

Solutions of the Exercises
in
Entropy in Dynamical Systems

T. Downarowicz

June, 2011

A few words about the Exercises

This file contains solutions of almost all exercises included in the book. There are three exceptions: 3.10, 8.6 and 12.5. The first one instructs to write a computer program. I think, even if I presented such a program, nobody would really read the code or want to see it work. The only way to enjoy this exercise is to actually do it. Exercise 8.6 is completely trivial. As to the last skipped exercise, although I am sure what I claim there is true, I have never done it before. I see a number of obstacles where the standard approach could break and some ingenuity might be needed. I decided to leave this challenge open; it might turn out worth a separate article. I should have formulated this as a question rather than an exercise.

I confess, there are several exercises that I never bothered to actually do before I put them in the book. I just had a rough idea how to proceed. When working on this file, it happened more than once that I encountered unexpected difficulties. Some solutions turned into pieces of work comparable to writing a small article (7.3, 8.9, 8.14, 9.5, 12.2). In some cases I needed to slightly alter the formulation of the exercise or add an assumption. In such cases the alternation is clearly indicated at the beginning of the solution (search for “Attention!”). In exercise 6.3 I extrapolated a property typical for smooth interval map to all systems, which is an evident symptom of tiredness. I apologize for all these errors. Some of them result from the fact that many exercises have been added after completing and proofreading the main body of the book.

Moreover, this work lead me to discovering a few more imprecisions in the book. All resulting corrections are included in the file “Errata”.

I encourage the readers of the book (and of the solutions, if anyone bothers) to send me information of any discovered further errors, or any comments, via e-mail. They will be welcome. The book cannot be corrected, but the errata can always be updated.

Tomasz Downarowicz

Part 1

1 Exercises in Chapter 1

Exercise 1.1.

We have

$$x = \frac{y}{x+y} \cdot 0 + \frac{x}{x+y} \cdot (x+y),$$

hence, by concavity,

$$f(x) \geq \frac{y}{x+y} \cdot f(0) + \frac{x}{x+y} \cdot f(x+y).$$

Analogously,

$$f(y) \geq \frac{x}{x+y} \cdot f(0) + \frac{y}{x+y} \cdot f(x+y).$$

Summing both sides, we get $f(x) + f(y) \geq f(0) + f(x+y) \geq f(x+y)$.

Exercise 1.2.

Convergence in ℓ^1 obviously implies coordinatewise convergence for any vectors in ℓ^1 . The converse holds only with a constraint, for example for probability vectors. Let $\mathbf{p} = (p_1, p_2, \dots)$ be a probability vector and let i_0 be so large, that $\sum_{i>i_0} p_i < \varepsilon/4$. Let $\mathbf{p}' = (p'_1, p'_2, \dots)$ be another probability vector such that $|p_i - p'_i| < \varepsilon/4i_0$ for all $i \leq i_0$. Then

$$\sum_{i>i_0} p'_i = 1 - \sum_{i \leq i_0} p'_i = 1 - \sum_{i \leq i_0} p_i + \sum_{i \leq i_0} (p_i - p'_i) \leq \sum_{i>i_0} p_i + \sum_{i \leq i_0} |p_i - p'_i| < \frac{\varepsilon}{2}$$

and thus

$$\sum_{i \geq 1} |p_i - p'_i| = \sum_{i \leq i_0} |p_i - p'_i| + \sum_{i>i_0} p_i + \sum_{i>i_0} p'_i < \varepsilon.$$

Exercise 1.3.

$$\begin{aligned} H(a \vee b|c) &= \\ H(a \vee b \vee c) - H(c) &= H(a \vee b \vee c) - H(b \vee c) + H(b \vee c) - H(c) = \\ &= H(a|b \vee c) + H(b|c), \end{aligned}$$

and

$$H(a \vee b|c) = H(a \vee b \vee c) - H(c) \geq H(a \vee c) - H(c) = H(a|c).$$

With the further assumptions,

$$\begin{aligned} H(a \vee b|c) &= H(a|b \vee c) + H(b|c) \leq H(a|c) + H(b|c); \\ H(a \vee b) &= H(a \vee b \vee e) = H(a \vee b|e) + H(e) \leq H(a|e) + H(b|e) + H(e) \\ &= H(a) + H(b) - H(e) \leq H(a) + H(b); \\ H(a \vee a'|b \vee b') &\leq H(a|b \vee b') + H(a'|b \vee b') \leq H(a|b) + H(a'|b'); \\ H(a|c) &\leq H(a \vee b|c) \leq H(a|b \vee c) + H(b|c) \leq H(a|b) + H(b|c). \end{aligned} \quad (*)$$

Suppose $H(a|c) \geq H(b|c)$. Then $|H(a|c) - H(b|c)| = H(a|c) - H(b|c)$ and (*) implies $|H(a|c) - H(b|c)| \leq H(a|b) \leq \max\{H(a|b), H(b|a)\}$. The other case is symmetric.

Similarly, suppose $H(a|c) \geq H(a|b)$. Then $|H(a|b) - H(a|c)| = H(a|c) - H(a|b)$ and (*) implies $|H(a|b) - H(a|c)| \leq H(b|c) \leq \max\{H(b|c), H(c|b)\}$. The other case is symmetric.

Finally, $|H(a) - H(b)| = |H(a|e) + H(e) - H(b|e) - H(e)| = |H(a|e) - H(b|e)| \leq \max\{H(a|b), H(b|a)\}$.

Exercise 1.4.

This is a very crude estimate. Suppose p_1 is the minimal term in \mathbf{p} , let $q = 1 - p_1$, and define two new probability vectors $\mathbf{r} = (p_1, q)$ and $\mathbf{q} = (\frac{p_2}{q}, \frac{p_3}{q}, \dots, \frac{p_l}{q})$. Then

$$\begin{aligned} H(\mathbf{p}) &= -\sum_{i=1}^l p_i \log p_i = -p_1 \log p_1 - q \sum_{i=2}^l \frac{p_i}{q} (\log \frac{p_i}{q} + \log q) = \\ &= -p_1 \log p_1 - q \log q + qH(\mathbf{q}) = H(\mathbf{r}) + qH(\mathbf{q}) \leq 1 + (1 - p_1) \log l. \end{aligned}$$

Exercise 1.5.

Pick m_1 so large that p'_1 is close enough to 1 to satisfy $\eta(p_1) < \varepsilon/2$ and $\eta(1 - p_1) < \varepsilon/4$. No matter how we pick m_2 we will have $p_2 \leq 1 - p_1$ and since η increases near zero, we will have $\eta(p_2) < \varepsilon/4$. We pick m_2 large enough to make $p_1 + p_2$ so close to 1 that $\eta(1 - p_1 - p_2) < \varepsilon/8$. No matter how we pick m_3 , we will have $\eta(p_3) < \varepsilon/8$. And so on. Eventually, we get

$$H(\mathbf{p}') = \sum_{i \geq 1} \eta(p'_i) < \sum_{i \geq 1} \varepsilon/2^i = \varepsilon.$$

Exercise 1.6.

Take the partitions \mathcal{P}, \mathcal{Q} and \mathcal{R} as in the proof of Fact 1.9.1 (they produce the vector $(1, 1, 1, 2, 2, 2, 2)$). Then $I(\mathcal{P} \vee \mathcal{Q}; \mathcal{R}) = H(\mathcal{P} \vee \mathcal{Q}) + H(\mathcal{R}) - H(\mathcal{P} \vee \mathcal{Q} \vee \mathcal{R}) = 2 + 1 - 2 = 1$, while both $I(\mathcal{P}; \mathcal{R})$ and $I(\mathcal{Q}; \mathcal{R})$ are zeros, because the partitions are pairwise independent (see Fact 1.8.2).

Exercise 1.7.

If $a = 0$ or $b = 0$ or $c = a + b$, the problem is trivial. Otherwise, let $p \in (0, 1)$ be such that $H(p, 1 - p) < \min\{a, b, a + b - c\}$. Divide the space in two parts A and B of measures p and $1 - p$, respectively. Then let \mathcal{R} be a partition of A with (relative) entropy $\frac{1}{p}(a + b - c - H(p, 1 - p))$. Let \mathcal{P}' and \mathcal{Q}' be two independent partitions of B with (relative) entropies $\frac{1}{1-p}(c - b)$ and $\frac{1}{1-p}(c - a)$, respectively. The partition \mathcal{P} be defined as \mathcal{R} on A and \mathcal{P}' on B , and analogously, \mathcal{Q} is defined as \mathcal{R} on A and \mathcal{Q}' on B . Letting $\mathcal{R}_0 = \{A, B\}$ we evaluate each entropy conditioning it on \mathcal{R}_0 , e.g., $H(\mathcal{P}) = H(\mathcal{P} \vee \mathcal{R}_0) = H(\mathcal{P}|\mathcal{R}_0) + H(\mathcal{R}_0)$. And so, using (1.4.4),

$$\begin{aligned} H(\mathcal{P}) &= pH_A(\mathcal{R}) + (1 - p)H_B(\mathcal{P}') + H(p, 1 - p) = a, \\ H(\mathcal{Q}) &= pH_A(\mathcal{R}) + (1 - p)H_B(\mathcal{Q}') + H(p, 1 - p) = b, \\ H(\mathcal{P} \vee \mathcal{Q}) &= pH_A(\mathcal{R}) + (1 - p)(H_B(\mathcal{P}') + H_B(\mathcal{Q}')) + H(p, 1 - p) = c, \end{aligned}$$

where, in the last case, we have used relative independence of \mathcal{Q}' and \mathcal{P}' on B and Fact 1.6.16.

2 Exercises in Chapter 2

Exercise 2.1.

It is clear that π is onto (both for the unilateral and bilateral shift space) and each (x_n) has the same image under π as (x'_n) , where $x'_n = x_n + 1$ (addition is modulo 2). So the mapping π is exactly 2 to 1. The preimage by π of a block $B = (b_0, \dots, b_{n-1})$ equals $C \cup C'$, where

$$C = (0, b_0, b_0 + b_1, b_0 + b_1 + b_2, \dots, b_0 + b_1 + \dots + b_{n-1}) \text{ and}$$

$$C' = (1, 1 + b_0, 1 + b_0 + b_1, 1 + b_0 + b_1 + b_2, \dots, 1 + b_0 + b_1 + \dots + b_{n-1}).$$

Each of these blocks (as cylinder) has measure 2^{-n-1} and since they are disjoint, their union has measure 2^{-n} , the same as B . We have proved that π sends the Bernoulli measure to itself. In other words, the factor process is the same Bernoulli shift and so the identity map (not π) provides an isomorphism between the process and its factor.

Exercise 2.2.

Note that for $n \in \mathbb{N}$, the past of the power process $(X, \mathcal{P}^n, \mu, T^n, \mathbb{S})$ equals the past \mathcal{P}^- of the original process. By the power rule we have

$$H(\mathcal{P}^n | \mathcal{P}^-) = h(\mu, T^n, \mathcal{P}^n) = nh(\mu, T, \mathcal{P}) (= nh(\mathcal{P})).$$

Exercise 2.3.

In the Bernoulli shift on two symbols with equal measures $1/2, 1/2$ let \mathcal{R} denote the zero-coordinate partition. Consider $\mathcal{P} = \mathcal{R}^{\{1,3\}}$ and $\mathcal{Q} = \mathcal{R}^{\{0,2\}}$. These partitions are independent, so $H(\mathcal{P}|\mathcal{Q}) = H(\mathcal{P}) = 2$. Next, $\mathcal{P}^2 = \mathcal{R}^{\{1,2,3,4\}}$ and $\mathcal{Q}^2 = \mathcal{R}^{\{0,1,2,3\}}$ and only one coordinate in the definition of \mathcal{P} does not occur in \mathcal{Q} , so $H(\mathcal{P}^2|\mathcal{Q}^2) = 1$. It is seen that $H(\mathcal{P}^n|\mathcal{Q}^n) = 1$ for all $n \geq 2$. The sequence $2, 1, 1, 1, \dots$ is not increasing, the increments are $(-1, 0, 0, 0, \dots)$ and do not decrease.

To have increments increasing for a longer time take e.g. $\mathcal{P} = \mathcal{R}^{\{1,3,4,7,8,9\}}$ and $\mathcal{Q} = \mathcal{R}^{\{0,2,5,6,10\}}$. Then the sequence $(H(\mathcal{P}^n|\mathcal{Q}^n))_n$ equals $(6, 3, 1, 0, 0, 0, \dots)$ and the increments are $(-3, -2, -1, 0, 0, 0, \dots)$.

Exercise 2.4.

Although the sequence $a_n = H(\mathcal{P}^n|\mathcal{Q}^n \vee \mathfrak{B})$ need not have decreasing increments, it has decreasing n ths and any sequence (a_n) with convergent n ths satisfies

$$\lim \frac{1}{n} a_n = \lim \frac{1}{n} \sum_{i=1}^n (a_i - a_{i-1}) \leq \limsup (a_{n+1} - a_n).$$

In our case we have

$$\begin{aligned} a_{n+1} - a_n &= H(\mathcal{P}^{n+1}|\mathcal{Q}^{n+1} \vee \mathfrak{B}) - H(\mathcal{P}^n|\mathcal{Q}^n \vee \mathfrak{B}) = \\ &= H(\mathcal{P}^{n+1}|\mathcal{Q}^{n+1} \vee \mathfrak{B}) - H(T^{-1}(\mathcal{P}^n)|T^{-1}(\mathcal{Q}^n) \vee T^{-1}(\mathfrak{B})). \end{aligned}$$

We have $T^{-1}(\mathcal{Q}^n) \preceq \mathcal{Q}^{n+1}$ and, by subinvariance, $T^{-1}(\mathfrak{B}) \preceq \mathfrak{B}$, hence the right hand side does not exceed

$$H(\mathcal{P}^{n+1}|\mathcal{Q}^{n+1} \vee \mathfrak{B}) - H(T^{-1}(\mathcal{P}^n)|\mathcal{Q}^{n+1} \vee \mathfrak{B}) = H(\mathcal{P}|\mathcal{P}^{[1,n]} \vee \mathcal{Q}^{n+1} \vee \mathfrak{B}).$$

The expressions on the right decrease to $H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{N}_0} \vee \mathfrak{B})$, so we have proved that

$$h(\mathcal{P}|\mathcal{Q}, \mathfrak{B}) = \lim \frac{1}{n} H(\mathcal{P}^n|\mathcal{Q}^n \vee \mathfrak{B}) \leq H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{N}_0} \vee \mathfrak{B}).$$

Exercise 2.5.

