Chapter 4
Existence of Generators

We have seen that the existence of a finite (or a finite entropy) generator is helpful in several settings. In the Kolmogorov–Sinaĭ theorem (Theorem 1.21) a generator allows us to calculate the entropy of a dynamical system using only the generator. Moreover, in Theorem 2.27 the Pinsker σ-algebra of a dynamical system is obtained as the tail σ-algebra of a generator. In this chapter we reverse the question and ask whether it is always possible to find a finite (or finite entropy) generator.

4.1 Countable Generators

The main result of this section is the following theorem of Rokhlin [173], which we will strengthen under the additional hypothesis of ergodicity in the next section. We recall that $(X, \mathscr{B}, \mu, T)$ is *aperiodic* if $\mu\left(\{x \in X \mid T^k x = x\}\right) = 0$ for all $k \in \mathbb{Z} \smallsetminus \{0\}$.

Theorem 4.1 (Finite entropy generator). *Let $(X, \mathscr{B}, \mu, T)$ be an invertible aperiodic measure-preserving system on a Borel probability space. If $h_\mu(T) < \infty$, then there exists a countable generator ξ with finite entropy. In fact, for every $\varepsilon > 0$ there is a countable generator ξ with $H_\mu(\xi) < h_\mu(T) + \varepsilon$.*

4.1.1 Entropy-free Construction

We introduce in this subsection a method for the construction of generators by giving a simpler result, whose proof will serve as a template for the proof of Theorem 4.1.

Proposition 4.2 (Existence of generators). *Let $(X, \mathscr{B}, \mu, T)$ be an aperiodic invertible measure-preserving system on a Borel probability space. Then there exists a countable partition ξ with $\mathscr{B} \underset{\mu}{=} \xi_{-\infty}^{\infty}$.*

In the proofs of this chapter we will identify partitions with measurable functions (called *codes*) whose range or *alphabet* is countable. In the proof below we will use some *ad hoc* language for the code, but in the proof of Theorem 4.1 we will choose the language more carefully.

PROOF OF PROPOSITION 4.2. As $(X, \mathscr{B}, \mu)$ is a Borel probability space, there exists an increasing sequence of finite partitions $(\xi(k))_{k\geqslant 1}$ with $\xi(k) \nearrow \mathscr{B}$ as $k \to \infty$. Let $(N_k)_{k\geqslant 1}$ be any sequence of natural numbers with $N_k \geqslant k$ for all $k \geqslant 1$. As $(X, \mathscr{B}, \mu, T)$ is aperiodic, for any $k \geqslant 1$ there exists by the Rokhlin–Halmos lemma (Theorem A.13) and Exercise A.6.2 a set Q_k, which we will refer to as a *marker* set, with

$$\mu\left(\bigcup_{n\in\mathbb{Z}} T^{-n} Q_k\right) = 1$$

and with the property that the first return of a point $x \in Q_k$ to Q_k is larger than N_k (that is, $x \in Q_k$ implies that $Tx, T^2x, \dots, T^{N_k}x \notin Q_k$). In particular, we therefore have

$$\mu(Q_k) \leqslant \tfrac{1}{N_k} \leqslant \tfrac{1}{k}. \tag{4.1}$$

We claim that we can find these sets in such a way that in addition we have

$$Q_1 \supseteq Q_2 \supseteq \cdots .$$

In fact we may simply apply the Rokhlin–Halmos lemma to find Q_1. Then consider the induced transformation

$$T_{Q_1} : Q_1 \to Q_1,$$

which is again aperiodic and invertible. Applying the Rokhlin–Halmos lemma again to T_{Q_1} gives the set $Q_2 \subseteq Q_1$ such that the first return under T_{Q_1} (and hence also under T) is larger than N_2. Applying this argument inductively gives the claim.

Notice that

$$\mu\left(\bigcap_{k=1}^{\infty} Q_k\right) = 0$$

by (4.1). We will ignore the orbit of this set, which is a null set. We now define our code by

$$\mathbf{c}(x) = \begin{cases} 0 & x \notin Q_1, \text{ and} \\ \left(k, \left(n_1(x), [x]_{\xi(1)_0^{n_1(x)-1}}\right), \dots, \left(n_k(x), [x]_{\xi(k)_0^{n_k(x)-1}}\right)\right) & x \in Q_k \smallsetminus Q_{k+1}, \end{cases}$$

where $n_k : Q_k \to \mathbb{N}$ is the first return time function of $x \in Q_k$ to Q_k.

