# 12

## Invariance and Tightness

In one of our heuristic descriptions of stability, in Section 1.3, we outlined a picture of a chain settling down to a stable regime independent of its initial starting point: we will show in Part III that positive Harris chains do precisely this, and one role of $\pi$ is to describe the final stochastic regime of the chain, as we have seen.

It is equally possible to approach the problem from the other end: if we have a limiting measure for $P^n$, then it may well generate a stationary measure for the chain. We saw this described briefly in (10.5): and our main goal now is to consider chains on topological spaces which do not necessarily enjoy the property of $\psi$-irreducibility, and to show how we can construct invariant measures for such chains through such limiting arguments, rather than through regenerative and splitting techniques.

We will develop the consequences of the following slightly extended form of boundedness in probability, introduced in Chapter 6.

Tightness and Boundedness in Probability on Average

A sequence of probabilities $\{\mu_k : k \in \mathbb{Z}_+\}$ is called *tight* if for each $\varepsilon > 0$, there exists a compact subset $C \subset \mathsf{X}$ such that

$$\liminf_{k\to\infty} \mu_k(C) \geq 1 - \varepsilon. \qquad (12.1)$$

The chain $\boldsymbol{\Phi}$ will be called *bounded in probability on average* if for each initial condition $x \in \mathsf{X}$ the sequence $\{\overline{P}_k(x,\,\cdot\,) : k \in \mathbb{Z}_+\}$ is tight, where we define

$$\overline{P}_k(x,\,\cdot\,) := \frac{1}{k}\sum_{i=1}^{k} P^i(x,\,\cdot\,). \qquad (12.2)$$

We have the following highlights of the consequences of these definitions.

**Theorem 12.0.1** **(i)** *If $\boldsymbol{\Phi}$ is a weak Feller chain which is bounded in probability on average then there exists at least one invariant probability measure.*

**(ii)** *If $\boldsymbol{\Phi}$ is an e-chain which is bounded in probability on average, then there exists a weak Feller transition function $\Pi$ such that for each $x$ the measure $\Pi(x,\cdot)$ is invariant, and*

$$\overline{P}_n(x,f) \to \Pi(x,f), \qquad \text{as } n \to \infty,$$

*for all bounded continuous functions $f$, and all initial conditions $x \in \mathsf{X}$.*

Proof We prove (i) in Theorem 12.1.2, together with a number of consequents for weak Feller chains. The proof of (ii) essentially occupies Section 12.4, and is concluded in Theorem 12.4.1. □

We will see that for Feller chains, and even more powerfully for e-chains, this approach based upon tightness and weak convergence of probability measures provides a quite different method for constructing an invariant probability measure. This is exemplified by the linear model construction which we have seen in Section 10.5.4.

From such constructions we will show in Section 12.4 that (V2) implies a form of positivity for a Feller chain. In particular, for e-chains, if (V2) holds for a compact set $C$ and an everywhere finite function $V$ then the chain is bounded in probability on average, so that there is a collection of invariant measures as in Theorem 12.0.1 (ii).

In this chapter we also develop a class of kernels, introduced by Neveu in [196], which extend the definition of the kernels $U_A$. This involves extending the definition of a stopping time to randomized stopping times. These operators have very considerable intuitive appeal and demonstrate one way in which the results of Section 10.4 can be applied to non-irreducible chains.

Using this approach, we will also show that (V1) gives a criterion for the existence of a $\sigma$-finite invariant measure for a Feller chain.

## 12.1 Chains bounded in probability

### 12.1.1 Weak and vague convergence

It is easy to see that for any chain, being bounded in probability on average is a stronger condition than being non-evanescent.

**Proposition 12.1.1** *If $\boldsymbol{\Phi}$ is bounded in probability on average then it is non-evanescent.*

Proof We obviously have

$$\mathsf{P}_x\{\bigcup_{j=n}^{\infty} \mathbb{1}(\Phi_j \in C)\} \geq P^n(x,C); \tag{12.3}$$

if $\boldsymbol{\Phi}$ is evanescent then for some $x$ there is an $\varepsilon > 0$ such that for every compact $C$,

$$\limsup_{n\to\infty} \mathsf{P}_x\{\bigcup_{j=n}^{\infty} \mathbb{1}(\Phi_j \in C)\} \leq 1 - \varepsilon$$

and so the chain is not bounded in probability on average. □

The consequences of an assumption of tightness are well-known (see Billingsley [24]): essentially, tightness ensures that we can take weak limits (possibly through a subsequence) of the distributions $\{\overline{P}_k(x,\,\cdot\,) : k \in \mathbb{Z}_+\}$ and the limit will then be a probability measure. In many instances we may apply Fatou's Lemma to prove that this limit is subinvariant for $\boldsymbol{\Phi}$; and since it is a probability measure it is in fact invariant.

We will then have, typically, that the convergence to the stationary measure (when it occurs) is in the weak topology on the space of all probability measures on $\mathcal{B}(\mathsf{X})$ as defined in Section D.5.

### 12.1.2 Feller chains and invariant probability measures

For weak Feller chains, boundedness in probability gives an effective approach to finding an invariant measure for the chain, even without irreducibility.

We begin with a general existence result which gives necessary and sufficient conditions for the existence of an invariant probability. From this we will find that the test function approach developed in Chapter 11 may be applied again, this time to establish the existence of an invariant probability measure for a Feller Markov chain.

Recall that the geometrically sampled Markov transition function, or resolvent, $K_{a_\varepsilon}$ is defined for $\varepsilon < 1$ as $K_{a_\varepsilon} = (1-\varepsilon)\sum_{k=0}^{\infty} \varepsilon^k P^k$

**Theorem 12.1.2** *Suppose that $\boldsymbol{\Phi}$ is a Feller Markov chain. Then*

**(i)** *If an invariant probability does not exist then for any compact set $C \subset \mathsf{X}$,*

$$\overline{P}_n(x,C) \to 0 \qquad \text{as } n \to \infty \tag{12.4}$$
$$K_{a_\varepsilon}(x,C) \to 0 \qquad \text{as } \varepsilon \uparrow 1 \tag{12.5}$$

*uniformly in $x \in \mathsf{X}$.*

**(ii)** *If $\boldsymbol{\Phi}$ is bounded in probability on average then it admits at least one invariant probability.*

PROOF We prove only (12.4), since the proof of (12.5) is essentially identical. The proof is by contradiction: we assume that no invariant probability exists, and that (12.4) does not hold.

Fix $f \in \mathcal{C}_c(\mathsf{X})$ such that $f \geq 0$, and fix $\delta > 0$. Define the open sets $\{A_k : k \in \mathbb{Z}_+\}$ by

$$A_k = \Big\{x \in \mathsf{X} : \overline{P}_k f > \delta\Big\}.$$

If (12.4) does not hold then for some such $f$ there exists $\delta > 0$ and a subsequence $\{N_i : i \in \mathbb{Z}_+\}$ of $\mathbb{Z}_+$ with $A_{N_i} \neq \emptyset$ for all $i$. Let $x_i \in A_{N_i}$ for each $i$, and define

$$\lambda_i := \overline{P}_{N_i}(x_i,\,\cdot\,)$$

We see from Proposition D.5.6 that the set of sub-probabilities is sequentially compact with respect to vague convergence. Let $\lambda_\infty$ be any vague limit point: $\lambda_{n_i} \xrightarrow{\mathrm{v}} \lambda_\infty$ for some subsequence $\{n_i : i \in \mathbb{Z}_+\}$ of $\mathbb{Z}_+$. The sub-probability $\lambda_\infty \neq 0$ because, by the definition of vague convergence, and since $x_i \in A_{N_i}$,

$$
\begin{aligned}
\int f\, d\lambda_\infty &\geq \liminf_{i\to\infty} \int f\, d\lambda_i \\
&= \liminf_{i\to\infty} \overline{P}_{N_i}(x_i, f) \\
&\geq \delta > 0. 
\end{aligned}
\tag{12.6}
$$

But now $\lambda_\infty$ is a non-trivial invariant measure. For, letting $g \in \mathcal{C}_c(\mathsf{X})$ satisfy $g \geq 0$, we have by continuity of $Pg$ and (D.6),

$$
\begin{aligned}
\int g\, d\lambda_\infty &= \lim_{i\to\infty} \overline{P}_{N_{n_i}}(x_{n_i}, g) \\
&= \lim_{i\to\infty}[\overline{P}_{N_{n_i}}(x_{n_i}, g) + N_i^{-1}(P^{N_{n_i}+1}(x_{n_i}, g) - Pg)] \\
&= \lim_{i\to\infty} \overline{P}_{N_{n_i}}(x_{n_i}, Pg) \\
&\geq \int (Pg)\, d\lambda_\infty
\end{aligned}
\tag{12.7}
$$

By regularity of finite measures on $\mathcal{B}(\mathsf{X})$ (cf Theorem D.3.2) this implies that $\lambda_\infty \geq \lambda_\infty P$, which is only possible if $\lambda_\infty = \lambda_\infty P$. Since we have assumed that no invariant probability exists it follows that $\lambda_\infty = 0$, which contradicts (12.6). Thus we have that $A_k = \emptyset$ for sufficiently large $k$.