Let $(X, \mathcal{P}, \mu, T, \mathbb{S})$ be any process with positive entropy h . Take \mathcal{Q} to be the trivial partition and set $\mathfrak{B} = \mathcal{P}^+$. Then $h(\mathcal{P}|\mathcal{Q}, \mathfrak{B}) = \lim \frac{1}{n} H(\mathcal{P}^n|\mathcal{P}^+) = \lim \frac{1}{n} H(\mathcal{P}|\mathcal{P}^+) = \lim \frac{1}{n} h = 0$, while $H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{N}_0} \vee \mathfrak{B}) = H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{P}^+) = H(\mathcal{P}|\mathcal{P}^+) = h > 0$.

Exercise 2.6.

Since $h(\mathcal{Q}) \leq h(\mathcal{Q} \vee \mathcal{P}) = h(\mathcal{Q}|\mathcal{P}) + h(\mathcal{P})$, it suffices to show that $h(\mathcal{Q}|\mathcal{P}) = 0$. By (2.3.11) (for trivial \mathfrak{B}), and since $\mathcal{Q} \preceq \mathcal{P}^{\mathbb{Z}}$, we do have $h(\mathcal{Q}|\mathcal{P}) = H(\mathcal{Q}|\mathcal{Q}^+ \vee \mathcal{P}^{\mathbb{Z}}) = 0$.

Exercise 2.7 ([Downarowicz–Serafin, 2002, Example 1]).

Let $X \subset \{0, 1, 2\}^{\mathbb{S}}$ consist of sequences in which 0 appears every other position (and not in between, e.g 0101020102... or 1010201020...). Let \mathcal{P} denote the zero-coordinate partition and let μ be the (shift-invariant) measure determined by saying that cylinders of even length $2n$ have equal masses 2^{-n-1} . The partition $\mathcal{Q} = \{0, 1 \cup 2\}$ is shift-invariant, so it determines a two-point factor $(Y, \mathcal{Q}, \nu, S, \mathbb{S})$ of $(X, \mathcal{P}, \mu, T, \mathbb{S})$. For ergodicity of $(X, \mathcal{P}, \mu, T, \mathbb{S})$ notice that T^2 is ergodic (in fact Bernoulli) on both 0 and $1 \cup 2$ so every T -invariant function (being T^2 -invariant) is constant on either set. Now T exchanges these sets, so the two constants must match. Obviously $h(\mu|\nu) = h(\mu) = 1/2$. On the other hand, the fiber entropy is not constant on Y : we have $h(\mathcal{P}|0) = \lim H(\mu_0, \mathcal{P}|\mathcal{P}^{[1,n]}) = 0$ (because \mathcal{P} is trivial on the fiber of 0). Now, since $h(\mu|\nu) = \frac{1}{2}(h(\mathcal{P}|0) + h(\mathcal{P}|1 \cup 2))$ it must be that $h(\mathcal{P}|1 \cup 2) = 1$.

Exercise 2.8.

By the Ergodic Theorem, the probability vectors $\mathbf{p}_{n, B_m(x)}$ assigning values to the elements of \mathcal{P}^n converge almost surely to the vector $\mathbf{p}(\mu, \mathcal{P}^n)$. For finite partition \mathcal{P} the assertion now follows from the definitions $H_n(B_m(x)) = \frac{1}{n} H(\mathbf{p}_{n, B_m(x)})$, $H(\mu, \mathcal{P}^n) = H(\mathbf{p}(\mu, \mathcal{P}^n))$ combined with continuity of static entropy on finite-dimensional probability vectors.

There is slight difficulty with countable partitions. By lower semicontinuity of the static entropy (Fact 1.1.8) we only have $\lim_m H_n(B_m(x)) \geq \frac{1}{n} H(\mu, \mathcal{P}^n)$. Note however, that $\mathbf{p}_{n, B_m(x)}$ is the same as the vector of masses assigned to the cells of \mathcal{P}^n by the

average measure $M_{m-n}\delta_x$. Integrating these probability vectors with respect to the invariant measure μ we get

$$\int \mathbf{p}_{n,B_m(x)} d\mu(x) = \mathbf{p}(\mu, \mathcal{P}^n).$$

Now we invoke the supharmonic property of H (Fact 1.1.10):

$$H(\mathbf{p}(\mu, \mathcal{P}^n)) \geq \int H(\mathbf{p}_{n,B_m(x)}) d\mu(x),$$

i.e., after dividing by n , $\frac{1}{n}H(\mu, \mathcal{P}^n) \geq \int H_n(B_m(x)) d\mu(x)$, for every m , and, by Fatou's Lemma,

$$\frac{1}{n}H(\mu, \mathcal{P}^n) \geq \int \liminf_m H_n(B_m(x)) d\mu(x).$$

This, together with the μ -almost sure converse inequality implies equality μ -almost everywhere.

3 Exercises in Chapter 3

Exercise 3.1.

Given $\varepsilon > 0$, we have, for large n : $\frac{1}{n}H(\mathcal{P}^n) \leq h(\mathcal{P}) + \varepsilon$, i.e.,

$$H(\mathcal{P}^n) \leq nh(\mathcal{P}) + n\varepsilon = h(\mu, T^n, \mathcal{P}^n) + n\varepsilon,$$

which is exactly the desired $n\varepsilon$ -entropy independence. It is not trivial whenever $\varepsilon < h(\mu, T, \mathcal{P})$ (then $n\varepsilon < h(\mu, T^n, \mathcal{P}^n)$). Since n is selected after fixing ε , the "error term" $n\varepsilon$ need not be small, so this does not translate to genuine δ -independence.

Exercise 3.2.

In an independent process $(X, \mathcal{P}, \mu, T, \mathbb{S})$ let $\mathcal{Q} = T^{-1}(\mathcal{P})$. Then $H(\mathcal{P}^n) = nH(\mathcal{P})$ and $H(\mathcal{P}^n|\mathcal{Q}^n) = H(\mathcal{P})$, so $\frac{1}{n}(H(\mathcal{P}^n) - H(\mathcal{P}^n|\mathcal{Q}^n)) = (1 - \frac{1}{n})H(\mathcal{P})$ which *increases* to its limit $H(\mathcal{P})$.

Note that $H(\mathcal{P}^n) - H(\mathcal{P}^n|\mathcal{Q}^n) = I(\mathcal{P}^n; \mathcal{Q}^n)$. Exercise 1.6 shows lack of subadditivity: $I(\mathcal{P}_1 \vee \mathcal{P}_2; \mathcal{Q}_1 \vee \mathcal{Q}_2)$ may exceed $I(\mathcal{P}_1; \mathcal{Q}_1) + I(\mathcal{P}_2; \mathcal{Q}_2)$ (Exercise 1.6 shows failure for $\mathcal{Q}_1 = \mathcal{Q}_2$). So, we cannot expect the sequence examined in this exercise to be subadditive. Now we have shown it need not even have descending n ths.

Exercise 3.3.

This follows directly from the Kolmogorov 0-1 law.

Exercise 3.4.

We have families Λ_k consisting of r_k blocks of lengths n_k . We have recursive relations $n_{k+1} = r_k n_k$ and $r_{k+1} = r_k!$. The key observation is that there is in fact a unique shift-invariant measure μ on our system X . It is so, because each block $B \in \Lambda_k$

appears exactly once in every block from Λ_{k+1} . Since every $x \in X$ is an infinite concatenation of the blocks from Λ_{k+1} , the ergodic theorem implies $\mu(B) = \frac{1}{n_{k+1}}$ for any ergodic measure. This determines the measure to be unique.

In order to compute the entropy, we will anticipate a bit and use the variational principle, which allows, in uniquely ergodic systems, to replace $h(\mu)$ by the easier topological entropy. Thus, we need to prove that the sequence $\frac{1}{n_k} \log \#\mathcal{B}_{n_k}$ has a positive limit, where \mathcal{B}_{n_k} is the family of all blocks of length n_k appearing in X . Clearly, $\#\mathcal{B}_{n_k} \geq \#\Lambda_k = r_k$, so it suffices to examine the sequence $\frac{1}{n_k} \log r_k$. It starts with $A = \frac{1}{2} \log \#\Lambda$, which we may assume much larger than 1. Then, by Stirling's formula,

$$\frac{1}{n_{k+1}} \log r_{k+1} \approx \frac{r_k \log r_k - r_k}{r_k n_k} = \frac{1}{n_k} \log r_k - \frac{1}{n_k}.$$

So, we have (roughly) $\lim \frac{1}{n_k} \log r_k \approx A - \sum_{k=1}^{\infty} \frac{1}{n_k}$. The sequence $1/n_k$ starts with $1/2$ and decreases much faster than exponentially, so its sum is smaller than 1. This shows that $\frac{1}{n_k} \log r_k$ has indeed a positive limit.

Exercise 3.5.

In any process $(X, \mathcal{P}, \mu, T, \mathbb{S})$ with positive entropy take $\mathfrak{B}_k = \mathcal{P}^{[k, \infty)}$. Then $\mathfrak{B} = \bigcap_k \mathfrak{B}_k$ is the Pinsker sigma-algebra and $h(\mathcal{P}|\mathfrak{B}) = h(\mathcal{P}) > 0$. On the other hand, for each k , $h(\mathcal{P}|\mathfrak{B}_k) = \lim_n \frac{1}{n} H(\mathcal{P}^n|\mathcal{P}^{[k, \infty)}) \leq \lim_n \frac{1}{n} H(\mathcal{P}^k) = 0$.

Exercise 3.6.

Let $(\Lambda^{\mathbb{S}}, \mu, \sigma, \mathbb{S})$ be a Bernoulli shift, i.e., $\mu = \mathbf{p}^{\mathbb{S}}$, where \mathbf{p} is a probability distribution on Λ . On the probability space (X, μ) consider the sequence of random variables $X_n(x) = -\log \mu(A_{\sigma^n x})$, where $A_{\sigma^n x}$ is the cell of \mathcal{P}_{Λ} containing $\sigma^n x$. They are independent and identically distributed with expected value $H(\mathbf{p})$ (which equals $h(\mu)$). The Law of Large Numbers asserts that the averages $\frac{1}{n} \sum_{i=0}^{n-1} X_i$ converge almost everywhere to this expected value. It suffices to note that $\sum_{i=0}^{n-1} X_i(x)$ equals $-\log \mu(A_x^n)$, i.e., the information function $I_{\mathcal{P}_{\Lambda}^n}(x)$.

Exercise 3.7.

Notice that the map $x \mapsto T^{\mathbf{R}_n(x)}(x)$ preserves the measure (it sends each cylinder of length n to itself and on every such cylinder it is just the induced map, which preserves the conditional measure). Let us abbreviate this map by S . Thus, for each $i \geq 1$ the variables $x \mapsto \mathbf{R}_n(S^i x)$ have the same distributions as \mathbf{R}_n , in particular the Ornstein–Weiss's assertion is fulfilled: $\lim_n \frac{1}{n} \log \mathbf{R}_n(S^i x) \rightarrow h(\mathcal{P})$ μ -almost everywhere. Given k , the variables

$$m_n(x) = \max\{\frac{1}{n} \log \mathbf{R}_n(S^i x) : 1 \leq i \leq k\} = \frac{1}{n} \log \max\{\mathbf{R}_n(S^i x) : 1 \leq i \leq k\}$$

also converge to $h(\mathcal{P})$ μ -a.e. Now

$$\mathbf{R}_n^{(k)}(x) = \mathbf{R}_n(x) + \mathbf{R}_n(Sx) + \mathbf{R}_n(S^2x) + \cdots + \mathbf{R}_n(S^{k-1}x)$$

lies between $\max\{\mathbf{R}_n(S^i x) : 1 \leq i \leq k\}$ and $k \max\{\mathbf{R}_n(S^i x) : 1 \leq i \leq k\}$. Thus $\frac{1}{n} \log \mathbf{R}_n^{(k)}(x)$ lies between $m_n(x)$ and $\frac{1}{n} \log k + m_n(x)$, which implies the desired convergence.

Exercise 3.8.

Attention! In the formulation the word “eventually” is missing: ... the cardinality of blocks of length n eventually exceeds $2^{n(\log l - \varepsilon)}$.

The numbers $p_k = \frac{1}{(l-1)l^{k-2}}$ decrease to zero as $k \rightarrow \infty$, so the following convex combinations of $l - 1$ and l

$$l_k = p_k(l - 1) + (1 - p_k)l$$

increase to l . Let k be so large that $\log l_k > \log l - \varepsilon$.

Let 0 and 1 denote two selected symbols from Λ . Let $W = 10000 \dots 0$ (with $k - 1$ zeros). Note that any two occurrences of W in an element of $\Lambda^{\mathbb{N}_0}$ are over disjoint intervals of coordinates. Let \mathcal{C}_n be the family of all blocks of length n over Λ , ending with W and in which W does not occur otherwise. It is obvious that $\bigcup_n \mathcal{C}_n$ is a prefix-free family. We need to estimate $c_n = \#\mathcal{C}_n$. A block $C \in \mathcal{C}_n$ is essentially a block C' of length $n - k$ in which W does not occur, with W appended on the right. So, c_n equals the cardinality of the family \mathcal{C}'_n of such blocks C' (of length $n - k$).

Let $n > 2k - 1$. Some of the blocks in \mathcal{C}'_n end (on the right) with $W' = 1000 \dots 0$ (with $k - 2$ zeros). We can change this ending, and as long as we do not replace the 1 by 0, we are sure that the new block still belongs to \mathcal{C}'_n . Thus, there are at least $(l - 1)l^{k-2} - 1$ such changed blocks made from one block ending with W' . This implies that the ratio between cardinalities of blocks ending with W' and not ending with W' in \mathcal{C}'_n is at most $p_k/(1 - p_k)$. Every block ending with W' can be prolonged to the right to a block belonging to \mathcal{C}'_{n+1} in $l - 1$ possible ways (we must not add 0 at the end). Every block not ending with W' can be prolonged arbitrarily (i.e., in l possible ways) and the prolonged block always belongs to \mathcal{C}'_{n+1} . Thus

$$c_{n+1} \geq c_n(p_k(l - 1) + (1 - p_k)l) = c_n l_k.$$

Since this is true for $n \geq 2k - 1$, we have proved that $c_n \geq l_k^{n-2k} = 2^{n \frac{n-2k}{n} (\log l_k)}$. Because $\log l_k$ is strictly larger than $\log l - \varepsilon$, so is $\frac{n-2k}{n} (\log l_k)$ for large enough n , hence c_n is eventually larger than $2^{n(\log l - \varepsilon)}$, as needed.

Remark. This family is better than prefix-free. A prefix-free family allows to determine the cutting places in every *unilateral* concatenation, but we must know where to start cutting. In infinite bilateral concatenations we may not be able to determine the cuts. As it was mentioned in the exercise, a prefix-free family is just a family of disjoint cylinders, so for example the family of all blocks of a given length n is prefix-free. Any sequence is now a bilateral concatenation and can be cut back in n different ways.