It is clear by construction that $\mathbf{c}$ has countable range. We claim that the associated countable partition is a generator. Suppose therefore that

$$\mathbf{c}(T^n x) = \mathbf{c}(T^n y)$$

for all $n \in \mathbb{Z}$. We need to show that this almost surely implies that $x = y$.

Let $k \geqslant 1$ be arbitrary. If $x \in Q_k$, then $\mathbf{c}(x) = \mathbf{c}(y)$ implies that $y \in Q_k$ and that $[x]_{\xi(k)} = [y]_{\xi(k)}$. If $x \notin Q_k$, then there exists a minimal $n \geqslant 1$ with

$$T^{-n}x \in Q_k$$

(and, equivalently, with $T^{-n}y \in Q_k$). As n is minimal we have $n_k\left(T^{-n}x\right) > n$. This implies that

$$[T^{-n}x]_{\xi(k)_0^{n_1(x)-1}} = [T^{-n}x]_{\xi(k)_0^{n_1(x)-1}}$$

since $\mathbf{c}(T^{-n}x) = \mathbf{c}(T^{-n}y)$. In particular, we have $[x]_{\xi(k)} = [y]_{\xi(k)}$. As this holds for all $k \geqslant 1$ and $\xi(k) \nearrow \mathscr{B}$ as $k \to \infty$, we obtain $x = y$ almost surely, as required. □

4.1.2 Outline of the Proof of Theorem 4.1

We use the notation and assumptions from the statement of the theorem. Let $\xi(k) \nearrow \mathscr{B}$ be an increasing sequence of finite partitions. Much as in Section 4.1.1, we wish to construct a generator using the sequence $(\xi(k))_{k\geqslant 1}$ via a sequence of marker sets $(Q_k)_{k\geqslant 1}$ and a sequence of codes $\mathbf{c}_k$ defined on the marker sets with values in a countable alphabet. Clearly this time we need to concern ourselves with the entropy of the constructed partition at each stage. For the main estimate of the entropy we will use Section 1.2, and therefore will work with the countable alphabet $\bigcup_{\ell\geqslant 1}\{0,1\}^{\ell}$.

Let $\xi(0) = \{X\}$ and $\mathscr{A}_k = \xi(k)_{-\infty}^{\infty}$ for $k \geqslant 0$. Notice that additivity of dynamical entropy (Proposition 2.19(2)) gives

$$\begin{aligned} h_\mu\bigl(T, \xi(k)\bigr) &= h_\mu\bigl(T, \xi(k) \vee \xi(k-1)\bigr) \\ &= h_\mu\bigl(T, \xi(k-1)\bigr) + h_\mu\left(T, \xi(k)\middle|\xi(k-1)_{-\infty}^{\infty}\right), \end{aligned}$$

for $k \geqslant 1$, and so

$$h_\mu(T) = \sum_{k=1}^{\infty} h_\mu\bigl(T, \xi(k)\bigm|\mathscr{A}_{k-1}\bigr)$$

by the Kolmogorov–Sinaĭ theorem (Theorem 2.20).

We use this to split our argument into countably many constructions of codes $\mathbf{c} = \mathbf{c}_k$, where at each stage we will work with a fixed finite partition

$$\xi = \xi(k)$$

and a fixed T-invariant σ-algebra

$$\mathscr{A} = \mathscr{A}_{k-1}.$$

Our generator will be easily constructed from the partitions obtained along the sequence.

4.1.3 Entropy of the First Return Time Partition

In the construction of the codes $\mathbf{c}_k$ we will need marker sets Q_k and the derived partitions ρ_Q, whose entropy we now estimate more generally.

Lemma 4.3. *Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space, and let $Q \in \mathscr{B}$ be a set with positive measure. We define the first return time partition of Q by*

$$\rho_Q = \Big\{ X \smallsetminus Q,\, Q_1 = Q \cap T^{-1}Q,\, Q_2 = Q \cap T^{-2}Q \smallsetminus T^{-1}Q,\, \dots, \\ Q_m = Q \cap T^{-m}Q \smallsetminus (T^{-1}Q \cup \dots \cup T^{-(m-1)}Q), \dots \Big\}.$$

Then $H_\mu(\rho_Q) < \infty$. Moreover, for every $\varepsilon > 0$ there exists some N_ε such that $H_\mu(\rho_Q) < \varepsilon$ if the set Q is such that the first return from Q to Q is always greater than or equal to N_ε.

