure ϕ with the property that for some m, $\varepsilon < 1, \delta > 0$
>
> $$\phi(A) > \varepsilon \Longrightarrow P^m(x, A) \geq \delta$$
>
> for every $x \in \mathsf{X}$.

From the equivalences in Theorem 16.2.1 and Theorem 16.2.2, we are now in a position to give a very simple proof of the equivalence of uniform ergodicity and this condition.

Theorem 16.2.3 *An aperiodic ψ-irreducible chain $\boldsymbol{\Phi}$ satisfies Doeblin's Condition if and only if $\boldsymbol{\Phi}$ is uniformly ergodic.*

PROOF Let C be any petite set with $\phi(C) > \varepsilon$ and consider the test function

$$V(x) = 1 + \mathbb{1}_{C^c}(x).$$

Then from Doeblin's Condition

$$\begin{aligned} P^m V(x) - V(x) = P^m(x, C^c) - \mathbb{1}_{C^c}(x) &\leq 1 - \delta - \mathbb{1}_{C^c}(x) \\ &= -\delta + \mathbb{1}_C(x) \\ &\leq -\tfrac{1}{2}\delta V(x) + \mathbb{1}_C(x). \end{aligned}$$

Hence V is a bounded solution to (16.10) for the m-skeleton, and it is thus the case that the m-skeleton and the original chain are uniformly ergodic by the contraction property of the total variation norm.

Conversely, we have from uniform ergodicity in the form (16.7) that for any $\varepsilon > 0$, if $\pi(A) \geq \varepsilon$ then

$$P^n(x, A) \geq \varepsilon - R\rho^n \geq \varepsilon/2$$

for all n large enough that $R\rho^n \leq \varepsilon/2$, and Doeblin's Condition holds with $\phi = \pi$. □

Thus we have proved the final equivalence in Theorem 16.0.2. We conclude by exhibiting the one situation where the bounds on convergence are simply calculated.

Theorem 16.2.4 *If a chain* $\boldsymbol{\Phi}$ *satisfies*

$$P^m(x, A) \geq \nu_m(A) \tag{16.20}$$

for all $x \in \mathsf{X}$ *and* $A \in \mathcal{B}(\mathsf{X})$ *then*

$$\|P^n(x, \cdot\,) - \pi\| \leq \rho^{n/m} \tag{16.21}$$

where $\rho = 1 - \nu_m(\mathsf{X})$.

PROOF This can be shown using an elegant argument based on the assumption (16.20) that the whole space is small which relies on a coupling method closely connected to the way in which the split chain is constructed.

Write (16.20) as

$$P^m(x, A) \geq (1 - \rho)\nu(A) \tag{16.22}$$

where $\nu = \nu_m/(1 - \rho)$ is a probability measure.

Assume first for simplicity that $m = 1$. Run two copies of the chain, one from the initial distribution concentrated at x and the other from the initial distribution π. At every time point either

(a) with probability $1-\rho$, choose for both chains the same next position from the distribution ν, after which they will be coupled and then can be run with identical sample paths; or

(b) with probability ρ, choose for each chain an independent position, using the distribution (as in the split chain construction) $[P(x, \cdot\,) - (1 - \rho)\nu(\,\cdot\,)]/\rho$, where x is the current position of the chain.

This is possible because of the minorization in (16.22). The marginal distributions of these chains are identical with the original distributions, for every n. If we let T denote the first time that the chains are chosen using the first option (a), then we have

$$\|P^n(x,\,\cdot\,) - \pi\| \le \mathsf{P}(T > n) \le \rho^n \tag{16.23}$$

which is (16.21).

When $m > 1$ we can use the contraction property as in Proposition 16.1.3 to give (16.21) in the general case. □

The optimal use of these many equivalent conditions for uniform ergodicity depends of course on the context of use. In practice, this last theorem, since it identifies the exact rate of convergence, is perhaps the most powerful, and certainly gives substantial impetus to identifying the actual minorization measure which renders the whole space a small set.

It can also be of importance to use these conditions in assessing when uniform convergence does not hold: for example, in the forward recurrence time chain $\mathbf{V}_\delta^+$ it is immediate from Theorem 16.2.2 (iii) that, since the mean return time to $[0, \delta]$ from x is of order x, the chain cannot be uniformly ergodic unless the state space can be reduced to a compact set.

Similar remarks apply to random walk on the half line: we see this explicitly in the simple random walk of Section 15.5, but it is a rather deeper result [47] that for general random walk on $[0, \infty)$, $\mathsf{E}_x[\tau_0] \sim cx$ so such chains are never uniformly ergodic.

16.2.2 Geometric convergence of given moments

It is instructive to note that, although the concept of uniform ergodicity is a very strong one for convergence of distributions, it need not have any implications for the convergence of moments or other unbounded functionals of the chain at a geometric rate.

This is obviously true in a trivial sense: an i.i.d. sequence Φ_n converges in a uniformly ergodic manner, regardless of whether $\mathsf{E}[\Phi_n]$ is finite or not.

But rather more subtly, we now show that it is possible for us to construct a uniformly ergodic chain with convergence rate ρ such that $\pi(f) < \infty$, so that we know $\mathsf{E}_x[f(\Phi_n)] \to \pi(f)$, but where not only does this convergence not take place at rate ρ, it actually does not take place at any geometric rate at all.

For convenience of exposition we construct this chain on a countable ladder space $\mathsf{X} = \mathbb{Z}_+ \times \mathbb{Z}_+$, even though the example is essentially one-dimensional.

Fix $\beta < 1/4$, and define for the i^{th} rung of the ladder the indices

$$\ell^m(i) := \lfloor (\frac{i-1}{i\beta})^m \rfloor, \qquad i \ge 1, m \ge 0.$$

Note that for $i = 1$ we have $\ell^m(1) = 0$ for all m, but for $i > 1$

$$(\frac{i-1}{i\beta})^{m+1} - (\frac{i-1}{i\beta})^m = (\frac{i-1}{i\beta})^m (\frac{i-1-i\beta}{i\beta}) \ge 1$$

since $(i-1-i\beta)/i\beta \ge (3i-1)/i \ge 2$. Hence from the second rung up, this sequence $\ell^m(i)$ forms a strictly monotone increasing set of states along the rung.

The transition mechanism we consider provides a chain satisfying the Doeblin Condition. We suppose P is given by

$$\begin{array}{rcll} P(i,\ell^m(i);i,\ell^{m+1}(i)) &=& \beta, & i=1,2,\ldots,\ m=1,2,\ldots \\ P(i,\ell^m(i);0,0) &=& 1-\beta, & i=1,2,\ldots,\ m=1,2,\ldots \\ P(i,k;0,0) &=& 1, & i=1,2,\ldots,k\neq\ell^m(i),\ m=1,2,\ldots \\ P(0,0;i,j) &=& \alpha_{ij}, & i,j\in\mathsf{X} \\ P(0,k;0,0) &=& 1, & k>0, \end{array} \tag{16.24}$$

where the α_{ij} are to be determined, with $\alpha_{00}>0$.