To prove (ii), let $\boldsymbol{\Phi}$ be bounded in probability on average. Since we can find $\varepsilon > 0$, $x \in \mathsf{X}$ and a compact set $C$ such that $\overline{P}^j(x, C) > 1 - \varepsilon$ for all sufficiently large $j$ by definition, (12.4) fails and so the chain admits an invariant probability. □

The following corollary easily follows: notice that the condition (12.8) is weaker than the obvious condition of Lemma D.5.3 for boundedness in probability on average.

**Proposition 12.1.3** *Suppose that the Markov chain $\boldsymbol{\Phi}$ has the Feller property, and that a norm-like function $V$ exists such that for some initial condition $x \in \mathsf{X}$,*

$$
\liminf_{k\to\infty} \mathsf{E}_x[V(\Phi_k)] < \infty. \tag{12.8}
$$

*Then an invariant probability exists.* □

These results require minimal assumptions on the chain. They do have two drawbacks in practice.

Firstly, there is no guarantee that the invariant probability is unique. Currently, known conditions for uniqueness involve the assumption that the chain is $\psi$-irreducible. This immediately puts us in the domain of Chapter 10, and if the measure $\psi$ has an open set in its support, then in fact we have the full T-chain structure immediately available, and so we would avoid the weak convergence route.

Secondly, and essentially as a consequence of the lack of uniqueness of the invariant measure $\pi$, we do not generally have guaranteed that

$$
P^n(x, \cdot\,) \xrightarrow{\mathrm{w}} \pi.
$$

However, we do have the result

**Proposition 12.1.4** *Suppose that the Markov chain $\boldsymbol{\Phi}$ has the Feller property, and is bounded in probability on average.*

*If the invariant measure $\pi$ is unique then for every $x$*

$$
\overline{P}_n(x, \cdot\,) \xrightarrow{\mathrm{w}} \pi. \tag{12.9}
$$

PROOF Since for every subsequence $\{n_k\}$ the set of probabilities $\{\overline{P}_{n_k}(x, \cdot)\}$ is sequentially compact in the weak topology, then as in the proof of Theorem 12.1.2, from boundedness in probability we have that there is a further subsequence converging weakly to a non-trivial limit which is invariant for $P$. Since all these limits coincide by the uniqueness assumption on $\pi$ we must have (12.9). □

Recall that in Proposition 6.4.2 we came to a similar conclusion. In that result, convergence of the distributions to a unique invariant probability, in a manner similar to (12.9), is given as a condition under which a Feller chain $\boldsymbol{\Phi}$ is an e-chain.

## 12.2 Generalized sampling and invariant measures

In this section we generalize the idea of sampled chains in order to develop another approach to the existence of invariant measures for $\boldsymbol{\Phi}$. This relies on an identity called the resolvent equation for the kernels $U_B$, $B \in \mathcal{B}(\mathsf{X})$. The idea of the generalized resolvent identity is taken from the theory of continuous time processes, and we shall see that even in discrete time it unifies several concepts which we have used already, and which we shall use in this chapter to give a different construction method for $\sigma$-finite invariant measures for a Feller chain, even without boundedness in probability.

To state the resolvent equation in full generality we introduce randomized first entrance times. These include as special cases the ordinary first entrance time $\tau_A$, and also random times which are completely independent of the process: the former have of course been used extensively in results such as the identification of the structure of the unique invariant measure for $\psi$-irreducible chains, whilst the latter give us the sampled chains with kernel $K_{a_\varepsilon}$.

The more general version involves a function $h$ which will usually be continuous with compact support when the chain is on a topological space, although it need not always be so.

Let $0 \le h \le 1$ be a function on $\mathsf{X}$. The random time $\tau_h$ which we associate with the function $h$ will have the property that $\mathsf{P}_x\{\tau_h \ge 1\} = 1$, and for any initial condition $x \in \mathsf{X}$ and any time $k \ge 1$,

$$\mathsf{P}_x\{\tau_h = k \mid \tau_h \ge k, \mathcal{F}^{\Phi}_\infty\} = h(\Phi_k). \tag{12.10}$$

A probabilistic interpretation of this equation is that at each time $k \ge 1$ a weighted coin is flipped with the probability of heads equal to $h(\Phi_k)$. At the first instance $k$ that a head is finally achieved we set $\tau_h = k$. Hence we must have, for any $k \ge 1$,

$$\mathsf{P}_x\{\tau_h = k \mid \mathcal{F}^{\Phi}_\infty\} = \prod_{i=1}^{k-1}(1 - h(\Phi_i))h(\Phi_k) \tag{12.11}$$

$$\mathsf{P}_x\{\tau_h \ge k \mid \mathcal{F}^{\Phi}_\infty\} = \prod_{i=1}^{k-1}(1 - h(\Phi_i)) \tag{12.12}$$

where the product is interpreted as one when $k = 1$.

For example, if $h = \mathbb{1}_B$ then we see that $\tau_h = \tau_B$. If $h = \frac{1}{2}\mathbb{1}_B$ then a fair coin is flipped on each visit to $B$, so that $\Phi_{\tau_h} \in B$, but with probability one half, the random time $\tau_h$ will be greater then $\tau_B$.

Note that this is very similar to the Athreya-Ney randomized stopping time construction of an atom, mentioned in Section 5.1.3.

By enlarging the probability space on which $\boldsymbol{\Phi}$ is defined, and adjoining an i.i.d. process $\mathbf{Y} = \{Y_k, k \in \mathbb{Z}_+\}$ to $\boldsymbol{\Phi}$, we now show that we can explicitly construct the random time $\tau_h$ so that it is an ordinary stopping time for the bivariate chain

$$\Psi_k = \begin{pmatrix} \Phi_k \\ Y_k \end{pmatrix}, \qquad k \in \mathbb{Z}_+.$$

Suppose that $\mathbf{Y}$ is i.i.d. and independent of $\boldsymbol{\Phi}$, and that each $Y_k$ has distribution $\mathbf{u}$, where $\mathbf{u}$ denotes the uniform distribution on $[0,1]$. Then for any sets $A \in \mathcal{B}(\mathsf{X})$, $B \in \mathcal{B}([0,1])$,

$$\mathsf{P}_x\{\Psi_1 \in A \times B \mid \Phi_0 = x, Y_0 = u\} = P(x, A)\mathbf{u}(B)$$

With this transition probability, $\boldsymbol{\Psi}$ is a Markov chain whose state space is equal to $\mathsf{Y} = \mathsf{X} \times [0,1]$.

Let $A_h \in \mathcal{B}(\mathsf{Y})$ denote the set

$$A_h = \{(x,u) \in \mathsf{Y} : h(x) \geq u\}$$

and define the random time $\tau_h = \min(k \geq 1 : \Psi_k \in A_h)$. Then $\tau_h$ is a stopping time for the bivariate chain.

We see at once from the definition and the fact that $Y_k$ is independent of $(\boldsymbol{\Phi}, Y_1, \ldots, Y_{k-1})$ that $\tau_h$ satisfies (12.10). For given any $k \geq 1$,

$$\begin{aligned} \mathsf{P}_x\{\tau_h = k \mid \tau_h \geq k, \mathcal{F}_\infty^\Phi\} &= \mathsf{P}_x\{h(\Phi_k) \geq Y_k \mid \tau_h \geq k, \mathcal{F}_\infty^\Phi\} \\ &= \mathsf{P}_x\{h(\Phi_k) \geq Y_k \mid \mathcal{F}_\infty^\Phi\} \\ &= h(\Phi_k), \end{aligned}$$

where in the second equality we used the fact that the event $\{\tau_h \geq k\}$ is measurable with respect to $\{\boldsymbol{\Phi}, Y_1, \ldots, Y_{k-1}\}$, and in the final equality we used independence of $\mathbf{Y}$ and $\boldsymbol{\Phi}$.