Proof. By construction, the sets

$$Q_1,\, Q_2,\, TQ_2,\, \dots,\, Q_m,\, TQ_m,\, \dots,\, T^{m-1}Q_m,\, \dots$$

are all disjoint, so

$$\sum_{m=1}^{\infty} m\mu(Q_m) \leqslant 1, \tag{4.2}$$

which will give us the entropy estimate. Fix $\varepsilon > 0$ and split the sum defining the entropy of ρ_Q as

$$\begin{aligned}H_\mu(\rho_Q) &= -\mu(X\smallsetminus Q)\log\mu(X\smallsetminus Q) - \sum_{\mu(Q_m)\geqslant \mathrm{e}^{-m\varepsilon}} \mu(Q_m)\log\mu(Q_m)\\ &\qquad - \sum_{\mu(Q_m)<\mathrm{e}^{-m\varepsilon}} \mu(Q_m)\log\mu(Q_m)\\ &\leqslant -\mu(X\smallsetminus Q)\log\mu(X\smallsetminus Q) + \varepsilon\sum_{m\geqslant 1} m\mu(Q_m) + \varepsilon\sum_{m\geqslant 1} m\mathrm{e}^{-m\varepsilon} + C,\end{aligned}$$

where we have used monotonicity of $x \mapsto -\log x$ in the first sum, monotonicity of $x \mapsto -x\log x$ for small values of x in the second sum, and where C denotes an absolute constant to bound the finitely many terms for which the latter monotonicity may fail. Applying this bound with $\varepsilon = 1$ gives the first claim in the lemma.

Suppose now that the first return time to Q is larger than N. In this case (4.2) implies that $\mu(Q) < \frac{1}{N}$. By continuity of the function $x \mapsto -x\log x$ we deduce that $-\mu(X\smallsetminus Q)\log\mu(X\smallsetminus Q) < \varepsilon$ for large enough N. Also notice that in this case we can set $C = 0$ in the entropy estimate above if N is sufficiently large, giving

$$H_\mu(\rho_Q) < \varepsilon + \varepsilon + \varepsilon\sum_{m>N} m\mathrm{e}^{-m\varepsilon}.$$

As the last sum converges, we deduce that $H_\mu(\rho_Q) < 3\varepsilon$ for large enough N. □

4.1.4 Definition of the Code $\mathbf{c} = \mathbf{c}_{\xi,\mathscr{A}}$

Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space, let ξ be a finite partition, and let $\mathscr{A} = T^{-1}\mathscr{A}$ be an invariant σ-algebra. We define in this section the code $\mathbf{c} = \mathbf{c}_{\xi,\mathscr{A}}$ with values in the alphabet $\bigcup_{\ell\geqslant 1}\{0,1\}^\ell$. We wish to control the average length of the code (that is, the entropy of the associated partition) by the conditional entropy $h_\mu(T, \xi\big|\mathscr{A})$.

For this, let $\varepsilon > 0$ be arbitrary and choose M large enough to ensure that

$$\tfrac{1}{M} < \varepsilon \tag{4.3}$$

and

$$\tfrac{M+1}{M}\tfrac{1}{M} H_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right) < h_\mu(T, \xi\big|\mathscr{A}) + \varepsilon, \tag{4.4}$$

and then choose $N \geqslant N_\varepsilon$ large enough to ensure that

$$\frac{\lceil \log_2 M\rceil + 3M\lceil \log_2 |\xi|\rceil}{N} + \frac{M}{N} < \varepsilon, \tag{4.5}$$

where N_ε is chosen as in Lemma 4.3. Finally, choose a set $Q \in \mathscr{B}$ using the Rokhlin–Halmos lemma with

$$\mu\left(\bigcup_{n\in\mathbb{Z}} T^{-n}Q\right) = 1$$

and with the property that the first return time to Q is bigger than N almost surely.

We define the code $\mathbf{c}(x) = \mathbf{c}_{\xi,\mathscr{A}}(x)$ for $x \in Q$ by the tuple

$$\mathbf{c}(x) = \Big(\big[j(x)\big]_2, \big[\xi_0^{j(x)-1}(x)\big]_{\text{uo}}, \big[\xi_{j(x)}^{M+j(x)-1}(x)\big|\mathscr{A}\big]_{\text{Sh}}, \big[\xi_{M+j(x)}^{2M+j(x)-1}(x)\big|\mathscr{A}\big]_{\text{Sh}}, \dots, \big[\xi_{(L(x)-1)M+j(x)}^{L(x)M+j(x)-1}(x)\big|\mathscr{A}\big]_{\text{Sh}}, \big[\xi_{L(x)M+j(x)}^{n(x)-1}(x)\big]_{\text{uo}} \Big)$$

as explained below, but note that we do not claim that $\mathbf{c}$ is prefix-free. In the definition of $\mathbf{c}(x)$ we are using the following notation:

