5

Pseudo-atoms

Much Markov chain theory on a general state space can be developed in complete analogy with the countable state situation when X contains an *atom* for the chain $\boldsymbol{\Phi}$.

> Atoms
>
> A set $\boldsymbol{\alpha} \in \mathcal{B}(\mathsf{X})$ is called an *atom* for $\boldsymbol{\Phi}$ if there exists a measure ν on $\mathcal{B}(\mathsf{X})$ such that
> $$P(x, A) = \nu(A), \quad x \in \boldsymbol{\alpha}.$$
> If $\boldsymbol{\Phi}$ is ψ-irreducible and $\psi(\boldsymbol{\alpha}) > 0$ then $\boldsymbol{\alpha}$ is called an *accessible atom.*

A single point $\boldsymbol{\alpha}$ is always an atom. Clearly, when X is countable and the chain is irreducible then every point is an accessible atom.

On a general state space, accessible atoms are less frequent. For the random walk on a half line as in (RWHL1), the set $\{0\}$ is an accessible atom when $\Gamma(-\infty, 0) > 0$: as we have seen in Proposition 4.3.1, this chain has $\psi(\{0\}) > 0$. But for the random walk on $\mathbb{R}$ when Γ has a density, accessible atoms do not exist.

It is not too strong to say that the single result which makes general state space Markov chain theory as powerful as countable space theory is that there exists an "artificial atom" for φ-irreducible chains, even in cases such as the random walk with absolutely continuous increments. The highlight of this chapter is the development of this result, and some of its immediate consequences.

Atoms are found for "strongly aperiodic" chains by constructing a "split chain" $\check{\boldsymbol{\Phi}}$ evolving on a split state space $\check{\mathsf{X}} = \mathsf{X}_0 \cup \mathsf{X}_1$, where X_0 and X_1 are copies of the state space X, in such a way that

(i) the chain $\boldsymbol{\Phi}$ is the marginal chain of $\check{\boldsymbol{\Phi}}$, in the sense that $\mathsf{P}(\Phi_k \in A) = \mathsf{P}(\check{\Phi}_k \in A_0 \cup A_1)$ for appropriate initial distributions, and

(ii) the "bottom level" X_1 is an accessible atom for $\check{\boldsymbol{\Phi}}$.

The existence of a splitting of the state space in such a way that the bottom level is an atom is proved in the next section. The proof requires the existence of so-called "small sets" C, which have the property that there exists an $m > 0$, and a minorizing measure ν on $\mathcal{B}(\mathsf{X})$ such that for any $x \in C$,

$$P^m(x, B) \geq \nu(B). \tag{5.1}$$

In Section 5.2, we show that, provided the chain is ψ-irreducible

$$\mathsf{X} = \bigcup_1^\infty C_i$$

where each C_i is small: thus we have that the splitting is always possible for such chains.

Another non-trivial consequence of the introduction of small sets is that on a general space we have a finite cyclic decomposition for ψ-irreducible chains: there is a cycle of sets $D_i, i = 0, 1, \dots, d-1$ such that

$$\mathsf{X} = N \cup \bigcup_0^{d-1} D_i$$

where $\psi(N) = 0$ and $P(x, D_i) \equiv 1$ for $x \in D_{i-1}$ (mod d). A more general and more tractable class of sets called petite sets are introduced in Section 5.5: these are used extensively in the sequel, and in Theorem 5.5.7 we show that every petite set is small if the chain is aperiodic.

5.1 Splitting φ-Irreducible Chains

Before we get to these results let us first consider some simpler consequences of the existence of atoms.

As an elementary first step, it is clear from the proof of the existence of a maximal irreducibility measure in Proposition 4.2.2 that we have an easy construction of ψ when X contains an atom.

Proposition 5.1.1 *Suppose there is an atom $\boldsymbol{\alpha}$ in X such that $\sum_n P^n(x, \boldsymbol{\alpha}) > 0$ for all $x \in \mathsf{X}$. Then $\boldsymbol{\alpha}$ is an accessible atom and $\boldsymbol{\Phi}$ is ν-irreducible with $\nu = P(\boldsymbol{\alpha}, \cdot)$.*

Proof We have, by the Chapman-Kolmogorov equations, that for any $n \geq 1$

$$\begin{aligned} P^{n+1}(x, A) &\geq \int_\alpha P^n(x, dy) P(y, A) \\ &= P^n(x, \boldsymbol{\alpha}) \nu(A) \end{aligned}$$

which gives the result by summing over n. □

The uniform communication relation "$\rightsquigarrow A$" introduced in Section 4.2.3 is also simplified if we have an atom in the space: it is no more than the requirement that there is a set of paths to A of positive probability, and the uniformity is automatic.

Proposition 5.1.2 *If $L(x, A) > 0$ for some state $x \in \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is an atom, then $\boldsymbol{\alpha} \rightsquigarrow A$.* □

In many cases the "atoms" in a state space will be real atoms: that is, single points which are reached with positive probability.

Consider the level in a dam in any of the storage models analyzed in Section 4.3.2. It follows from Proposition 4.3.1 that the single point $\{0\}$ forms an accessible atom satisfying the hypotheses of Proposition 5.1.1, even when the input and output processes are continuous.

However, our reason for featuring atoms is not because some models have singletons which can be reached with probability one: it is because even in the completely general ψ-irreducible case, by suitably extending the probabilistic structure of the chain, we are able to artificially construct sets which have an atomic structure and this allows much of the critical analysis to follow the form of the countable chain theory.

This unexpected result is perhaps the major innovation in the analysis of general Markov chains in the last two decades. It was discovered in slightly different forms, independently and virtually simultaneously, by Nummelin [200] and by Athreya and Ney [12].

Although the two methods are almost identical in a formal sense, in what follows we will concentrate on the Nummelin Splitting, touching only briefly on the Athreya-Ney random renewal time method as it fits less well into the techniques of the rest of this book.

5.1.1 Minorization and splitting

To construct the artificial atom or regeneration point involves a probabilistic "splitting" of the state space in such a way that atoms for a "split chain" become natural objects.

In order to carry out this construction we need to consider sets satisfying the following

Minorization Condition

For some $\delta > 0$, some $C \in \mathcal{B}(\mathsf{X})$ and some probability measure ν with $\nu(C^c) = 0$ and $\nu(C) = 1$

$$P(x, A) \geq \delta \mathbb{1}_C(x)\nu(A), \qquad A \in \mathcal{B}(\mathsf{X}),\ x \in \mathsf{X}. \tag{5.2}$$

The form (5.2) ensures that the chain has probabilities uniformly bounded below by multiples of ν for every $x \in C$. The crucial question is, of course, whether any chains ever satisfy the Minorization Condition. This is answered in the positive in Theorem 5.2.2 below: for φ-irreducible chains "small sets" for which the Minorization Condition holds exist, at least for the m-skeleton. The existence of such small sets is a deep and difficult result: by indicating first how the Minorization Condition provides the promised atomic structure to a split chain, we motivate rather more strongly the development of Theorem 5.2.2.

In order to construct a split chain, we split both the space and all measures that are defined on $\mathcal{B}(\mathsf{X})$.

We first split the space X itself by writing $\check{\mathsf{X}} = \mathsf{X} \times \{0,1\}$, where $\mathsf{X}_0 := \mathsf{X} \times \{0\}$ and $\mathsf{X}_1 := \mathsf{X} \times \{1\}$ are thought of as copies of X equipped with copies $\mathcal{B}(\mathsf{X}_0)$, $\mathcal{B}(\mathsf{X}_1)$ of the σ-field $\mathcal{B}(\mathsf{X})$

We let $\mathcal{B}(\check{\mathsf{X}})$ be the σ-field of subsets of $\check{\mathsf{X}}$ generated by $\mathcal{B}(\mathsf{X}_0)$, $\mathcal{B}(\mathsf{X}_1)$: that is, $\mathcal{B}(\check{\mathsf{X}})$ is the smallest σ-field containing sets of the form $A_0 := A \times \{0\}$, $A_1 := A \times \{1\}$, $A \in \mathcal{B}(\mathsf{X})$.

We will write $x_i, i = 0,1$ for elements of $\check{\mathsf{X}}$, with x_0 denoting members of the upper level X_0 and x_1 denoting members of the lower level X_1. In order to describe more easily the calculations associated with moving between the original and the split chain, we will also sometimes call X_0 the *copy* of X, and we will say that $A \in \mathcal{B}(\mathsf{X})$ is a *copy* of the corresponding set $A_0 \subseteq \mathsf{X}_0$.

If λ is any measure on $\mathcal{B}(\mathsf{X})$, then the next step in the construction is to *split* the measure λ into two measures on each of X_0 and X_1 by defining the measure λ^* on $\mathcal{B}(\check{\mathsf{X}})$ through

$$\left.\begin{array}{rcl} \lambda^*(A_0) &=& \lambda(A \cap C)[1-\delta] + \lambda(A \cap C^c), \\ \lambda^*(A_1) &=& \lambda(A \cap C)\delta, \end{array}\right\} \tag{5.3}$$

where δ and C are the constant and the set in (5.2). Note that in this sense the splitting is dependent on the choice of the set C, and although in general the set chosen is not relevant, we will on occasion need to make explicit the set in (5.2) when we use the split chain.

It is critical to note that λ is the marginal measure induced by λ^*, in the sense that for any A in $\mathcal{B}(\mathsf{X})$ we have

$$\lambda^*(A_0 \cup A_1) = \lambda(A). \tag{5.4}$$

In the case when $A \subseteq C^c$, we have $\lambda^*(A_0) = \lambda(A)$; only subsets of C are really effectively split by this construction.