The family constructed in the solution (according to the hint) has the stronger property that the cutting places are determined in any bilateral concatenation (it suffices to find the blocks W). This is important in coding algorithms applicable to bilateral sequences.

Exercise 3.9.

Attention! The formulation of Theorem 3.5.1 contains errors. It should be mentioned that Λ is finite and that $\mathcal{P} = \mathcal{P}_\Lambda$. The second sentence should be deleted and the assertion should read:

Then, for every $\varepsilon > 0$, the joint measure of all blocks B of length m and compression rate smaller than $(h - \varepsilon)/\log \#\Lambda$ tends to zero with m .

Fix an $\varepsilon > 0$. Consider the blocks B of length m and compression rate smaller than $(h - \varepsilon)/\log \#\mathcal{P}$. Their compressed images are binary blocks of lengths smaller than $m(h - \varepsilon)$, so there at most $2^{m(h - \varepsilon)}$ such blocks. In particular, the conditional entropy of \mathcal{P}^m on any subset of their union, denoted A_m , is smaller than $m(h - \varepsilon)$. Next, fix n so large that $\frac{1}{n}H(\mathcal{P}^n) < h + \delta$, where δ is some small positive number. In order for a block B to satisfy $H_n(B) \leq h + \delta$ we need $H_n(B)$ to be closer to $\frac{1}{n}H(\mathcal{P}^n)$ than the (positive) difference $h + \delta - \frac{1}{n}H(\mathcal{P}^n)$. By continuity of entropy on finite-dimensional vectors, it suffices that the vector $\mathbf{p}_{n,B}$ (of frequencies in B of blocks of length n) is very close to the vector $\mathbf{p}(\mu, \mathcal{P}^n)$. By the Ergodic Theorem, for m sufficiently large, this is satisfied for blocks B of length m covering a set X_m of measure at least $1 - \delta$. By Lemma 2.8.2, if m is large enough, the cardinality of blocks in X_m does not exceed $2^{m(h + \delta)}$. In particular, the conditional entropy of \mathcal{P}^m on any subset of X_m is smaller than $m(h + \delta)$. We divide X into three parts: $X_m \cap A_m$ (treated as a subset of A_m), $X_m \setminus A_m$ (a subset of X_m) and the rest, $X \setminus X'$, whose measures are estimated from above by $\mu(A_m)$, $1 - \mu(A_m)$ and δ , respectively. We have

$$mh \leq H(\mathcal{P}^m) \leq \mu(A_m)m(h - \varepsilon) + (1 - \mu(A_m))m(h + \delta) + \delta m \log \#\mathcal{P} + \log 3.$$

This easily implies

$$\mu(A_m) \leq \frac{\delta}{\varepsilon + \delta} (1 + \log \#\mathcal{P} + \frac{\log 3}{m\delta}).$$

Passing with m to infinity we get $\limsup_m \mu(A_m) \leq \frac{\delta}{\varepsilon + \delta} (1 + \log \#\mathcal{P})$ and since δ is arbitrarily small, we conclude that $\lim_m \mu(A_m) = 0$.

Exercise 3.10.

We skip the solution of this exercise.

4 Exercises in Chapter 4

Exercise 4.1.

In fact, it suffices to assume that the sequence \mathcal{P}_k generates *under the action*, that is, the double sequence of partitions \mathcal{P}_k^n (in case $\mathbb{S} = \mathbb{N}_0$) or $\mathcal{P}_k^{[-n, n]}$ (in case $\mathbb{S} = \mathbb{Z}$) ((k, n) range over $\mathbb{N}_0 \times \mathbb{N}_0$) generates \mathfrak{A} .

Given any finite partition \mathcal{P} of cardinality m , let k and n be so large that there exists a partition $\mathcal{P}' \preceq \mathcal{P}_k^n$ (in case $\mathbb{S} = \mathbb{N}_0$; $\mathcal{P}' \preceq \mathcal{P}_k^{[-n, n]}$ for $\mathbb{S} = \mathbb{Z}$) also of cardinality m , with $d_{\mathbb{R}}(\mathcal{P}, \mathcal{P}') < \varepsilon$ (recall that in \mathfrak{P}_m the metric $d_{\mathbb{R}}$ is uniformly equivalent to d_1). Then, by (2.4.10), regardless of the conditioning sigma-algebra, $|h(\mathcal{P}|\mathfrak{B}) - h(\mathcal{P}'|\mathfrak{B})| < \varepsilon$, which implies $h(\mathcal{P}|\mathfrak{B}) < h(\mathcal{P}_k^{[2n+1]}|\mathfrak{B}) + \varepsilon = h(\mathcal{P}_k|\mathfrak{B}) + \varepsilon$ (we have also used Fact 2.4.1). Taking the (increasing) limit over k on the right and then supremum on the left, we get $h(\mathfrak{A}|\mathfrak{B}) \leq \lim_k h(\mathcal{P}_k|\mathfrak{B}) + \varepsilon$. Now we can remove ε and combine the result with the obvious converse inequality.

Exercise 4.2.

This is now an immediate consequence of the preceding exercise and Theorem 2.5.1.

Exercise 4.3.

We will prove the formula $h(\mu, T^n | \mathfrak{B}) = |n|h(\mu, T | \mathfrak{B})$ assuming \mathfrak{B} to be invariant only for negative n , otherwise it works for subinvariant \mathfrak{B} (regardless of \mathbb{S}). We copy the proof of Fact 4.1.14 almost verbatim. By Fact 2.4.19 (the conditional version; this is where invariance of \mathfrak{B} may be needed), we have, for every partition \mathcal{P} : $h(\mu, T^n, \mathcal{P} | \mathfrak{B}) \leq h(\mu, T^n, \mathcal{P}^{|n|} | \mathfrak{B}) = |n|h(\mu, T, \mathcal{P} | \mathfrak{B})$. Now we apply $\sup_{\mathcal{P}}$ and get

$$h(\mu, T^n | \mathfrak{B}) \leq |n|h(\mu, T | \mathfrak{B}) = \sup_{\mathcal{Q}=\mathcal{P}^{|n|}} h(\mu, T^n, \mathcal{Q} | \mathfrak{B}) \leq h(\mu, T^n | \mathfrak{B}).$$

Exercise 4.4.

We will actually prove Theorem 4.2.9 without using Remark 4.2.7 and then prove that remark using the theorem. Given a finite partition \mathcal{Q} measurable with respect to Π_μ , we can approximate it up to ε in d_1 (equivalently in d_R) by a partition \mathcal{Q}' of the same cardinality as \mathcal{Q} , measurable with respect to a finite join $\bigvee_{i=1}^k \Pi_{\mathcal{P}_i}$ for some partitions \mathcal{P}_i . We do not need Remark 4.2.7 for that; the partitions \mathcal{P}_i need not be members of any *a priori* fixed sequence of partitions. The difficulty lies in understanding the join of possibly uncountably many sigma-algebras. In fact, Π_μ is, by definition, the smallest sigma-algebra containing $\bigcup \Pi_{\mathcal{P}}$ (union over all finite partitions \mathcal{P}). But every set in Π_μ is obtained via countably many set operations involving at most countably many sets in that union, so it is contained in $\bigvee_{i=1}^\infty \Pi_{\mathcal{P}_i}$ for some sequence of finite partitions \mathcal{P}_i . Now we can use the usual approximation within this countable join. The rest of proof of Theorem 4.2.9 is (almost) unchanged. We remark, that without assuming the partitions \mathcal{P}_i to be linearly ordered by \preceq , it is no longer true that the join $\bigvee_{i=1}^k \Pi_{\mathcal{P}_i}$ equals $\Pi_{\mathcal{P}}$, where $\mathcal{P} = \bigvee_{i=1}^k \mathcal{P}_i$; it is only refined (which is very easy to see). There are examples where the converse refining fails (perhaps this should be better emphasized in the book). Anyway, the valid direction is sufficient in the proof of Theorem 4.2.9.

We shall now prove Remark 4.2.7 in a stronger version, which includes the case of one generating partition. Namely, we will only assume that the refining sequence \mathcal{P}_k generates under the action (as we did in the solution of Exercise 4.1).

Clearly, $\bigvee_{k=1}^\infty \Pi_{\mathcal{P}_k} \preceq \Pi_\mu$. We need to prove the converse. Let \mathcal{Q} be a partition measurable with respect to Π_μ . Then, by Theorem 4.2.9, $h(\mu, \mathcal{Q}, T) = 0$ and hence, by the power rule (Fact 2.4.19), $h(\mu, T^n, \mathcal{Q}) \leq h(\mu, T^n, \mathcal{Q}^n) = 0$ for every $n \geq 1$. Now we approximate \mathcal{Q} by \mathcal{Q}' up to ε in d_R , where $\mathcal{Q}' \preceq \mathcal{P}_k$ for some k (here it is important that \mathcal{P}_k is a refining sequence). By (2.4.10) (for T^n and trivial \mathfrak{B}), we have $|h(\mu, T^n, \mathcal{Q}') - h(\mu, T^n, \mathcal{Q})| < \varepsilon$, i.e., $h(\mu, T^n, \mathcal{Q}') < \varepsilon$ for all $n \geq 1$. Further, notice that $h(\mu, T^n, \mathcal{Q}') = H(\mathcal{Q}' | \mathcal{Q}'^{\{n, 2n, 3n, \dots\}}) \geq H(\mathcal{Q}' | \mathcal{Q}'^{[n, \infty)})$. As a consequence, $H(\mathcal{Q}' | \mathcal{P}_k^{[n, \infty)}) < \varepsilon$ for every n . Now we invoke (1.7.14) and get $H(\mathcal{Q}' | \Pi_{\mathcal{P}_k}) < \varepsilon$, all the more $H(\mathcal{Q}' | \bigvee_k \Pi_{\mathcal{P}_k}) < \varepsilon$. We use (1.6.36) and get $H(\mathcal{Q} | \bigvee_k \Pi_{\mathcal{P}_k}) < 2\varepsilon$. Since this is true for every ε , $H(\mathcal{Q} | \bigvee_k \Pi_{\mathcal{P}_k}) = 0$ and (1.6.28) implies $\mathcal{Q} \preceq \bigvee_k \Pi_{\mathcal{P}_k}$.

Exercise 4.5.

We have $h(\mathfrak{A}|\mathfrak{B}) = \sup_{\mathcal{P}} h(\mathcal{P}|\mathfrak{B}) > 0$, where \mathcal{P} ranges over all finite \mathfrak{A} -measurable partitions of X . Thus, there exists a finite partition \mathcal{P} of X such that $h(\mathcal{P}|\mathfrak{B}) > 0$. We restrict our attention to the system generated jointly by \mathcal{P} and $(Y, \mathfrak{B}, \nu, S, \mathbb{S})$. We will prove that almost every y has an infinite preimage already in this system. In other words, we can assume that $\mathfrak{A} = \mathcal{P}^{\mathbb{S}} \vee \mathfrak{B}$.

Suppose that the set of points y with finite preimages $\pi^{-1}(y)$ has positive measure ν . For y in this set, the points in the preimage of y are distinguished by their \mathcal{P} -names, so there is a minimal n_y such that all these points are in different cells of $\mathcal{P}^{[-n_y, n_y]}$ (or just \mathcal{P}^{n_y} for $\mathbb{S} = \mathbb{N}_0$). Some value of n_y , say n_0 must occur with positive measure ν , say on a set A_0 . By ergodicity, the value of the time of the first visit in A_0 is finite almost everywhere, so it is bounded by some n_1 on a set A_1 of measure ν larger than $\delta = \frac{h(\mathfrak{A}|\mathfrak{B})}{\log \#\mathcal{P}}$. That means that relatively on A_1 , $\mathcal{P}^{\mathbb{S}} \vee \mathfrak{B} = \mathcal{R} \vee \mathfrak{B}$, where \mathcal{R} is a finite partition (\mathcal{R} can be $\mathcal{P}^{[-n, n]}$ joined with the partition determined by the entry times to D trimmed at n_1). Now, for every n , we can write

$$H(\mathcal{P}^n|\mathfrak{B}) \leq H(\mathcal{P}^n|\mathfrak{B} \vee \varkappa) + H(\varkappa) \leq (1 - \delta)H_{A_1}(\mathcal{R}|\mathfrak{B}) + \delta \log \#\mathcal{P}^n + H(\delta, 1 - \delta),$$

where \varkappa is the partition into A_1 and its complement. After dividing by n and passing with n to infinity, the left hand side converges to $h(\mathfrak{A}|\mathfrak{B})$. The first and last terms on the right hand side (divided by n) decrease to zero, while the middle term becomes, by the choice of δ , strictly smaller than $h(\mathfrak{A}|\mathfrak{B})$, a contradiction.

Exercise 4.6.

By (2.3.5) and Exercise 2.4 (with trivial \mathfrak{B}), $h(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{S}})$ in any case of \mathbb{S} . Further,

$$h(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{S}}) \leq H(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^{\mathbb{S}} \vee \mathcal{Q}) = \sum_{B \in \mathcal{Q}} \mu(B) H_B(\mathcal{P}|\mathcal{P}^+ \vee \mathcal{Q}^+).$$

The full future $\mathcal{P}^+ \vee \mathcal{Q}^+$ of the joint process, restricted to B , obviously contains the future of the process generated by \mathcal{P} with respect to the induced map on B , denoted by $\mathcal{P}^{B,+}$ (all we need to determine the forward $\mathcal{P}^{B,+}$ -name of an $x \in B$ is its forward \mathcal{P} -name and the return times to B , which are \mathcal{Q}^+ -measurable). So, we get

$$h(\mathcal{P}|\mathcal{Q}) \leq \sum_{B \in \mathcal{Q}} \mu(B) H_B(\mathcal{P}|\mathcal{P}^{B,+}) = \sum_{B \in \mathcal{Q}} \mu(B) h(\mu_B, T_B, \mathcal{P}).$$

To see an example with sharp inequality, suppose $h(\mathcal{P}_0) = h > 0$, while \mathcal{Q} generates a periodic factor with some period p . Set $\mathcal{P} = \mathcal{P}_0^p$. Then $h(\mathcal{P}|\mathcal{Q}) = h(\mathcal{P}_0^p|\mathcal{Q}) = h$. Notice that for each $B \in \mathcal{Q}$, $T_B = T^p$. Since $\mu = \sum_{B \in \mathcal{Q}} \mu(B)\mu_B$ and all these measures are T^p -invariant, by affinity of the dynamical entropy and the power rule, we have

$$\sum_{B \in \mathcal{Q}} \mu(B) h(\mu_B, T^p, \mathcal{P}) = h(\mu, T^p, \mathcal{P}) = h(\mu, T^p, \mathcal{P}_0^p) = ph(\mu, T, \mathcal{P}_0) = ph > h.$$

Exercise 4.7.