- $n(x)$ is the first return time of $x \in Q_{n(x)} \subseteq Q$ to Q, $L(x) = \lfloor \frac{n(x)}{M} \rfloor - 1$;
- $j(x) \in \{0, \dots, M-1\}$ is chosen as the offset to minimize the partial ergodic sum

$$\frac{1}{L(x)} \sum_{\ell=0}^{L(x)-1} I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^{\ell M + j(x)}(x)), \tag{4.6}$$

  for the transformation T^M;
- $[j(x)]_2$ denotes the binary expansion of $j(x)$ with added leading zeros to ensure that the total length of the expansion is precisely $\lceil \log_2 M \rceil$ (which makes it prefix-free);
- $[\xi_0^{j(x)-1}(x)]_{\text{uo}}$ denotes an arbitrary (unoptimized) prefix-free code everywhere of length $M\lceil \log_2 |\xi| \rceil$, describing the fragment $x, Tx, \dots, T^{j(k)-1}x$ of the orbit, say by using a fixed enumeration of ξ and again using the binary expansion of the corresponding number padded out with trailing zeros;
- $[\xi_{\ell M+j(x)}^{(\ell+1)M+j(x)-1}(x)\big|\mathscr{A}]_{\text{Sh}} = [\xi_0^{M-1}(T^{\ell M+j(x)}x)\big|\mathscr{A}]_{\text{Sh}}$ denotes the Shannon code for the partition element of ξ_0^{M-1} containing the point $T^{\ell M+j(x)}x$ using the conditional measure $\mu_x^{\mathscr{A}}$ in the definition of the Shannon code (and using some fixed lexicographical order to break potential ties in the definition of the Shannon code for $\mu_x^{\mathscr{A}}$ in a well-defined manner);

- and finally, $[\xi_{L(x)M+j(x)}^{n(x)-1}(x)]_{\mathrm{uo}}$ denotes an arbitrary (unoptimized) prefix-free code of length $2M\lceil\log_2|\xi|\rceil$ describing the remaining fragment of the orbit of x until the next return to Q, again padded out with trailing zeros.

4.1.5 Properties of the Code $\mathbf{c}_{\xi,\mathscr{A}}$

We start with some observations about the code $\mathbf{c}_{\xi,\mathscr{A}}$ constructed above.

- If $\mathscr{A}=\mathscr{N}$ (which it will be at the beginning of the induction), then $\mathbf{c}$ defines a prefix-free code on each set $Q_n\in\rho_Q$. Here n needs to be fixed to ensure the prefix-free property.
- In general not even that may be true since $n(x)$ does not determine the Shannon code used to encode $[\xi_{j(x)}^{M+j(x)-1}(x)\big|\,\mathscr{A}]_{\mathrm{Sh}}$ (which depends on the measure $\mu_x^{\mathscr{A}}$). We will fix this by a rather simple procedure in Lemma 4.4 below. For example, we may insert after every digit a 0 except at the very end we put a 1 to signal the end of the code (which also lifts the above restriction that x needs to belong to a particular Q_n).
- By construction and the properties of the Shannon code (see Lemma 1.11) the length of $\mathbf{c}(x)$ for $x\in Q$ is bounded by

$$
\begin{aligned}
\lceil\log_2 M\rceil+3M\lceil\log_2|\xi|\rceil+\sum_{\ell=0}^{L(x)-1}\frac{1}{\log 2}I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^{\ell M+j(x)}x)+L(x)&\\
<\varepsilon N+\frac{1}{M}\sum_{j=0}^{M-1}\sum_{\ell=0}^{L(x)-1}\frac{1}{\log 2}I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^{\ell M+j}x)+\frac{n(x)}{M}&\\
<\sum_{n=0}^{n(x)-1}\left(\frac{1}{M\log 2}I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^n x)+2\varepsilon\right),&
\end{aligned}
$$

  where we have used (4.5), the choice of $j(x)$ to minimize (4.6), the definition of $L(x)$ in terms of $n(x)$, the fact that $n(x)>N$, and (4.3).
- Let $\eta_{\xi,\mathscr{A}}$ be the partition generated by ρ_Q and the partition defined by the code $\mathbf{c}_{\xi,\mathscr{A}}$, where the latter contains $X\smallsetminus Q$ and the various level sets of the code. Then

$$
\xi_{-\infty}^{\infty}\underset{\mu}{\subseteq}\mathscr{A}\vee(\eta_{\xi,\mathscr{A}})_{-\infty}^{\infty}. \tag{4.7}
$$