Now the third, and most subtle, step in the construction is to split the chain $\boldsymbol{\Phi}$ to form a chain $\check{\boldsymbol{\Phi}}$ which lives on $(\check{\mathsf{X}}, \mathcal{B}(\check{\mathsf{X}}))$. Define the split kernel $\check{P}(x_i, A)$ for $x_i \in \check{\mathsf{X}}$ and $A \in \mathcal{B}(\check{\mathsf{X}})$ by

$$\check{P}(x_0, \cdot\,) = P(x, \cdot\,)^*, \qquad x_0 \in \mathsf{X}_0 \backslash C_0; \tag{5.5}$$

$$\check{P}(x_0, \cdot\,) = [1-\delta]^{-1}[P(x, \cdot\,)^* - \delta\nu^*(\,\cdot\,)], \qquad x_0 \in C_0; \tag{5.6}$$

$$\check{P}(x_1, \cdot\,) = \nu^*(\,\cdot\,), \qquad x_1 \in \mathsf{X}_1. \tag{5.7}$$

where C, δ and ν are the set, the constant and the measure in the Minorization Condition.

Outside C the chain $\{\check{\Phi}_n\}$ behaves just like $\{\Phi_n\}$, moving on the "top" half X_0 of the split space. Each time it arrives in C, it is "split"; with probability $1-\delta$ it remains in C_0, with probability δ it drops to C_1. We can think of this splitting of the chain as tossing a δ-weighted coin to decide which level to choose on each arrival in the set C where the split takes place.

When the chain remains on the top level its next step has the modified law (5.6). That (5.6) is always non-negative follows from (5.2). This is the sole use of the Minorization Condition, although without it this chain cannot be defined.

Note here the whole point of the construction: the bottom level X_1 is an atom, with $\varphi^*(\mathsf{X}_1) = \delta\varphi(C) > 0$ whenever the chain $\boldsymbol{\Phi}$ is φ-irreducible. By (5.3) we have $\check{P}^n(x_i, \mathsf{X}_1 \backslash C_1) = 0$ for all $n \ge 1$ and all $x_i \in \check{\mathsf{X}}$, so that the atom $C_1 \subseteq \mathsf{X}_1$ is the only part of the bottom level which is reached with positive probability. We will use the notation

$$\check{\boldsymbol{\alpha}} := C_1 \tag{5.8}$$

when we wish to emphasize the fact that all transitions out of C_1 are identical, so that C_1 is an atom in $\check{\mathsf{X}}$.

5.1.2 Connecting the split and original chains

The splitting construction is valuable because of the various properties that $\check{\boldsymbol{\Phi}}$ inherits from, or passes on to, $\boldsymbol{\Phi}$. We give the first of these in the next result.

Theorem 5.1.3 **(i)** *The chain $\boldsymbol{\Phi}$ is the marginal chain of $\{\check{\Phi}_n\}$: that is, for any initial distribution λ on $\mathcal{B}(\mathsf{X})$ and any $A \in \mathcal{B}(\mathsf{X})$,*

$$\int_{\mathsf{X}} \lambda(dx) P^k(x, A) = \int_{\check{\mathsf{X}}} \lambda^*(dy_i) \check{P}^k(y_i, A_0 \cup A_1). \tag{5.9}$$

(ii) *The chain $\boldsymbol{\Phi}$ is φ-irreducible if $\check{\boldsymbol{\Phi}}$ is φ^*-irreducible; and if $\boldsymbol{\Phi}$ is φ-irreducible with $\varphi(C) > 0$ then $\check{\boldsymbol{\Phi}}$ is ν^*-irreducible, and $\check{\boldsymbol{\alpha}}$ is an accessible atom for the split chain.*

PROOF **(i)** From the linearity of the splitting operation we only need to check the equivalence in the special case of $\lambda = \delta_x$, and $k = 1$. This follows by direct computation. We analyze two cases separately.

Suppose first that $x \in C^c$. Then

$$\int_{\check{\mathsf{X}}} \delta_x^*(dy_i) \check{P}(y_i, A_0 \cup A_1) = \check{P}(x_0, A_0 \cup A_1) = P(x, A),$$

by (5.5) and (5.4). On the other hand suppose $x \in C$. Then

$$\begin{aligned}
&\int_{\check{\mathsf{X}}} \delta_x^*(dy_i) \check{P}(y_i, A_0 \cup A_1) \\
&\qquad = (1-\delta)\check{P}(x_0, A_0 \cup A_1) + \delta \check{P}(x_1, A_0 \cup A_1) \\
&\qquad = (1-\delta)\Big[[1-\delta]^{-1}[P^*(x, A_0 \cup A_1) - \delta\nu^*(A_0 \cup A_1)]\Big] + \delta\nu^*(A_0 \cup A_1) \\
&\qquad = P(x, A)
\end{aligned}$$

from (5.6), (5.7) and (5.4) again.

(ii) If the split chain is φ^*-irreducible it is straightforward that the original chain is φ-irreducible from (i). The converse follows from the fact that $\check{\boldsymbol{\alpha}}$ is an accessible atom if $\varphi(C) > 0$, which is easy to check, and Proposition 5.1.1. □

The following identity will prove crucial in later development. For any measure μ on $\mathcal{B}(\mathsf{X})$ we have

$$\int_{\check{\mathsf{X}}} \mu^*(dx_i)\check{P}(x_i, \cdot\,) = \Big(\int_{\mathsf{X}} \mu(dx)P(x, \cdot\,)\Big)^* \tag{5.10}$$

or, using operator notation, $\mu^*\check{P} = (\mu P)^*$. This follows from the definition of the * operation and the transition function $\check{P}$, and is in effect a restatement of Theorem 5.1.3 (i).

Since it is only the marginal chain $\boldsymbol{\Phi}$ which is really of interest, we will usually consider only sets of the form $\check{A} = A_0 \cup A_1$, where $A \in \mathcal{B}(\mathsf{X})$, and we will largely restrict ourselves to functions on $\check{\mathsf{X}}$ of the form $\check{f}(x_i) = f(x_i)$, where f is some function on X; that is, $\check{f}$ is identical on the two copies of X. By (5.9) we have for any k, any initial distribution λ, and any function $\check{f}$ identical on X_0 and X_1

$$\mathsf{E}_\lambda[f(\Phi_k)] = \check{\mathsf{E}}_{\lambda^*}[\check{f}(\check{\Phi}_k)].$$

To emphasize this identity we will henceforth denote $\check{f}$ by f, and $\check{A}$ by A in these special instances. The context should make clear whether A is a subset of X or $\check{\mathsf{X}}$, and whether the domain of f is X or $\check{\mathsf{X}}$.

The Minorization Condition ensures that the construction in (5.6) gives a probability law on $\check{\mathsf{X}}$. A similar construction can also be carried out under the seemingly more general minorization requirement that there exists a function $h(x)$ with $\int h(x)\varphi(dx) > 0$, and a measure $\nu(\cdot)$ on $\mathcal{B}(\mathsf{X})$ such that

$$P(x, A) \geq h(x)\nu(A), \quad x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X}). \tag{5.11}$$

The details are, however, slightly less easy than for the approach we give above although there are some other advantages to the approach through (5.11): the interested reader should consult Nummelin [202] for more details.

The construction of a split chain is of some value in the next several chapters, although much of the analysis will be done directly using the small sets of the next section. The Nummelin Splitting technique will, however, be central in our approach to the asymptotic results of Part III.

5.1.3 A random renewal time approach

There is a second construction of a "pseudo-atom" which is formally very similar to that above. This approach, due to Athreya and Ney [12], concentrates, however, not on a "physical" splitting of the space but on a random renewal time.

If we take the existence of the minorization (5.2) as an assumption, and if we also assume

$$L(x, C) \equiv 1, \quad x \in \mathsf{X} \tag{5.12}$$

we can then construct an almost surely finite random time $\tau \geq 1$ on an enlarged probability space such that $\mathsf{P}_x(\tau < \infty) = 1$ and for every A

$$\mathsf{P}_x(\Phi_n \in A, \tau = n) = \nu(C \cap A)\mathsf{P}_x(\tau = n). \tag{5.13}$$

To construct τ, let $\boldsymbol{\Phi}$ run until it hits C; from (5.12) this happens eventually with probability one. The time and place of first hitting C will be, say, k and x. Then with probability δ distribute Φ_{k+1} over C according to ν; with probability $(1-\delta)$ distribute Φ_{k+1} over the whole space with law $Q(x,\cdot)$, where

$$Q(x,A) = [P(x,A) - \delta\nu(A \cap C)]/(1-\delta);$$

from (5.2) Q is a probability measure, as in (5.6). Repeat this procedure each time $\boldsymbol{\Phi}$ enters C; since this happens infinitely often from (5.12) (a fact yet to be proven in Chapter 9), and each time there is an independent probability δ of choosing ν, it is intuitively clear that sooner or later this version of Φ_k is chosen. Let the time when it occurs be τ. Then $\mathsf{P}_x(\tau < \infty) = 1$ and (5.13) clearly holds; and (5.13) says that τ is a regeneration time for the chain.

The two constructions are very close in spirit: if we consider the split chain construction then we can take the random time τ as $\tau_{\check{\alpha}}$, which is identical to the hitting time on the bottom level of the split space.

There are advantages to both approaches, but the Nummelin Splitting does not require the recurrence assumption (5.12), and more pertinently, it exploits the rather deep fact that some m-skeleton always obeys the Minorization Condition when ψ-irreducibility holds, as we now see.

5.2 Small Sets

In this section we develop the theory of small sets. These are sets for which the Minorization Condition holds, at least for the m-skeleton chain. From the splitting construction of Section 5.1.1, then, it is obvious that the existence of small sets is of considerable importance, since they ensure the splitting method is not vacuous.

Small sets themselves behave, in many ways, analogously to atoms, and in particular the conclusions of Proposition 5.1.1 and Proposition 5.1.2 hold. We will find also many cases where we exploit the "pseudo-atomic" properties of small sets without directly using the split chain.

Small Sets

A set $C \in \mathcal{B}(\mathsf{X})$ is called a *small set* if there exists an $m > 0$, and a non-trivial measure ν_m on $\mathcal{B}(\mathsf{X})$, such that for all $x \in C$, $B \in \mathcal{B}(\mathsf{X})$,

$$P^m(x,B) \geq \nu_m(B). \tag{5.14}$$

When (5.14) holds we say that C is ν_m-small.