In effect this chain moves only on the states $(0,0)$ and the sequences $\ell^m(i)$, and the whole space is small with

$$P(i,k;\,\cdot\,)\geq\min(1-\beta,\alpha_{00})\delta_{00}(\,\cdot\,).$$

Thus the chain is clearly uniformly and hence geometrically ergodic.

Now consider the function f defined by $f(i,k)=k$; that is, f denotes the distance of the chain along the rung independent of the rung in question. We show that the chain is f-ergodic but not f-geometrically ergodic, under suitable choice of the distribution α_{ij}.

First note that we can calculate

$$\begin{array}{rcl} \mathsf{E}_{i,1}[\sum_0^{\tau_{0,0}-1}f(\Phi_n)] &=& (1-\beta)\sum_{n=0}^\infty\beta^n\sum_{m=0}^n\ell^m(i) \\ &\leq& (1-\beta)\sum_{n=0}^\infty\beta^n\sum_{m=0}^n(\frac{i-1}{i\beta})^m \\ &=& i; \\ \mathsf{E}_{i,\ell^m(i)}[\sum_0^{\tau_{0,0}-1}f(\Phi_n)] &\leq& (\frac{i-1}{i\beta})^m i, \qquad m=1,2,\ldots; \\ \mathsf{E}_{i,k}[\sum_0^{\tau_{0,0}-1}f(\Phi_n)] &=& k, \qquad k\neq\ell^m(i),\ m=1,2,\ldots. \end{array}$$

Now let us choose

$$\begin{array}{rcll} \alpha_{ik} &=& c2^{-i-k}, & k\neq\ell^m(i),\ m=1,2,\ldots; \\ \alpha_{ik} &=& c\sum_{m=0}^\infty 2^{-i-\ell^m(i)}, & k=1, \end{array}$$

and all other values except α_{00} as zero, and where c is chosen to ensure that the α_{ik} form a probability distribution.

With this choice we have

$$\begin{array}{rcl} \mathsf{E}_{0,0}[\sum_0^{\tau_{0,0}-1}f(\Phi_n)] &\leq& 1+\sum_{i\geq1}\sum_{k\neq\ell^m(i),m\geq0}k2^{-i-k}+\sum_{i\geq1}[\sum_{m=0}^\infty 2^{-i-\ell^m(i)}]i \\ &\leq& 1+2\sum_{i\geq1}i2^{-i}<\infty \end{array}$$

so that the chain is certainly f-ergodic by Theorem 14.0.1. However for any $r\in(1,\beta^{-1})$,

$$\begin{aligned}\mathsf{E}_{i,1}[\sum_0^{\tau_{0,0}-1} f(\Phi_n)r^n] &= (1-\beta)\sum_{n=0}^\infty \beta^n r^n \sum_{m=0}^n \ell^m(i)\\ &\geq (1-\beta)\sum_{n=0}^\infty (\beta r)^n \sum_{m=0}^n [(\tfrac{i-1}{i\beta})^m - 1]\\ &= -(\tfrac{1-\beta}{1-\beta r}) + \sum_{n=0}^\infty (\beta r)^n [\tfrac{[(i-1)/i\beta]^{n+1}-1}{[(i-1)/i\beta]-1}]\end{aligned}$$

which is infinite if

$$\beta r[\frac{i-1}{i\beta}] > 1;$$

that is, for those rungs i such that $i > r/(r-1)$. Since there is positive probability of reaching such rungs in one step from $(0,0)$ it is immediate that

$$\mathsf{E}_{0,0}[\sum_0^{\tau_{0,0}-1} f(\Phi_n)r^n] = \infty$$

for all $r > 1$, and hence from Theorem 15.4.2 for all $r > 1$

$$\sum_n r^n \|P^n(0,0;\,\cdot\,) - \pi\|_f = \infty.$$

Since $\{0,0\} \in \mathcal{B}^+(\mathsf{X})$, this implies that $\|P^n(x;\,\cdot\,) - \pi\|_f$ is not $o(\rho^n)$ for any x or any $\rho < 1$.

We have thus demonstrated that the strongest rate of convergence in the simple total variation norm may not be inherited, even by the simplest of unbounded functions; and that one really needs, when considering such functions, to use criteria such as (V4) to ensure that these functions converge geometrically.

16.2.3 Uniform ergodicity: T-chains on compact spaces

For T-chains, we have an almost trivial route to uniform ergodicity, given the results we now have available.

Theorem 16.2.5 *If $\boldsymbol{\Phi}$ is a ψ-irreducible and aperiodic T-chain, and if the state space* X *is compact, then $\boldsymbol{\Phi}$ is uniformly ergodic.*

Proof If $\boldsymbol{\Phi}$ is a ψ-irreducible T-chain, and if the state space X is compact, then it follows directly from Theorem 6.0.1 that X is petite. Applying the equivalence of (i) and (ii) given in Theorem 16.2.2 gives the result. □

One specific model, the nonlinear state space model, is also worth analyzing in more detail to show how we can identify other conditions for uniform ergodicity.

The NSS($\boldsymbol{F}$) model In a manner similar to the proof of Theorem 16.2.5 we show that the the NSS(F) model defined by (NSS1) and (NSS2) is uniformly ergodic, provided that the associated control model CM(F) is stable in the sense of Lagrange, so that in effect the state space is reduced to a compact invariant subset.

> Lagrange Stability The CM(F) model is called *Lagrange stable* if $\overline{A_+(x)}$ is compact for each $x \in \mathsf{X}$.

Typically in applications, when the CM(F) model is Lagrange stable the input sequence will be constrained to lie in a bounded subset of $\mathbb{R}^p$. We stress however that no conditions on the input are made in the general definition of Lagrange stability.

The key to analyzing the NSS(F) corresponding to a Lagrange stable control model lies in the following lemma:

Lemma 16.2.6 *Suppose that the CM(F) model is forward accessible, Lagrange stable, M-irreducible and aperiodic, and suppose that for the NSS(F) model conditions (NSS1) - (NSS3) are satisfied.*

Then for each $x \in \mathsf{X}$ the set $\overline{A_+(x)}$ is closed, absorbing, and small.

PROOF By Lagrange stability it is sufficient to show that any compact and invariant set $C \subset \mathsf{X}$ is small. This follows from Theorem 7.3.5 (ii), which implies that compact sets are small under the conditions of the lemma. □

Using Lemma 16.2.6 we now establish geometric convergence of the expectation of functions of $\boldsymbol{\Phi}$:

Theorem 16.2.7 *Suppose the NSS(F) model satisfies Conditions (NSS1)-(NSS3) and that the associated control model CM(F) is forward accessible, Lagrange stable, M-irreducible and aperiodic.*

Then a unique invariant probability π exists, and the chain restricted to the absorbing set $\overline{A_+(x)}$ is uniformly ergodic for each initial condition.

Hence also for every function $f\colon \mathsf{X} \to \mathbb{R}$ which is uniformly bounded on compact sets, and every initial condition,

$$\mathsf{E}_y[f(\Phi_k)] \to \int f\, d\pi$$

at a geometric rate.