Just take any endomorphism of finite entropy that does not admit a unilateral generator. For example, T can be invertible with positive entropy, yet we consider only the action of \mathbb{N}_0 . If you want a genuine (not invertible) endomorphism, consider the direct product of a bilateral Bernoulli shift with a unilateral Bernoulli shift. If there existed a unilateral generator \mathcal{P} , the bilateral Bernoulli factor would be, by invariance, measurable with respect to $\mathcal{P}^{(n, \infty)}$ for every positive n , and hence with respect to the Pinsker sigma-algebra, which is impossible due to positive entropy.

Exercise 4.8.

For any pair of partitions \mathcal{P} and \mathcal{Q} measurable with respect to \mathfrak{A} and \mathfrak{B} , correspondingly, we have

$$H((\mathcal{P} \vee \mathcal{Q})^n | \mathcal{C}') = H(\mathcal{P}^n \vee \mathcal{Q}^n | \mathcal{C}') = H(\mathcal{P}^n | \mathcal{C}') + H(\mathcal{Q}^n | \mathcal{C}').$$

Dividing by n and passing to the limit we get $h(\mathcal{P} \vee \mathcal{Q} | \mathcal{C}') = h(\mathcal{P} | \mathcal{C}') + h(\mathcal{Q} | \mathcal{C}')$. It suffices to take supremum over all such pairs of partitions, and notice that the joined partitions $\mathcal{P} \vee \mathcal{Q}$ generate the joined sigma-algebra $\mathfrak{A} \vee \mathfrak{B}$, to get the desired equality $h(\mathfrak{A} \vee \mathfrak{B} | \mathcal{C}') = h(\mathfrak{A} | \mathcal{C}') + h(\mathfrak{B} | \mathcal{C}')$.

Exercise 4.9.

Attention! In the formulation, the ergodicity assumption is obviously missing (otherwise we may have no generator at all).

By Theorem 4.5.1 (Sinai), our system has a partition \mathcal{P} which generates an independent process with full entropy h . By Theorems 4.4.7, there is a partition \mathcal{Q} which generates a process of entropy smaller than $\varepsilon/2$ and such that $\mathcal{P} \vee \mathcal{Q}$ generates everything. We will prove that we can replace the partition \mathcal{Q} by another, \mathcal{Q}' , which generates the same factor as \mathcal{Q} and has static entropy smaller than ε . This will end the proof, as then $\mathcal{P} \vee \mathcal{Q}'$ is a generator and $H(\mathcal{P} \vee \mathcal{Q}') < H(\mathcal{P}) + \varepsilon = h + \varepsilon$, i.e., the corresponding process is ε -independent.

Let r be so large that $H(\mathcal{Q}^r) < r\varepsilon/2$. By a standard modification of the Rokhlin Lemma, there is a set $A \in \mathcal{Q}^{\mathbb{Z}}$ such that the return time to A assumes only two values: $r-1$ and r . Let \varkappa be the partition $\{A, T(A), \dots, T^{r-2}(A), B\}$, where B is the remaining set (contained in $T^{r-1}(A)$). We have

$$\frac{r\varepsilon}{2} > H(\mathcal{Q}^r) \geq H(\mathcal{Q}^r | \varkappa) \geq \sum_{n=0}^{r-2} \mu(T^n(A)) H_{T^n(A)}(\mathcal{Q}^r) \geq \frac{1}{r} \sum_{n=0}^{r-2} H_{T^n(A)}(\mathcal{Q}^r).$$

We multiply both sides by $r/(r-1)$ and get

$$\frac{1}{r-1} \sum_{n=0}^{r-2} H_{T^n(A)}(\mathcal{Q}^r) < \frac{\varepsilon}{2} \frac{r^2}{r-1}$$

which implies that for at least one index $n_0 \in \{0, r-2\}$, $H_{T^{n_0}(A)}(\mathcal{Q}^r) < \frac{\varepsilon}{2} \frac{r^2}{r-1}$.

Let \mathcal{Q}' be defined as \mathcal{Q}^r intersected with $A' = T^{n_0}(A)$ and the complement of A' in one piece. Notice that the partition \mathcal{Q}' generates the same process as \mathcal{Q} : meaningful symbols in the \mathcal{Q}' -name of a point x occur at coordinates n for which $T^n x \in A'$ and they encode the blocks of length r starting at n in the \mathcal{Q} -name of x , while the next such symbol is not further than r positions forward. Denoting by \mathcal{R} the partition into A' and its complement, the static entropy of \mathcal{Q}' can be estimated as follows

$$H(\mathcal{Q}') \leq H(\mathcal{Q}'|\mathcal{R}) + H(\mathcal{R}) \leq \mu(A')H_{A'}(\mathcal{Q}^r) + 0 + H(\mathcal{R}) \leq \frac{\varepsilon}{2} \frac{r^2}{(r-1)^2} + H\left(\frac{1}{r-1}, 1 - \frac{1}{r-1}\right) < \varepsilon$$

if r is chosen large enough. This concludes the construction.

Exercise 4.10.

Attention! Again, in the formulation, the ergodicity assumption is obviously missing. First of all, recall that the original proof of the Sinai Theorem is valid also for endomorphisms, this is why the exercise is formulated for both cases of \mathbb{S} . Now, it suffices to take for \mathcal{P} a generator of a Bernoulli factor of full entropy, which can be finite in the finite entropy case.

Part 2

6 Exercises in Chapter 6

Exercise 6.1.

Consider the (unilateral or bilateral) subshift of finite type on three symbols $\{0, 1, 2\}$, where we prohibit repetitions 00, 11, and 22. Let the cover \mathcal{U} depend on the zero coordinate and consist of the unions of two symbols each: $\mathcal{U} = \{0 \cup 1, 1 \cup 2, 0 \cup 2\}$. There are 6 admitted words of length 2: 01, 02, 12, 10, 20 and 21 and they are covered by two sets (for example) $(0 \cup 1) \times (1 \cup 2)$ and $(1 \cup 2) \times (1 \cup 0)$ (in the order they are written). So, $\frac{1}{2}\mathbf{H}(\mathcal{U}^2) = \frac{1}{2} \log 2 = \log \sqrt{2}$. To keep $\frac{1}{3}\mathbf{H}(\mathcal{U}^3)$ not increased, we need $N(\mathcal{U}^3)$ not larger than $2\sqrt{2}$ (strictly smaller than 3), which means that we would have to cover all admitted blocks of length 3 by only two elements of \mathcal{U}^3 . This is impossible, because there are 12 admitted words of length 3, while each element of \mathcal{U}^3 contains at most 4 of them; in the definition of $U \in \mathcal{U}^3$ we must specify 3 pairs of different symbols, so at least two symbols must be used twice, which means that U (containing a priori 8 words) contains at least 4 forbidden words, hence at most 4 admitted words.

Exercise 6.2.

This is a direct consequence of $(\mathcal{U}^n)^m = \mathcal{U}^{n+m}$ and the convergence $\frac{m+n}{m} \xrightarrow{m} 1$.

Exercise 6.3.

Attention! The statement is in general false. Any system (X, T, \mathbb{S}) is topologically conjugate to the subsystem of the unilateral shift on $X^{\mathbb{N}_0}$ consisting of the forward orbits. In the product metric

$$d_p((x_n), (y_n)) = \sum_{n=0}^{\infty} \frac{1}{2^n} d(x_n, y_n)$$

the shift map is Lipschitz with the constant $c = 2$, while it can have arbitrarily large topological entropy (the same as (X, T, \mathbb{S})).

The statement does hold if $c \leq 1$ (an important application of that is, that all isometries have entropy zero). In such case all the metrics d^n are equal to d hence the number of (n, ε) -separated points does not grow with n .

For arbitrary Lipschitz constants c the statement is valid for C^1 interval maps in the standard metric, which follows e.g. from the Margulis-Ruelle Inequality (9.4.1) (and the Variational Principle).

Exercise 6.4.

Let $\mathcal{P} = \{I_1, I_2, \dots, I_N\}$ be the partition of $[0, 1]$ into the branches of monotonicity (intervals on which T is monotone) $I_1 = [0, a_1), I_2 = [a_1, a_2), \dots, I_N = [a_{N-1}, 1]$. Notice that for each $n \geq 1$ the cells $J \in \mathcal{P}^n$ are in fact intervals on which all the iterates T, T^2, \dots, T^{n-1} are monotone. We will estimate the number of (n, ε) -separated points

contained in J . Consider the intervals between neighboring (n, ε) -separated points in J . Each of them must be stretched to at least the length ε by one of the functions T, T^2, \dots, T^{n-1} . By monotonicity, each of these functions can stretch at most $1/\varepsilon$ of these intervals, because their images are disjoint. Together, at most n/ε intervals can be stretched, which limits the cardinality of the points to $n/\varepsilon + 1$. Now, the total number of (n, ε) -separated points is at most $N^n(\frac{n}{\varepsilon} + 1)$ (where N^n bounds the number of cells J). The desired estimate of $\mathbf{h}(T)$ is obtained by taking the logarithm, dividing by n , passing with n to infinity and then letting ε tend to zero.

Exercise 6.5.

Let $T \vee S$ denote the joining of T and S within the common extension. Fact (6.4.13) applied twice (first to $T \vee S$ and S , then to $T \vee S$ and T) and the triangle inequality yield

$$|\mathbf{h}^*(S) - \mathbf{h}^*(T)| \leq \mathbf{h}(T \vee S|S) + \mathbf{h}(T \vee S|T).$$

By (6.5.8), $\mathbf{h}(T \vee S|S) = \mathbf{h}(T|S)$ and $\mathbf{h}(T \vee S|T) = \mathbf{h}(S|T)$.

Exercise 6.6.

First take $n \geq 0$. Notice that $(\mathcal{U}^n)^m$, where the exponent m refers to the action of T^m , equals \mathcal{U}^{nm} (in the action of T). So,

$$\mathbf{h}(T^n, \mathcal{U}^n|\mathcal{V}^n) = \lim_m \frac{1}{m} \mathbf{H}(\mathcal{U}^{nm}|\mathcal{V}^{nm}) = n\mathbf{h}(T, \mathcal{U}|\mathcal{V})$$

(since the limit defining $\mathbf{h}(T, \mathcal{U}|\mathcal{V})$ exists, it is achieved along the subsequence nm). Further,

$$\begin{aligned} \mathbf{h}(T^n|\mathcal{V}^n) &= \sup_{\mathcal{W}} \mathbf{h}(T^n, \mathcal{W}|\mathcal{V}^n) \geq \sup_{\mathcal{W}=\mathcal{U}^n} \mathbf{h}(T^n, \mathcal{U}^n|\mathcal{V}^n) = \\ & \sup_{\mathcal{U}} n\mathbf{h}(T, \mathcal{U}|\mathcal{V}) = n\mathbf{h}(T|\mathcal{V}). \end{aligned}$$

On the other hand, since $\mathcal{U}^n \succcurlyeq \mathcal{U}$, we also have

$$\mathbf{h}(T^n|\mathcal{V}^n) = \sup_{\mathcal{U}} \mathbf{h}(T^n, \mathcal{U}|\mathcal{V}^n) \leq \sup_{\mathcal{U}} \mathbf{h}(T^n, \mathcal{U}^n|\mathcal{V}^n) = n\mathbf{h}(T|\mathcal{V}).$$

We have proved the equality $\mathbf{h}(T^n|\mathcal{V}^n) = n\mathbf{h}(T|\mathcal{V})$. We proceed similarly with the conditioning covers:

$$\mathbf{h}^*(T^n) = \inf_{\mathcal{W}} \mathbf{h}(T^n|\mathcal{W}) \leq \inf_{\mathcal{W}=\mathcal{V}^n} \mathbf{h}(T^n|\mathcal{V}^n) = \inf_{\mathcal{V}} n\mathbf{h}(T|\mathcal{V}) = n\mathbf{h}^*(T),$$

and, since $\mathcal{V}^n \succcurlyeq \mathcal{V}$,

$$\mathbf{h}^*(T^n) = \inf_{\mathcal{V}} \mathbf{h}(T^n|\mathcal{V}) \geq \inf_{\mathcal{V}} \mathbf{h}(T^n|\mathcal{V}^n) = n\mathbf{h}^*(T).$$

It remains to show that for homeomorphisms, $\mathbf{h}^*(T^{-1}) = \mathbf{h}^*(T)$. Note that \mathcal{U}^n , where the exponent refers to the action of T^{-1} , equals $\mathcal{U}^{[-n+1, 0]}$ (in the notation referring to T). Since T is a homeomorphism, we have $\mathbf{H}(\mathcal{U}^{[-n+1, 0]}|\mathcal{V}^{[-n+1, 0]}) = \mathbf{H}(\mathcal{U}^n|\mathcal{V}^n)$. This equality passes via all intermediate definitions leading to $\mathbf{h}^*(T^{-1}) = \mathbf{h}^*(T)$.

Exercise 6.7.

For $n \geq 0$ we use again that $(\mathcal{U}^n)^m$, where the exponent m refers to the action of T^n , equals \mathcal{U}^{nm} . We have

$$\mathbf{h}(T^n, \mathcal{U}^n|y) = \limsup_{m \rightarrow \infty} \frac{1}{m} \mathbf{H}(\mathcal{U}^{nm}|y) = n \cdot \limsup_{m \rightarrow \infty} \frac{1}{mn} \mathbf{H}(\mathcal{U}^{nm}|y) \leq n\mathbf{h}(T, \mathcal{U}|y).$$

This time we have no subadditivity to deduce the existence of the limit. Instead, to prove the converse inequality, we will use monotonicity. Every positive integer can be written as $i = m_i n - r_i$ with $m_i \geq 1$ and $0 \leq r_i \leq n - 1$. Then $\mathbf{H}(\mathcal{U}^i|y) \leq \mathbf{H}(\mathcal{U}^{m_i n}|y)$, so

$$n\mathbf{h}(T, \mathcal{U}|y) \leq n \limsup_{i \rightarrow \infty} \frac{1}{i} \mathbf{H}(\mathcal{U}^i|y) \leq \frac{m_i n}{i} \frac{1}{m_i} \mathbf{H}(\mathcal{U}^{m_i n}|y) = \mathbf{h}(T^n, \mathcal{U}^n|y),$$

because $\frac{m_i n}{i} \rightarrow 1$. Further, the definition of $\mathbf{h}(T|y)$ involves the supremum over \mathcal{U} , which is handled identically as in the preceding exercise (we skip rewriting). If we replace the conditioning y by a measure ν , the only essential difference in the definitions is the presence of \inf rather than \limsup . So now we first derive $\mathbf{h}(T^n, \mathcal{U}^n|\nu) \geq n\mathbf{h}(T, \mathcal{U}|\nu)$, and we use monotonicity for the converse inequality (we represent i as $m_i n + r_i$, so that $i \geq m_i n$).