The central result (Theorem 5.2.2 below), on which a great deal of the subsequent development rests, is that for a ψ-irreducible chain, every set $A \in \mathcal{B}^+(\mathsf{X})$ contains a small set in $\mathcal{B}^+(\mathsf{X})$. As a consequence, every ψ-irreducible chain admits some m-skeleton which can be split, and for which the atomic structure of the split chain can be exploited.

In order to prove this result, we need for the first time to consider the densities of the transition probability kernels. Being a probability measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ for each individual x and each n, the transition probability kernel $P^n(x, \cdot)$ admits a Lebesgue decomposition into its absolutely continuous and singular parts, with respect to any finite non-trivial measure ϕ on $\mathcal{B}(\mathsf{X})$: we have for any fixed x and $B \in \mathcal{B}(\mathsf{X})$

$$P^n(x, B) = \int_B p^n(x, y)\phi(dy) + P_\perp(x, B). \tag{5.15}$$

where $p^n(x, y)$ is the density of $P^n(x, \cdot)$ with respect to ϕ and $P_\perp$ is orthogonal to ϕ.

Theorem 5.2.1 *Suppose ϕ is a σ-finite measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$. Suppose A is any set in $\mathcal{B}(\mathsf{X})$ with $\phi(A) > 0$ such that*

$$\phi(B) > 0,\ B \subseteq A \ \Rightarrow\ \sum_{k=1}^{\infty} P^k(x, B) > 0, \quad x \in A.$$

Then, for every n, the function p^n defined in (5.15) can be chosen to be a measurable function on X^2, and there exists $C \subseteq A$, $m > 1$, and $\delta > 0$ such that $\phi(C) > 0$ and

$$p^m(x, y) > \delta, \quad x, y \in C. \tag{5.16}$$

Proof We include a detailed proof because of the central place small sets hold in the development of the theory of ψ-irreducible Markov chains. However, the proof is somewhat complex, and may be omitted without interrupting the flow of understanding at this point.

It is a standard result that the densities $p^n(x, y)$ of $P^n(x, \cdot)$ with respect to ϕ exist for each $x \in \mathsf{X}$, and are unique except for definition on ϕ-null sets. We first need to verify that

(i) the densities $p^n(x, y)$ can be chosen jointly measurable in x and y, for each n;

(ii) the densities $p^n(x, y)$ can be chosen to satisfy an appropriate form of the Chapman-Kolmogorov property, namely for n, $m \in \mathbb{Z}_+$, and all x, z

$$p^{n+m}(x, z) \geq \int_{\mathsf{X}} p^n(x, y) p^m(y, z)\phi(dy). \tag{5.17}$$

To see (i), we appeal to the fact that $\mathcal{B}(\mathsf{X})$ is assumed countably generated. This means that there exists a sequence $\{\mathcal{B}_i; i \geq 1\}$ of finite partitions of X, such that $\mathcal{B}_{i+1}$ is a refinement of $\mathcal{B}_i$, and which generate $\mathcal{B}(\mathsf{X})$. Fix $x \in \mathsf{X}$, and let $B_i(x)$ denote the element in $\mathcal{B}_i$ with $x \in B_i(x)$.

For each i, the functions

$$p_i^1(x, y) = \begin{cases} 0 & \phi(B_i(y)) = 0 \\ P(x, B_i(y))/\phi(B_i(y)), & \phi(B_i(y)) > 0 \end{cases}$$

are non-negative, and are clearly jointly measurable in x and y. The Basic Differentiation Theorem for measures (cf. Doob [68], Chapter 7, Section 8) now assures us that for y outside a ϕ-null set N,

$$p^1_\infty(x,y) = \lim_{i\to\infty} p^1_i(x,y) \tag{5.18}$$

exists as a jointly measurable version of the density of $P(x,\cdot)$ with respect to ϕ.

The same construction gives the densities $p^n_\infty(x,y)$ for each n, and so jointly measurable versions of the densities exist as required.

We now define inductively a version $p^n(x,y)$ of the densities satisfying (5.17), starting from $p^n_\infty(x,y)$. Set $p^1(x,y) = p^1_\infty(x,y)$ for all x, y; and set, for $n \geq 2$ and any x, y,

$$p^n(x,y) = p^n_\infty(x,y) \bigvee \max_{1\leq m\leq n-1} \int P^m(x,dw)p^{n-m}(w,y).$$

One can now check (see Orey [208] p 6) that the collection $\{p^n(x,y), x, y \in \mathsf{X}, n \in \mathbb{Z}_+\}$ satisfies both (i) and (ii).

We next verify (5.16). The constraints on ϕ in the statement of Theorem 5.2.1 imply that

$$\sum_{n=1}^\infty p^n(x,y) > 0, \quad x \in A, \qquad \text{a.e}\, y \in A\ [\phi];$$

and thus we can find integers n, m such that

$$\int_A \int_A \int_A p^n(x,y)p^m(y,z)\phi(dx)\phi(dy)\phi(dz) > 0.$$

Now choose $\eta > 0$ sufficiently small that, writing

$$A_n(\eta) := \{(x,y) \in A \times A : p^n(x,y) \geq \eta\}$$

and ϕ^3 for the product measure $\phi \times \phi \times \phi$ on $\mathsf{X} \times \mathsf{X} \times \mathsf{X}$, we have

$$\phi^3\left(\{(x,y,z) \in A \times A \times A : (x,y) \in A_n(\eta), (y,z) \in A_m(\eta)\}\right) > 0.$$

We suppress the notational dependence on η from now on, since η is fixed for the remainder of the proof.

For any x, y, set $B_i(x,y) = B_i(x) \times B_i(y)$, where $B_i(x)$ is again the element containing x of the finite partition $\mathcal{B}_i$ above. By the Basic Differentiation Theorem as in (5.18), this time for measures on $\mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{X})$, there are ϕ^2-null sets $N_k \subseteq \mathsf{X} \times \mathsf{X}$ such that for any k and $(x,y) \in A_k \backslash N_k$,

$$\lim_{i\to\infty} \phi^2(A_k \cap B_i(x,y))/\phi^2(B_i(x,y)) = 1. \tag{5.19}$$

Now choose a fixed triplet (u,v,w) from the set

$$\{(x,y,z) : (x,y) \in A_n\backslash N_n, (y,z) \in A_m\backslash N_m\}.$$

From (5.19) we can find j large enough that

$$\begin{aligned} \phi^2(A_n \cap B_j(u,v)) &\geq (3/4)\phi^2(B_j(u,v)) \\ \phi^2(A_m \cap B_j(v,w)) &\geq (3/4)\phi^2(B_j(v,w)). \end{aligned} \tag{5.20}$$

Let us write $A_n(x) = \{y \in A : (x,y) \in A_n\}$, $A^*_m(z) = \{y \in A : (y,z) \in A_m\}$ for the sections of A_n and A_m in the different directions. If we define

$$E_n = \{x \in A_n \cap B_j(u) : \phi(A_n(x) \cap B_j(v)) \ge (3/4)B_j(v)\} \tag{5.21}$$

$$D_m = \{z \in A_m \cap B_j(w) : \phi(A^*_m(z) \cap B_j(v)) \ge (3/4)B_j(v)\}, \tag{5.22}$$

then from (5.20) we have that $\phi(E_n) > 0$, $\phi(D_m) > 0$. This then implies, for any pair $(x,z) \in E_n \times D_m$,

$$\phi(A_n(x) \cap A^*_m(z)) \ge (1/2)\phi(B_j(v)) > 0 \tag{5.23}$$

from (5.21) and (5.22).

Our pieces now almost fit together. We have, from (5.17), that for $(x,z) \in E_n \times D_m$

$$\begin{aligned} p^{n+m}(x,z) &\ge \int_{A_n(x)\cap A^*_m(z)} p^n(x,y)p^m(y,z)\phi(dy) \\ &\ge \eta^2 \phi(A_n(x) \cap A^*_m(z)) \\ &\ge [\eta^2/2]\phi(B_j(v)) \\ &\ge \delta_1, \text{ say }. \end{aligned} \tag{5.24}$$

To finish the proof, note that since $\phi(E_n) > 0$, there is an integer k and a set $C \subseteq D_m$ with $P^k(x, E_n) > \delta_2 > 0$, for all $x \in C$. It then follows from the construction of the densities above that for all $x, z \in C$

$$\begin{aligned} p^{k+n+m}(x,z) &\ge \int_{E_n} P^k(x,dy)p^{n+m}(y,z) \\ &\ge \delta_1\delta_2, \end{aligned}$$

and the result follows with $\delta = \delta_1\delta_2$ and $M = k + n + m$. □

The key fact proven in this theorem is that we can define a version of the densities of the transition probability kernel such that (5.16) holds uniformly over $x \in C$. This gives us

Theorem 5.2.2 *If $\boldsymbol{\Phi}$ is ψ-irreducible, then for every $A \in \mathcal{B}^+(\mathsf{X})$, there exists $m \ge 1$ and a ν_m-small set $C \subseteq A$ such that $C \in \mathcal{B}^+(\mathsf{X})$ and $\nu_m\{C\} > 0$.*

Proof When $\boldsymbol{\Phi}$ is ψ-irreducible, every set in $\mathcal{B}^+(\mathsf{X})$ satisfies the conditions of Theorem 5.2.1, with the measure $\phi = \psi$. The result then follows immediately from (5.16). □

As a direct corollary of this result we have

Theorem 5.2.3 *If $\boldsymbol{\Phi}$ is ψ-irreducible, then the Minorization Condition holds for some m-skeleton, and for every K_{a_ε}-chain, $0 < \varepsilon < 1$.* □

Any $\boldsymbol{\Phi}$ which is ψ-irreducible is well-endowed with small sets from Theorem 5.2.1, even though it is far from clear from the initial definition that this should be the case. Given the existence of just one small set from Theorem 5.2.2, we now show that it is further possible to cover the whole of X with small sets in the ψ-irreducible case.