PROOF When CM(F) is forward accessible, M-irreducible and aperiodic, we have seen in Theorem 7.3.5 that the Markov chain $\boldsymbol{\Phi}$ is ψ-irreducible and aperiodic.

The result then follows from Lemma 16.2.6: the chain restricted to $\overline{A_+(x)}$ is uniformly ergodic by Theorem 16.0.2. □

16.3 Geometric ergodicity and increment analysis

16.3.1 Strengthening ergodicity to geometric ergodicity

It is possible to give a "generic" method of establishing that (V4) holds when we have already used the test function approach to establishing simple (non-geometric) ergodicity through Theorem 13.0.1. This method builds on the specific technique for random walks, shown in Section 16.1.3 above, and is an increment-based method similar to that in Section 9.5.1.

Suppose that V is a test function for regularity. We assume that V takes on the "traditional" form due to Foster: V is finite-valued, and for some petite set C and some constant $b < \infty$, we have

$$\int P(x,dy)V(y) \leq \begin{cases} V(x)-1 & \text{for } x \in C^c; \\ b & \text{for } x \in C \end{cases} \tag{16.25}$$

Recall that $V_C(x) = \mathsf{E}_x[\sigma_C]$ is the minimal solution to (16.25) from Theorem 11.3.5.

Theorem 16.3.1 *If $\boldsymbol{\Phi}$ is a ψ-irreducible ergodic chain and V is a test function satisfying (16.25), and if P satisfies, for some $c, d < \infty$ and $\beta > 0$, and all $x \in \mathsf{X}$,*

$$\int_{V(y)\geq V(x)} P(x,dy)\exp\{\beta(V(y)-V(x))\} \leq c \tag{16.26}$$

and

$$\int_{V(y)<V(x)} P(x,dy)(V(y)-V(x))^2 \leq d \tag{16.27}$$

then $\boldsymbol{\Phi}$ is V^*-uniformly ergodic, where $V^*(y) = e^{\delta V(y)}$ for some $\delta < \beta$.

Proof For positive $\delta < \beta$ we have

$$\begin{aligned} [V^*(x)]^{-1}\int P(x,dy)V^*(y) &= \int P(x,dy)\exp\{\delta(V(y)-V(x))\} \\ &= \int P(x,dy)\Big\{1+\delta(V(y)-V(x)) \\ &\qquad +\tfrac{\delta^2}{2}(V(y)-V(x))^2\exp\{\delta\theta_x(V(y)-V(x))\}\Big\} \end{aligned} \tag{16.28}$$

for some $\theta_x \in [0,1]$, by using a second order Taylor expansion. Since V satisfies (16.25), the right hand side of (16.28) is bounded for $x \in C^c$ by

$$\begin{aligned} &1-\delta+\tfrac{\delta^2}{2}\Big\{\textstyle\int_{V(y)<V(x)} P(x,dy)(V(y)-V(x))^2 \\ &\quad +\textstyle\int_{V(y)\geq V(x)} P(x,dy)((V(y)-V(x))^2\exp\{\delta(V(y)-V(x))\}\Big\} \\ &\qquad \leq 1-\delta+\tfrac{\delta^2}{2}d+\tfrac{\delta^{2-\xi}}{2}\textstyle\int_{V(y)\geq V(x)} P(x,dy)\exp\{(\delta+\delta^{\xi/2})(V(y)-V(x))\} \\ &\qquad \leq 1-\delta+\tfrac{\delta^{2-\xi}}{2}(d+c), \end{aligned} \tag{16.29}$$

for some $\xi \in (0,1)$ such that $\delta + \delta^{\xi/2} < \beta$ by virtue of (16.26) and (16.27), and the fact that x^2 is bounded by e^x on $\mathbb{R}_+$. This proves the theorem, since we have

$$1-\delta+\frac{\delta^{2-\xi}}{2}(d+c) < 1$$

for sufficiently small $\delta > 0$, and thus (V4) holds for V^*. □

The typical example of this behavior, on which this proof is modeled, is the random walk in Section 16.1.3. In that case $V(x) = x$, and (16.26) is the requirement that $\Gamma \in \mathcal{G}^+(\gamma)$. In this case we do not actually need (16.27), which may not in fact hold.

It is often easier to verify the conditions of this theorem than to evaluate directly the existence of a test function for geometric ergodicity, as we shall see in the next section.

How necessary are the conditions of this theorem on the "tails" of the increments? By considering for example the forward recurrence time chain, we see that for some chains $\Gamma \in \mathcal{G}^+(\gamma)$ may indeed be necessary for geometric ergodicity. However, geometric tails are certainly not always necessary for geometric ergodicity: to demonstrate this simply consider any i.i.d. process, which is trivially uniformly ergodic, regardless of its "increment" structure.

It is interesting to note, however, that although they seem somewhat "proof-dependent", the uniform bounds (16.26) and (16.27) on P that we have imposed cannot be weakened in general when moving from ergodicity to geometric ergodicity.

We first show that we can ensure lack of geometric ergodicity if the drift to the right is not uniformly controlled in terms of V as in (16.26), even for a chain satisfying all our other conditions. To see this we consider a chain on $\mathbb{Z}_+$ with transition matrix given by, for each $i \in \mathbb{Z}_+$,

$$\begin{aligned} P(0,i) &= \alpha_i > 0, \\ P(i,i-1) &= \gamma_i > 0, \\ P(i,i+n) &= [1-\gamma_i][1-\beta_i]\beta_i^n, \qquad n \in \mathbb{Z}_+. \end{aligned} \tag{16.30}$$

where $\sum \alpha_i = 1$ and γ_i, β_i are less than unity for all i.

Provided $\sum i\alpha_i < \infty$ and we choose γ_i sufficiently large that

$$[1-\gamma_i]\beta_i/[1-\beta_i] - \gamma_i \le -\varepsilon$$

for some $\varepsilon > 0$, then the chain is ergodic since $V(x) = x$ satisfies (V2): this can be done if we choose, for example,

$$\gamma_i \ge \beta_i + \varepsilon[1-\beta_i].$$

And now if we choose $\beta_j \to 1$ as $j \to \infty$ we see that the chain is not geometrically ergodic: we have for any j

$$\mathsf{P}_j(\tau_0 > n) \ge [1-\gamma_j][1-\beta_j]\beta_j^n$$

so $\mathsf{P}_0(\tau_0 > n)$ does not decrease geometrically quickly, and the chain is not geometrically ergodic from Theorem 15.4.2 (or directly from Theorem 15.1.1).

In this example we have bounded variances for the left tails of the increment distributions, and exponential tails of the right increments: it is the lack of uniformity in these tails that fails along with the geometric convergence.

To show the need for (16.27), consider the chain on $\mathbb{Z}_+$ with the transition matrix (15.20) given for all $j \in \mathbb{Z}_+$ by $P(0,0) = 0$ and

$$P(0,j) = \gamma_j > 0, \qquad P(j,j) = \beta_j, \qquad P(j,0) = 1-\beta_j,$$

where $\sum_j \gamma_j = 1$. We saw in Section 15.1.4 that if $\beta_j \to 1$ as $n \to \infty$, the chain cannot be geometrically ergodic regardless of the structure of the distribution $\{\gamma_j\}$.