Now define the kernel $U_h$ on $\mathsf{X} \times \mathcal{B}(\mathsf{X})$ by

$$U_h(x, B) = \mathsf{E}_x\Big[\sum_{k=1}^{\tau_h} \mathbb{1}_B(\Phi_k)\Big]. \tag{12.13}$$

where the expectation is understood to be on the enlarged probability space. We have

$$U_h(x,B) = \sum_{k=1}^{\infty} \mathsf{E}_x[\mathbb{1}_B(\Phi_k)\mathbb{1}\{\tau_h \geq k\}]$$

and hence from (12.12)

$$U_h(x,B) = \sum_{k=0}^{\infty} P(I_{1-h}P)^k\,(x,B) \tag{12.14}$$

where $I_{1-h}$ denotes the kernel which gives multiplication by $1-h$. This final expression for $U_h$ defines this kernel independently of the bivariate chain.

In the special cases $h \equiv 0$, $h = \mathbb{1}_B$, and $h \equiv 1$ we have, respectively,

$$U_h = U, \qquad U_h = U_B, \qquad U_h = P.$$

When $h = \frac{1}{2}$ so that $\tau_h$ is completely independent of $\boldsymbol{\Phi}$ we have

$$U_{\frac{1}{2}} = \sum_{k=1}^{\infty} (\tfrac{1}{2})^{k-1} P^k = K_{a_{\frac{1}{2}}}.$$

For general functions $h$, the expression (12.14) defining $U_h$ involves only the transition function $P$ for $\boldsymbol{\Phi}$ and hence allows us to drop the bivariate chain if we are only interested in properties of the kernel $U_h$. However the existence of the bivariate chain and the construction of $\tau_h$ allows a transparent proof of the following resolvent equation.

**Theorem 12.2.1 (Resolvent Equation)** *Let $h \le 1$ and $g \le 1$ be two functions on* $\mathsf{X}$ *with $h \ge g$. Then the resolvent equation holds:*

$$U_g = U_h + U_h I_{h-g} U_g = U_h + U_g I_{h-g} U_h.$$

PROOF To prove the theorem we will consider the bivariate chain $\boldsymbol{\Psi}$. We will see that the resolvent equation formalizes several relationships between the stopping times $\tau_g$ and $\tau_h$ for $\boldsymbol{\Psi}$. Note that since $h \ge g$, we have the inclusion $A_g \subseteq A_h$ and hence $\tau_g \ge \tau_h$.

To prove the first resolvent equation we write

$$\sum_{k=1}^{\tau_g} f(\Phi_k) = \sum_{k=1}^{\tau_h} f(\Phi_k) + \mathbb{1}\{\tau_g > \tau_h\} \sum_{k=\tau_h+1}^{\tau_g} f(\Phi_k)$$

so by the strong Markov property for the process $\boldsymbol{\Psi}$,

$$U_g(x,f) = U_h(x,f) + \mathsf{E}_x[\mathbb{1}\{g(\Phi_{\tau_h}) < U_{\tau_h}\} U_g(\Phi_{\tau_h}, f)]. \tag{12.15}$$

The latter expectation can be computed using (12.12). We have

$$\begin{aligned}
&\mathsf{E}_x[\mathbb{1}\{g(\Phi_{\tau_h}) < Y_{\tau_h}\} U_g(\Phi_{\tau_h}, f) \mathbb{1}\{\tau_h = k\} \mid \mathcal{F}_\infty^{\Phi}] \\
&= \mathsf{E}_x[\mathbb{1}\{g(\Phi_k) < Y_k\} U_g(\Phi_k, f) \mathbb{1}\{\tau_h = k\} \mid \mathcal{F}_\infty^{\Phi}] \\
&= \mathsf{E}_x[\mathbb{1}\{g(\Phi_k) < Y_k\} \mathbb{1}\{h(\Phi_k) \ge Y_k\} U_g(\Phi_k, f) \mathbb{1}\{\tau_h \ge k\} \mid \mathcal{F}_\infty^{\Phi}] \\
&= \mathsf{E}_x[\mathbb{1}\{g(\Phi_k) < Y_k \le h(\Phi_k)\} U_g(\Phi_k, f) \mathbb{1}\{\tau_h \ge k\} \mid \mathcal{F}_\infty^{\Phi}] \\
&= [h(\Phi_k) - g(\Phi_k)] U_g(\Phi_k, f) \prod_{i=1}^{k-1} [1 - h(\Phi_i)].
\end{aligned}$$

Taking expectations and summing over $k$ gives

$$\begin{aligned}
&\mathsf{E}_x[\mathbb{1}\{g(\Phi_{\tau_h}) < Y_{\tau_h}\} U_g(\Phi_{\tau_h}, f)] \\
&\quad = \sum_{k=1}^{\infty} \mathsf{E}_x\Big[\prod_{i=1}^{k-1} [1 - h(\Phi_i)][h(\Phi_k) - g(\Phi_k)] U_g(\Phi_k, f)\Big] \\
&\quad = \sum_{k=0}^{\infty} (P I_{1-h})^k P I_{h-g} U_g\,(x,f).
\end{aligned}$$

This together with (12.15) gives the first resolvent equation.

To prove the second, break the sum to $\tau_g$ into the pieces between consecutive visits to $A_h$:

$$\sum_{k=1}^{\tau_g} f(\Phi_k) = \sum_{k=1}^{\tau_h} f(\Phi_k) + \sum_{k=1}^{\tau_g} \mathbb{1}\{\Psi_k \in \{A_h \setminus A_g\}\}\theta^k\Big\{\sum_{i=1}^{\tau_h} f(\Phi_i)\Big\}.$$

Taking expectations gives

$$\begin{aligned} U_g(x,f) &= U_h(x,f) \\ &\quad + \mathsf{E}_x\Big[\sum_{k=1}^{\tau_g} \mathbb{1}\{g(\Phi_k) < Y_k \le h(\Phi_k)\}\theta^k\Big\{\sum_{i=1}^{\tau_h} f(\Phi_i)\Big\}\Big]. \end{aligned} \tag{12.16}$$

The expectation can be transformed, using the Markov property for the bivariate chain, to give

$$\begin{aligned} &\mathsf{E}_x\Big[\sum_{k=1}^{\tau_g} \mathbb{1}\{g(\Phi_k) < Y_k \le h(\Phi_k)\}\theta^k\Big\{\sum_{i=1}^{\tau_h} f(\Phi_i)\Big\}\Big] \\ &= \sum_{k=1}^{\infty} \mathsf{E}_x\Big[\mathbb{1}\{g(\Phi_k) < Y_k \le h(\Phi_k)\}\mathbb{1}\{\tau_g \ge k\}\mathsf{E}_{\Psi_k}\Big[\sum_{i=1}^{\tau_h} f(\Phi_i)\Big]\Big] \\ &= \sum_{k=1}^{\infty} \mathsf{E}_x\Big[[h(\Phi_k) - g(\Phi_k)]\mathbb{1}\{\tau_g \ge k\}U_h(\Phi_k, f)\Big] \\ &= U_g I_{h-g} U_h \end{aligned}$$

which together with (12.16) proves the second resolvent equation. □

When $\tau_h$ is a.s. finite for each initial condition the kernel $P_h$ defined as

$$P_h(x,A) = U_h I_h\,(x,A)$$

is a Markov transition function. This follows from (12.11), which shows that

$$\begin{aligned} P_h(x,\mathsf{X}) = U_h(x,h) &= \sum_{k=1}^{\infty} \mathsf{E}_x\Big[\prod_{i=1}^{k-1}(1-h(\Phi_i))h(\Phi_k)\Big] \\ &= \sum_{k=1}^{\infty} \mathsf{P}_x\{\tau_h = k\} \end{aligned} \tag{12.17}$$

and hence $P_h(x,\mathsf{X}) = 1$ if $\mathsf{P}_x\{\tau_h < \infty\} = 1$.

It is natural to seek conditions which will ensure that $\tau_h$ is finite, since this is of course analogous to the concept of Harris recurrence, and indeed identical to it for $h = \mathbb{1}_C$. The following result answers this question as completely as we will find necessary.

Define $L(x,h) = U_h(x,h)$ and $Q(x,h) = \mathsf{P}_x\{\sum_{k=1}^{\infty} h(\Phi_k) = \infty\}$. Theorem 12.2.2 now shows that these functions are extensions of the the functions $L$ and $Q$ which we have used extensively: in the special case where $h = \mathbb{1}_B$ for some $B \in \mathcal{B}(\mathsf{X})$ we have $Q(x,\mathbb{1}_B) = Q(x,B)$ and $L(x,\mathbb{1}_B) = L(x,B)$.