Proposition 5.2.4 **(i)** *If $C \in \mathcal{B}(\mathsf{X})$ is ν_n-small, and for any $x \in D$ we have $P^m(x,C) \ge \delta$, then D is ν_{n+m}-small, where ν_{n+m} is a multiple of ν_n.*

(ii) *Suppose Φ is ψ-irreducible. Then there exists a countable collection C_i of small sets in $\mathcal{B}(\mathsf{X})$ such that*

$$\mathsf{X} = \bigcup_{i=0}^{\infty} C_i. \tag{5.25}$$

(iii) *Suppose Φ is ψ-irreducible. If $C \in \mathcal{B}^+(\mathsf{X})$ is ν_n-small, then we may find $M \in \mathbb{Z}_+$ and a measure ν_M such that C is ν_M-small, and $\nu_M\{C\} > 0$.*

PROOF **(i)** By the Chapman-Kolmogorov equations, for any $x \in D$,

$$\begin{aligned} P^{n+m}(x,B) &= \int_{\mathsf{X}} P^n(x,dy)P^m(y,B) \\ &\geq \int_C P^n(x,dy)P^m(y,B) \\ &\geq \delta\nu_n(B). \end{aligned} \tag{5.26}$$

(ii) Since Φ is ψ-irreducible, there exists a ν_m-small set $C \in \mathcal{B}^+(\mathsf{X})$ from Theorem 5.2.2. Moreover from the definition of ψ-irreducibility the sets

$$\bar{C}(n,m) := \{y : P^n(y,C) \geq m^{-1}\} \tag{5.27}$$

cover X and each $\bar{C}(n,m)$ is small from (i).

(iii) Since $C \in \mathcal{B}^+(\mathsf{X})$, we have $K_{a_{\frac{1}{2}}}(x,C) > 0$ for all $x \in \mathsf{X}$. Hence $\nu K_{a_{\frac{1}{2}}}(C) > 0$, and it follows that for some $m \in \mathbb{Z}_+$,

$$\nu_M(C) := \nu P^m(C) > 0.$$

To complete the proof observe that, for all $x \in C$,

$$P^{n+m}(x,B) = \int_{\mathsf{X}} P^n(x,dy)P^m(y,B) \geq \nu P^m(B) = \nu_M(B),$$

which shows that C is ν_M-small, where $M = n + m$. □

5.3 Small Sets for Specific Models

5.3.1 Random walk on a half line

Random walks on a half line provide a simple example of small sets, regardless of the structure of the increment distribution.

It follows as in the proof of Proposition 4.3.1 that every set $[0,c], c \in \mathbb{R}_+$ is small, provided only that $\Gamma(-\infty,0) > 0$: in other words, whenever the chain is ψ-irreducible, every compact set is small. Alternatively, we could derive this result by use of Proposition 5.2.4 (i) since $\{0\}$ is, by definition, small.

This makes the analysis of queueing and storage models very much easier than more general models for which there is no atom in the space. We now move on to identify conditions under which these have identifiable small sets.

5.3.2 "Spread-out" random walks

Let us again consider a random walk $\boldsymbol{\Phi}$ of the form

$$\Phi_n = \Phi_{n-1} + W_n,$$

satisfying (RW1). We showed in Section 4.3 that, if Γ has a density γ with respect to Lebesgue measure μ^{Leb} on $\mathbb{R}$ with

$$\gamma(x) \geq \delta > 0, \qquad |x| < \beta,$$

then $\boldsymbol{\Phi}$ is ψ-irreducible: re-examining the proof shows that in fact we have demonstrated that $C = \{x : |x| \leq \beta/2\}$ is a small set.

Random walks with nonsingular distributions with respect to μ^{Leb}, of which the above are special cases, are particularly well adapted to the ψ-irreducible context. To study them we introduce so-called "spread-out" distributions.

Spread-Out Random Walks

(RW2) We call the random walk *spread-out* (or equivalently, we call Γ spread out) if some convolution power Γ^{n*} is non-singular with respect to μ^{Leb}.

For spread out random walks, we find that small sets are in general relatively easy to find.

Proposition 5.3.1 *If $\boldsymbol{\Phi}$ is a spread-out random walk, with Γ^{n*} non-singular with respect to μ^{Leb} then there is a neighborhood $C_\beta = \{x : |x| \leq \beta\}$ of the origin which is ν_{2n}-small, where $\nu_{2n} = \varepsilon\mu^{\mathrm{Leb}}\mathbb{1}_{[s,t]}$ for some interval $[s,t]$, and some $\varepsilon > 0$.*

PROOF Since Γ is spread out, we have for some bounded non-negative function γ with $\int \gamma(x)\, dx > 0$, and some $n > 0$,

$$P^n(0, A) \geq \int_A \gamma(x)\, dx, \qquad A \in \mathcal{B}(\mathbb{R}).$$

Iterating this we have

$$P^{2n}(0, A) \geq \int_A \int_{\mathbb{R}} \gamma(y)\gamma(x-y)\, dy\, dx = \int_A \gamma * \gamma(x)\, dx : \tag{5.28}$$

but since from Lemma D.4.3 the convolution $\gamma * \gamma(x)$ is continuous and not identically zero, there exists an interval $[a,b]$ and a δ with $\gamma * \gamma(x) \geq \delta$ on $[a,b]$. Choose $\beta =$

$[b-a]/4$, and $[s,t]=[a+\beta,b-\beta]$, to prove the result using the translation invariant properties of the random walk. □

For spread out random walks, a far stronger irreducibility result will be provided in Chapter 6 : there we will show that if $\boldsymbol{\Phi}$ is a random walk with spread-out increment distribution Γ, with $\Gamma(-\infty,0)>0, \Gamma(0,\infty)>0$, then $\boldsymbol{\Phi}$ is μ^{Leb}-irreducible, and every compact set is a small set.

5.3.3 Ladder chains and the GI/G/I queue

Recall from Section 3.5 the Markov chain constructed on $\mathbb{Z}_+ \times \mathbb{R}$ to analyze the GI/G/1 queue, defined by

$$\Phi_n = (N_n, R_n), \quad n \geq 1$$

where N_n is the number of customers at T'_n- and R_n is the residual service time at T'_n+.

This has the transition kernel

$$\begin{aligned} P(i,x;j\times A) &= 0, && j>i+1 \\ P(i,x;j\times A) &= \Lambda_{i-j+1}(x,A), && j=1,\dots,i+1 \\ P(i,x;0\times A) &= \Lambda_i^*(x,A), \end{aligned}$$

where

$$\Lambda_n(x,[0,y]) = \int_0^\infty P_n^t(x,y)G(dt), \tag{5.29}$$

$$\Lambda_n^*(x,[0,y]) = \Big[\sum_{n+1}^{\infty} \Lambda_j(x,[0,\infty))\Big] H[0,y], \tag{5.30}$$

$$P_n^t(x,y) = \mathsf{P}(S'_n \leq t < S'_{n+1}, R_t \leq y \mid R_0 = x); \tag{5.31}$$

here, $R_t = S'_{N(t)+1} - t$, where $N(t)$ is the number of renewals in $[0,t]$ of a renewal process with inter-renewal time H, and if $R_0 = x$ then $S'_1 = x$.

At least one collection of small sets for this chain can be described in some detail.

Proposition 5.3.2 *Let $\boldsymbol{\Phi} = \{N_n, R_n\}$ be the Markov chain at arrival times of a GI/G/1 queue described above. Suppose $G(\beta) < 1$ for all $\beta < \infty$. Then the set $\{0 \times [0,\beta]\}$ is ν_1-small for $\boldsymbol{\Phi}$, with $\nu_1(\,\cdot\,)$ given by $G(\beta,\infty)H(\,\cdot\,)$.*

Proof We consider the bottom "rung" $\{0 \times \mathbb{R}\}$. By construction

$$\Lambda_0^*(x,[0,\,\cdot\,]) = H[0,\,\cdot\,][1-\Lambda_0(x,[0,\infty])],$$

and since

$$\begin{aligned} \Lambda_0(x,[0,\infty)] &= \int G(dt)\mathsf{P}(0 \leq t < \sigma_1 \mid R_0 = x) \\ &= \int G(dt)\mathbb{1}\{t < x\} \\ &= G(-\infty,x], \end{aligned}$$

we have

$$\Lambda_0^*(x,[0,\,\cdot\,]) = H[0,\,\cdot\,]G(x,\infty).$$

The result follows immediately, since for $x<\beta, \Lambda_0^*(x,[0,\,\cdot\,]) \geq H[0,\,\cdot\,]G(\beta,\infty)$. □

5.3.4 The forward recurrence time chain

Consider the forward recurrence time δ-skeleton $\mathbf{V}_\delta^+ = V^+(n\delta), n \in \mathbb{Z}_+$, which was defined in Section 3.5.3: recall that

$$V^+(t) := \inf(Z_n - t : Z_n \geq t), \qquad t \geq 0$$

where $Z_n := \sum_{i=0}^n Y_i$ for $\{Y_1, Y_2, \dots\}$ a sequence of independent and identical random variables with distribution Γ, and Y_0 a further independent random variable with distribution Γ_0.