If we consider the minimal solution to (16.25), namely

$$V_0(j) = \mathsf{E}_j[\sigma_0] = [1-\beta_j]^{-1}, \qquad j > 0,$$

then clearly the right hand increments are uniformly bounded in relation to V for $j > 0$: but we find that

$$\sum P(i,j)(V_0(j) - V_0(i))^2 = P(i,0)[1-\beta_i]^{-2} = [1-\beta_i]^{-1} \to \infty, \qquad i \to \infty.$$

Hence (16.27) is necessary in this model for the conclusion of Theorem 16.3.1 to be valid.

16.3.2 Geometric ergodicity and the structure of π

The relationship between spatial and temporal geometric convergence in the previous section is largely a result of the spatial homogeneity we have assumed when using increment analysis.

We now show that this type of relationship extends to the invariant probability measure π also, at least in terms of the "natural" ordering of the space induced by petite sets and test functions.

Let us we write, for any function g,

$$A_{g,n}(x) = \{y : g(y) \le g(x) - n\}.$$

We say that the chain is "g-skip-free to the left" if there is some $k \in \mathbb{Z}_+$, such that for all $x \in \mathsf{X}$,

$$P(x, A_{g,k}(x)) = 0, \tag{16.31}$$

so that the chain can only move a limited amount of "distance" through the sub-level sets of g in one step. Note that such skip-free behavior precludes the Doeblin Condition if g is unbounded off petite sets, and requires a more random-walk like behavior.

Theorem 16.3.2 *Suppose that Φ is geometrically ergodic. Then there exists $\beta > 0$ such that*

$$\int \pi(dy) e^{\beta V_C(y)} < \infty \tag{16.32}$$

where $V_C(y) = \mathsf{E}_y[\sigma_C]$ for any petite set $C \in \mathcal{B}^+(\mathsf{X})$.

If Φ is g-skip-free to the left for a function g which is unbounded off petite sets, then for some $\beta' > 0$

$$\int \pi(dy) e^{\beta' g(y)} < \infty. \tag{16.33}$$

PROOF From geometric ergodicity, we have from Theorem 15.2.4 that for any petite set $C \in \mathcal{B}^+(\mathsf{X})$ there exists $r > 1$ such that $V(y) = G_C^{(r)}(y, \mathsf{X})$ satisfies (V4). It follows from Theorem 14.3.7 that $\pi(V) < \infty$. Using the interpretation (15.29) we have that

$$\infty > \pi(V) \ge \int \pi(dy) \mathsf{E}_y[r^{\sigma_C}]. \tag{16.34}$$

Now the function $f(j) = z^j$ is convex in $j \in \mathbb{Z}_+$, so that $\mathsf{E}_x[r^{\sigma_C}] \ge r^{\mathsf{E}_x[\sigma_C]}$ by Jensen's inequality. Thus we have (16.32) as desired.

Now suppose that g is such that the chain is g-skip-free to the left, and fix b so that the petite set $C = \{y : g(y) \leq b\}$ is in $\mathcal{B}^+(\mathsf{X})$. Because of the left skip-free property (16.31), for $g(x) \geq nk + b$, we have $\mathsf{P}_x(\sigma_C \leq n) = 0$ so that $\mathsf{E}_x[r^{\sigma_C}] \geq r^{(g(x)-b)/k}$.

As $\int \pi(dx)\mathsf{E}_x[r^{\sigma_C}] < \infty$ by virtue of (16.34), we have thus proved the second part of the theorem for $e^\beta = \sqrt[k]{r}$. □

This result shows two things; firstly, if we think of V_C (or equivalently $G_C(x, \mathsf{X})$) as providing a natural scaling of the space in some way, then geometrically ergodic chains do have invariant measures with geometric "tails" in this scaling.

Secondly, and in practice more usefully, we have an identifiable scaling for such tails in terms of a "skip-free" condition, which is frequently satisfied by models in queueing applications on $\mathbb{N}^n$ in particular. For example, if we embed a model at the departure times in such applications, and a limited number of customers leave each time, we get a skip-free condition holding naturally. Indeed, in all of the queueing models of the next section this condition is satisfied, so that this theorem can be applied there.

To see that geometric ergodicity and conditions on π such as (16.33) are not always linked in the given topology on the space, however, again consider any i.i.d. chain. This is always uniformly ergodic, regardless of π: the rescaling through g_C here is too trivial to be useful.

In the other direction, consider again the chain on $\mathbb{Z}_+$ with the transition matrix given for all $j \in \mathbb{Z}_+$ by

$$P(0, j) = \gamma_j, \qquad P(j, j) = \beta_j, \qquad P(j, 0) = 1 - \beta_j,$$

where $\sum_j \gamma_j = 1$: we know that if $\beta_j \to 1$ as $n \to \infty$, the chain is not geometrically ergodic. But for this chain, since we know that $\pi(j)$ is proportional to

$$\mathsf{E}_0[\text{Number of visits to } j \text{ before return to } 0]$$

we have

$$\pi(j) \propto \gamma_j[1 - \beta_j]^{-1}$$

and so for suitable choice of γ_j we can clearly ensure that the tails of π are geometric or otherwise in the given topology, regardless of the geometric ergodicity of P.

16.4 Models from queueing theory

We further illustrate the use of these theorems through the analysis of three queueing systems.

These are all models on $\mathbb{Z}_+^n$ and their analysis consists of showing that there exists $\varepsilon_1, \varepsilon_2 > 0$, such that $\varepsilon_1|i|_1 \leq V(i) \leq \varepsilon_2|i|_1$, where V is the minimal solution to (16.25) and $|i|_1$ is the ℓ_1-norm on $\mathbb{Z}_+^n$; we then find that $\boldsymbol{\Phi}$ is V^*-uniformly ergodic for $V^*(i) = e^{\delta V(i)}$, so that in particular we conclude that V^* is bounded above and below by exponential functions of $|i|_1$ for these models.

Typically in all of these examples the key extra assumption needed to ensure geometric ergodicity is a geometric tail on the distributions involved: that is, the increment distributions are in $\mathcal{G}^+(\gamma)$ for some γ. Recall that this was precisely the condition used for regenerative models in Section 16.1.3.

16.4.1 The embedded M/G/1 queue N_n

The M/G/1 queue exemplifies the steps needed to apply Theorem 16.3.1 in queueing models.

Theorem 16.4.1 *If $\boldsymbol{\Phi}$ the Markov chain N_n defined by (Q4) is ergodic, then $\boldsymbol{\Phi}$ is also geometrically ergodic provided the service time distributions are in $\mathcal{G}^+(\gamma)$ for some $\gamma > 0$.*

PROOF We have seen in Section 11.4 that $V(i) = i$ is a solution to (16.25) with $C = \{0\}$.