We remark that for an invariant measure ν we can alternatively use Corollary 6.7.4 (c) and (d), and power rules pass to $\mathbf{h}(T, \mathcal{U}|\nu)$ and $\mathbf{h}(T|\nu)$ via integration.

Attention! For negative n the equalities $\mathbf{h}(T^n, \mathcal{U}^{|n|}|y) = |n|\mathbf{h}(T, \mathcal{U}|y)$ and $\mathbf{h}(T^n|y) = |n|\mathbf{h}(T|y)$ may actually fail. An easy example is the subshift of finite type on three symbols $\Lambda = \{0, 1, 2\}$ where we prohibit the blocks 10 and 20, the cover \mathcal{U} equal to the zero-coordinate partition \mathcal{P}_Λ , and the factor map that glues together the symbols 1 and 2 (to a symbol denoted in the factor as 1). The fiber of $y = \dots 00\underline{0}111 \dots$ (we underline the coordinate 0) intersects 2^{n-1} elements of \mathcal{U}^n , while only one element of $\mathcal{U}^{\lfloor -n+1, 0 \rfloor}$. Hence $\mathbf{h}(T^n, \mathcal{U}|y) = \log 2 \neq 0 = \mathbf{h}(T^{-1}, \mathcal{U}|y)$. It is not hard to see that in fact $\mathbf{h}(T^{-1}, \mathcal{U}|y) = 0$ for any other finite cover \mathcal{U} , so $\mathbf{h}(T|y) \geq \log 2 \neq 0 = \mathbf{h}(T^{-1}|y)$. Nevertheless, the equalities $\mathbf{h}(T^n, \mathcal{U}|\nu) = |n|\mathbf{h}(T, \mathcal{U}|\nu)$ and $\mathbf{h}(T^n|\nu) = |n|\mathbf{h}(T|\nu)$ do hold whenever ν is invariant. It is so because $\mathbf{H}(\mathcal{U}^n|y) = \mathbf{H}(\mathcal{U}^{\lfloor -n+1, 0 \rfloor}|T^{-n+1}y)$. The change of the variable vanishes after integrating with respect to an invariant measure, so $\mathbf{H}(\mathcal{U}^n|\nu) = \mathbf{H}(\mathcal{U}^{\lfloor -n+1, 0 \rfloor}|\nu)$, which easily implies the above two equalities for $n = -1$ (and hence for all $n < 0$).

Exercise 6.8.

See Exercise 6.2.

Exercise 6.9.

This is nontrivial only when $\mathbf{h}(T) < \infty$. Then all involved measure-theoretic entropies are finite. By the Inner Variational Principle (Theorem 6.8.4), then Fact 4.1.6, and finally the Variational Principle, we have

$$\mathbf{h}(T|\nu) + h(\nu) = \sup_{\mu \in \pi^{-1}(\nu)} h(\mu|\nu) + h(\nu) = \sup_{\mu \in \pi^{-1}(\nu)} h(\mu) \leq \mathbf{h}(T).$$

Exercise 6.10.

Let $\pi = \psi\phi$. We have, using consecutively the Inner Variational Principle, the Variational Principle and the Conditional Variational Principle,

$$\begin{aligned} \mathbf{h}(T|\xi) &= \sup_{\mu \in \pi^{-1}(\xi)} h(\mu|\xi) = \sup_{\mu \in \pi^{-1}(\xi)} (h(\mu|\phi(\mu)) + h(\phi(\mu)|\xi)) \leq \\ &\sup_{\mu \in \mathcal{M}_T(X)} h(\mu|\phi(\mu)) + \sup_{\nu \in \psi^{-1}(\xi)} h(\nu|\xi) = \mathbf{h}(T|S) + \mathbf{h}(S|\xi). \end{aligned}$$

Exercise 6.11.

Let both spaces be $\{0, 1, 2\}$, let both covers be $\mathcal{U} = \mathcal{V} = \{\{0, 1\}, \{0, 2\}, \{1, 2\}\}$. We have $N(\mathcal{U}) = N(\mathcal{V}) = 2$, while $N(\mathcal{U} \otimes \mathcal{V}) = 3 < 4$ (the product space is covered for example by $\{0, 1\} \times \{0, 1\}$, $\{1, 2\} \times \{1, 2\}$ and $\{0, 2\} \times \{0, 2\}$).

Exercise 6.12.

The inequality 2. is easy and can be derived using only the Outer Variational Principle and (6.5.8), as follows:

$$\mathbf{h}(T|\xi) \leq \mathbf{h}(T|R) = \mathbf{h}(S \vee R|R) \leq \mathbf{h}(S).$$

(An alternative way is via the Inner Variational Principle and the second inequality in Fact 4.4.3. Yet another alternative is to first prove 1. and then use Corollary 6.7.4 (d).)

The inequality 1. is a bit harder. By Definition 6.5.2, Fact 6.5.9 (and its proof), we can think of (X, T, \mathbb{S}) as a subsystem of $(Y, S, \mathbb{S}) \times (Z, R, \mathbb{S})$ and we can restrict our attention to product covers $\mathcal{U} \otimes \mathcal{V}$. Note that $(\mathcal{U} \otimes \mathcal{V})^n$ (the exponent refers to $T = S \times R$) equals $\mathcal{U}^n \otimes \mathcal{V}^n$ (exponents refer to S and R , respectively). At any point $z \in Z$ we have $\mathbf{H}(\mathcal{U}^n \otimes \mathcal{V}^n|z) \leq \mathbf{H}(\mathcal{U}^n)$. We divide both sides by n , pass to \limsup_n , then apply supremum over all pairs of covers \mathcal{U} and \mathcal{V} .

7 Exercises in Chapter 7

Exercise 7.1.

Denote our subshift by (X, T, \mathbb{S}) . The inequality $\limsup_k \frac{1}{p_k} \log \#\mathcal{B}_k \leq \mathbf{h}(T)$ is obvious; \mathcal{B}_k contains only blocks of length p_k appearing in our subshift (usually not all of them).

To derive the converse inequality consider the set A_k of points having a p_k -periodic marker at the coordinate zero. Clearly, A_k is compact, T^{p_k} -invariant and with the action of T^{p_k} it is conjugate to a subshift over the alphabet \mathcal{B}_k . Thus the topological entropy of $(A_k, T^{p_k}, \mathbb{S})$ does not exceed $\log \#\mathcal{B}_k$. The space X contains the disjoint union X' of sets $A_k, T(A_k), \dots, T^{p_k-1}(A_k)$, and the systems $(T^i(A_k), T^{p_k}, \mathbb{S})$ are factors of $(A_k, T^{p_k}, \mathbb{S})$ via T^i , so their topological entropies are not larger than $\log \#\mathcal{B}_k$. In addition, in the unilateral case, there maybe points not belonging to X' but all such point fall into X' after less than p_k iterates. This proves that $\mathbf{h}(T^{p_k}) \leq \log \#\mathcal{B}_k$ on the entire space X and, by the power rule for topological entropy (Fact 6.2.3), $\mathbf{h}(T) \leq \frac{1}{p_k} \log \#\mathcal{B}_k$, for every k . This implies the existence of the limit and the desired equality.

Exercise 7.2.

This is a standard exercise in topological dynamics, having nothing to do with entropy. In the surjective case the natural extension is conjugate to the bilateral subshift with the same *language* as the given unilateral subshift, i.e., with the same finite blocks occurring. In the non-surjective case there are blocks occurring in the unilateral shift which do not extend to the left within the language. We call these block “dead ends”. We append a new symbol (say $*$) to the alphabet and we enhance the language by blocks of the form $***\cdots*$ (of any finite length) and $***\cdots*B$, where B is a dead end. The natural extension is conjugate to the bilateral shift with this enhanced language. We skip the tedious but easy verification of the conjugacies.

Exercise 7.3.

Again, this is an exercise in topological dynamics and has little to do with entropy. We begin with the remark that the Marker Lemma 7.5.4 applies in fact to continuous maps, not necessarily homeomorphisms. If we replace the starting clopen cover \mathcal{U} by $\mathcal{U}' = T^{-nm}(\mathcal{U})$, then each $U \in \mathcal{U}'$ has clopen forward images through nm iterates and all the sets F_j constructed in the proof (including the marker set F) are clopen together with their n backward and forward images. We skip further details here (see [Downarowicz, 2008]).

In any zero-dimensional system without periodic points we can mimic the odometer factor. The only difference is that the analogs of the p_k -periodic markers will not appear periodically, yet with gaps ranging between p_k and some $p'_k > p_k$. Here is how we do. We fix a sequence (p_k) and the associated quotients $q_k \geq 2$ just as in Definition A.3.1. We find a p_1 -marker set F_1 . Since there are no periodic points in X , all orbits visit F_1 and the gaps between the visits range between p_1 and $2p_1 - 1$. In the induced system $(F_1, T_{F_1}, \mathbb{S})$ we find a q_1 -marker F_2 . Since the induced system has no periodic points, every T_{F_1} -orbit visits F_2 with gaps ranging between q_1 and $2q_1 - 1$, which implies that every T -orbit in X visits F_2 with gaps ranging between $p_1q_1 = p_2$ and $p'_2 = (2p_1 - 1)(2q_1 - 1) > p_2$. Proceeding inductively we construct a decreasing sequence of marker sets F_k . Abusing slightly our convention, we will call them k -markers (they are in fact p_k -markers). If we visualize the k -markers in the array-name representation of our system (in form of vertical bars in the k th row) then in every array x we see the $(k+1)$ -markers only at coordinates where k -markers occur, there are at least q_k k -markers between two consecutive $(k+1)$ -markers, while the distances between two consecutive $(k+1)$ -markers are bounded. The blocks appearing in row k between two neighboring k -markers will be called k -blocks. The (finite) collection of the k -blocks appearing in the system will be denoted by \mathcal{B}_k (this time the blocks in \mathcal{B}_k have various but bounded lengths). In injective systems the arrays are bilateral, so the k th row of every x is a concatenation of the k -blocks (there is no problem with truncated k -blocks at the left end).

We are ready to encode our system using a countable alphabet. The alphabet is going to be $\Lambda = \bigcup_{k=1}^{\infty} \mathcal{B}_k \cup \{*\}$, where $*$ is added as the topological accumulation point, so that Λ is homeomorphic to the one-point compactification of \mathbb{N}_0 .

We now define the map ϕ from X into the shift over Λ by describing the image $y = \phi(x) \in \Lambda^{\mathbb{Z}}$ of every $x \in X$. We will encode x “row after row”. We encode the

first row by placing in y , at the positions of all 1-markers in x , the symbols from \mathcal{B}_1 representing the 1-blocks that follow these markers in x . Since $p_1 \geq 2$, every sector in y between the positions of two consecutive 1-markers has at least one unfilled position. We now encode the second row of x by placing in y , in the first empty slot between two 2-markers of x , the symbol from \mathcal{B}_2 representing the block sitting there in the second row of x . Since $q_1 \geq 2$, after this step every sector in y between two 2-markers has at least one empty slot. We continue in this manner through all rows. All eventually unfilled positions in y we fill with the stars. (The situation resembles that on Figure 7.2, except that the cuts are not exactly at equal distances and that in every step the information is stored in only one symbol per “period”, so in the end there will be much more unfilled space in y .) It is clear that so defined map $x \mapsto y$ is continuous: every symbol (except the star) in y is determined by a bounded rectangle in x (i.e., its preimage is clopen). The star alone is not an open set, while any open neighborhood of the star is a complement of finitely many other symbols, so its preimage is also a clopen set. It is evident that so defined map ϕ commutes with the shift transformation. To see that it is injective, note that we can easily reconstruct from y the consecutive rows of x . For $k = 1$, we locate in y the symbols belonging to \mathcal{B}_1 . Their positions determine the 1-markers and the symbols themselves provide information about the contents of the corresponding 1-blocks in x . We continue inductively: Suppose the k th row of x has been reconstructed (together with the k -markers). We locate in y all symbols belonging to \mathcal{B}_{k+1} , and then we “unload” their contents each time starting at the nearest k -marker to the left, where we also place a $(k+1)$ -marker. So, the map ϕ is a topological conjugacy of X with its image.

To see that periodic points are an obstacle, take the identity map on the Cantor set. Every point is a fixpoint, so in any subshift it must be represented by a sequence filled with one symbol. Thus, uncountably many symbols are needed to encode all points.

To see how the above fails in non-injective systems, consider an odometer plus a Cantor set which is sent by T to one point (say x) in the odometer. No matter how we encode the system as a unilateral shift, the sequence representing x must admit uncountably many shift-preimages, that is one-coordinate prolongations to the left. So, an uncountable alphabet is needed.

Exercise 7.4.

Let $\mathbf{h} = \mathbf{h}(T)$. Let \mathcal{B}_n denote the family of all blocks of length n occurring in X . Let Λ be an alphabet of cardinality $\lfloor 2^{\mathbf{h}} \rfloor + 1$. It is important that $\log \#\Lambda > \mathbf{h}$, so we can invoke our Exercise 3.8 (and the remark following the solution): there exists a “better than prefix-free” family \mathcal{C} of blocks over Λ such that denoting by \mathcal{C}_n the family of blocks of length n contained in \mathcal{C} , we have $\#\mathcal{C}_n \geq 2^{n(\mathbf{h}+\varepsilon)}$ for some $\varepsilon > 0$ and n sufficiently large. On the other hand, we know that $\log \#\mathcal{B}_n < 2^{n(\mathbf{h}+\varepsilon)}$ for large n . So, we can find an n_0 such that for every $n \geq n_0$, $\#\mathcal{B}_n \leq \#\mathcal{C}_n$. Then there exists an injective length-preserving map $\Phi : \bigcup_{n \geq n_0} \mathcal{B}_n \rightarrow \bigcup_{n \geq n_0} \mathcal{C}_n$. Now we apply the Marker Lemma and find an n_0 -marker. The code ϕ from X into $\Lambda^{\mathbb{Z}}$ is constructed as follows: we cut every x at the markers into blocks of lengths at least n_0 (and bounded). Then we replace every such block B by $\Phi(B)$. Because Φ is length-preserving, this is a shift-invariant procedure, and by boundedness of the blocks, it is continuous. Since

the image of Φ is contained in a “better than prefix-free” family, the cutting places (i.e., the markers) can be reconstructed in every $\phi(x)$, and because Φ is injective, we can then reconstruct x completely. Thus, the code is a topological conjugacy of X with its image.