We shall prove

Proposition 5.3.3 *When Γ is spread out then for δ sufficiently small the set $[0,\delta]$ is a small set for $\mathbf{V}_\delta^+$.*

PROOF As in (5.28), since Γ is spread out there exists $n \in \mathbb{Z}_+$, an interval $[a,b]$ and a constant $\beta > 0$ such that

$$\Gamma^{n*}(du) \geq \beta\mu^{\mathrm{Leb}}(du), \qquad du \subseteq [a,b].$$

Hence if we choose small enough δ then we can find $k \in \mathbb{Z}_+$ such that

$$\Gamma^{n*}(du) \geq \beta \mathbb{1}_{[k\delta,(k+4)\delta]}(u)\mu^{\mathrm{Leb}}(du), \qquad du \subseteq [a,b]. \tag{5.32}$$

Now choose $m \geq 1$ such that $\Gamma[m\delta,(m+1)\delta) = \gamma > 0$; and set $M = k+m+2$. Then for $x \in [0,\delta)$, by considering the occurrence of the n^{th} renewal where n is the index so that (5.32) holds we find

$$\begin{aligned} &\mathsf{P}_x(V^+(M\delta) \in du \cap [0,\delta)) \\ &\quad\geq \mathsf{P}_0(x + Z_{n+1} - M\delta \in du \cap [0,\delta), Y_{n+1} \geq \delta) \\ &\quad= \int_{y\in[\delta,\infty)} \Gamma(dy)\mathsf{P}_0(x + y - M\delta + Z_n \in du \cap [0,\delta)) \\ &\quad\geq \int_{y\in[m\delta,(m+1)\delta)} \Gamma(dy)\mathsf{P}_0(Z_n \in du \cap \{[0,\delta) - x - y + M\delta\}). \end{aligned} \tag{5.33}$$

Now when $y \in [m\delta,(m+1)\delta)$ and $x \in [0,\delta)$, we must have

$$\{[0,\delta) - x - y + M\delta\} \subseteq [k\delta,(k+3)\delta) \tag{5.34}$$

and therefore from (5.33)

$$\begin{aligned} \mathsf{P}_x(V^+(M\delta) \in du \cap [0,\delta)) &\geq \beta\mathbb{1}_{[0,\delta)}(u)\mu^{\mathrm{Leb}}(du)\Gamma(m\delta,(m+1)\delta) \\ &\geq \beta\gamma\mathbb{1}_{[0,\delta)}(u)\mu^{\mathrm{Leb}}(du). \end{aligned} \tag{5.35}$$

Hence $[0,\delta)$ is a small set, and the measure ν can be chosen as a multiple of Lebesgue measure over $[0,\delta)$. $\square$

In this proof we have demanded that (5.32) holds for $u \in [k\delta,(k+4)\delta]$ and in (5.34) we only used the fact that the equation holds for $u \in [k\delta,(k+3)\delta]$. This is not an oversight: we will use the larger range in showing in Proposition 5.4.5 that the chain is also aperiodic.

5.3.5 Linear state space models

For the linear state space LSS(F,G) model we showed in Proposition 4.4.3 that in the Gaussian case when (LSS3) holds, for every initial condition $x_0 \in \mathsf{X} = \mathbb{R}^n$,

$$P^k(x_0, \cdot\,) = N(F^k x_0, \sum_{i=0}^{k-1} F^i G G^\top F^{i\top}); \tag{5.36}$$

and if (F, G) is controllable then from (4.18) the n-step transition function possesses a smooth density $p_n(x, y)$ which is continuous and everywhere positive on $\mathbb{R}^{2n}$. It follows from continuity that for any pair of bounded open balls B_1 and $B_2 \subset \mathbb{R}^n$, there exists $\varepsilon > 0$ such that

$$p_n(x, y) \geq \varepsilon, \qquad (x, y) \in B_1 \times B_2.$$

Letting ν_n denote the normalized uniform distribution on B_2 we see that B_1 is ν_n-small.

This shows that for the controllable, Gaussian LSS(F,G) model, all compact subsets of the state space are small.

5.4 Cyclic Behavior

5.4.1 The cycle phenomenon

In the previous sections of this chapter we concentrated on the communication structure between states. Here we consider the set of time-points at which such communication is possible; for even within a communicating class, it is possible that the chain returns to given states only at specific time points, and this certainly governs the detailed behavior of the chain in any longer term analysis.

A highly artificial example of cyclic behavior on the finite set $\mathsf{X} = \{1, 2, 3, \dots, d\}$ is given by the transition probability matrix

$$P(x, x+1) = 1, \quad x \in \{1, 2, 3, \dots, d-1\}, \qquad P(d, 1) = 1.$$

Here, if we start in x then we have $P^n(x, x) > 0$ if and only if $n = 0, d, 2d, \dots$, and the chain $\boldsymbol{\Phi}$ is said to *cycle* through the states of X.

On a continuous state space the same phenomenon can be constructed equally easily: let $\mathsf{X} = [0, d)$, let U_i denote the uniform distribution on $[i, i+1)$, and define

$$P(x, \cdot) := \mathbb{1}_{[i-1,i)}(x) U_i(\cdot), \quad i = 0, 1, \dots, d-1 \pmod{d}.$$

In this example, the chain again cycles through a fixed finite number of sets. We now prove a series of results which indicate that, no matter how complex the behavior of a ψ-irreducible chain, or a chain on an irreducible absorbing set, the finite cyclic behavior of these examples is typical of the worst behavior to be found.

5.4.2 Cycles for a countable space chain

We discuss this structural question initially for a countable space X.

Let $\boldsymbol{\alpha}$ be a specific state in X, and write

$$d(\boldsymbol{\alpha}) = g.c.d.\{n \ge 1 : P^n(\boldsymbol{\alpha}, \boldsymbol{\alpha}) > 0\}. \tag{5.37}$$

This does not guarantee that $P^{md(\alpha)}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) > 0$ for all m, but it does imply $P^n(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 0$ unless $n = md(\boldsymbol{\alpha})$, for some m.

We call $d(\boldsymbol{\alpha})$ the *period* of $\boldsymbol{\alpha}$. The result we now show is that the value of $d(\boldsymbol{\alpha})$ is common to all states y in the class $C(\boldsymbol{\alpha}) = \{y : \boldsymbol{\alpha} \leftrightarrow y\}$, rather than taking a separate value for each y.

Proposition 5.4.1 *Suppose $\boldsymbol{\alpha}$ has period $d(\boldsymbol{\alpha})$: then for any $y \in C(\boldsymbol{\alpha})$, $d(\boldsymbol{\alpha}) = d(y)$.*

PROOF Since $\boldsymbol{\alpha} \leftrightarrow y$, we can find m and n such that $P^m(\boldsymbol{\alpha}, y) > 0$ and $P^n(y, \boldsymbol{\alpha}) > 0$. By the Chapman-Kolmogorov equations, we have

$$P^{m+n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \ge P^m(\boldsymbol{\alpha}, y)P^n(y, \boldsymbol{\alpha}) > 0, \tag{5.38}$$

and so by definition, $(m+n)$ is a multiple of $d(\boldsymbol{\alpha})$. Choose k such that k is not a multiple of $d(\boldsymbol{\alpha})$. Then $(k+m+n)$ is not a multiple of $d(\boldsymbol{\alpha})$: hence, since

$$P^m(\boldsymbol{\alpha}, y)P^k(y, y)P^n(y, \boldsymbol{\alpha}) \le P^{k+m+n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 0,$$

we have $P^k(y, y) = 0$, which proves $d(y) \ge d(\boldsymbol{\alpha})$. Reversing the role of $\boldsymbol{\alpha}$ and y shows $d(\boldsymbol{\alpha}) \ge d(y)$, which gives the result. □

This result leads to a further decomposition of the transition probability matrix for an irreducible chain; or, equivalently, within a communicating class.

Proposition 5.4.2 *Let $\boldsymbol{\Phi}$ be an irreducible Markov chain on a countable space, and let d denote the common period of the states in X. Then there exist disjoint sets $D_1 \ldots D_d \subseteq \mathsf{X}$ such that*

$$\mathsf{X} = \bigcup_{i=1}^{d} D_k,$$

and

$$P(x, D_{k+1}) = 1, \qquad x \in D_k, \quad k = 0, \ldots, d-1 \pmod{d}. \tag{5.39}$$

PROOF The proof is similar to that of the previous proposition. Choose $\boldsymbol{\alpha} \in \mathsf{X}$ as a distinguished state, and let y be another state, such that for some M

$$P^M(y, \boldsymbol{\alpha}) > 0.$$

Let k be any other integer such that $P^k(\boldsymbol{\alpha}, y) > 0$. Then $P^{k+M}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) > 0$, and thus $k + M = jd$ for some j; equivalently, $k = jd - M$. Now M is fixed, and so we must have $P^k(\boldsymbol{\alpha}, y) > 0$ only for k in the sequence $\{r, r+d, r+2d, \ldots\}$, where the integer $r = r(y) \in \{1, \ldots, d\}$ is uniquely defined for y.

Call D_r the set of states which are reached with positive probability from $\boldsymbol{\alpha}$ only at points in the sequence $\{r, r+d, r+2d, \ldots\}$ for each $r \in \{1, 2 \ldots d\}$. By definition $\boldsymbol{\alpha} \in D_d$, and $P(\boldsymbol{\alpha}, D_1^c) = 0$ so that $P(\boldsymbol{\alpha}, D_1) = 1$. Similarly, for any $y \in D_r$ we have $P(y, D_{r+1}^c) = 0$, giving our result. □

The sets $\{D_i\}$ covering X and satisfying (5.39) are called *cyclic classes*, or a *d-cycle*, of $\boldsymbol{\Phi}$. With probability one, each sample path of the process $\boldsymbol{\Phi}$ "cycles" through values in the sets $D_1, D_2, \ldots D_d, D_1, D_2, \ldots$.

Diagrammatically, we have shown that we can write an irreducible transition probability matrix in "super-diagonal" form

$$P = \begin{bmatrix} 0 & P_1 & & & \\ 0 & 0 & P_2 & & 0 \\ \vdots & \ddots & 0 & P_3 & \\ \vdots & \vdots & \ddots & 0 & \ddots \\ P_d & \dots & \dots & \dots & 0 \end{bmatrix}$$

where each block P_i is a square matrix whose dimension may depend upon i.

> Aperiodicity
>
> An irreducible chain on a countable space X is called
>
> **(i)** *aperiodic*, if $d(x) \equiv 1$, $x \in \mathsf{X}$;
>
> **(ii)** *strongly aperiodic*, if $P(x,x) > 0$ for some $x \in \mathsf{X}$.