Let us now assume that the service time distribution $H \in \mathcal{G}^+(\gamma)$. We prove that (16.26) and (16.27) hold. Application of Theorem 16.3.1 then proves V^*-uniform ergodicity of the embedded Markov chain where $V^*(i) = e^{\delta i}$ for some $\delta > 0$.

Let a_k denote the probability of k arrivals within one service. Note that (16.27) trivially holds, since $\sum_{j \le k} P(k,j)(j-k)^2 \le a_0$. For $l \ge 0$ we have

$$P(k, k+l) = a_{l+1} = \frac{1}{(l+1)!} \int_0^\infty e^{-\lambda t} (\lambda t)^{l+1} dH(t).$$

Let $\delta > 0$, so that

$$\sum_{l \ge 0} e^{\delta(l+1)} P(k, k+l) \le \int_0^\infty \exp\{(e^\delta - 1)\lambda t\} dH(t)$$

which is assumed to be finite for $(e^\delta - 1)\lambda < \gamma$. Thus we have the result. □

16.4.2 A gated-limited polling system

We next consider a somewhat more complex multidimensional queueing model. Consider a system consisting of K infinite capacity queues and a single server.

The server visits the queues in order (hence the name polling system) and during a visit to queue k the server serves $\min(x, \ell_k)$ customers, where x is the number of customers present at queue k at the instant the server arrives there: thus ℓ_k is the "gate-limit".

To develop a Markovian representation, this system is observed at each instant the server arrives back at queue 1: the queue lengths at the respective queues are then recorded. We thus have a K-dimensional state description $\boldsymbol{\Phi}_n = \boldsymbol{\Phi}_n^k$, where $\boldsymbol{\Phi}_n^k$ stands for the number of customers in queue k at the server's n^{th} visit to queue 1.

The arrival stream at queue k is assumed to be a Poisson stream with parameter λ_k; the amount of service given to a queue k customer is drawn from a general distribution with mean μ_k^{-1}.

To make the process $\boldsymbol{\Phi}$ a Markov chain we assume that the sequence of service times to queue k are i.i.d. random variables. Moreover, the arrival streams and service times are assumed to be independent of each other.

Theorem 16.4.2 *The gated-limited polling model $\boldsymbol{\Phi}$ described above is geometrically ergodic provided*

$$1 > \rho := \sum_k \lambda_k / \mu_k \tag{16.35}$$

and the service-time distributions are in $\mathcal{G}^+(\gamma)$ for some γ.

PROOF It is straightforward to show that $\boldsymbol{\Phi}$ is ergodic for the gated-limited service discipline when (16.35) holds, by identifying a drift function that is linear in the number of customers in the respective queues: specifically $V(i) = \sum_{k=1}^{K} i_k/\mu_k$ where i is a K-dimensional vector with k^{th} component i_k, can easily be shown to satisfy (16.25).

To apply the results in this section, observe that for this embedded chain there are only finitely many different possible one-step increments , depending on whether $\boldsymbol{\Phi}_n^k$ exceeds ℓ_k or equals $x < \ell_k$. Combined with the linearity of V, we conclude that both sums

$$\{ \sum_{j:V(j)\geq V(i)} P(i,j)e^{\lambda(V(j)-V(i))} : i \in \mathsf{X}\}$$

and

$$\{ \sum_{j:V(j)<V(i)} P(i,j)(V(j)-V(i))^2 : i \in \mathsf{X}\}$$

have only finitely many non-zero elements. We must ensure that these expressions are all finite, but it is straightforward to check as in Theorem 16.4.1 that convergence of the Laplace-Stieltjes transforms of the service-time distributions in a neighborhood of 0 is sufficient to achieve this, and the theorem follows. □

16.4.3 A queue with phase-type service times

In many cases of ergodic chains there are no closed form expressions for the drift function, even though it follows from Chapter 11 that such functions exist. However, once ergodicity has been established, we do know by minimality that the function $V_C(x) = \mathsf{E}_x[\sigma_C]$ is a finite solution to (16.25). We now consider a queueing model for which we can study properties of this function without explicit calculation: this is the single server queue with phase-type service time distribution.

Jobs arrive at a service facility according to a Poisson process with parameter λ. With probability p_k any job requires k independent exponentially distributed phases of service each with mean ν. The sum of these phases is the "phase-type" service time distribution, with mean service time $\mu^{-1} = \sum_{k=1}^{\infty} kp_k/\nu$.

This process can be viewed as a continuous time Markov process on the state space

$$\mathsf{X} = \{i = (i_1, i_2) \mid i_1, i_2 \in \mathbb{Z}_+\}$$

where i_1 stands for the number of jobs in the queue and i_2 for the remaining number of phases of service the job currently in service is to receive.

We consider an approximating discrete time Markov chain, which has the following transition probabilities for $h < (\lambda+\nu)^{-1}$ and $e_1 = (1,0), e_2 = (0,1)$:

$$\begin{array}{rcl} P(0, 0+e_2) & = & \lambda p_l h, \\ P(i, i+e_1) & = & \lambda h, \;\; i_1, i_2 > 0 \\ P(i, i-e_2) & = & \nu h, \;\; i_1 > 0, i_2 > 1 \\ P(i, i-e_1+le_2) & = & \nu p_l h, \;\; i_1 > 0, i_2 = 1 \\ P(i,i) & = & 1 - \sum_{j\neq i} P(i,j). \end{array} \qquad (16.36)$$

We call this the h-approximation to the M/PH/1 queue.

Although we do not evaluate a drif criterion explicitly for this chain, we will use a coupling argument to show for $V_0(i) = \mathsf{E}_i[\sigma_0]$ that when $i \neq 0$

$$\begin{aligned} V_0(i+e_2) - V_0(i) &= c, & (16.37)\\ V_0(i+e_1) - V_0(i) &= c' := c\sum_{l=1}^{\infty} l p_l & (16.38)\end{aligned}$$

for some constant $c > 0$, so that $V_0(i) = c' i_1 + c i_2$ is thus linear in both components of the state variable for $i \neq 0$.

Theorem 16.4.3 *The h-approximation of the M/PH/1 queue as in (16.36) is geometrically ergodic whenever it is ergodic, provided the phase-distribution of the service times is in $\mathcal{G}^+(\gamma)$ for some $\gamma > 0$.*

In particular if there are a finite number of phases ergodicity is equivalent to geometric ergodicity for the h-approximation.

PROOF To develop the coupling argument, we first generate sample paths of $\boldsymbol{\Phi}$ drawing from two i.i.d. sequences $U^1 = \{U^1_n\}_n$, $U^2 = \{U^2_n\}_n$ of random variables having a uniform distribution on $(0,1]$. The first sequence generates arrivals and phase-completions, the second generates the number of phases of service that will be given to a customer starting service. The procedure is as follows. If $U^1_n \in (0, \lambda h]$ an arrival is generated in $(nh, (n+1)h]$; if $U^1_n \in (\lambda h, \lambda h + \nu h]$ a phase completion is generated, and otherwise nothing happens. Similarly, if $U^2_n \in (\sum_{l=0}^{k-1} p_l, \sum_{l=0}^{k} p_l]$ k phases will be given to the n^{th} job starting service. This stochastic process has the same probabilistic behavior as $\boldsymbol{\Phi}$.