  The proof of (4.7) is very similar to the argument in Section 4.1.1. Let

$$
\mathscr{C}=\mathscr{A}\vee(\eta_{\xi,\mathscr{A}})_{-\infty}^{\infty}
$$

  and $x,y\in X$ with $[x]_{\mathscr{C}}=[y]_{\mathscr{C}}$. If $x\in Q$ then $y\in Q$, and since the partition element $[x]_{\rho_Q}=[y]_{\rho_Q}\in\rho_Q$ determines $n(x)$, we obtain

$$n(x) = n(y).$$

Also, since $[x]_{\mathscr{A}} = [y]_{\mathscr{A}}$ the same measure $\mu_x^{\mathscr{A}}$ is used to encode the orbit $\xi_0^{n(x)-1}(x)$, which, together with $\mathbf{c}_{\xi,\mathscr{A}}(x) = \mathbf{c}_{\xi,\mathscr{A}}(y)$, implies that $\xi_0^{n(x)-1}(x) = \xi_0^{n(y)-1}(y)$. As almost every point has infinitely many returns to Q in the past and in the future, this gives $[x]_{\xi_{-\infty}^{\infty}} = [y]_{\xi_{-\infty}^{\infty}}$ almost surely, and hence the claim.

4.1.6 Entropy of the Partition Associated to $\mathbf{c}_{\xi,\mathscr{A}}$

Lemma 4.4. *Let $\eta_{\xi,\mathscr{A}}$ be the partition generated by ρ_Q and the partition defined by the code $\mathbf{c}_{\xi,\mathscr{A}}$ as above. Then $H_\mu\bigl(\eta_{\xi,\mathscr{A}}\bigr) < h_\mu(T,\xi\big|\mathscr{A}) + 10\varepsilon$.*

Proof. As mentioned above, the code $\mathbf{c}_{\xi,\mathscr{A}}$ may not be prefix-free. We wish to modify the code to make it prefix-free. For this, we first add extra trailing 0s (at most $M-1$ of them) to the code until the length of the new sequence is divisible by M. Then we insert after every block of M digits a 0 (to signify that the code continues) and instead of the 0 after the last block add a 1 (to signify the end of the code). Finally, we add another block of length M comprising leading 0s and trailing 1s, where the leading 0s indicate which digits of the previous block existed in $\mathbf{c}_{\xi,\mathscr{A}}$ and the 1s indicate the added digits. In this way we obtain a new prefix-free code $\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}$ with the property that the length of $\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}(x)$ for $x \in Q_{n(x)} \subseteq Q$ is bounded by

$$\begin{aligned}\frac{M+1}{M}&\left(\sum_{n=0}^{n(x)-1}\left(\frac{1}{M\log 2}I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^n x) + 2\varepsilon\right) + 2M\right)\\ &\leqslant \frac{M+1}{M}\left(\sum_{n=0}^{n(x)-1}\left(\frac{1}{M\log 2}I_\mu\left(\xi_0^{M-1}\big|\mathscr{A}\right)(T^n x) + 4\varepsilon\right)\right) \qquad (4.8)\end{aligned}$$

by (4.5) and $n(x) \geqslant N$. Switching from $\mathbf{c}_{\xi,\mathscr{A}}$ to $\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}$ does not change the partition $\eta_{\xi,\mathscr{A}}$ but allows us to use Section 1.2 in the argument below.

Let us write $\mu_Q = \frac{1}{\mu(Q)}\mu|_Q$ for the normalized restriction to Q. Recall that $H_\mu(\rho_Q) < \varepsilon$ by Lemma 4.3 and our choice of $N \geqslant N_\varepsilon$. Therefore,

$$\begin{aligned}H_\mu(\eta_{\xi,\mathscr{A}}) &= H_\mu(\rho_Q) + H_\mu(\eta_{\xi,\mathscr{A}}\big|\rho_Q)\\ &< \varepsilon + H_\mu(\eta_{\xi,\mathscr{A}}\big|\{Q, X\smallsetminus Q\})\\ &= \varepsilon + \mu(Q)H_{\mu_Q}(\eta_{\xi,\mathscr{A}})\\ &\leqslant \varepsilon + \log 2\int_Q |\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}(x)|\,\mathrm{d}\mu(x)\end{aligned}$$

by Lemma 1.10, where we write $|\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}(x)|$ for the length of the code for $x \in Q$. Let us assume $\varepsilon \in (0,1]$ and recall that $\frac{1}{M} < \varepsilon$ by (4.3), which gives

$$\tfrac{M+1}{M} < 1 + \varepsilon < 2.$$

Using the upper bound (4.8) for $|\widetilde{\mathbf{c}}_{\xi,\mathscr{A}}(x)|$ we obtain