To see how the above fails in non-injective systems, consider a unilateral subshift X over a finite alphabet Λ and with entropy smaller than $\log \#\Lambda$. We enhance the subshift by adding points of the form ax , where $x \in X$ and a is a single symbol belonging to a strictly larger alphabet $\Lambda' \supset \Lambda$. The enhanced system is a subshift over the alphabet Λ' . Since all its points fall into X' after one iterate, the topological entropy of the enhanced subshift is the same as that on X . Nonetheless, in any unilateral subshift representation, each point from X has $\#\Lambda'$ shift-preimages, so an alphabet of at least such cardinality is needed.

Exercise 7.5.

In every bilateral subshift any element of a periodic orbit of period n has the form $\dots BBB\dots$, where B has length n . Moreover, at most n different blocks B produce elements of the same orbit. Because there are l^n blocks of length n , using an alphabet of cardinality l we can produce at most l^n and at least l^n/n different periodic orbits with period n . Let X be the union of l^n/n disjoint orbits of period n modeled as a subshift over l symbols. The entropy of such a primitive system is zero, so, if Exercise 7.4 worked, X should admit a representation over two symbols. But with two symbols we can model at most 2^n periodic orbits with period n . For l large enough, $l^n/n > 2^n$, a contradiction.

8 Exercises in Chapter 8

Exercise 8.1.

We have $E\mathcal{H} \equiv \infty$ and $\alpha_0 = \aleph_0$. The best way to see this is by examining the transfinite sequence. Notice that the k th tail θ_k equals 1 on the dense set $\{x_{k+1}, x_{k+2}, \dots\}$, so $\bar{\theta}_k \equiv 1$. This implies $u_1 \equiv 1$. Now adding u_1 to the tails only shifts the picture up by a unit, hence $u_2 \equiv 2$, and, inductively, $u_\alpha \equiv \alpha$, for natural α . This clearly implies $u_{\aleph_0} \equiv \infty$ and this is where the transfinite procedure stops for the first time.

Exercise 8.2 (cf. [Boyle–Downarowicz, 2004, Proposition 3.10]).

We proceed by induction on $\text{ord}(x)$. By Theorem 8.1.14, $u_{\mathcal{H}}(x) = 0$ at any isolated point, so the statement holds if $\text{ord}(x) = 0$. Suppose we have proved it for some $\text{ord}(x) = r \geq 0$. Let x be a point of order $r + 1$. Define u as $u_{\mathcal{H}}$ except at x where we set $u(x) = (r + 1)u_1(x)$. We complete the proof by showing that u is a repair function for \mathcal{H} . Since we have altered $u_{\mathcal{H}}$ only at x , it suffices to verify that the defects of $u + \theta_k$ converge to zero at x . Note that x is surrounded by points x' of order at most

r , at which the inductive hypothesis holds. For each k we have

$$\begin{aligned} (u + \theta_k)(x) &\leq \ddot{u}(x) + \ddot{\theta}_k(x) = \limsup_{x' \rightarrow x} u(x') - u(x) + \tilde{\theta}_k(x) - \theta_k(x) \leq \\ &\limsup_{x' \rightarrow x} ru_1(x') - (r+1)u_1(x) + \tilde{\theta}_k(x) - \theta_k(x) = \\ &r \left(\limsup_{x' \rightarrow x} u_1(x') - u_1(x) \right) - u_1(x) + \tilde{\theta}_k(x) - \theta_k(x). \end{aligned}$$

The first term equals r times the defect of u_1 , which is zero, because u_1 is upper semicontinuous. (All functions u_α in the transfinite sequence are upper semicontinuous – this is obvious from the definition, but perhaps not sufficiently emphasized in the book). Now we let k tend to infinity, and then $\tilde{\theta}_k(x)$ decreases to $u_1(x)$ and $\theta_k(x)$ decreases to zero, so the entire expression tends to zero, as required.

Exercise 8.3.

Each entry in the matrix, say $M_{n,r}$, representing $u_n(x)$ at points of order r can be verbalized as

$$M_{n,r} = \begin{cases} a_0 + a_1 + \cdots + a_{r-1}, & \text{for } r \leq n \\ \text{“the maximal sum of } n \text{ different terms indexed up to } r-1\text{”}, & \text{for } r \geq n, \end{cases}$$

where by “terms” we mean the numbers a_i . Notice that the maximal sum of n different terms from a set of nonnegative numbers dominates all shorter sums from this set, so the above second case description can be written as “the maximal sum of up to n different terms indexed up to $r-1$ ”. This phrasing includes the first case, because for $r \leq n$ the maximal such sum is clearly the sum of all terms indexed up to $r-1$. Thus,

$$M_{n,r} = \text{“the maximal sum of up to } n \text{ different terms indexed up to } r-1\text{”},$$

is the general form, including also $M_{0,r}$ for all r (any sum of 0 terms is 0). We will verify that $u_n(x) = M_{n, \text{ord}(x)}$ by induction on n .

For $n = 0$ the formula holds. Assume it holds for some $n \geq 0$. We need to evaluate u_{n+1} . Take a point x and denote $r = \text{ord}(x)$. For k sufficiently large $\theta_k(x) = 0$ and then $(u_n + \theta_k)(x) = u_n(x) = M_{n,r}$. Every neighborhood of x contains (in spite of x) only points x' of orders $r' \leq r-1$, moreover, for every such r' it contains infinitely many points of order r' . Thus, no matter how large k , the function $(u_n + \theta_k)$ assumes within this neighborhood the value $M_{n,r'} + a_{r'}$, i.e.,

$$\text{“the maximal sum of up to } n \text{ different terms indexed up to } r'-1\text{”} + a_{r'}$$

which is the same as

$$\text{“the maximal sum of up to } n+1 \text{ different terms indexed up to } r' \text{ including } a_{r'}\text{”}.$$

Alltogether, $u_{n+1}(x)$ equals the maximum over $r' \leq r-1$ of the above maximal sums and $M_{n,r}$. It is hence clear that $u_{n+1}(x)$ does not exceed “the maximal sum of up to $n+1$ different terms indexed up to $r-1$ ”, i.e., $M_{n+1,r}$.

But every “sum of up to $n + 1$ different terms indexed up to $r - 1$ ” has its maximal index, some $r' \leq r - 1$, and then this sum is a “sum of up to $n + 1$ different terms indexed up to r' including $a_{r'}$ ”, and is taken into account in the maximum defining $u_{n+1}(x)$. This implies that, $u_{n+1}(x) = M_{n+1,r}$.

Exercise 8.4.

We need to show that $u_{\mathcal{H}} = u_1$. Since always $u_1 \leq u_{\mathcal{H}}$, we focus on the converse inequality. Because $u_{\mathcal{H}} = E\mathcal{H} - \tilde{h}$ and we assume that $E\mathcal{H} = \tilde{h}$, the desired inequality becomes $u_1 \geq \tilde{h} - h$. We have

$$u_1 = \lim_k \widetilde{\theta}_k = \lim_k \widetilde{h - h_k} \geq \tilde{h} - \lim_k \widetilde{h_k}$$

(we have used the inequality $\widetilde{f + g} \leq \widetilde{f} + \widetilde{g}$ for $f = h - h_k$ and $g = h_k$). Since all functions h_k are assumed upper semicontinuous, the last limit equals $\lim_k h_k = h$.

Exercise 8.5.

Attention! The formulation of the exercise contains a misprint. It should say not about superenvelopes only about repair functions.

The Tarski-Knaster Theorem asserts that any order-preserving operator $P : L \rightarrow L$ defined on a complete lattice L has its smallest fixpoint. Recall that a *complete lattice* is a partially ordered set in which every subset has its *infimum* (greatest lower bound) and *supremum* (smallest upper bound). In our case, the lattice will be the collection of all nonnegative upper semicontinuous functions on the domain \mathfrak{X} , where we include the constant infinity function. The infimum of a subset A of L is simply the pointwise infimum $\inf\{h : h \in A\}$, while the supremum equals $\sup\{h : h \in A\}$ or the infinity function when the supremum is unbounded. Given an increasing sequence $\mathcal{H} = (h_k)$ of nonnegative functions on \mathfrak{X} , tending to a finite limit h (hence we also have the tails $\theta_k = h - h_k$) we define the operator $P : L \rightarrow L$ by

$$P(f) = \lim_k \downarrow \widetilde{(f + \theta_k)}.$$

It is immediate to see that the operator is well defined (the image functions belong to L) and preserves the order. So far, we have just recalled what was given in the formulation of the exercise. We need to verify that fixpoints of P are exactly the repair functions of the tails of \mathcal{H} . Then the smallest fixpoint will coincide with the smallest repair function $u_{\mathcal{H}}$. To this end we write

$$\begin{aligned} P(u) = u &\iff \lim_k \widetilde{u + \theta_k} = u \iff \lim_k (\widetilde{u + \theta_k} - u - \theta_k) = 0 \iff \\ &\lim_k \overset{\dots\dots\dots}{(u + \theta_k)} = 0 \iff u \text{ is a repair function.} \end{aligned}$$

Exercise 8.6.

This is completely elementary and will be skipped.

Exercise 8.7.

The derivation of the first statement is based on the implication $f \geq g \implies \tilde{f} \geq \tilde{g}$. The converse need not hold, for example if $\mathfrak{X} = [0, 1]$ with $\theta_k \equiv \frac{1}{k}$ and $\theta'_k = \frac{1}{k} \mathbb{1}_{\mathbb{Q}}$, where \mathbb{Q} denotes the set of rational numbers in $[0, 1]$.

Exercise 8.8.

On a compact domain, say, \mathcal{K} , our upper semicontinuous function f attains its maximum y_0 at some point x_0 . Since \mathcal{K} is convex, the Choquet Theorem asserts that there exists a probability distribution ξ supported by $\text{ex}\mathcal{K}$ with $\text{bar}(\xi) = x_0$. But f is also convex, so by Fact A.2.10 it is supharmonic, and thus

$$y_0 = f(x_0) = f(\text{bar}(\xi)) \leq \int f(x) d\xi.$$

Since $f(x) \leq y_0$ at all points, this inequality is only possible when it is an equality and $f = y_0$ ξ -almost everywhere. In particular, $f(x) = y_0$ at at least one point in $\text{ex}\mathcal{K}$.

Exercise 8.9.

Attention! We do not show that $u_\alpha^{\mathcal{H}}|_{\text{ex}\mathcal{K}} = u_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}}$, only that both determine $u_\alpha^{\mathcal{H}}$ via the same operations. We do not invoke Lemma 8.2.13 directly, only the same proving methods.

This is a hard exercise. We claim the following

$$u_\alpha^{\mathcal{H}} = \left(\left(u_\alpha^{\mathcal{H}}|_{\text{ex}\mathcal{K}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} = \left(\left(u_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]}.$$

The statement obviously holds for $\alpha = 0$. Suppose it holds for all $\beta < \alpha$. Recall that $v_\alpha^{\mathcal{H}}$ stands for $\sup_{\beta < \alpha} u_\beta^{\mathcal{H}}$. We now write a sequence of (in)equalities and then we will explain why each of them is true.

$$\begin{aligned} u_\alpha^{\mathcal{H}} &\stackrel{(1)}{\geq} \left(\left(u_\alpha^{\mathcal{H}}|_{\text{ex}\mathcal{K}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \stackrel{(2)}{\geq} \left(\left(u_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \stackrel{(3)}{=} \\ &\left(\left(\lim_{\kappa} \left(\widetilde{v_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} + \theta_\kappa|_{\text{ex}\mathcal{K}}} \right) \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \stackrel{(4)}{=} \left(\lim_{\kappa} \left(\left(\widetilde{v_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} + \theta_\kappa|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right) \right)^{[\mathcal{K}]} \stackrel{(5)}{=} \\ \lim_{\kappa} \left(\left(\left(\widetilde{v_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} + \theta_\kappa|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \right) &\stackrel{(6)}{\geq} \lim_{\kappa} \left[\left(\left(\widetilde{v_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} + \theta_\kappa|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \right] \stackrel{(7)}{\geq} \\ \lim_{\kappa} \left[\left(\left(\widetilde{v_\alpha^{\mathcal{H}|_{\text{ex}\mathcal{K}}} \right)^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} + \theta_\kappa \right] &\stackrel{(8)}{\geq} \lim_{\kappa} \left(\widetilde{v_\alpha^{\mathcal{H}}} + \theta_\kappa \right) \stackrel{(9)}{=} u_\alpha^{\mathcal{H}}. \end{aligned}$$

At first notice that since each h_κ is harmonic, it is affine, so h is affine (although not necessarily harmonic), hence each θ_κ is affine. This makes $\tilde{\theta}_\kappa$ concave (Fact A.2.5), and by an easy induction, all functions $u_\alpha^{\mathcal{H}}$ are concave. On the other hand, for any

distribution ξ we have $\int h d\xi \geq \int h_\kappa d\xi = h_\kappa(\text{bar}(\xi))$ for every κ , hence $\int h d\xi \geq h(\text{bar}(\xi))$ and h is shown to be subharmonic. Also each $\theta_\kappa = h - h_\kappa$ is subharmonic.

(1) is derived as follows: $(u_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}})^{\text{har}_{\mathcal{M}}}$ is upper semicontinuous, so

$$\left((u_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}})^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} = (u_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}})^{\text{har}_{\mathcal{M}}}(\xi) = \int u_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}} d\xi = \int u_\alpha^\mathcal{H} d\xi \leq u_\alpha^\mathcal{H}(x),$$

where ξ is some distribution on $\overline{\text{ex}\mathcal{K}}$ with barycenter at x , and the last inequality is a consequence of $u_\alpha^\mathcal{H}$ being concave and upper semicontinuous, hence supharmonic.

(2) results from the fact that throughout the transfinite induction leading to $u_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}}$ (at a point in $\overline{\text{ex}\mathcal{K}}$) the “tildes” are taken in the context of $\overline{\text{ex}\mathcal{K}}$, so they produce not larger functions than the “tildes” taken in the wider context of \mathcal{K} , leading to $u_\alpha^\mathcal{H}$ at this point.

(3) is just the transfinite definition applied to the restriction $\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}}$.

In (4) we pull the decreasing limit outside the harmonic extension. Since the harmonic extension relies on integrals, we need a kind of Lebesgue Theorem. The functions are bounded from some index on (otherwise the case is trivial, as we have infinity on the right), so if the net is actually a sequence, we can use the Lebesgue Dominated Theorem. For nets, however, we must invoke a stronger result: for a decreasing net of upper semicontinuous functions, the integral commutes with the limit (see e.g. (A7) in the Appendix of [Downarowicz-Serafin, 2002]).

In (5) we exchange the limit with the push-down. Since the fibers are compact, the functions are upper semicontinuous and decrease, this can be done by virtue of the exchanging suprema and infima Fact A.1.24.