To prove (16.37) we compare two sample paths, say $\phi^k = \{\phi^k_n\}_n$, $k = 1, 2$, with $\phi^1_1 = i$ and $\phi^2_1 = i + e_2$, generated by one realization of U^1 and U^2. Clearly $\phi^2_n = \phi^1_n + e_2$, until the first moment that ϕ^1 hits 0, say at time n^*. But then $\phi^2_{n^*} = (0,1)$. This holds for all realizations ϕ^1 and ϕ^2 and we conclude that $V_0(i+e_2) = \mathsf{E}_{i+e_2}[\sigma_0] = \mathsf{E}_i[\sigma_0] + \mathsf{E}_{e_2}[\sigma_0] = V_0(i) + c$, for $c = \mathsf{E}_{e_2}[\sigma_0]$.

If ϕ^2 starts in $i + e_1$ then $\phi^2_{n^*} = (0, l)$ with probability p_l, so that $V_0(i+e_2) = V_0(i) + \sum_l p_l \mathsf{E}_{le_2}[\sigma_0] = V_0(i) + c\sum_l p_l l$.

Hence, (16.38) and (16.37) hold, and the combination of (16.38) and (16.37) proves (16.26) if we assume that the service time distribution is in $\mathcal{G}^+(\gamma)$ for some $\gamma > 0$, again giving sufficiency of this condition for geometric ergodicity. □

16.5 Autoregressive and state space models

As we saw briefly in Section 15.5.2, models with some autoregressive character may be geometrically ergodic without the need to assume that the innovation distribution is in $\mathcal{G}^+(\gamma)$. We saw this occur for simple linear models, and for scalar bilinear models.

We now consider rather more complex versions of such models and see that the phenomenon persists, even with increasing complexity of space and structure, if there is a multiplicative constant essentially driving the movement of the chain.

16.5.1 Multidimensional RCA models

The model we consider next is a multidimensional version of the RCA model. The process of n-vector observations $\boldsymbol{\Phi}$ is generated by the Markovian system

$$\Phi_{k+1} = (A + \Gamma_{k+1})\Phi_k + W_{k+1} \tag{16.39}$$

where A is an $n \times n$ non-random matrix, $\boldsymbol{\Gamma}$ is a sequence of random $(n \times n)$ matrices, and $\mathbf{W}$ is a sequence of random p-vectors.

Such models are developed in detail in [198], and we will assume familiarity with the Kronecker product "$\otimes$" and the "vec" operations, used in detail there. In particular we use the basic identities

$$\begin{aligned} \mathrm{vec}\,(ABC) &= (C^\top \otimes A)\mathrm{vec}\,(B) \\ (A \otimes B)^\top &= (A^\top \otimes B^\top). \end{aligned} \tag{16.40}$$

To obtain a Markov chain and then establish ergodicity we assume:

Random Coefficient Autoregression

(RCA1) The sequences $\boldsymbol{\Gamma}$ and $\mathbf{W}$ are i.i.d. and also independent of each other.

(RCA2) The following expectations exist, and have the prescribed values:

$$\begin{array}{llll} \mathsf{E}[W_k] = 0 & & \mathsf{E}[W_k W_k^\top] = G & (n \times n), \\ \mathsf{E}[\Gamma_k] = 0 & (n \times n) & \mathsf{E}[\Gamma_k \otimes \Gamma_k] = C & (n^2 \times n^2), \end{array}$$

and the eigenvalues of $A \otimes A + C$ have moduli less than unity.

(RCA3) The distribution of $\binom{\Gamma_k}{W_k}$ has an everywhere positive density with respect to μ^{Leb} on $\mathbb{R}^{n^2+p}$.

Theorem 16.5.1 *If the assumptions (RCA1)-(RCA3) hold for the Markov chain defined in (16.39), then $\boldsymbol{\Phi}$ is V-uniformly ergodic, where $V(x) = |x|^2$. Thus these assumptions suffice for a second-order stationary version of $\boldsymbol{\Phi}$ to exist.*

Proof Under the assumptions of the theorem the chain is weak Feller and we can take ψ as μ^{Leb} on $\mathbb{R}^n$. Hence from Theorem 6.2.9 the chain is an irreducible T-chain, and compact subsets of the state space are petite. Aperiodicity is immediate from the density assumption (RCA3). We could also apply the techniques of Chapter 7 to conclude that $\boldsymbol{\Phi}$ is a T-chain, and this would allow us to weaken (RCA3).

To prove $|x|^2$-uniform ergodicity we will use the following two results, which are proved in [198]. Suppose that (RCA1) and (RCA2) hold, and let N be any $n \times n$ positive definite matrix.

(i) If M is defined by

$$\mathrm{vec}\,(M) = (I - A^\top \otimes A^\top - C)^{-1}\mathrm{vec}\,(N) \tag{16.41}$$

then M is also positive definite.

(ii) For any x,

$$\mathsf{E}[\Phi_k^\top (A+\Gamma_{k+1})^\top M(A+\Gamma_{k+1})\Phi_k \mid \Phi_k = x] = x^\top Mx - x^\top Nx. \tag{16.42}$$

Now let N be any positive definite $(n \times n)$-matrix and define M as in (16.41). Then with $V(x) := x^\top Mx$,

$$\begin{aligned}\mathsf{E}[V(\Phi_{k+1}) \mid \Phi_k = x] &= \mathsf{E}[\Phi_k^\top (A+\Gamma_{k+1})^\top M(A+\Gamma_{k+1})\Phi_k \mid \Phi_k = x] \\ &\quad + \mathsf{E}[W_{k+1}^\top MW_{k+1}]\end{aligned} \tag{16.43}$$

on applying (RCA1) and (RCA2).

From (16.42) we also deduce that

$$PV(x) = V(x) - x^\top Nx + \mathrm{tr}\,(VG) < \lambda V(x) + L \tag{16.44}$$

for some $\lambda < 1$ and $L < \infty$, from which we see that (V4) follows, using Lemma 15.2.8.

Finally, note that for some constant c we must have $c^{-1}|x|^2 \le V(x) \le c|x|^2$ and the result is proved. □

16.5.2 Adaptive control models

In this section we return to the simple adaptive control model defined by (SAC1)–(SAC2) whose associated Markovian state process $\boldsymbol{\Phi}$ is defined by (2.24).

We showed in Proposition 12.5.2 that the distributions of the state process $\boldsymbol{\Phi}$ for this adaptive control model are tight whenever stability in the mean square sense is possible, for a certain class of initial distributions. Here we refine the stability proof to obtain V-uniform ergodicity for the model.

Once these stability results are obtained we can further analyze the system equations and find that we can bound the steady state variance of the output process by the mean square tracking error $\mathsf{E}_\pi[|\tilde\theta_0|^2]$ and the disturbance intensity σ_w^2.