$$\begin{aligned} H_\mu(\eta_{\xi,\mathscr{A}}) &\leqslant \varepsilon + \tfrac{M+1}{M} \sum_{Q_m \subseteq Q} \int_{Q_m} \sum_{n=0}^{m-1} \left(\tfrac{1}{M} I_\mu \left(\xi_0^{M-1} \middle| \mathscr{A} \right) (T^n x) + 4\varepsilon \right) \mathrm{d}\mu \\ &= \tfrac{M+1}{M} \tfrac{1}{M} H_\mu \left(\xi_0^{M-1} \middle| \mathscr{A} \right) + 9\varepsilon \\ &< h_\mu \left(T, \xi \middle| \mathscr{A} \right) + 10\varepsilon \end{aligned}$$

where we used the consequence

$$\bigsqcup_{m=1}^{\infty} \bigsqcup_{n=0}^{m-1} T^n Q_m \underset{\mu}{=} X$$

of the definition of Q_m for $m \geqslant 1$, and the choice of M in (4.4). □

4.1.7 The Inductive Proof of Theorem 4.1

Using the arguments above countably many times, it is now relatively easy to construct a countable generator with finite entropy.

Proof of Theorem 4.1. Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space with finite entropy, and let $\xi(k) \nearrow \mathscr{B}$ be an increasing sequence of finite partitions, so that

$$h_\mu(T) = \sum_{k=1}^{\infty} h_\mu \left(T, \xi(k) \middle| \mathscr{A}_{k-1} \right),$$

where $\mathscr{A}_0 = \mathscr{N}$ and $\mathscr{A}_k = \sigma(\xi(k)_{-\infty}^{\infty})$ for $k \geqslant 1$.

Now apply the construction from Sections 4.1.3–4.1.6 for the partition $\xi(k)$, the σ-algebra $\mathscr{A}_{k-1}$, and the number $\frac{\varepsilon}{2^k}$ for all $k \geqslant 1$. In this way we obtain a sequence $\left(\eta_k = \eta_{\xi(k),\mathscr{A}_{k-1}}\right)$ of countable partitions with

$$H_\mu(\eta_k) < h_\mu \left(T, \xi(k) \middle| \mathscr{A}_{k-1} \right) + 10 \frac{\varepsilon}{2^k}$$

by Lemma 4.4. Therefore,

$$H_\mu \left(\bigvee_{k=1}^{K} \eta_k \right) < \sum_{k=1}^{K} h_\mu \left(T, \xi(k) \middle| \mathscr{A}_{k-1} \right) + 10\varepsilon,$$

and letting $K \to \infty$ gives

$$H_\mu\left(\bigvee_{k=1}^{\infty} \eta_k\right) < h_\mu(T) + 10\varepsilon$$

by monotone convergence. However, this implies that there exists a *countable partition* η with

$$\sigma(\eta) = \bigvee_{k=1}^{\infty} \eta_k.$$

We claim that $\eta_{-\infty}^{\infty} \underset{\mu}{=} \mathscr{B}$. Notice that

$$\eta_{-\infty}^{\infty} \supseteq (\eta_1)_{-\infty}^{\infty} \supseteq \xi(1)_{-\infty}^{\infty} = \mathscr{A}_1$$

by (4.7) with $\xi = \xi(1)$, $\mathscr{A} = \mathscr{N}$, and $\eta_{\xi,\mathscr{A}} = \eta_1$. Similarly, (4.7) applied with $\xi = \xi(k)$, $\mathscr{A} = \mathscr{A}_{k-1}$, and $\eta_{\xi,\mathscr{A}} = \eta_k$ now inductively gives

$$\eta_{-\infty}^{\infty} \underset{\mu}{\supseteq} \mathscr{A}_k = \xi(k)_{-\infty}^{\infty}.$$

Therefore, $\eta_{-\infty}^{\infty} \underset{\mu}{=} \mathscr{B}$. □

4.2 Existence of a Finite Generator

If $(X, \mathscr{B}, \mu, T)$ is an invertible measure-preserving system on a Borel probability space and ξ is a finite generator, then

$$h_\mu(T) = h_\mu(T, \xi) \leqslant H_\mu(\xi) \leqslant \log|\xi| \tag{4.9}$$

by the Kolmogorov–Sinaĭ theorem (Theorem 1.21). Also notice that equality in (4.9) implies that each $P \in \xi$ has $\mu(P) = \frac{1}{|\xi|}$ (by the trivial bound in Propositions 1.17(1) and 1.5) and that the partitions

$$T^n\xi, \dots, T\xi, \xi, T^{-1}\xi, \dots, T^{-n}\xi$$

are mutually independent. In other words, equality in (4.9) implies that the system $(X, \mathscr{B}, \mu, T)$ is measurably isomorphic to the Bernoulli shift with $|\xi|$ symbols of equal weight. This shows (apart from the ergodicity assumption) that the following result due to Krieger [110], [111] is the best possible one could hope to obtain for general systems.[(21)]