Whilst cyclic behavior can certainly occur, as illustrated in the examples at the beginning of this section, and the periodic behavior of the control systems in Theorem 7.3.3 below, most of our results will be given for aperiodic chains. The justification for using such chains is contained in the following, whose proof is obvious.

Proposition 5.4.3 *Suppose $\boldsymbol{\Phi}$ is an irreducible chain on a countable space X, with period d and cyclic classes $\{D_1 \dots D_d\}$. Then for the Markov chain $\boldsymbol{\Phi}_d = \{\Phi_d, \Phi_{2d}, \dots\}$ with transition matrix P^d, each D_i is an irreducible absorbing set of aperiodic states.*

5.4.3 Cycles for a general state space chain

The existence of small sets enables us to show that, even on a general space, we still have a finite periodic breakup into cyclic sets for ψ-irreducible chains.

Suppose that C is any ν_M-small set, and assume that $\nu_M(C) > 0$, as we may without loss of generality by Proposition 5.2.4.

We will use the set C and the corresponding measure ν_M to define a cycle for a general irreducible Markov chain. To simplify notation we will suppress the subscript on ν. Hence we have $P^M(x, \cdot\,) \geq \nu(\,\cdot\,)$, $x \in C$, and $\nu(C) > 0$, so that, when the chain starts in C, there is a positive probability that the chain will return to C at time M. Let

$$E_C = \{n \geq 1 : \text{ the set } C \text{ is } \nu_n\text{-small, with } \nu_n = \delta_n \nu \text{ for some } \delta_n > 0.\} \tag{5.40}$$

be the set of timepoints for which C is a small set with minorizing measure proportional to ν. Notice that for $B \subseteq C$, $n, m \in E_C$ implies

$$\begin{aligned} P^{n+m}(x,B) &\geq \int_C P^m(x,dy)P^n(y,B) \\ &\geq [\delta_m \delta_n \nu(C)]\nu(B), \quad x \in C; \end{aligned}$$

so that E_C is closed under addition. Thus there is a natural "period" for the set C, given by the greatest common divisor of E_C; and from Lemma D.7.4, C is ν_{nd}-small for all large enough n.

We show that this value is in fact a property of the whole chain $\boldsymbol{\Phi}$, and is independent of the particular small set chosen, in the following analogue of Proposition 5.4.2.

Theorem 5.4.4 *Suppose that $\boldsymbol{\Phi}$ is a ψ-irreducible Markov chain on X. Let $C \in \mathcal{B}(\mathsf{X})^+$ be a ν_M-small set and let d be the greatest common divisor of the set E_C. Then there exist disjoint sets $D_1 \ldots D_d \in \mathcal{B}(\mathsf{X})$ (a "d-cycle") such that*

(i) *for $x \in D_i$, $P(x, D_{i+1}) = 1$, $i = 0 \ldots d-1 \pmod d$;*

(ii) *the set $N = [\bigcup_{i=1}^d D_i]^c$ is ψ-null.*

The d-cycle $\{D_i\}$ is maximal in the sense that for any other collection $\{d', D'_k, k = 1, \ldots d'\}$ satisfying (i)-(ii), we have d' dividing d; whilst if $d = d'$, then, by reordering the indices if necessary, $D'_i = D_i$ a.e. ψ.

Proof For $i = 0, 1 \ldots d-1$ set

$$D_i^* = \left\{ y : \sum_{n=1}^{\infty} P^{nd-i}(y,C) > 0 \right\} :$$

by irreducibility, $\mathsf{X} = \cup D_i^*$.

The D_i^* are in general not disjoint, but we can show that their intersection is ψ-null. For suppose there exists i,k such that $\psi(D_i^* \cap D_k^*) > 0$. Then for some fixed $m, n > 0$, there is a subset $A \subseteq D_i^* \cap D_k^*$ with $\psi(A) > 0$ such that

$$\begin{aligned} P^{md-i}(w,C) &\geq \delta_m > 0, \qquad w \in A \\ P^{nd-k}(w,C) &\geq \delta_n > 0, \qquad w \in A \end{aligned} \tag{5.41}$$

and since ψ is the maximal irreducibility measure, we can also find r such that

$$\int_C \nu(dy) P^r(y,A) = \delta_c > 0. \tag{5.42}$$

Now we use the fact that C is a ν_M-small set: for $x \in C$, $B \subseteq C$, from (5.41), (5.42),

$$\begin{aligned} P^{2M+md-i+r}(x,B) &\geq \int_C P^M(x,dy) \int_A P^r(y,dw) \int_C P^{md-i}(w,dz) P^M(z,B) \\ &\geq [\delta_c \delta_m]\nu(B), \end{aligned}$$

so that $[2M+md+r]-i \in E_C$. By identical reasoning, we also have $[2M+nd+r]-k \in E_C$. This contradicts the definition of d, and we have shown that $\psi(D_i^* \cap D_k^*) = 0$, $i \neq k$.

Let $N = \cup_{i,j}(D_i^* \cap D_k^*)$, so that $\psi(N) = 0$. The sets $\{D_i^* \backslash N\}$ form a disjoint class of sets whose union is full. By Proposition 4.2.3, we can find an absorbing set D such that $D_i = D \cap (D_i^* \backslash N)$ are disjoint and $D = \cup D_i$. By the Chapman-Kolmogorov equations again, if $x \in D$ is such that $P(x, D_j) > 0$, then we have $x \in D_{j-1}$, by definition, for $j = 0, \ldots, d-1$ (mod d). Thus $\{D_i\}$ is a d-cycle.

To prove the maximality and uniqueness result, suppose $\{D_i'\}$ is another cycle with period d', with $N = [\cup D_i']^c$ such that $\psi(N) = 0$. Let k be any index with $\nu(D_k' \cap C) > 0$: since $\psi(N) = 0$ and $\psi \succ \nu$, such a k exists. We then have, since C is a ν_M-small set, $P^M(x, D_k' \cap C) \geq \nu(D_k' \cap C) > 0$ for every $x \in C$. Since $(D_k' \cap C)$ is non-empty, this implies firstly that M is a multiple of d'; since this happens for any $n \in E_C$, by definition of d we have d' divides d as required. Also, we must have $C \cap D_j'$ empty for any $j \neq k$: for if not we would have some $x \in C$ with $P^M(x, C \cap D_k') = 0$, which contradicts the properties of C.

Hence we have $C \subseteq (D_k' \cup N)$, for some particular k. It follows by the definition of the original cycle that each D_j' is a union up to ψ-null sets of (d/d_i) elements of D_i. □

It is obvious from the above proof that the cycle does not depend, except perhaps for ψ-null sets, on the small set initially chosen, and that any small set must be essentially contained inside one specific member of the cyclic class $\{D_i\}$.

Periodic and aperiodic chains

Suppose that $\boldsymbol{\Phi}$ is a φ-irreducible Markov chain.

The largest d for which a d-cycle occurs for $\boldsymbol{\Phi}$ is called the *period* of $\boldsymbol{\Phi}$.

When $d = 1$, the chain $\boldsymbol{\Phi}$ is called *aperiodic.*

When there exists a ν_1-small set A with $\nu_1(A) > 0$, then the chain is called *strongly aperiodic.*

As a direct consequence of these definitions and Theorem 5.2.3 we have

Proposition 5.4.5 *Suppose that $\boldsymbol{\Phi}$ is a ψ-irreducible Markov chain.*

(i) *If $\boldsymbol{\Phi}$ is strongly aperiodic, then the Minorization Condition (5.2) holds.*

(ii) *The resolvent, or K_{a_ε}-chain, is strongly aperiodic for all $0 < \varepsilon < 1$.*

(iii) *If $\boldsymbol{\Phi}$ is aperiodic then every skeleton is ψ-irreducible and aperiodic, and some m-skeleton is strongly aperiodic.*

□

This result shows that it is clearly desirable to work with strongly aperiodic chains. Regrettably, this condition is not satisfied in general, even for simple chains; and we will often have to prove results for strongly aperiodic chains and then use special methods to extend them to general chains through the m-skeleton or the K_{a_ε}-chain.

We will however concentrate almost exclusively on aperiodic chains. In practice this is not greatly restrictive, since we have as in the countable case

Proposition 5.4.6 *Suppose $\boldsymbol{\Phi}$ is a ψ-irreducible chain with period d and d-cycle $\{D_i, i = 1 \dots d\}$. Then each of the sets D_i is an absorbing ψ-irreducible set for the chain $\boldsymbol{\Phi}_d$ corresponding to the transition probability kernel P^d, and $\boldsymbol{\Phi}_d$ on each D_i is aperiodic.*

PROOF That each D_i is absorbing and irreducible for $\boldsymbol{\Phi}_d$ is obvious: that $\boldsymbol{\Phi}_d$ on each D_i is aperiodic follows from the definition of d as the largest value for which a cycle exists. □

5.4.4 Periodic and aperiodic examples: forward recurrence times

For the forward recurrence time chain on the integers it is easy to evaluate the period of the chain. For let p be the distribution of the renewal variables, and let

$$d = g.c.d.\{n : p(n) > 0\}.$$

It is a simple exercise to check that d is also the g.c.d. of the set of times $\{n : P^n(0,0) > 0\}$ and so d is the period of the chain.

Now consider the forward recurrence time δ-skeleton $\mathbf{V}_\delta^+ = V^+(n\delta)$, $n \in \mathbb{Z}_+$ defined in Section 3.5.3. Here, we can find explicit conditions for aperiodicity even though the chain has no atom in the space. We have

Proposition 5.4.7 *If F is spread out then $\mathbf{V}_\delta^+$ is aperiodic for sufficiently small δ.*

PROOF In Proposition 5.3.3 we showed that for sufficiently small δ, the set $[0, \delta)$ is a ν_M-small set, where ν is a multiple of Lebesgue measure restricted to $[0, \delta]$.