Let $y\colon \mathsf{X} \to \mathbb{R}$, $\tilde\theta\colon \mathsf{X} \to \mathbb{R}$, $\Sigma\colon \mathsf{X} \to \mathbb{R}$ denote the coordinate variables on X so that

$$Y_k = y(\Phi_k) \quad \tilde\theta_k = \tilde\theta(\Phi_k) \quad \Sigma_k = \Sigma(\Phi_k) \qquad k \in \mathbb{Z}_+,$$

and define the norm-like function V on X by

$$V(y, \tilde\theta, \Sigma) = \tilde\theta^4 + \varepsilon_0 \tilde\theta^2 y^2 + \varepsilon_0^2 y^2 \tag{16.45}$$

where $\varepsilon_0 > 0$ is a small constant which will be specified below.

Letting P denote the Markov transition function for $\boldsymbol{\Phi}$ we have by (2.22),

$$Py^2 = \tilde\theta^2 y^2 + \sigma_w^2. \tag{16.46}$$

This is far from (V4), but applying the operator P to the function $\tilde\theta^2 y^2$ gives

$$\begin{aligned}P\tilde\theta^2 y^2 &= \mathsf{E}\Big[\Big(\frac{\alpha\sigma_0^2\tilde\theta - \alpha\Sigma y W_1}{\sigma_0^2 + \Sigma y^2} + Z_1\Big)^2 (\tilde\theta y + W_1)^2\Big] \\ &= \sigma_z^2\tilde\theta^2 y^2 + \sigma_z^2\sigma_w^2 \\ &\quad + \Big(\frac{\alpha}{\sigma_0^2 + \Sigma y^2}\Big)^2 \mathsf{E}[(\sigma_0^2\tilde\theta - \Sigma y W_1)^2(\tilde\theta y + W_1)^2]\end{aligned}$$

and hence we may find a constant $K_1 < \infty$ such that

$$P\tilde{\theta}^2 y^2 \leq \sigma_z^2 \tilde{\theta}^2 y^2 + K_1(\tilde{\theta}^4 + \tilde{\theta}^2 + 1). \tag{16.47}$$

From (2.21) it is easy to show that for some constant $K_2 > 0$

$$P\tilde{\theta}^4 \leq \alpha^4 \tilde{\theta}^4 + K_2(\tilde{\theta}^2 + 1). \tag{16.48}$$

When $\sigma_z^2 < 1$ we combine equations (16.46-16.48) to find, for any $1 > \rho > \max(\sigma_z^2, \alpha^4)$, constants $R < \infty$ and $\varepsilon_0 > 0$ such that with V defined in (16.45), $PV \leq \rho V + R$. Applying Theorem 16.1.2 and Lemma 15.2.8 we have proved

Proposition 16.5.2 *The Markov chain $\boldsymbol{\Phi}$ is V-uniformly ergodic whenever $\sigma_z^2 < 1$, with V given by (16.45); and for all initial conditions $x \in \mathsf{X}$, as $k \to \infty$,*

$$\mathsf{E}_x[Y_k^2] \to \int y^2 \, d\pi \tag{16.49}$$

at a geometric rate. □

Hence the performance of the closed loop system is characterized by the unique invariant probability π.

From ergodicity of the model it can be shown that in steady state $\tilde{\theta}_k = \theta_k - \mathsf{E}[\theta_k \mid Y_0, \dots, Y_k]$, and $\Sigma_k = \mathsf{E}[\tilde{\theta}_k^2 \mid Y_0, \dots, Y_k]$. Using these identities we now obtain bounds on performance of the closed loop system by integrating the system equations with respect to the invariant measure.

Taking expectations in (2.22) and (2.23) under the probability P_π gives

$$\begin{aligned} \mathsf{E}_\pi[Y_0^2] &= \mathsf{E}_\pi[\Sigma_0 Y_0^2] + \sigma_w^2 \\ \sigma_z^2 \mathsf{E}_\pi[Y_0^2] &= \mathsf{E}_\pi[\Sigma_0 Y_0^2] - \alpha^2 \sigma_w^2 \mathsf{E}_\pi[\Sigma_0]. \end{aligned}$$

Hence, by subtraction, and using the identity $\mathsf{E}_\pi[|\tilde{\theta}_0|^2] = \mathsf{E}_\pi[\Sigma_0]$, we can evaluate the limit (16.49) as

$$\mathsf{E}_\pi[Y_0^2] = \frac{\sigma_w^2}{1 - \sigma_z^2}(1 + \alpha^2 \mathsf{E}_\pi[|\tilde{\theta}_0|^2]) \tag{16.50}$$

This shows precisely how the steady state performance is related to the disturbance intensity σ_w^2, the parameter variation intensity σ_z^2, and the mean square parameter estimation error $\mathsf{E}_\pi[|\tilde{\theta}_0|^2]$.

Using obvious bounds on $\mathsf{E}_\pi[\Sigma_0]$ we obtain the following bounds on the steady state performance in terms of the system parameters only:

$$\frac{\sigma_w^2}{1 - \sigma_z^2}(1 + \alpha^2 \sigma_z^2) \leq \mathsf{E}_\pi[Y_0^2] \leq \frac{\sigma_w^2}{1 - \sigma_z^2}(1 + \frac{\alpha^2 \sigma_z^2}{1 - \alpha^2}).$$

If it were possible to directly observe θ_{k-1} at time k then the optimal performance would be

$$\mathsf{E}_\pi[Y_0^2] = \frac{\sigma_w^2}{1 - \sigma_z^2}.$$

This shows that the lower bound in the previous chain of inequalities is non-trivial.

The performance of the closed loop system is illustrated in Chapter 2.

A sample path of the output $\mathbf{Y}$ of the controlled system is given in Figure 2.8, which is comparable to the noise sample path illustrated in Figure 2.7. To see how

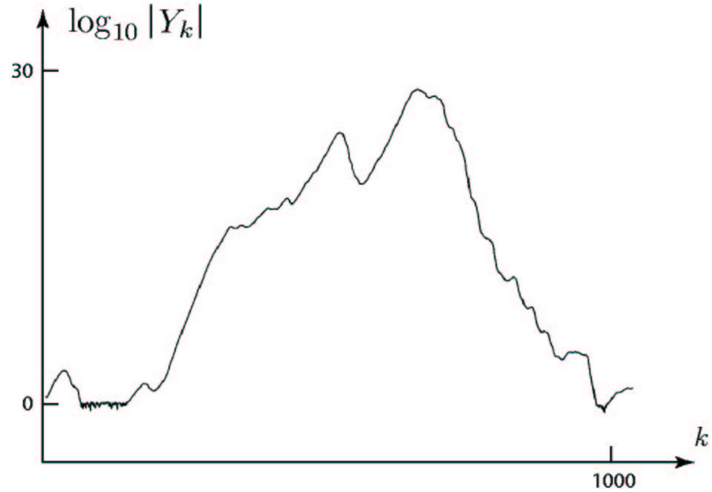

Fig. 16.1. The output of the simple adaptive control model when the control U_k is set equal to zero. The resulting process is equivalent to the dependent parameter bilinear model with $\alpha = 0.99$, $W_k \sim N(0, 0.01)$ and $Z_k \sim N(0, 0.04)$

this compares to the control-free system, a simulation of the simple adaptive control model with the control value U_k set equal to zero for all k is given in Figure 16.1. The resulting process $\binom{\boldsymbol{\theta}}{\mathbf{Y}}$ becomes a version of the dependent parameter bilinear model. Even though we will see in Chapter 17 that this process is bounded in probability, the sample paths fluctuate wildly, with the output process $\mathbf{Y}$ quickly exceeding 10^{100} in this simulation.