**Theorem 12.2.2** *For any $x \in \mathsf{X}$ and function $0 \le h \le 1$,*

**(i)** $\mathsf{P}_x\{\Psi_k \in A_h \quad \text{i.o.}\} = Q(x,h)$;

**(ii)** $\mathsf{P}_x\{\tau_h < \infty\} = L(x,h)$, *and hence* $L(x,h) \geq Q(x,h)$*;*

**(iii)** *If for some* $\varepsilon < 1$ *the function* $h$ *satisfies* $h(x) \leq \varepsilon$ *for all* $x \in \mathsf{X}$ *then* $L(x,h) = 1$ *if and only if* $Q(x,h) = 1$*.*

Proof **(i)** We have from the definition of $A_h$,

$$\mathsf{P}_x\{\Psi_k \in A_h \quad \text{i.o.} \mid \mathcal{F}_\infty^\Phi\} = \mathsf{P}_x\{Y_k \leq h(\Phi_k) \quad \text{i.o.} \mid \mathcal{F}_\infty^\Phi\}.$$

Conditioned on $\mathcal{F}_\infty^\Phi$, the events $\{Y_k \leq h(\Phi_k)\}$, $k \geq 1$, are mutually independent. Hence by the Borel-Cantelli Lemma,

$$\mathsf{P}_x\{\Psi_k \in A_h \quad \text{i.o.} \mid \mathcal{F}_\infty^\Phi\} = \mathbb{1}\Big\{\sum_{k=1}^\infty \mathsf{P}_x\{Y_k \leq h(\Phi_k) \mid \mathcal{F}_\infty^\Phi\} = \infty\Big\}.$$

Since $\mathsf{P}_x\{Y_k \leq h(\Phi_k) \mid \mathcal{F}_\infty^\Phi\} = h(\Phi_k)$, taking expectations of each side of this identity completes the proof of (i).

**(ii)** This follows directly from the definitions and (12.17).

**(iii)** Suppose that $h(x) \leq \varepsilon$ for all $x$, and suppose that $Q(x,h) < 1$ for some $x$. We will show that $L(x,h) < 1$ also.

If this is the case then by (i), for some $N < \infty$ and $\delta > 0$,

$$\mathsf{P}_x\{\ \Psi_k \in A_h^c \text{ for all } k > N\} = \delta.$$

But then by the fact that $\mathbf{Y}$ is i.i.d. and independent of $\boldsymbol{\Phi}$,

$$\begin{aligned} 1 - L(x,h) &\geq \mathsf{P}_x\{\ \Psi_k \in A_h^c \text{ for all } k > N, \text{ and } Y_k > \varepsilon \text{ for all } k \leq N\} \\ &= \mathsf{P}_x\{\ \Psi_k \in A_h^c \text{ for all } k > N\}\mathsf{P}_x\{\ Y_k > \varepsilon \text{ for all } k \leq N\} \\ &= \delta(1-\varepsilon)^N > 0. \end{aligned}$$

□

We now present an application of Theorem 12.2.2 which gives another representation for an invariant measure, extending the development of Section 10.4.2.

**Theorem 12.2.3** *Suppose that* $0 \leq h \leq 1$ *with* $Q(x,h) = 1$ *for all* $x \in \mathsf{X}$*.*

**(i)** *If* $\mu$ *is any* $\sigma$*-finite subinvariant measure then* $\mu$ *is invariant, and has the representation*

$$\mu(A) = \int \mu(dx)h(x)U_h(x,A)$$

**(ii)** *If* $\nu$ *is a finite measure satisfying, for some* $A \in \mathcal{B}(\mathsf{X})$,

$$\nu(B) = \nu U_h I_h(B), \qquad B \subseteq A$$

*then the measure* $\mu := \nu U_h$ *is invariant for* $\boldsymbol{\Phi}$*. The sets*

$$C_\varepsilon = \{x : K_{a_{\frac{1}{2}}}(x,h) > \varepsilon\}$$

*cover* $\mathsf{X}$ *and have finite* $\mu$*-measure for every* $\varepsilon > 0$*.*

PROOF We prove (i) by considering the bivariate chain $\boldsymbol{\Psi}$. The set $A_h \subset \mathsf{Y}$ is Harris recurrent and in fact $\mathsf{P}_x\{\boldsymbol{\Psi} \in A_h \quad \text{i.o.}\} = 1$ for all $x \in \mathsf{X}$ by Theorem 12.2.2. Now define the measure $\overline{\mu}$ on $\mathsf{Y}$ by

$$\overline{\mu}(A \times B) = \mu(A)\mathbf{u}(B), \qquad A \in \mathcal{B}(\mathsf{X}),\ B \in \mathcal{B}([0,1]). \tag{12.18}$$

Obviously $\overline{\mu}$ is an invariant measure for $\Psi$ and hence by Theorem 10.4.7,

$$\begin{aligned} \mu(A) = \overline{\mu}(A \times [0,1]) &= \int_{(x,y)\in A_h} \mu(dx)\mathbf{u}(dy)U_h(x,A) \\ &= \int \mu(dx)h(x)U_h(x,A) \end{aligned}$$

which is the first result.

To prove (ii) first extend $\nu$ to $\mathcal{B}(\mathsf{Y})$ as $\mu$ was extended in (12.18) to obtain a measure $\overline{\nu}$ on $\mathcal{B}(\mathsf{Y})$. Now apply Theorem 10.4.7. The measure $\overline{\mu}'$ defined as

$$\overline{\mu}'(A \times B) = \mathsf{E}_{\overline{\nu}}\Big[\sum_{k=1}^{\tau_h} \mathbb{1}\{\Psi_k \in A \times B\}\Big]$$

is invariant for $\boldsymbol{\Psi}$, and since the distribution of $\boldsymbol{\Phi}$ is the marginal distribution of $\boldsymbol{\Psi}$, the measure $\mu$ defined for $A \in \mathcal{B}(\mathsf{X})$ by $\mu(A) := \overline{\mu}'(A \times [0,1])$, $A \in \mathcal{B}(\mathsf{X})$, is invariant for $\boldsymbol{\Phi}$.

We now demonstrate that $\mu$ is $\sigma$-finite. From the assumptions of the theorem and Theorem 12.2.2 (ii) the sets $C_\varepsilon$ cover $\mathsf{X}$. We have from the representation of $\mu$,

$$\nu(\mathsf{X}) = \mu(h) = \mu K_{a_{\frac{1}{2}}}(h) \geq \varepsilon\mu(C_\varepsilon)$$

Hence for all $\varepsilon$ we have the bound $\mu(C_\varepsilon) \leq \mu(h)/\varepsilon < \infty$, which completes the proof of (ii). □

## 12.3 The existence of a $\sigma$-finite invariant measure

### 12.3.1 The smoothed chain on a compact set

Here we shall give a weak sufficient condition for the existence of a $\sigma$-finite invariant measure for a Feller chain. This provides an analogue of the results in Chapter 10 for recurrent chains. The construction we use mimics the construction mentioned in Section 10.4.2: here, though, a function on a compact set plays the part of the petite set $A$ used in the construction of the "process on $A$", and the fact that there is an invariant measure to play the part of the measure $\nu$ in Theorem 10.4.8 is an application of Theorem 12.1.2.

These results will again lead to a test function approach to establishing the existence of an invariant measure for a Feller chain, even without $\psi$-irreducibility.

We will, however, assume that some one compact set $C$ satisfies a strong form of Harris recurrence: that is, that there exists a compact set $C \subset \mathsf{X}$ with

$$L(x,C) = \mathsf{P}_x\{\boldsymbol{\Phi} \text{ enters } C\} \equiv 1, \qquad x \in \mathsf{X}. \tag{12.19}$$

Observe that by Proposition 9.1.1, (12.19) implies that $\boldsymbol{\Phi}$ visits $C$ infinitely often from each initial condition, and hence $\boldsymbol{\Phi}$ is at least non-evanescent.

To construct an invariant measure we essentially consider the chain $\boldsymbol{\Phi}^C$ obtained by sampling $\boldsymbol{\Phi}$ at consecutive visits to the compact set $C$. Suppose that the resulting sampled chain on $C$ had the Feller property. In this case, since the sampled chain evolves on the compact set $C$, we could deduce from Theorem 12.1.2 that an invariant probability existed for the sampled chain, and we would then need only a few further steps for an existence proof for the original chain $\boldsymbol{\Phi}$.

However, the transition function $P_C$ for the sampled chain is given by

$$P_C = \sum_{k=0}^{\infty} (PI_{C^c})^k PI_C = U_C I_C$$

which does not have the Feller property in general. To proceed, we must "smooth around the edges of the compact set $C$". The kernels $P_h$ introduced in the previous section allow us to do just that.

Let $N$ and $O$ be open subsets of $\mathsf{X}$ with compact closure for which $C \subset O \subset \bar{O} \subset N$, where $C$ satisfies (12.19) and let $h\colon \mathsf{X} \to \mathbb{R}$ be a continuous function such as

$$h(x) = \frac{d(x, N^c)}{d(x, N^c) + d(x, \bar{O})}$$

for which

$$\mathbb{1}_O(x) \le h(x) \le \mathbb{1}_N(x). \tag{12.20}$$

The kernel $P_h := U_h I_h$ is a Markov transition function since by (12.19) we have that $Q(x, h) \equiv 1$. Since $P_h(x, \bar{N}) = 1$ for all $x \in \mathsf{X}$, we will immediately have an invariant measure for $P_h$ by Theorem 12.1.2 if $P_h$ has the weak Feller property.