But since the bounds on the densities in (5.35) hold, not just for the range $[k\delta, (k+3)\delta)$ for which they were used, but by construction for the greater range $[k\delta, (k+4)\delta)$, the same proof shows that $[0, \delta)$ is a ν_{M+1}-small set also, and thus aperiodicity follows from the definition of the period of $\mathbf{V}_\delta^+$ as the g.c.d. in (5.40). □

5.5 Petite Sets and Sampled Chains

5.5.1 Sampling a Markov chain

A convenient tool for the analysis of Markov chains is the *sampled chain*, which extends substantially the idea of the m-skeleton or the resolvent chain.

Let $a = \{a(n)\}$ be a distribution, or probability measure, on $\mathbb{Z}_+$, and consider the Markov chain $\boldsymbol{\Phi}_a$ with probability transition kernel

$$K_a(x, A) := \sum_{n=0}^{\infty} P^n(x, A) a(n), \qquad x \in \mathsf{X}, A \in \mathcal{B}(\mathsf{X}). \tag{5.43}$$

It is obvious that K_a is indeed a transition kernel, so that $\boldsymbol{\Phi}_a$ is well-defined by Theorem 3.4.1.

We will call $\boldsymbol{\Phi}_a$ the *K_a-chain*, with *sampling distribution* a. Probabilistically, $\boldsymbol{\Phi}_a$ has the interpretation of being the chain $\boldsymbol{\Phi}$ "sampled" at time-points drawn successively according to the distribution a, or more accurately, at time-points of an independent renewal process with increment distribution a as defined in Section 2.4.1.

There are two specific sampled chains which we have already invoked, and which will be used frequently in the sequel. If $a = \delta_m$ is the Dirac measure with $\delta_m(m) = 1$, then the K_{δ_m}-chain is the m-skeleton with transition kernel P^m. If a_ε is the geometric distribution with

$$a_\varepsilon(n) = [1-\varepsilon]\varepsilon^n, \qquad n \in \mathbb{Z}_+$$

then the kernel K_{a_ε} is the resolvent K_ε which was defined in Chapter 3. The concept of sampled chains immediately enables us to develop useful conditions under which one set is uniformly accessible from another. We say that a set $B \in \mathcal{B}(\mathsf{X})$ is *uniformly accessible using* a from another set $A \in \mathcal{B}(\mathsf{X})$ if there exists a $\delta > 0$ such that

$$\inf_{x \in A} K_a(x, B) > \delta; \tag{5.44}$$

and when (5.44) holds we write $A \overset{a}{\rightsquigarrow} B$.

Lemma 5.5.1 *If $A \overset{a}{\rightsquigarrow} B$ for some distribution a then $A \rightsquigarrow B$.*

Proof Since $L(x, B) = \mathsf{P}_x(\tau_B < \infty) = \mathsf{P}_x(\Phi_n \in B \text{ for some } n \in \mathbb{Z}_+)$ and $K_a(x, B) = \mathsf{P}_x(\Phi_\eta \in B)$ where η has the distribution a, it follows that

$$L(x, B) \geq K_a(x, B) \tag{5.45}$$

for any distribution a, and the result follows. □

The following relationships will be used frequently.

Lemma 5.5.2 **(i)** *If a and b are distributions on $\mathbb{Z}_+$ then the sampled chains with transition laws K_a and K_b satisfy the generalized Chapman-Kolmogorov equations*

$$K_{a*b}(x, A) = \int K_a(x, dy) K_b(y, A) \tag{5.46}$$

where $a * b$ denotes the convolution of a and b.

(ii) *If $A \overset{a}{\rightsquigarrow} B$ and $B \overset{b}{\rightsquigarrow} C$, then $A \overset{a*b}{\rightsquigarrow} C$.*

(iii) *If a is a distribution on $\mathbb{Z}_+$ then the sampled chain with transition law K_a satisfies the relation*

$$U(x, A) \geq \int U(x, dy) K_a(y, A) \tag{5.47}$$

PROOF To see (i), observe that by definition and the Chapman-Kolmogorov equation

$$\begin{aligned}
K_{a*b}(x,A) &= \sum_{n=0}^{\infty} P^n(x,A)\, a*b(n) \\
&= \sum_{n=0}^{\infty} P^n(x,A) \sum_{m=0}^{n} a(m)b(n-m) \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{n} \int P^m(x,dy)P^{n-m}(y,A)a(m)b(n-m) \\
&= \int \sum_{m=0}^{\infty} P^m(x,dy)a(m) \sum_{n=m}^{\infty} P^{n-m}(y,A)b(n-m) \\
&= \int K_a(x,dy)K_b(yA), \qquad (5.48)
\end{aligned}$$

as required.

The result (ii) follows directly from (5.46) and the definitions.

For (iii), note that for fixed m,n,

$$P^{m+n}(x,A)a(n) = \int P^m(x,dy)P^n(y,A)a(n)$$

so that summing over m gives

$$U(x,A)a(n) \geq \sum_{m>n} P^m(x,A)a(n) = \int U(x,dy)P^n(y,A)a(n);$$

a second summation over n gives the result since $\sum_n a(n) = 1$. □

The probabilistic interpretation of Lemma 5.5.2 (i) is simple: if the chain is sampled at a random time $\eta = \eta_1 + \eta_2$, where η_1 has distribution a and η_2 has independent distribution b, then since η has distribution $a*b$, it follows that (5.46) is just a Chapman-Kolmogorov decomposition at the intermediate random time.

5.5.2 The property of petiteness

Small sets always exist in the ψ-irreducible case, and provide most of the properties we need. We now introduce a generalization of small sets, petite sets, which have even more tractable properties, especially in topological analyses.

> Petite Sets
>
> We will call a set $C \in \mathcal{B}(\mathsf{X})$ ν_a-*petite* if the sampled chain satisfies the bound
>
> $$K_a(x,B) \geq \nu_a(B),$$
>
> for all $x \in C$, $B \in \mathcal{B}(\mathsf{X})$, where ν_a is a non-trivial measure on $\mathcal{B}(\mathsf{X})$.

From the definitions we see that a small set is petite, with the sampling distribution a taken as δ_m for some m. Hence the property of being a small set is in general stronger than the property of being petite. We state this formally as

Proposition 5.5.3 *If $C \in \mathcal{B}(\mathsf{X})$ is ν_m-small then C is ν_{δ_m}-petite.* □

The operation "$\overset{a}{\leadsto}$" interacts usefully with the petiteness property. We have

Proposition 5.5.4 **(i)** *If $A \in \mathcal{B}(\mathsf{X})$ is ν_a-petite, and $D \overset{b}{\leadsto} A$ then D is ν_{b*a}-petite, where ν_{b*a} can be chosen as a multiple of ν_a.*

(ii) *If $\mathbf{\Phi}$ is ψ-irreducible and if $A \in \mathcal{B}^+(\mathsf{X})$ is ν_a-petite, then ν_a is an irreducibility measure for $\mathbf{\Phi}$.*

PROOF To prove (i) choose $\delta > 0$ such that for $x \in D$ we have $K_b(x, A) \geq \delta$. By Lemma 5.5.2 (i),

$$\begin{aligned} K_{b*a}(x, B) &= \int_{\mathsf{X}} K_b(x, dy) K_a(y, B) \\ &\geq \int_A K_b(x, dy) K_a(y, B) \qquad (5.49) \\ &\geq \delta \nu_a(B). \end{aligned}$$

To see (ii), suppose A is ν_a-petite and $\nu_a(B) > 0$. For $x \in \overline{A}(n, m)$ as in (5.27) we have

$$P^n K_a(x, B) \geq \int_A P^n(x, dy) K_a(y, B) \geq m^{-1} \nu_a(B) > 0$$

which gives the result. □

Proposition 5.5.4 provides us with a prescription for generating an irreducibility measure from a petite set A, even if all we know for general $x \in \mathsf{X}$ is that the single petite set A is reached with positive probability. We see the value of this in the examples later in this chapter

The following result illustrates further useful properties of petite sets, which distinguish them from small sets.

Proposition 5.5.5 *Suppose $\mathbf{\Phi}$ is ψ-irreducible.*

(i) *If A is ν_a-petite, then there exists a sampling distribution b such that A is also ψ_b-petite where ψ_b is a maximal irreducibility measure.*

(ii) *The union of two petite sets is petite.*

(iii) *There exists a sampling distribution c, an everywhere strictly positive, measurable function $s : \mathsf{X} \to \mathbb{R}$, and a maximal irreducibility measure ψ_c such that*

$$K_c(x, B) \geq s(x) \psi_c(B), \qquad x \in \mathsf{X},\ B \in \mathcal{B}(\mathsf{X})$$

Thus there is an increasing sequence $\{C_i\}$ of ψ_c-petite sets, all with the same sampling distribution c and minorizing measure equivalent to ψ, with $\cup C_i = \mathsf{X}$.

Proof To prove (i) we first show that we can assume without loss of generality that ν_a is an irreducibility measure, even if $\psi(A) = 0$.

From Proposition 5.2.4 there exists a ν_b-petite set C with $C \in \mathcal{B}^+(\mathsf{X})$. We have $K_{a_\varepsilon}(y, C) > 0$ for any $y \in \mathsf{X}$ and any $\varepsilon > 0$, and hence for $x \in A$,

$$K_{a*a_\varepsilon}(x, C) \geq \int \nu_a(dy) K_{a_\varepsilon}(y, C) > 0.$$

This shows that $A \overset{a*a_\varepsilon}{\rightsquigarrow} C$, and hence from Proposition 5.5.4 we see that A is $\nu_{a*a_\varepsilon*b}$-petite, where $\nu_{a*a_\varepsilon*b}$ is a constant multiple of ν_b. Now, from Proposition 5.5.4 (ii), the measure $\nu_{a*a_\varepsilon*b}$ is an irreducibility measure, as claimed.