In (6), for each κ we delay the application of “tilde” till after the harmonic extension and the push-down. (“Tilde” commutes with the harmonic extension because we are on a Bauer simplex, so it is not important in what order we interpret them applied.) For the push-down the inequality is obvious, because the function on the left is upper semicontinuous (see Fact A.1.26) and dominates the function without the “tilde” on the right.

When evaluating the push-down on the left hand side of (7) at some $x \in \mathcal{K}$ we must integrate the sum of two functions with respect to all measures supported by $\overline{\text{ex}\mathcal{K}}$ with barycenter at x . Since θ_κ is subharmonic, the integral of θ_κ with respect to such a measure will be always at least $\theta_\kappa(x)$. So the integral of the sum will always be at least the integral of the first function plus $\theta_\kappa(x)$. Now we can take the supremum over all such measures.

For (8) note that $\left((v_\alpha^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}})^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]} \geq \left((u_\beta^\mathcal{H}|_{\overline{\text{ex}\mathcal{K}}})^{\text{har}_{\mathcal{M}}} \right)^{[\mathcal{K}]}$ for every $\beta < \alpha$. By the inductive assumption, we replace the latter by $u_\beta^\mathcal{H}$, and then we apply supremum over all $\beta < \alpha$.

(9) is just the transfinite definition.

The claim about the order of accumulation is now obvious.

Exercise 8.10.

This is a direct consequence of two inequalities: $H(\mu, \mathcal{U} \vee \mathcal{V}) \leq H(\mu, \mathcal{U}) + H(\mu, \mathcal{V})$ and $H(\mu, T^{-n}(\mathcal{U})) \leq H(\mu, \mathcal{U})$. The first one holds since whenever $\mathcal{P} \succcurlyeq \mathcal{U}$ and $\mathcal{Q} \succcurlyeq \mathcal{V}$

then $\mathcal{P} \vee \mathcal{Q} \succcurlyeq \mathcal{U} \vee \mathcal{Q}$ and $H(\mu, \mathcal{P} \vee \mathcal{Q}) \leq H(\mu, \mathcal{P}) + H(\mu, \mathcal{Q})$. The second is true by invariance of μ and since whenever $\mathcal{P} \succcurlyeq \mathcal{U}$ then $T^{-n}(\mathcal{P}) \succcurlyeq T^{-n}(\mathcal{U})$.

Exercise 8.11.

For the first inequality note that any \mathcal{P} inscribed in \mathcal{U} has diameter at most $\text{diam}(\mathcal{U})$, for the second – that any \mathcal{P} of diameter smaller than $\text{Leb}(\mathcal{U})$ is inscribed in \mathcal{U} .

Exercise 8.12.

Just note that the cover \mathcal{U}^n is inscribed in the cover constituted by the $(n, \text{diam}(\mathcal{U}))$ -balls and that the cover by the $(n, \text{Leb}(\mathcal{U}))$ -balls is inscribed in \mathcal{U} . Then use the monotonicity (6.3.5).

Exercise 8.13.

Given a cover \mathcal{U} and a set A , the smallest cardinality of a subfamily of \mathcal{U}^n covering A is at least equal to the maximal cardinality of $(n, \text{diam}(\mathcal{U}))$ -separated set. This easily implies that $\mathbf{h}(T, \mathcal{U}|F, \mathcal{V}) \geq \mathbf{h}(T, \text{diam}(\mathcal{U})|F, \mathcal{V})$. On the other hand, any maximal $(n, \text{Leb}(\mathcal{U}))$ -separated set E in A is also $(n, \text{Leb}(\mathcal{U}))$ -spanning in A . Each element of E is contained, together with its $(n, \text{Leb}(\mathcal{U}))$ -ball, in an element of \mathcal{U}^n . In this manner we select a subfamily of \mathcal{U}^n which covers A and has at most the cardinality of E . This implies $\mathbf{h}(T, \mathcal{U}|F, \mathcal{V}) \leq \mathbf{h}(T, \text{Leb}(\mathcal{U})|F, \mathcal{V})$. Now we can apply the above to a refining sequence of covers \mathcal{U}_k (then both $\text{diam}(\mathcal{U}_k)$ and $\text{Leb}(\mathcal{U}_k)$ tend to zero).

Exercise 8.14.

Attention! Implicitly, \mathcal{V} is assumed finite. Otherwise I don't know how to proceed.

This is an extremely unpleasant exercise. The reason I put it is to illuminate how convenient it is to have entropy structure defined as a uniform equivalence class. We may afford not to care much about measurability of functions in one particular entropy structure because in the same class there are other sequences of functions known to be measurable (even upper semicontinuous). In [Downarowicz, 2005a], I simply used the upper integral to extend the function $h(T|\mu, \mathcal{V})$ to nonergodic measures.

The strategy is to assume that X is zero-dimensional, and (1) approximate \mathcal{V} by a sequence of *clopen covers* \mathcal{V}_k , i.e., having clopen (not necessarily disjoint) cells and then (2) prove the assertion for such clopen covers. In step (3) we will apply principal extensions to get rid of the zero-dimensionality assumption.

(1) We fix a finite open cover \mathcal{V} (of our zero-dimensional space X) and temporarily we also fix an ergodic measure μ . We will exploit the following variant of Lemma 8.3.20: If \mathcal{W} is another cover then

$$h(T|\mu, \mathcal{W}) \leq h(T|\mu, \mathcal{V}) + \liminf_{\sigma \rightarrow 1} \{h(T, \mathcal{V}|F, \mathcal{W}) : \mu(F) > \sigma\}.$$

The proof is exactly the same as that of Lemma 8.3.20 except that it uses the full version of (6.3.10) (i.e., we keep F in the last term).

For $k \in \mathbb{N}$ and $V \in \mathcal{V}$ there exists a clopen set V_k contained in V and containing $\{x : d(x, V^c) \geq 1/k\}$ (because the latter set is compact). We can easily arrange the

sets V_k to grow with k . It is easy to see than if $1/k < \text{Leb}(\mathcal{V})/2$ then the collection $\mathcal{V}_k = \{V_k : V \in \mathcal{V}\}$ is a (clopen) cover of X . Since $\mathcal{V}_k \supseteq \mathcal{V}_{k'}$ whenever $k' \geq k$, the sequence of functions $h(T|\mu, \mathcal{V}_k)$ increases. Moreover, since for each k the sets V_k grow to V , the measure of the set $A_k = \bigcup_{V \in \mathcal{V}} (V \setminus V_k)$ is smaller than ε for large enough k (this is where we need \mathcal{V} to be finite). By the Ergodic Theorem, given $\sigma > 0$ there exists $n_{k,\sigma} \in \mathbb{N}$ and a set $F_{k,\sigma}$ of measure larger than σ such that for all $n \geq n_{k,\sigma}$ all n -orbits starting in $F_{k,\sigma}$ visit A_k at most $n\varepsilon$ times. By the last displayed formula, we have

$$h(T|\mu, \mathcal{V}) \leq h(T|\mu, \mathcal{V}_k) + \lim_{\sigma \rightarrow 1} h(T, \mathcal{V}_k|F_{k,\sigma}, \mathcal{V}).$$

We need to estimate the last term. Let $x \in F_{k,\sigma}$. For $n \geq n_{k,\sigma}$ choose a cell V_x^n of \mathcal{V}^n containing x , say $V_x^n = \bigcup_{i=1}^{n-1} T^{-i}V^{(i)}$ (each $V^{(i)} \in \mathcal{V}$). Then $T^i(x) \in V_k^{(i)}$ except for at most $n\varepsilon$ indices i . This implies that x belongs to one of at most

$$L_n = \binom{n}{n\varepsilon} \#\mathcal{V}^{n\varepsilon}$$

modifications of $\bigcup_{i=1}^{n-1} T^{-i}V_k^{(i)}$ in which at most $n\varepsilon$ terms are altered (i.e., $V_k^{(i)}$ is replaced by another cell of \mathcal{V}_k). We have covered $F_{k,\sigma} \cap V_x^n$ by at most L_n elements of \mathcal{V}_k^n . Thus

$$h(T, \mathcal{V}_k|F_{k,\sigma}, \mathcal{V}) < \lim_n \frac{1}{n} \log L_n \leq H(\varepsilon, 1 - \varepsilon)\varepsilon \log \#\mathcal{V},$$

regardless of σ . We have proved that the function $h(T|\mu, \mathcal{V})$ on ergodic measures equals the increasing limit of the functions $h(T|\mu, \mathcal{V}_k)$, where \mathcal{V}_k are clopen covers.

(2) We will check measurability of the function $h(T|\mu, \mathcal{V})$ for a finite clopen cover $\mathcal{V} = \{V_1, V_2, \dots, V_l\}$ of a zero-dimensional space X . In this setup consider the joining (X', T', \mathbb{S}) of our system with the subshift over the alphabet $\Lambda = \{1, 2, \dots, l\}$, such that every point $x \in X$ is joined with all sequences $(a_n) \in \Lambda^{\mathbb{S}}$ such that $T^n(x) \in V_{a_n}$ (the joining associates to each x all its possible \mathcal{V} -names). Every ergodic measure μ can be lifted to a measure μ' on the joining by the rule, that all possible Λ -words assigned to a finite piece of an orbit have equal probabilities (given x , at each coordinate we choose the available symbols with equal probabilities and independently of the choices made on other coordinates). It is not hard to see that the measure μ' is ergodic and that the assignment $\mu \mapsto \mu'$ is continuous on ergodic measures (we skip the standard arguments via estimating the frequencies of blocks). Using ergodic decomposition, we extend this assignment to a continuous map from $\mathcal{M}_T(X)$ into $\mathcal{M}_{T'}(X')$. Let \mathcal{P} denote a clopen partition (hence a cover) of X , let \mathcal{V}' and \mathcal{P}' denote the lifts of \mathcal{V} and \mathcal{P} to X' , respectively. Additionally, on X' we have the zero-coordinate clopen partition (and cover) \mathcal{Q} corresponding to the symbols in Λ . Notice that the cells of \mathcal{V}' are precisely the fiber saturations of the cells of \mathcal{Q} . Choose a closed set $G \in X$ and denote by F its projection to X , and let G' be the lift of F (so that G' is the fiber-saturation of G ; recall that both F and G' are closed). Because the cells of \mathcal{P}'^n are fiber-saturated, it does not matter whether we cover the sets $G \cap B$ (where $B \in \mathcal{Q}^n$) or their fiber saturations. Thus

$$h(T', \mathcal{P}'|G, \mathcal{Q}) = h(T', \mathcal{P}'|G', \mathcal{V}') = h(T, \mathcal{P}|F, \mathcal{V}).$$

In the expression on the left, we are dealing with two partitions and we count the cells of \mathcal{P}'^n needed to cover a cell in \mathcal{Q}^n (intersected with F), so the count will be exactly the same as if we counted the cells of $(\mathcal{P}' \vee \mathcal{Q})^n$ instead of \mathcal{P}'^n . This leads to

$$h(T', \mathcal{P}' \vee \mathcal{Q} | G, \mathcal{Q}) = h(T, \mathcal{P} | F, \mathcal{V}).$$

Now we temporarily fix an ergodic measure μ (and μ') and we repeat the argument used in the proof of Lemma 8.3.21 (with the correction in the definition of the set $G = G_\varepsilon$; see *Errata*): We fix a sequence of clopen partitions \mathcal{P}_k refining in X , and we choose a set G of measure larger than σ of points satisfying, up to ε , the Shannon-McMillan-Breiman Theorem applied to the ergodic measure μ' and each partition $\mathcal{P}'_k \vee \mathcal{Q}$ and \mathcal{Q} (for each partition with perhaps different threshold length). Now we let $k \rightarrow \infty$, and then the right hand of the last displayed equality simply converges to $h(T | F, \mathcal{V})$, while the left side remains within the range $h(\mu', \mathcal{P}'_k | \mathcal{Q}) \pm \varepsilon$, ($h(\mu', \mathcal{P}'_k | \mathcal{Q}) = h(\mu', \mathcal{P}'_k \vee \mathcal{Q} | \mathcal{Q})$ is the usual measure-theoretic conditional entropy involving two partitions; this we do exactly as in the proof of Lemma 8.3.21). Since every subset of X of measure larger than σ contains sets F (images of G as described above) for arbitrarily small ε , the application of the infimum over such sets and then supremum over σ leads to

$$h(T, |\mu, \mathcal{V}) = \lim_k \uparrow h(\mu', \mathcal{P}_k | \mathcal{Q}).$$

This equality extends to all measures μ via integrating over the ergodic decomposition (the function on the right is harmonic, the one on the left is harmonic by definition). The function $\mu \mapsto h(\mu', \mathcal{P}_k | \mathcal{Q})$ is now a composition of the continuous map $\mu \mapsto \mu'$ with the conditional entropy function for two clopen covers, which, as we know very well, is upper semicontinuous. So, $h(T, |\mu, \mathcal{V})$ is of Young class LU.

(3) It remains to extend the result to general systems. We change the meaning of the notation: from now on (X, T, \mathbb{S}) will denote a general topological dynamical system, while (X', T', \mathbb{S}) will be its principal zero-dimensional extension (see Theorem 7.6.1). We have a continuous surjection $\pi : \mathcal{K}' \rightarrow \mathcal{K}$ between Choquet simplices $\mathcal{K}' = \mathcal{M}_{T'}(X')$ and $\mathcal{K} = \mathcal{M}_T(X)$. We fix a finite open cover \mathcal{V} of X and we let \mathcal{V}' denote its lift. As we have shown in the preceding step, the function $f(\mu') = h(T' | \mu', \mathcal{V}')$ is an increasing limit of some upper semicontinuous and affine (hence harmonic) functions, say $f_k(\mu')$. The pushed-down functions $f_k^{[\mathcal{K}]}$ maintain these two properties (see Fact A.2.22). It is an elementary observation, that the operation push-down preserves increasing limits (it is a matter of exchanging two suprema). Thus $f^{[\mathcal{K}]} = \lim_k \uparrow f_k^{[\mathcal{K}]}$. This monotone limit is obviously of Young class LU, and, by the Lebesgue Monotone Theorem, it is a harmonic function. Lemma 8.3.18 implies that $f^{[\mathcal{K}]}(\mu)$ coincides with $h(T | \mu, \mathcal{V})$ on ergodic measures, and, since both functions are harmonic, they coincide everywhere.

9 Exercises in Chapter 9

Exercise 9.1.

Example 9.3.6 is the one. Here $E\mathcal{H} = \tilde{h}$, hence $\mathbf{h}_{\text{sex}}(T) = \mathbf{h}(T)$ thus $\mathbf{h}_{\text{res}}(T) = 0$. On the other hand, the measure $\mu_0 = \sum_k 2^{-k} \mu_{(B_{k+1})}$ (playing the role of the point b in Example 8.2.17) has obviously entropy zero, while $E\mathcal{H}$ at this measure is 1. So, $h_{\text{res}}(\mu_0) = 1 > \mathbf{h}_{\text{res}}(T)$.