**Proposition 12.3.1** *Suppose that the transition function $P$ is weak Feller. If $0 \le h \le 1$ is continuous and if $Q(x, h) \equiv 1$, then $P_h$ is also weak Feller.*

Proof By the Feller property, the kernel $(PI_{1-h})^n PI_h$ preserves positive lower semicontinuous functions. Hence if $f$ is positive and lower semicontinuous, then

$$P_h f = \sum_{k=0}^{\infty} (PI_{1-h})^n PI_h f$$

is lower semicontinuous, being the increasing limit of a sequence of lower semicontinuous functions.

Suppose now that $f$ is bounded and continuous, and choose a constant $L$ so large that $L + f$ and $L - f$ are both positive. Then the functions

$$L + f \qquad L - f \qquad P_h(L + f) \qquad P_h(L - f)$$

are all positive and lower semicontinuous, from which it follows that $P_h f$ is continuous. Hence $P_h$ is weak Feller as required. □

We now prove using the generalized resolvent operators

**Theorem 12.3.2** *If $\boldsymbol{\Phi}$ is Feller and (12.19) is satisfied then there exists at least one invariant measure which is finite on compact sets.*

PROOF From Theorem 12.1.2 an invariant probability $\nu$ exists which is invariant for $P_h = U_h I_h$. Hence from Theorem 12.2.3, the measure $\mu = \nu U_h$ is invariant for $\boldsymbol{\Phi}$ and is finite on the sets $\{x : K_{a_{\frac{1}{2}}}(x,h) > \varepsilon\}$. Since $K_{a_{\frac{1}{2}}}(x,h)$ is a continuous function of $x$, and is strictly positive everywhere by (12.19), it follows that $\mu$ is finite on compact sets. □

### 12.3.2 Drift criteria for the existence of invariant measures

We conclude this section by proving that the test function which implies Harris recurrence or regularity for a $\psi$-irreducible T-chain may also be used to prove the existence of $\sigma$-finite invariant measures or invariant probability measures for Feller chains.

**Theorem 12.3.3** *Suppose that $\boldsymbol{\Phi}$ is Feller and that (V1) is satisfied with a compact set $C \subset \mathsf{X}$. Then an invariant measure exists which is finite on compact subsets of $\mathsf{X}$.*

PROOF If $L(x,C) = 1$ for all $x \in \mathsf{X}$, then the proof follows from Theorem 12.3.2.

Consider now the only other possibility, where $L(x,C) \neq 1$ for some $x$. In this case the adapted process $\{V(\Phi_k)\mathbb{1}\{\tau_C > k\}, \mathcal{F}_k^{\Phi}\}$ is a convergent supermartingale, as in the proof of Theorem 9.4.1, and since by assumption $\mathsf{P}_x\{\tau_C = \infty\} > 0$, this shows that

$$\mathsf{P}_x\{\limsup_{k\to\infty} V(\Phi_k) < \infty\} \geq 1 - L(x,C) > 0.$$

By Theorem 12.1.2, it follows that an invariant *probability* exists, and this completes the proof. □

Finally we prove that in the weak Feller case, the drift condition (V2) again provides a criterion for the existence of an invariant probability measure.

**Theorem 12.3.4** *Suppose that the chain $\boldsymbol{\Phi}$ is weak Feller. If (V2) is satisfied with a compact set $C$ and a positive function $V$ which is finite at one $x_0 \in \mathsf{X}$ then an invariant probability measure $\pi$ exists.*

PROOF Iterating (V2) $n$ times gives

$$\frac{1}{n}\sum_{k=0}^{n} 1 \leq \frac{1}{n}V(x_0) + b\frac{1}{n}\sum_{k=0}^{n} P^k(x_0, C).$$

Letting $n \to \infty$ we see that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} P^k(x_0,C) \geq \frac{1}{b}. \tag{12.21}$$

Theorem 12.3.4 then follows directly from Theorem 12.1.2 (i). □

## 12.4 Invariant Measures for e-Chains

### 12.4.1 Existence of an invariant measure for e-chains

Up to now we have shown under very mild conditions that an invariant probability measure exists for a Feller chain, based largely on arguments using weak convergence of $P^n$.

As we have seen, such weak limits will depend in general on the value of $x$ chosen, unless as in Proposition 12.1.4 there is a unique invariant measure. In this section we will explore the properties of the collection of such limiting measures.

Suppose that the chain is weak Feller and we can prove that a Markov transition function $\Pi$ exists which is itself weak Feller, such that for any $f \in \mathcal{C}(\mathsf{X})$,

$$\lim_{k\to\infty} P^k f\,(x) = \Pi f\,(x), \qquad x \in \mathsf{X}. \tag{12.22}$$

In this case, it follows as in Proposition 6.4.2 from Ascoli's Theorem D.4.2 that $\{P^k f : k \in \mathbb{Z}_+\}$ is equicontinuous on compact subsets of $\mathsf{X}$ whenever $f \in \mathcal{C}(\mathsf{X})$, and so it is necessary that the chain $\boldsymbol{\Phi}$ be an e-chain, in the sense of Section 6.4, whenever we have convergence in the sense of (12.22).

The key to analyzing e-chains lies in the following result:

**Theorem 12.4.1** *Suppose that $\boldsymbol{\Phi}$ is an e-chain. Then*

**(ii)** *There exists a substochastic kernel $\Pi$ such that*

$$\overline{P}_k(x,\,\cdot\,) \xrightarrow{\mathrm{v}} \Pi(x,\,\cdot\,) \qquad \text{as } k \to \infty \tag{12.23}$$

$$K_{a_\varepsilon}(x,\,\cdot\,) \xrightarrow{\mathrm{v}} \Pi(x,\,\cdot\,) \qquad \text{as } \varepsilon \uparrow 1 \tag{12.24}$$

*for all $x \in \mathsf{X}$.*

**(ii)** *For each $j,\ k,\ \ell \in \mathbb{Z}_+$ we have*

$$P^j \Pi^k P^\ell = \Pi, \tag{12.25}$$

*and hence for all $x \in \mathsf{X}$ the measure $\Pi(x,\,\cdot\,)$ is invariant with $\Pi(x,\mathsf{X}) \le 1$.*

**(iii)** *The Markov chain is bounded in probability on average if and only if $\Pi(x,\mathsf{X}) = 1$ for all $x \in \mathsf{X}$.*

PROOF We prove the result (12.23), the proof of (12.24) being similar. Let $\{f_n\} \subset \mathcal{C}_c(\mathsf{X})$ denote a fixed dense subset. By Ascoli's theorem and a diagonal subsequence argument, there exists a subsequence $\{k_i\}$ of $\mathbb{Z}_+$ and functions $\{g_n\} \subset \mathcal{C}(\mathsf{X})$ such that

$$\lim_{i\to\infty} \overline{P}_{k_i} f_n\,(x) = g_n(x) \tag{12.26}$$

uniformly for $x$ in compact subsets of $\mathsf{X}$ for each $n \in \mathbb{Z}_+$. The set of all subprobabilities on $\mathcal{B}(\mathsf{X})$ is sequentially compact with respect to vague convergence, and any vague limit $\nu$ of the probabilities $\overline{P}_{k_i}(x,\,\cdot\,)$ must satisfy $\int f_n\,d\nu = g_n(x)$ for all $n \in \mathbb{Z}_+$. Since the functions $\{f_n\}$ are dense in $\mathcal{C}_c(\mathsf{X})$, this shows that for each $x$ there is exactly one vague limit point, and hence a kernel $\Pi$ exists for which

$$\overline{P}_{k_i}(x,\,\cdot\,) \xrightarrow{\mathrm{v}} \Pi(x,\,\cdot\,) \qquad \text{as } i \to \infty$$

for each $x \in \mathsf{X}$.

Observe that by equicontinuity, the function $\Pi f$ is continuous for every function $f \in \mathcal{C}_c(\mathsf{X})$. It follows that $\Pi f$ is positive and lower semicontinuous whenever $f$ has these properties.

By the Dominated Convergence Theorem we have for all $k, j \in \mathbb{Z}_+$,

$$P^j \Pi^k = \Pi.$$

Next we show that $\Pi P = \Pi$, and hence that

$$\Pi^k P^j = \Pi, \qquad k, j \in \mathbb{Z}_+.$$

Let $f \in \mathcal{C}_c(\mathsf{X})$ be a continuous positive function with compact support. Then, since the function $Pf$ is also positive and continuous, (D.6) implies that

$$\begin{aligned} \Pi(Pf) &\leq \liminf_{i \to \infty} \overline{P}_{k_i}(Pf) \\ &= \Pi f, \end{aligned}$$

which shows that $\Pi P = \Pi$.