16.6 Commentary

This chapter brings together some of the oldest and some of the newest ergodic theorems for Markov chains.

Initial results on uniform ergodicity for countable chains under, essentially, Doeblin's Condition date to Markov [162]: transition matrices with a column bounded from zero are often called *Markov matrices*. For general state space chains use of the condition of Doeblin is in [65]. These ideas are strengthened in Doob [68], whose introduction and elucidation of Doeblin's condition as Hypothesis D (p. 192 of [68]) still guides the analysis of many models and many applications, especially on compact spaces.

Other areas of study of uniformly ergodic (sometimes called strongly ergodic, or quasi-compact) chains have a long history, much of it initiated by Yosida and Kakutani [286] who considered the equivalence of (iii) and (v) in Theorem 16.0.2, as did Doob [68]. Somewhat surprisingly, even for countable spaces the hitting time criterion of Theorem 16.2.2 for uniformly ergodic chains appears to be as recent as the work of Huang and Isaacson [100], with general-space extensions in Bonsdorff [26]; the obvious value of a bounded drift function is developed in Isaacson and Tweedie

[104] in the countable space case. Nummelin ([202], Chapters 5.6 and 6.6) gives a discussion of much of this material.

There is a large subsequent body of theory for quasi-compact chains, exploiting operator-theoretic approaches. Revuz ([223], Chapter 6) has a thorough discussion of uniformly ergodic chains and associated quasi-compact operators when the chain is not irreducible. He shows that in this case there is essentially a finite decomposition into recurrent parts of the space: this is beyond the scope of our work here.

We noted in Theorem 16.2.5 that uniform ergodicity results take on a particularly elegant form when we are dealing with irreducible T-chains: this is first derived in a different way in [269]. It is worth noting that for reducible T-chains there is an appealing structure related to the quasi-compactness above. It is shown by Tuominen and Tweedie [269] that, even for chains which are not necessarily irreducible, if the space is compact then for any T-chain there is also a finite decomposition

$$\mathsf{X} = \bigcup_{k=0}^{n} H_k \cup E$$

where the H_i are disjoint absorbing sets and $\boldsymbol{\Phi}$ restricted to any H_k is uniformly ergodic, and E is uniformly transient.

The introduction to uniform ergodicity that we give here appears brief given the history of such theory, but this is a largely a consequence of the fact that we have built up, for ψ-irreducible chains, a substantial set of tools which makes the approach to this class of chains relatively simple.

Much of this simplicity lies in the ability to exploit the norm $\| \cdot \|_V$. This is a very new approach. Although Kartashov [124, 125] has some initial steps in developing a theory of general space chains using the norm $\| \cdot \|_V$, he does not link his results to the use of drift conditions, and the appearance of V-uniform results are due largely to recent observations of Hordijk and Spieksma [252, 99] in the countable space case.

Their methods are substantially different from the general state space version we use, which builds on Chapter 15: the general space version was first developed in [178] for strongly aperiodic chains. This approach shows that for V-uniformly ergodic chains, it is in fact possible to apply the same quasi-compact operator theory that has been exploited for uniformly ergodic chains, at least within the context of the space L_V^∞. This is far from obvious: it is interesting to note Kendall himself ([131], p 183) saying that " ... the theory of quasi-compact operators is completely useless" in dealing with geometric ergodicity, whilst Vere-Jones [284] found substantial difficulty in relating standard operator theory to geometric ergodicity. This appears to be an area where reasonable further advances may be expected in the theory of Markov chains.

It is shown in Athreya and Pantula [14] that an ergodic chain is always strong mixing. The extension given in Section 16.1.2 for V-uniformly ergodic chains was proved for bounded functions in [64], and the version given in Theorem 16.1.5 is essentially taken from Meyn and Tweedie [178].

Verifying the V-uniform ergodicity properties is usually done through test functions and drift conditions, as we have seen. Uniform ergodicity is generally either a trivial or a more difficult property to verify in applications. Typically one must either take the state space of the chain to be compact (or essentially compact), or be able to apply the Doeblin or small set conditions, in order to gain uniform ergodicity. The identification of the rate of convergence in this last case is a powerful incentive to use such an approach. The delightful proof in Theorem 16.2.4 is due to Rosenthal [231], following the strong stopping time results of Aldous and Diaconis [2, 62], although the result itself is inherent in Theorem 6.15 of Nummelin [202]. An application of this result to Markov chain Monte Carlo methods is given by Tierney [264].

However, as we have shown, V-uniform ergodicity can often be obtained for some V under much more readily obtainable conditions, such as a geometric tail for any i.i.d. random variables generating the process. This is true for queues, general storage models, and other random-walk related models, as the application of the increment analysis of Section 16.3 shows. Such chains were investigated in detail by Vere-Jones [281] and Miller [185].

The results given in Section 16.3 and Section 16.3.2 are new in the case of general X, but are based on a similar approach for countable spaces in Spieksma and Tweedie [253], which also contains a partial converse to Theorem 16.3.2. There are some precursors to these conditions: one obvious way of ensuring that P has the characteristics in (16.26) and (16.27) is to require that the increments from any state are of bounded range, with the range allowed depending on V, so that for some b

$$|V(j) - V(k)| \geq b \Rightarrow P(k,j) = 0 : \tag{16.51}$$

and in [159] it is shown that under the bounded range condition (16.51) an ergodic chain is geometrically ergodic.

A detailed description of the polling system we consider here can be found in [3]. Note that in [3] the system is modeled slightly differently, with arrivals of the server at each gate defining the times of the embedded process. The coupling construction used to analyze the h-approximation to the phase-service model is based on [236] and clearly is ideal for our type of argument. Further examples are given in [253].

For the adaptive control and linear models, as we have stressed, V-uniform ergodicity is often actually equivalent to simple ergodicity: the examples in this chapter are chosen to illustrate this. The analysis of the bilinear and the vector RCA model given here is taken from Feigin and Tweedie [74]; the former had been previously analyzed by Tong [266]. In a more traditional approach to RCA models through time series methods, Nicholls and Quinn [198] also find (RCA2) appropriate when establishing conditions for strict stationarity of $\boldsymbol{\Phi}$, and also when treating asymptotic results of estimators.

The adaptive model was introduced in [165] and a stability analysis appeared in [172] where the performance bound (16.50) was obtained. Related results appeared in [251, 91, 171, 83]. The stability of the multidimensional adaptive control model was only recently resolved in Rayadurgam et al [221].