We now assume that ν_a is an irreducibility measure, which is justified by the discussion above, and use Proposition 5.5.2 (i) to obtain the bound, valid for any $0 < \varepsilon < 1$,

$$K_{a*a_\varepsilon}(x, B) = K_a K_{a_\varepsilon}(x, B) \geq \nu_a K_{a_\varepsilon}(B), \qquad x \in A, \qquad B \in \mathcal{B}(\mathsf{X}).$$

Hence A is ψ_b-petite with $b = a_\varepsilon * a$ and $\psi_b = \nu_a K_{a_\varepsilon}$. Proposition 4.2.2 (iv) asserts that, since ν_a is an irreducibility measure, the measure ψ_b is a maximal irreducibility measure.

To see (ii), suppose that A_1 is ψ_{a_1}-petite, and that A_2 is ψ_{a_2}-petite. Let $A_0 \in \mathcal{B}^+(\mathsf{X})$ be a fixed petite set and define the sampling measure a on $\mathbb{Z}_+$ as $a(i) = \frac{1}{2}[a_1(i) + a_2(i)]$, $i \in \mathbb{Z}_+$.

Since both ψ_{a_1} and ψ_{a_2} can be chosen as maximal irreducibility measures, it follows that for $x \in A_1 \cup A_2$

$$K_a(x, A_0) \geq \tfrac{1}{2} \min(\psi_{a_1}(A_0), \psi_{a_2}(A_0)) > 0$$

so that $A_1 \cup A_2 \overset{a}{\rightsquigarrow} A_0$. From Proposition 5.5.4 we see that $A_1 \cup A_2$ is petite.

For (iii), first apply Theorem 5.2.2 to construct a ν_n-small set $C \in \mathcal{B}^+(\mathsf{X})$. By (i) above we may assume that C is ψ_b-petite with ψ_b a maximal irreducibility measure. Hence $K_b(y, \cdot\,) \geq \mathbb{1}_C(y)\psi_b(\,\cdot\,)$ for all $y \in \mathsf{X}$.

By irreducibility and the definitions we also have $K_{a_\varepsilon}(x, C) > 0$ for all $0 < \varepsilon < 1$, and all $x \in \mathsf{X}$. Combining these bounds gives for any $x \in \mathsf{X}$, $B \in \mathcal{B}(\mathsf{X})$,

$$K_{b*a_\varepsilon}(x, B) \geq \int_C K_{a_\varepsilon}(y, dz) K_b(z, B) \geq K_{a_\varepsilon}(x, C)\psi_b(B)$$

which shows that (iii) holds with $c = b * a_\varepsilon$, $s(x) = K_{a_\varepsilon}(x, C)$ and $\psi_c = \psi_b$.

The petite sets forming the countable cover can be taken as $C_m := \{x \in \mathsf{X} : s(x) \geq m^{-1}\}$, $m \geq 1$. □

Clearly the result in (ii) is best possible, since the whole space is a countable union of small (and hence petite) sets from Proposition 5.2.4, yet is not necessarily petite itself.

Our next result is interesting of itself, but is more than useful as a tool in the use of petite sets.

Proposition 5.5.6 *Suppose that $\boldsymbol{\Phi}$ is ψ-irreducible and that C is ν_a-petite.*

(i) *Without loss of generality we can take a to be either a uniform sampling distribution $a_m(i) = 1/m$, $1 \leq i \leq m$, or a to be the geometric sampling distribution a_ε. In either case, there is a finite mean sampling time*

$$m_a = \sum_i i a(i).$$

(ii) *If Φ is strongly aperiodic then the set $C_0 \cup C_1 \subseteq \check{\mathsf{X}}$ corresponding to C is ν_a^*-petite for the split chain $\check{\Phi}$.*

Proof To see (i), let $A \in \mathcal{B}^+(\mathsf{X})$ be ν_n-small. By Proposition 5.5.5 (i) we have

$$K_b(x, A) \geq \psi_b(A) > 0, \qquad x \in C$$

where ψ_b is a maximal irreducibility measure. Hence $\sum_{k=1}^{N} P^k(x, A) \geq \frac{1}{2}\psi_b(A)$, $x \in C$, for some N sufficiently large.

Since A is ν_n-small, it follows that for any $B \in \mathcal{B}(\mathsf{X})$,

$$\sum_{k=1}^{N+n} P^k(x, B) \geq \sum_{k=1}^{N} P^{k+n}(x, B) \geq \tfrac{1}{2}\psi_b(A)\nu_n(B)$$

for $x \in C$. This shows that C is ν_a-petite with $a(k) = (N+n)^{-1}$ for $1 \leq k \leq N+n$. Since for all ε and m there exists some constant c such that $a_\varepsilon(j) \geq c a_m(j)$, $j \in \mathbb{Z}_+$, this proves (i).

To see (ii), suppose that the chain is split with the small set $A \in \mathcal{B}^+(\mathsf{X})$. Then $A_0 \cup \mathsf{X}_1$ is also petite: for X_1 is small, and A_0 is also small since $\check{P}(x, \mathsf{X}_1) \geq \delta$ for $x_0 \in A_0$, and we know that the union of petite sets is petite, by Proposition 5.5.5.

Since when $x_0 \in A_0^c$ we have for $n \geq 1$, $\check{P}^n(x_0, A_0 \cup \mathsf{X}_1) = \check{P}^n(x_0, A_0 \cup A_1) = P^n(x, A)$ it follows that

$$\check{K}_a(x_0, A_0 \cup \mathsf{X}_1) = \sum_{j=0}^{\infty} a(j)\check{P}^j(x_0, A_0 \cup X_1)$$

is uniformly bounded from below for $x_0 \in C_0 \setminus A_0$, which shows that $C_0 \setminus A_0$ is petite. Since the union of petite sets is petite, $C_0 \cup \mathsf{X}_1$ is also petite. □

5.5.3 Petite sets and aperiodicity

If A is a petite set for a ψ-irreducible Markov chain then the corresponding minorizing measure can always be taken to be equal to a maximal irreducibility measure, although the measure ν_m appropriate to a small set is not as large as this.

We now prove that in the ψ-irreducible aperiodic case, every petite set is also small for an appropriate choice of m and ν_m.

Theorem 5.5.7 *If Φ is irreducible and aperiodic then every petite set is small.*

Proof Let A be a petite set. From Proposition 5.5.5 we may assume that A is ψ_a-petite, where ψ_a is a maximal irreducibility measure.

Let C denote the small set used in (5.40). Since the chain is aperiodic, it follows from Theorem 5.4.4 and Lemma D.7.4 that for some $n_0 \in \mathbb{Z}_+$, the set C is ν_k-small, with $\nu_k = \delta\nu$ for some $\delta > 0$, for all $n_0/2 - 1 \leq k \leq n_0$.

Since $C \in \mathcal{B}^+(\mathsf{X})$, we may also assume that n_0 is so large that

$$\sum_{k=\lfloor n_0/2 \rfloor}^{\infty} a(k) \leq \tfrac{1}{2}\psi_a(C).$$

With n_0 so fixed, we have for all $x \in A$ and $B \in \mathcal{B}(\mathsf{X})$,

$$
\begin{aligned}
P^{n_0}(x,B) &\geq \sum_{k=0}^{\lceil n_0/2 \rceil} \Big\{ \int_C P^k(x,dy) P^{n_0-k}(y,B) \Big\} a(k) \\
&\geq \Big(\sum_{k=0}^{\lceil n_0/2 \rceil} P^k(x,C) a(k) \Big) \Big(\delta \nu(B) \Big) \\
&\geq \Big(\tfrac{1}{2} \psi_a(C) \Big) \Big(\delta \nu(B) \Big)
\end{aligned}
$$

which shows that A is ν_{n_0}-small, with $\nu_{n_0} = (\frac{1}{2}\delta\psi_a(C))\nu$. □

This somewhat surprising result, together with Proposition 5.5.5, indicates that the class of small sets can be used for different purposes, depending on the choice of sampling distribution we make: if we sample at a fixed finite time we may get small sets with their useful fixed time-point properties; and if we extend the sampling as in Proposition 5.5.5, we develop a petite structure with a maximal irreducibility measure. We shall use this duality frequently.

5.6 Commentary

We have already noted that the split chain and the random renewal time approaches to regeneration were independently discovered by Nummelin [200] and Athreya and Ney [12]. The opportunities opened up by this approach are exploited with growing frequency in later chapters.

However, the split chain only works in the generality of φ-irreducible chains because of the existence of small sets, and the ideas for the proof of their existence go back to Doeblin [67], although the actual existence as we have it here is from Jain and Jamison [106]. Our proof is based on that in Orey [208], where small sets are called C-sets. Nummelin [202] Chapter 2 has a thorough discussion of conditions equivalent to that we use here for small sets; Bonsdorff [26] also provides connections between the various small set concepts.

Our discussion of cycles follows that in Nummelin [202] closely. A thorough study of cyclic behavior, expanding on the original approach of Doeblin [67], is given also in Chung [48].

Petite sets as defined here were introduced in Meyn and Tweedie [178]. The "small sets" defined in Nummelin and Tuominen [204] as well as the *petits ensembles* developed in Duflo [69] are also special instances of petite sets, where the sampling distribution a is chosen as $a(i) = 1/N$ for $1 \leq i \leq N$, and $a(i) = (1-\alpha)\alpha^i$ respectively. To a French speaker, the term "petite set" might be disturbing since the gender of *ensemble* is masculine: however, the nomenclature does fit normal English usage since [21] the word "petit" is likened to "puny", while "petite" is more closely akin to "small".

It might seem from Theorem 5.5.7 that there is little reason to consider both petite sets and small sets. However, we will see that the two classes of sets are useful in distinct ways. Petite sets are easy to work with for several reasons: most particularly, they span periodic classes so that we do not have to assume aperiodicity, they are always closed under unions for irreducible chains (Nummelin [202] also finds that unions of small sets are small under aperiodicity), and by Proposition 5.5.5 we may assume that the petite measure is a maximal irreducibility measure whenever the chain is irreducible.

Perhaps most importantly, when in the next chapter we introduce a class of Markov chains with desirable topological properties, we will see that the structure of these chains is closely linked to petiteness properties of compact sets.