Theorem 4.5 (Krieger's finite generator theorem). *Let $(X, \mathscr{B}, \mu, T)$ be an ergodic invertible measure-preserving system on a Borel probability space.*

Suppose that $h_\mu(T) < \log k$ for some $k \geqslant 1$. Then there exists a generator for $(X, \mathscr{B}, \mu, T)$ with k atoms.

For the proof we will use both the Shannon–McMillan–Breiman theorem (Theorem 3.1) as well as the existence of a countable generator (Theorem 4.1). In fact, given the countable generator ξ from Theorem 4.1 we wish to construct a map $\phi : X \to \{1, 2, \dots, k\}^{\mathbb{Z}}$ with the properties

$$\phi(\sigma(x)) = \sigma(\phi(x)) \tag{4.10}$$

and

$$\phi(x) = \phi(y) \Longrightarrow x = y \tag{4.11}$$

for $x, y \in X'$, where $X' \in \mathscr{B}$ is a set of full measure and σ denotes the left shift map. The generator with the property claimed in the theorem is then defined by

$$\eta = \{\phi^{-1}([0]), \dots, \phi^{-1}([k-1])\},$$

where

$$[j] = \{x \in \{0, \dots, k-1\}^{\mathbb{Z}} \mid x_0 = j\}$$

is the cylinder set at time zero. The map ϕ is defined by using a marker set $Q \subseteq X$ and a marker word $w_Q \in \{1, 2, \dots, k\}^M$ chosen so that every appearance of w_Q in $\phi(x)$ at location n corresponds to a visit of the orbit of x to Q at time n (that is, to the statement $T^n x \in Q$). The words in between two consecutive appearances of w_Q, at 0 and at N say, will encode the partition elements of ξ that x belongs to between the corresponding two consecutive visits of x to Q. To be able to do this we need to analyze the number of possible trajectories between those times, and this is where the Shannon–McMillan–Breiman theorem will be useful.

Proof of Theorem 4.5. Let $(X, \mathscr{B}, \mu, T)$ and $k > 1$ be as in Theorem 4.5. Applying Theorem 4.1 we find a countable generator ξ with finite entropy.

Let $h = h_\mu(T) = h_\mu(T, \xi)$. By the Shannon–McMillan–Breiman theorem (Theorem 3.1) we find for every $\varepsilon > 0$ some measurable subset $X_\varepsilon \subseteq X$ with measure greater than $\frac{3}{4}$ and some N_ε such that

$$\mathrm{e}^{-(h+\varepsilon)n} < \mu\bigl([x]_{\xi_0^{n-1}}\bigr) < \mathrm{e}^{-(h-\varepsilon)n}$$

for $x \in X_\varepsilon$ and $n \geqslant N_\varepsilon$. In particular, we deduce that if $n \geqslant N_\varepsilon$ then

$$X_\varepsilon \text{ is covered by less than } \mathrm{e}^{(h+\varepsilon)n} \text{ elements of } \xi_0^{n-1}. \tag{4.12}$$

We define w_Q as the word

$$10^M = 1\underbrace{0\cdots0}_{M}.$$

Let W'_ℓ be the set of words in the alphabet $A = \{1, 2, \dots, k\}$ with length ℓ that do not contain w_Q as a subword, where we say that w_Q is a subword of w if there exist words $w_1, w_2 \in \bigcup_{\ell \geqslant 0} A^\ell$ with the property that $w = w_1 w_Q w_2$ is the concatenation of w_1, w_Q, and w_2 in turn. We claim that

$$|W'_\ell| \geqslant k^{\frac{M-1}{M}\ell} = \mathrm{e}^{\left(\frac{M-1}{M}\log k\right)\ell}.$$

To prove the claim, let $\ell = aM + b$ for $a \geqslant 0$ and $b \in \{0, 1, \dots, M-1\}$. Let

$$v \in A^{a(M-1)+b}$$

be an arbitrary word. Then we can construct a word $w(v) \in W'_\ell$ of length ℓ from v by inserting the symbol 1 after every $(M-1)$ letters. In this way we ensure that $w(v)$ does not contain a contiguous run 0^M of 0s of length M, so $w(v) \in W'_\ell$. Since the map $v \mapsto w(v)$ is injective by construction, the claim follows. Below we would like to always use elements of W'_ℓ after w_Q, so we define

$$W_n = \{(w_Q w) \mid w \in W'_{n-(M+1)}\}.$$

As above, this has

$$|W_n| \geqslant k^{\frac{M-1}{M}(n-M-1)} = \mathrm{e}^{(\log k)\frac{M-1}{M}\frac{n-M-1}{n}n} \tag{4.13}$$

many elements.