We now show that (12.23) holds. Suppose that $\overline{P}_N$ does not converge vaguely to $\Pi$. Then there exists a different subsequence $\{m_j\}$ of $\mathbb{Z}_+$, and a distinct kernel $\Pi'$ such that

$$\overline{P}_{m_j} \xrightarrow{\mathrm{v}} \Pi'(x, \cdot), \qquad j \to \infty.$$

However, for each positive function $f \in \mathcal{C}_c(\mathsf{X})$,

$$\begin{aligned} \Pi f &= \lim_{j \to \infty} \Pi \overline{P}_{m_j} f \\ &= \Pi \Pi' f \qquad \text{by the Dominated Convergence Theorem} \\ &\leq \liminf_{i \to \infty} \overline{P}_{k_i} \Pi' f \qquad \text{since } \Pi' f \text{ is continuous and positive} \\ &= \Pi' f. \end{aligned}$$

Hence by symmetry, $\Pi' = \Pi$, and this completes the proof of (i) and (ii).

The result (iii) follows from (i) and Proposition D.5.6. □

### 12.4.2 Hitting time and drift criteria for stability of e-chains

We now consider the stability of e-chains. First we show in Theorem 12.4.3 that if the chain hits a fixed compact subset of $\mathsf{X}$ with probability one from each initial condition, and if this compact set is positive in a well defined way, then the chain is bounded in probability on average. This is an analogue of the rather more powerful regularity results in Chapter 11.

This result is then applied to obtain a drift criterion for boundedness in probability using (V2).

To characterize boundedness in probability we use the following weak analogue of Kac's Theorem 10.2.2, connecting positivity of $K_{a_\varepsilon}(x, C)$ with finiteness of the mean return time to $C$.

**Proposition 12.4.2** *For any compact set $C \subset \mathsf{X}$*

$$\liminf_{\varepsilon \uparrow 1} K_{a_\varepsilon}(x, C) \geq \left(\sup_{y \in C} \mathsf{E}_y[\tau_C]\right)^{-1}, \qquad x \in C.$$

Proof For the first entrance time $\tau_C$ to the compact set $C$, let $\theta^{\tau_C}$ denote the $\tau_C$-fold shift on sample space, defined so that $\theta^{\tau_C} f(\Phi_k) = f(\Phi_{k+\tau_C})$ for any function $f$ on $\mathsf{X}$.

Fix $x \in C$, $0 < \varepsilon < 1$, and observe that by conditioning at time $\tau_C$ and using the strong Markov property we have for $x \in C$,

$$\begin{aligned} K_{a_\varepsilon}(x, C) &= (1-\varepsilon)\mathsf{E}_x\Big[\sum_{k=0}^{\infty} \varepsilon^k \mathbb{1}\{\Phi_k \in C\}\Big] \\ &= (1-\varepsilon)\mathsf{E}_x\Big[1 + \sum_{k=0}^{\infty} \varepsilon^{\tau_C + k}(\theta^{\tau_C}\mathbb{1}\{\Phi_k \in C\})\Big] \\ &= (1-\varepsilon) + (1-\varepsilon)\mathsf{E}_x\Big[\varepsilon^{\tau_C}\mathsf{E}_{\Phi_{\tau_C}}\Big[\sum_{k=0}^{\infty} \varepsilon^k \mathbb{1}\{\Phi_k \in C\}\Big]\Big] \\ &\geq (1-\varepsilon) + \mathsf{E}_x[\varepsilon^{\tau_C}] \inf_{y \in C} K_{a_\varepsilon}(y, C) \end{aligned}$$

Taking the infimum over all $x \in C$, we obtain

$$\inf_{y \in C} K_{a_\varepsilon}(y, C) \geq (1-\varepsilon) + \inf_{y \in C} \mathsf{E}_y[\varepsilon^{\tau_C}] \inf_{y \in C} K_{a_\varepsilon}(y, C) \tag{12.27}$$

By Jensen's inequality we have the bound $\mathsf{E}[\varepsilon^{\tau_C}] \geq \varepsilon^{\mathsf{E}[\tau_C]}$. Hence letting $M_C = \sup_{x \in C} \mathsf{E}_x[\tau_C]$ it follows from (12.27) that for $y \in C$,

$$K_{a_\varepsilon}(y, C) \geq \frac{1-\varepsilon}{1-\varepsilon^{M_C}}.$$

Letting $\varepsilon \uparrow 1$ we have for each $y \in C$,

$$\liminf_{\varepsilon \uparrow 1} K_{a_\varepsilon}(y, C) \geq \lim_{\varepsilon \uparrow 1}\left(\frac{1-\varepsilon}{1-\varepsilon^{M_C}}\right) = \frac{1}{M_C}.$$

□

We saw in Theorem 12.4.1 that $\boldsymbol{\Phi}$ is bounded in probability on average if and only if $\Pi(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$. Hence the following result shows that compact sets serve as test sets for stability: if a fixed compact set is reachable from all initial conditions, and if $\boldsymbol{\Phi}$ is reasonably well behaved from initial conditions on that compact set, then $\boldsymbol{\Phi}$ will be bounded in probability on average.

**Theorem 12.4.3** *Suppose $\boldsymbol{\Phi}$ is an e-chain. Then*

**(i)** $\max_{x \in \mathsf{X}} \Pi(x, \mathsf{X})$ *exists, and is equal to zero or one;*

**(ii)** *if* $\min_{x \in \mathsf{X}} \Pi(x, \mathsf{X})$ *exists, then it is equal to zero or one;*

**(iii)** *if there exists a compact set $C \subset \mathsf{X}$ such that*

$$\mathsf{P}_x\{\tau_C < \infty\} = 1 \qquad x \in \mathsf{X}$$

*then* $\min_{x \in \mathsf{X}} \Pi(x, \mathsf{X})$ *exists, and is attained on $C$, so that*

$$\inf_{x \in \mathsf{X}} \Pi(x, \mathsf{X}) = \min_{x \in C} \Pi(x, \mathsf{X});$$

**(iv)** *if $C \subset \mathsf{X}$ is compact, then*

$$\inf_{x \in C} \Pi(x, \mathsf{X}) \geq \Big(\sup_{x \in C} \mathsf{E}_x[\tau_C]\Big)^{-1}.$$

PROOF **(i)** If $\Pi(x,\mathsf{X}) > 0$ for some $x \in \mathsf{X}$, then an invariant probability $\pi$ exists. In fact, we may take $\pi = \Pi(x,\,\cdot\,)/\Pi(x,\mathsf{X})$.

From the definition of $\Pi$ and the Dominated Convergence Theorem we have that for any $f \in \mathcal{C}_c(\mathsf{X})$,

$$\pi(f) = \lim_{n\to\infty} [\pi \overline{P}_n(f)] = \pi \Pi(f)$$

which shows that $\pi = \pi\Pi$. Hence $1 = \pi(\mathsf{X}) = \int \pi(dx)\Pi(x,\mathsf{X})$. This shows that $\Pi(y,\mathsf{X}) = 1$ for a.e. $y \in \mathsf{X}$ $[\pi]$, proving (i) of the theorem.

**(ii)** Let $\rho = \inf_{x\in\mathsf{X}} \Pi(x,\mathsf{X})$, and let

$$S_\rho = \{x \in \mathsf{X} : \Pi(x,\mathsf{X}) = \rho\}.$$

By the assumptions of (ii), $S_\rho \neq \emptyset$. Letting $u(\,\cdot\,) := \Pi(\,\cdot\,,\mathsf{X})$, we have $Pu = u$, and this implies that the set $S_\rho$ is absorbing. Since $u$ is lower semicontinuous, the set $S_\rho$ is also a closed subset of $\mathsf{X}$.

Since $S_\rho$ is closed, it follows by vague convergence and (D.6) that for all $x \in \mathsf{X}$,

$$\liminf_{N\to\infty} \overline{P}_N(x, S_\rho^c) \geq \Pi(x, S_\rho^c),$$

and since $S_\rho$ is also absorbing, this shows that for all $x \in S_\rho$

$$\Pi(x, S_\rho^c) = 0. \tag{12.28}$$

Suppose now that $0 \leq \rho < 1$. As in the proof of (i),

$$\pi\{y \in \mathsf{X} : \Pi(y,\mathsf{X}) = 1\} = 1$$

for any invariant probability $\pi$, and hence

$$\Pi(x, S_\rho) \leq \Pi(x, \{y \in \mathsf{X} : \Pi(y,\mathsf{X}) < 1\}) = 0. \tag{12.29}$$

Equations (12.28) and (12.29) show that for any $x \in S_\rho$,

$$\rho = \Pi(x,\mathsf{X}) = \Pi(x, S_\rho) + \Pi(x, S_\rho^c) = 0,$$

and this proves (ii).