We want to split the difference between the two estimates (4.12) and (4.13). Specifically, we let $\varepsilon = \frac{\log k - h}{2}$, choose M so large that $(\log k)\frac{M-1}{M} > \log k - \varepsilon$ and choose N' so large that

$$(\log k)\frac{M-1}{M}\frac{n-M-1}{n} > \log k - \varepsilon = h + \varepsilon$$

for $n \geqslant N'$. With these choices we have

$$|W_n| \geqslant \mathrm{e}^{(h+\varepsilon)n} \tag{4.14}$$

for $n \geqslant N'$.

Having chosen ε, the set X_ε and the number N_ε are now also defined (by the discussion leading to (4.12)). We now define the marker set Q using the Kakutani–Rokhlin lemma. In fact by that lemma there exists a set $Y \subseteq X$ such that

$$Z = Y \sqcup TY \sqcup \cdots \sqcup T^{\max\{N_\varepsilon, N'\}}Y$$

is a disjoint union and $\mu(Z) > \frac{3}{4}$. It follows that $\mu(Z \cap X_\varepsilon) > \frac{1}{2}$ and so there exists some $0 \leqslant \ell \leqslant \max\{N_\varepsilon, N'\}$ such that $Q = T^\ell Y \cap X_\varepsilon$ has positive measure.

The so-defined marker set Q has the following convenient properties:

(1) for any $x \in Q$ the first return time to Q is greater than $\max\{N_\varepsilon, N'\}$;

(2) the subset

$$Q_n = Q \cap T^{-n}Q \setminus (T^{-1}Q \cup \dots \cup T^{-(n-1)}Q)$$

of Q with first return time equal to $n > \max\{N_\varepsilon, N'\}$ is covered by less than $\mathrm{e}^{(h+\varepsilon)n}$ many elements of the partition $\zeta_n = \xi_0^{n-1}$;

(3) for $n > \max\{N_\varepsilon, N'\}$ there are at least $\mathrm{e}^{(h+\varepsilon)n}$ many words in W_n that begin with w_Q but do not contain it later; and

(4) almost every point in X' has infinitely many visits to Q in the future and in the past.

We now define $\phi : X \to A^{\mathbb{Z}}$ using some fixed sequence of injective maps Φ_n from ζ_n as in (2) to W_n as in (3) for all $n > \max\{N_\varepsilon, N'\}$. In fact, if we let

$$V_Q(x) = \{m \mid T^m(x) \in Q\}$$

be the set of visits to Q and define $\phi(x)$ by the concatenations of

$$\Phi_n\left([T^m x]_{\xi_0^{n-1}}\right)$$

whenever

$$[m, m+n] \cap V_Q = \{m, m+n\}. \tag{4.15}$$

More precisely, if (4.15) holds, then

$$\Phi_n\left([T^m x]_{\xi_0^{n-1}}\right)$$

defines the coordinates $(\phi(x))_j$ for

$$j \in \{m, m+1, \dots, m+n-1\}.$$

For[†] $x \in X'$ as in (4), this defines a point $\phi(x) \in A^{\mathbb{Z}}$.

The construction ensures that the property (4.10) holds, so that ϕ defines a factor map. Suppose now that $x, y \in X'$ and $\phi(x) = \phi(y)$ as in (4.11). The properties of the set W_n for $n \geqslant N'$ then imply that $V_Q(x) = V_Q(y)$ and that

$$[x]_{\xi_{-\infty}^{\infty}} = [y]_{\xi_{-\infty}^{\infty}}$$

since we use injective maps Φ_n from ζ_n to W_n for $n \geqslant N'$. By the property of ξ this gives $x = y$ almost everywhere. This gives the theorem (see also Exercise 4.2.1). □

[†] For x in the null set $X' \setminus X$ we may set $\phi(x) = 1^\infty$ or leave ϕ undefined.

Exercises for Section 4.2

Exercise 4.2.1. Analyze $\phi^{-1}([j]_0)$ for any $j \in \{0, 1, \ldots, k-1\}$ for ϕ as in the proof of Theorem 4.5. Conclude from this that ϕ is measurable (as implicitly required for the result).