**(iii)** Since $u(x) := \Pi(x,\mathsf{X})$ is lower semicontinuous we have

$$\inf_{x\in C} u(x) = \min_{x\in C} u(x).$$

That is, the infimum is attained.

Since $Pu = u$, the sequence $\{u(\Phi_k), \mathcal{F}_k^\Phi\}$ is a martingale, which converges to a random variable $u_\infty$ satisfying $\mathsf{E}_x[u_\infty] = u(x)$, $x \in \mathsf{X}$. By Proposition 9.1.1, the assumption that $\mathsf{P}_x\{\tau_C < \infty\} \equiv 1$ implies that

$$\mathsf{P}_x\{\boldsymbol{\Phi} \in C \text{ i.o.}\} = 1, \qquad x \in \mathsf{X}. \tag{12.30}$$

If $\Phi_k \in C$ for some $k \in \mathbb{Z}_+$, then obviously $u(\Phi_k) \geq \min_{x\in C} u(x)$, which by (12.30) implies that

$$u_\infty = \lim_{k\to\infty} u(\Phi_k) \geq \min_{x\in C} u(x) \qquad \text{a.s.}$$

Taking expectations shows that $u(y) \geq \min_{x\in C} u(x)$ for all $y \in \mathsf{X}$, proving part (iii) of the theorem.

**(iv)** Letting $M_C = \sup_{x\in C} \mathsf{E}_x[\tau_C]$ it follows from Proposition 12.4.2 that

$$\inf_{y\in C} \liminf_{\varepsilon\uparrow 1} K_{a_\varepsilon}(y, C) \geq \frac{1}{M_C}.$$

This proves the result since $\limsup_{\varepsilon\uparrow 1} K_{a_\varepsilon}(y, C) \leq \Pi(y, C)$ by Theorem 12.4.1. □

We have immediately

**Proposition 12.4.4** *Let $\boldsymbol{\Phi}$ be an e-chain, and let $C \subset \mathsf{X}$ be compact. If $\mathsf{P}_x\{\tau_C < \infty\} = 1$, $x \in \mathsf{X}$, and $\sup_{x\in C} \mathsf{E}_x[\tau_C] < \infty$, then $\boldsymbol{\Phi}$ is bounded in probability on average.*

Proof From Theorem 12.4.3 (iii) we see that for all $x$,

$$\min_{x\in\mathsf{X}} \Pi(x, \mathsf{X}) = \min_{x\in C} \Pi(x, \mathsf{X}) \geq \Big(\sup_{x\in C} \mathsf{E}_x[\tau_C]\Big)^{-1} > 0.$$

Hence from Theorem 12.4.3 (ii) we have $\Pi(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$. Theorem 12.4.1 then implies that the chain is bounded in probability on average. □

The next result shows that the drift criterion for positive recurrence for $\psi$-irreducible chains also has an impact on the class of e-chains.

**Theorem 12.4.5** *Let $\boldsymbol{\Phi}$ be an e-chain, and suppose that condition (V2) holds for a compact set $C$ and an everywhere finite function $V$. Then the Markov chain $\boldsymbol{\Phi}$ is bounded in probability on average.*

Proof It follows from Theorem 11.3.4 that $\mathsf{E}_x[\tau_C] \leq V(x)$ for $x \in C^c$, so that *a fortiori* we also have $L(x, C) \equiv 1$. As in the proof of Theorem 12.3.4, for any $x \in \mathsf{X}$,

$$\Pi(x, \mathsf{X}) \geq \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} P^k(x, C) \geq \frac{1}{b}, \qquad x \in \mathsf{X}.$$

From this it follows from Theorem 12.4.3 (iii) and (ii) that $\Pi(x, \mathsf{X}) \equiv 1$, and hence $\boldsymbol{\Phi}$ is bounded in probability on average as claimed. □

## 12.5 Establishing boundedness in probability

Boundedness in probability is clearly the key condition needed to establish the existence of an invariant measure under a variety of continuity regimes. In this section we illustrate the verification of boundedness in probability for some specific models.

### 12.5.1 Linear state space models

We show first that the conditions used in Proposition 6.3.5 to obtain irreducibility are in fact sufficient to establish boundedness in probability for the linear state space model. Thus with no extra conditions we are able to show that a stationary version of this model exists.

Recall that we have already seen in Chapter 7 that the linear state space model is an e-chain when (LSS5) holds.

**Proposition 12.5.1** *Consider the linear state space model defined by (LSS1) and (LSS2). If the eigenvalue condition (LSS5) is satisfied then $\boldsymbol{\Phi}$ is bounded in probability. Moreover, if the nonsingularity condition (LSS4) and the controllability condition (LCM3) are also satisfied then the model is positive Harris.*

Proof Let us take

$$M := I + \sum_{i=1}^{\infty} F^{\top i} F^{i},$$

where $F^{\top}$ denotes the transpose of $F$. If Condition (LSS5) holds then by Lemma 6.3.4, the matrix $M$ is finite and positive definite with $I \leq M$, and for some $\alpha < 1$

$$|Fx|_M^2 \leq \alpha |x|_M^2 \tag{12.31}$$

where $|y|_M^2 := y^{\top} M y$ for $y \in \mathbb{R}^n$.

Let $m = \left(\sum_{i=0}^{\infty} F^i\right) G\, \mathsf{E}[W_1]$, and define

$$V(x) = |x-m|_M^2, \qquad x \in \mathsf{X}. \tag{12.32}$$

Then it follows from (LSS1) that

$$\begin{aligned} V(X_{k+1}) \;\; &= \;\; |F(X_k - m)|_M^2 + |G(W_{k+1} - \mathsf{E}[W_{k+1}])|_M^2 \\ &\quad + (X_k - m)^{\top} F^{\top} M G (W_{k+1} - \mathsf{E}[W_{k+1}]) \\ &\quad + (W_{k+1} - \mathsf{E}[W_{k+1}])^{\top} G^{\top} M F (X_k - m). \end{aligned} \tag{12.33}$$

Since $W_{k+1}$ and $X_k$ are independent, this together with (12.31) implies that

$$\mathsf{E}[V(X_{k+1}) \mid X_0, \dots, X_k] \leq \alpha V(X_k) + \mathsf{E}[|G(W_{k+1} - \mathsf{E}[W_{k+1}])|_M^2], \tag{12.34}$$

and taking expectations of both sides gives

$$\limsup_{k\to\infty} \mathsf{E}[V(X_k)] \leq \frac{\mathsf{E}[|G(W_{k+1} - \mathsf{E}[W_{k+1}])|_M^2]}{1-\alpha} < \infty.$$

Since $V$ is a norm-like function on $\mathsf{X}$, Lemma D.5.3 gives a direct proof that the chain is bounded in probability.

We note that (12.34) also ensures immediately that (V2) is satisfied. Under the extra conditions (LSS4) and (LCM3) we have from Proposition 6.3.5 that all compact sets are petite, and it immediately follows from Theorem 11.3.11 that the chain is regular and hence positive Harris. □

It may be seen that stability of the linear state space model is closely tied to the stability of the deterministic system $x_{k+1} = Fx_k$. For each initial condition $x_0 \in \mathbb{R}^n$ of this deterministic system, the resulting trajectory $\{x_k\}$ satisfies the bound

$$|x_k|_M \leq \alpha^k |x_0|_M$$

and hence is ultimately bounded in the sense of Section 11.2: in fact, in the dynamical systems literature such a system is called *globally exponentially stable*. It is precisely this stability for the deterministic "core" of the linear state space model which allows us to obtain boundedness in probability for the stochastic process $\boldsymbol{\Phi}$.

We now generalize the model (LSS1) to include random variation in the coefficients $F$ and $G$.

### 12.5.2 Bilinear models

Let us next consider the scalar example where $\boldsymbol{\Phi}$ is the bilinear state space model on $\mathsf{X} = \mathbb{R}$ defined in (SBL1)–(SBL2)

$$X_{k+1} = \theta X_k + bW_{k+1}X_k + W_{k+1} \tag{12.35}$$

where $\mathbf{W}$ is a zero-mean disturbance process. This is related closely to the linear model above, and the analysis is almost identical.

To obtain boundedness in probability by direct calculation, observe that

$$\mathsf{E}[|X_{k+1}| \mid X_k = x] \leq \mathsf{E}[|\theta + bW_{k+1}|]|x| + \mathsf{E}[|W_{k+1}|] \tag{12.36}$$

Hence for every initial condition of the process,

$$\limsup_{k\to\infty} \mathsf{E}[|X_k|] \leq \frac{\mathsf{E}[|W_{k+1}|]}{1 - \mathsf{E}[|\theta + bW_{k+1}|]}$$

provided that

$$\mathsf{E}[|\theta + bW_{k+1}|] < 1. \tag{12.37}$$

Since $|\cdot|$ is a norm-like function on $\mathsf{X}$, this shows that $\boldsymbol{\Phi}$ is bounded in probability provided that (12.37) is satisfied.

Again observe that in fact the bound (12.36) implies that the mean drift criterion (V2) holds.

### 12.5.3 Adaptive control models

Finally we consider the adaptive control model (2.21)-(2.23).

The closed loop system described by (2.24) is a Feller Markov chain, and thus an invariant probability exists if the distributions of the process are tight for some initial condition. We show here that the distributions of $\boldsymbol{\Phi}$ are tight when the initial conditions are chosen so that

$$\tilde{\theta}_k = \theta_k - \mathsf{E}[\theta_k \mid \mathcal{Y}_k], \quad \text{and} \quad \Sigma_k = \mathsf{E}[\tilde{\theta}_k^2 \mid \mathcal{Y}_k]. \tag{12.38}$$

For example, this is the case when $y_0 = \tilde{\theta}_0 = \Sigma_0 = 0$. If (12.38) holds then it follows from (2.22) that

$$\mathsf{E}[Y_{k+1}^2 \mid \mathcal{Y}_k] = \Sigma_k Y_k^2 + \sigma_w^2. \tag{12.39}$$

This identity will be used to prove the following result:

**Proposition 12.5.2** *For the adaptive control model satisfying (SAC1) and (SAC2), suppose that the process $\boldsymbol{\Phi}$ defined in (2.24) satisfies (12.38) and that $\sigma_z^2 < 1$. Then we have*

$$\limsup_{k\to\infty} \mathsf{E}[|\Phi_k|^2] < \infty$$

*so that distributions of the chain are tight, and hence $\boldsymbol{\Phi}$ is positive recurrent.*

PROOF We note first that since the sequence $\{\Sigma_k\}$ is bounded below and above by $\underline{\Sigma} = \sigma_z > 0$ and $\overline{\Sigma} = \sigma_z/(1-\alpha^2) < \infty$, and the process $\boldsymbol{\theta}$ clearly satisfies

$$\limsup_{k\to\infty} \mathsf{E}[\theta_k^2] = \frac{\sigma_z^2}{1-\alpha^2},$$

to prove the proposition it is enough to bound $\mathsf{E}[Y_k^2]$.

From (12.39) and (2.23) we have

$$\begin{aligned}
\mathsf{E}[Y_{k+1}^2 \Sigma_{k+1} \mid \mathcal{Y}_k] &= \Sigma_{k+1}\mathsf{E}[Y_{k+1}^2 \mid \mathcal{Y}_k] \\
&= \Sigma_{k+1}(\Sigma_k Y_k^2 + \sigma_w^2) \\
&= (\sigma_z^2 + \alpha^2\sigma_w^2\Sigma_k(\Sigma_k Y_k^2 + \sigma_w^2)^{-1})(\Sigma_k Y_k^2 + \sigma_w^2) \\
&= \sigma_z^2\Big(Y_k^2\Sigma_k\Big) + \Big(\sigma_w^2\sigma_z^2 + \alpha^2\sigma_w^2\Sigma_k\Big).
\end{aligned} \tag{12.40}$$

Taking total expectations of each side of (12.40), we use the condition $\sigma_z^2 < 1$ to obtain by induction, for all $k \in \mathbb{Z}_+$,

$$\underline{\Sigma}\mathsf{E}[Y_{k+1}^2] \le \mathsf{E}[Y_{k+1}^2\Sigma_{k+1}] \le \frac{\sigma_w^2\sigma_z^2 + \alpha^2\sigma_w^2\overline{\Sigma}}{1-\sigma_z^2} + \sigma_z^{2k}\mathsf{E}[Y_0^2\Sigma_0]. \tag{12.41}$$

This shows that the mean of $Y_k^2$ is uniformly bounded.

Since $\boldsymbol{\Phi}$ has the Feller property it follows from Proposition 12.1.3 that an invariant probability exists. Hence from Theorem 7.4.3 the chain is positive recurrent. □

In fact, we will see in Chapter 16 that not only is the process bounded in probability, but the conditional mean of $Y_k^2$ converges to the steady state value $\mathsf{E}_\pi[Y_0^2]$ at a geometric rate from *every* initial condition. These results require a more elaborate stability proof.

Note that equation (12.40) does *not* obviously imply that there is a solution to a drift inequality such as (V2): the conditional expectation is taken with respect to $\mathcal{Y}_k$, which is strictly smaller than $\mathcal{F}_k^{\Phi}$.

The condition that $\sigma_z^2 < 1$ cannot be omitted in this analysis: indeed, we have that if $\sigma_z^2 \ge 1$, then

$$\mathsf{E}[Y_k^2] \ge [\sigma_z^2]^k Y_0 + k\sigma_w^2 \to \infty$$

as $k$ increases, so that the chain is unstable in a mean square sense, although it may still be bounded in probability.

It is well worth observing that this is one of the few models which we have encountered where obtaining a drift inequality of the form (V2) is much more difficult than merely proving boundedness in probability. This is due to the fact that the dynamics of this model are extremely nonlinear, and so a direct stability proof is difficult. By exploiting equation (12.39) we essentially linearize a portion of the dynamics, which makes the stability proof rather straightforward. However the identity (12.39) only holds for a restricted class of initial conditions, so in general we are forced to tackle the nonlinear equations directly.

## 12.6 Commentary

The key result Theorem 12.1.2 is taken from Foguel [78]. Versions of this result have also appeared in papers by Beneš [18, 19] and Stettner [256] which consider processes in continuous time. For more results on Feller chains the reader is referred to Krengel [141], and the references cited therein.

For an elegant operator-theoretic proof of results related to Theorem 12.3.2, see Lin [154] and Foguel [80]. The method of proof based upon the use of the operator $P_h = U_h I_h$ to obtain a $\sigma$-finite invariant measure is taken from Rosenblatt [228]. Neveu in [196] promoted the use of the operators $U_h$, and proved the resolvent equation Theorem 12.2.1 using direct manipulations of the operators. The kernel $P_h$ is often called the *balayage operator* associated with the function $h$ (see Krengel [141] or Revuz [223]). In the Supplement to Krengel's text by Brunel ([141] pp. 301–309) a development of the recurrence structure of irreducible Markov chains is developed based upon these operators. This analysis and much of [223] exploits fully the resolvent equation, illustrating the power of this simple formula although because of our emphasis on $\psi$-irreducible chains and probabilistic methods, we do not address the resolvent equation further in this book.

Obviously, as with Theorem 12.1.2, Theorem 12.3.4 can be applied to an irreducible Markov chain on countable space to prove positive recurrence. It is of some historical interest to note that Foster's original proof of the sufficiency of (V2) for positivity of such chains is essentially that in Theorem 12.3.4. Rather than showing in any direct way that (V2) gives an invariant measure, Foster was able to use the countable space analogue of Theorem 12.1.2 (i) to deduce positivity from the "non-nullity" of a "compact" finite set of states as in (12.21). We will discuss more general versions of this classification of sets as positive or null further, but not until Chapter 18.

Observe that Theorem 12.3.4 only states that an invariant probability exists. Perhaps surprisingly, it is not known whether the hypotheses of Theorem 12.3.4 imply that the chain is bounded in probability when $V$ is finite-valued except for e-chains as in Theorem 12.4.5.

The theory of e-chains is still being developed, although these processes have been the subject of several papers over the past thirty years, most notably by Jamison and Sine [109, 112, 243, 242, 241], Rosenblatt [227], Foguel [78] and the text by Krengel [141]. In most of the e-chain literature, however, the state space is assumed compact so that stability is immediate. The drift criterion for boundedness in probability on average in Theorem 12.4.5 is new. The criterion Theorem 12.3.4 for the existence of an invariant probability for a Feller chain was first shown in Tweedie [280].

The stability analysis of the linear state space model presented here is standard. For an early treatment see Kalman and Bertram [120], while Caines [39] contains a modern and complete development of discrete time linear systems. Snyders [250] treats linear models with a continuous time parameter in a manner similar to the presentation in this book. The bilinear model has been the subject of several papers: see for example Feigin and Tweedie [74], or the discussion in Tong [267]. The stability of the adaptive control model was first resolved in Meyn and Caines [172], and related stability results were described in Solo [251]. The stability proof given here is new, and is far simpler than any previous results.