Exercise 9.2.

This is well known. There is only one invariant measure, the normalized Lebesgue measure. Any partition \mathcal{P} of the circle in two arcs I_1, I_2 (no matter how we attach the endpoints) has the small boundary property and generates (via the dynamics). So, this one partition suffices to build the zero-dimensional principal extension, which becomes a subshift (the closure of the \mathcal{P} -names). With small boundary property, the standard zero-dimensional extension is not only principal but even isomorphic.

Exercise 9.3.

Although the system looks even more trivial than the preceding one, this exercise is a bit more intricate. Since this system has no small boundary property, we must first lift it to a product with something minimal of entropy zero. Let X denote the product of $[0, 1]$ with the unit circle (also viewed as $[0, 1]$, but with endpoints glued together). On this space we apply the product dynamics of the identity times some fixed irrational rotation (by some s): $T(t, x) = (t, x + s)$. This system is a principal extension of $([0, 1], \text{id}, \mathbb{S})$ (for both cases of \mathbb{S}). Take the partition \mathcal{P} into two sets separated by a skew line crossing all vertical sections (for instance $y(t) = \frac{1}{4} + \frac{t}{2}$) and the horizontal line $y = 0$. Label the bottom set by 0 and the top set by 1. The ergodic measures on X are $\delta_t \times \lambda$, and it is obvious that the boundary of \mathcal{P} (the dividing lines) has measure zero for all such measures. So, \mathcal{P} has small boundary. Moreover, this partition generates (via the dynamics) in X . So, this one partition suffices to build the zero-dimensional principal extension, which becomes a subshift Y (the closure of the \mathcal{P} -names). This extension is isomorphic to the product system (for every invariant measure), but obviously not to the base system $([0, 1], \text{Id}, \mathbb{S})$.

As a matter of fact, this principal symbolic extension is a disjoint union of Sturmian subshifts (over the same rotation, but different arc partitions), and the factor map $\pi : Y \rightarrow [0, 1]$ associates to every y such t that $\frac{1}{4} + \frac{t}{2}$ is the density of zeros in y . We skip proving this.

Exercise 9.4.

We must copy the construction of Example 9.3.5 except that there must be two ergodic measures supported by the first row and with k growing to infinity, the measures supported by the k th row must approach the average of the two measures in the first row. So, we take two bilateral uniquely ergodic subshifts X_0 and X_1 (say, over disjoint alphabets), each of entropy 1, and we denote their measures by μ_0 and μ_1 , respectively. We choose blocks B_k appearing in X_0 with lengths increasing with k , so that $\mu_{(B_k)} \rightarrow \mu_0$ and we choose C_k analogously in X_1 (the length of C_k should be the same as that of B_k). Now we let X consist of all symbolic arrays obeying the following rules:

1. The first row x_1 of x either belongs to X_0 or to X_1 or it has the form $x_{(k)} = \dots B_k C_k B_k C_k B_k C_k \dots$.
2. If the first row is $x_{(k)}$, then the k th row of x is an element of X_0 .
3. All other rows are filled with zeros.

Now there are two ergodic measures of entropy 1 (plus some periodic measures) supported by the first row and for each k there are finitely many measures $\mu_{k,i}$ supported by matrices with nontrivial k th row. All these measures are isomorphic to μ_0 joined with a periodic orbit, so all of them have entropy 1. These measures accumulate at the measure $\frac{1}{2}(\mu_0 + \mu_1)$, because for large k short blocks in the 1st row occur half of the time with the frequency as in B_k , and half of the time with the frequency as in C_k , while other rows are filled with zeros, except one very distant row, which we can ignore for the weak-star distance. The structure of Example 8.2.18 is now copied.

Exercise 9.5.

Of course, we could build a system whose simplex of invariant measures is a Bauer simplex spanned by the unit interval and the entropy structure restricted to ergodic measures copies the sequence (h_k) in Exercise 8.1 (the pick-up stick game on a dense sequence). Instead, we will describe the example proposed by Mike Boyle in the early 90's, before entropy structures were introduced, and the lack of symbolic extensions was proved using topological methods. This example triggered the development of the theory of symbolic extensions. Below it is adapted to the language of symbolic arrays.

Let X consist of 0-1 symbolic arrays obeying the following rules:

1. The 1st row x_1 of x is arbitrary. If x_1 is not periodic then all other rows are filled with zeros.
2. If x_1 is periodic with minimal period k_1 then we allow x_{1+k_1} to be arbitrary. If x_{1+k_1} is not periodic then all other rows are filled with zeros.
3. If x_{1+k_1} is periodic with minimal period k_2 then we allow $x_{1+k_1+k_2}$ to be arbitrary. If $x_{1+k_1+k_2}$ is not periodic then all other rows are filled with zeros.
4. and so on...

Every ergodic measure is supported by arrays with only one aperiodic row, hence its entropy is at most $\log 2$ and so is the topological entropy of the system.

Suppose Y is a symbolic extension via a factor map $\pi : Y \rightarrow X$. Then, for each k the composition π_k of π with the projection onto the subshift X_k visible in the k th row, as a factor map between two subshifts, is a sliding block code of some finite horizon r_k . Choose integers p_1 and $p_{i+1} = p_i q_i$ ($q_i \in \mathbb{N}$), and define $k_0 = 1$ and $k_i = 1 + p_1 + \dots + p_i$. By letting the numbers p_k grow fast enough we can easily arrange that $r_{k_i}/p_{i+1} \leq 1/3$.

Let us focus only the rectangular blocks R_j of length p_{j+1} extending over k_j rows and having, for each $i \leq j$, in row k_i , periodic repetitions of some block B_i of length p_{i+1} (and zeros in all other rows). Note that all such rectangles are admitted in our system

X. Every R_j has the following structure: in rows 1 through k_{i-1} it has q_j repetitions of one and the same rectangle R_{j-1} (of length p_j) and in row k_j it has a completely arbitrary block B_j . We will write $R_j = [R_{j-1}^{q_j}, B_j]$. For each rectangle R_j we let $\mathcal{C}(R_j)$ be the family of blocks of length p_{j+1} appearing in Y “above” R_j (i.e., in the preimage by π of the cylinder associated to R_j , at the same horizontal coordinates as R_j). We let L_j denote the minimal cardinality of $\mathcal{C}(R_j)$. Also, we let

$$\mathcal{C}(R_{j-1}^{q_j}) = \bigcup_{B_j} \mathcal{C}([R_{j-1}^{q_j}, B_j])$$

(the family of blocks admitted “above” $R_{j-1}^{q_j}$). Notice that if two rectangles R_j differ in the “central parts” (denoted B'_j) of B_j , of length $p_{j+1} - 2r_{k_j} \geq p_{j+1}/3$, then their families $\mathcal{C}(R_j)$ are disjoint (because the blocks B'_j are completely determined via the block code by the considered blocks in Y). Since there are at least $2^{p_{j+1}/3}$ different blocks B'_j , we obtain that, for any R_{j-1} ,

$$\#\mathcal{C}(R_{j-1}^{q_j}) \geq 2^{\frac{p_{j+1}}{3}} L_j.$$

On the other hand, given R_{j-1} , each block in the family $\mathcal{C}(R_{j-1}^{q_j})$ must be concatenated exclusively from blocks belonging to one family $\mathcal{C}(R_{j-1})$. So,

$$\#\mathcal{C}(R_{j-1}^{q_j}) \leq (\#\mathcal{C}(R_{j-1}))^{q_j},$$

which, combined with the preceding inequality (and the equality $p_{j+1}/q_j = p_j$) yields

$$\#\mathcal{C}(R_{j-1}) \geq 2^{\frac{p_j}{3}} L_j^{\frac{1}{q_j}}.$$

Because this holds for any R_{j-1} , we have obtained the inductive dependence

$$L_{j-1} \geq 2^{\frac{p_j}{3}} L_j^{\frac{1}{q_j}}.$$

Let us iterate this inequality two times:

$$L_{j-2} \geq 2^{\frac{p_{j-1}}{3}} L_{j-1}^{\frac{1}{q_{j-1}}} \geq 2^{\frac{p_{j-1}}{3}} 2^{\frac{p_{j-1}}{3}} L_j^{\frac{1}{q_j q_{j-1}}} = 2^{2\frac{p_{j-1}}{3}} L_j^{\frac{p_{j-1}}{p_{j+1}}}.$$

Iterating j times we get

$$L_0 \geq 2^j \frac{p_1}{3} L_j^{\frac{p_1}{p_{j+1}}} \geq 2^j \frac{p_1}{3}$$

(recall that the index 0 refers to the family of one-row rectangles R_0 in row $k_0 = 1$ of length p_1 , and hence L_0 is at most the cardinality of all blocks of length p_1 in Y). Because the above holds for every j , we have proved that the cardinality of all blocks of length p_1 in Y is unbounded. A contradiction.

Remark. In this example the simplex of ergodic measures and the entropy structure do not resemble those of Exercise 8.1. The picture is more like the one in Exercise 8.3, but with infinite order of accumulation (however, one has to take a quotient space to see such a structure). The measures of positive entropy are supported by arrays with finitely

many (say $j - 1$) periodic rows and one nonperiodic row. These measures (after suitable identification) correspond to the points with topological order of accumulation j . When the index of the last aperiodic row grows, the corresponding measures accumulate at measures with one row less (this resembles the situation in Example 9.3.5). But we also have “backward” accumulation points (when the order of accumulation grows). This corresponds to letting the number of nonzero rows grow. These accumulation points are measures supported by some odometers.

Part 3

11 Exercises in Chapter 11

Exercise 11.1.

Recall that $H_\mu(\mathcal{F}|\mathcal{G})$ is defined in (11.2.1) as $H_\mu(\mathcal{F} \sqcup \mathcal{G}) - H_\mu(\mathcal{G})$. Thus (11.2.2) is implied by $H_\mu(\mathcal{F} \sqcup \mathcal{G}) \geq H_\mu(\mathcal{F})$ (in presence of the trivial family \mathcal{O} , these two are in fact equivalent). From now on, this exercise appends Exercise 1.3 by reversing one of its implications (recall, we assumed $H(a \vee b) \geq H(a)$ and $H(a|b) \geq H(a|b \vee c)$, i.e., (11.2.2) and (11.2.3) and we were to derive $H(a \vee b|c) \leq H(a|b) + H(b|c)$ i.e., (11.2.10)).

We proceed as follows:

$$\begin{aligned} H_\mu(\mathcal{F}|\mathcal{G} \sqcup \mathcal{H}) &= H_\mu(\mathcal{F} \sqcup \mathcal{G} \sqcup \mathcal{H}) - H_\mu(\mathcal{G} \sqcup \mathcal{H}) = \\ H_\mu(\mathcal{F} \sqcup \mathcal{G} \sqcup \mathcal{H}) - H_\mu(\mathcal{G} \sqcup \mathcal{H}) - H_\mu(\mathcal{G}) + H_\mu(\mathcal{G}) &= H_\mu(\mathcal{F} \sqcup \mathcal{H}|\mathcal{G}) - H_\mu(\mathcal{H}|\mathcal{G}) \leq \\ H_\mu(\mathcal{F}|\mathcal{G}) + H_\mu(\mathcal{H}|\mathcal{G}) - H_\mu(\mathcal{H}|\mathcal{G}) &= H_\mu(\mathcal{F}|\mathcal{G}). \end{aligned}$$

Exercise 11.2.

We have

$$\mathbf{T}^l(\mathcal{F}^n) = (\mathbf{T}^l\mathcal{F})^n = \bigsqcup_{i=l}^{l+n-1} \mathbf{T}^i\mathcal{F}$$

thus

$$\mathcal{F}^{l+n} = \mathcal{F}^l \sqcup (\mathbf{T}^l\mathcal{F})^n$$

Using (11.2.2) with $\mathcal{H} = \mathcal{O}$, and (11.2.11), we obtain

$$H_\mu((\mathbf{T}^l\mathcal{F})^n) \leq H_\mu(\mathcal{F}^{l+n}) \leq H_\mu(\mathcal{F}^l) + H_\mu((\mathbf{T}^l\mathcal{F})^n).$$

We now divide both sides by $l+n$ and take \limsup as n tends to infinity. The middle expression becomes $h_\mu(\mathbf{T}, \mathcal{F})$ while both the left and right hand sides become $h_\mu(\mathbf{T}, \mathbf{T}^l\mathcal{F})$.

Exercise 11.3.

Notice that $(\mathcal{F}^n)^m$, where the exponent m refers to the action of \mathbf{T}^n equals \mathcal{F}^{nm} (in the notation referring to \mathbf{T}). Any natural k equals $nm - i$, where $0 \leq i \leq n - 1$ and then, as in the preceding exercise,

$$\mathcal{F}^{nm} = \mathcal{F}^i \sqcup (\mathbf{T}^i\mathcal{F})^k$$

and

$$H_\mu((\mathbf{T}^i\mathcal{F})^k) \leq H_\mu(\mathcal{F}^{nm}) \leq H_\mu(\mathcal{F}^i) + H_\mu((\mathbf{T}^i\mathcal{F})^k).$$

Now we divide both sides by k and apply \limsup as k tends to infinity. The extreme terms become $h_\mu(\mathbf{T}, \mathbf{T}^i\mathcal{F})$, which, by the preceding exercise, equals $h_\mu(\mathbf{T}, \mathcal{F})$. For

the middle term we note that in the limit $1/k$ can be replaced by $1/nm$ (where m tends to infinity) and we obtain $\frac{1}{n}h_\mu(\mathbf{T}^n, \mathcal{F}^n)$. This proves the first equality in Fact 11.2.6.

The second equality follows in a standard way from two facts: the supremum over all \mathcal{F} applied to $h_\mu(\mathbf{T}^n, \mathcal{F}^n)$ gives not more than $h_\mu(\mathbf{T}^n)$ because it takes into account only families of the form \mathcal{F}^n . On the other hand it gives not less, because $\mathcal{F} \subset \mathcal{F}^n$ and thus $h_\mu(\mathbf{T}^n, \mathcal{F}) \leq h_\mu(\mathbf{T}^n, \mathcal{F}^n)$.

Exercise 11.4.

This and the next exercises are general facts concerning doubly stochastic operators and have nothing to do with entropy.

Suppose \mathbf{T} be a doubly stochastic operator operator. As we know (see (11.2.21)),

$$\mathbf{T}(f \vee g) \geq \mathbf{T}f \vee \mathbf{T}g.$$

If \mathbf{T} is invertible, \mathbf{T}^{-1} is easily seen to be a doubly stochastic operator as well, so, applying \mathbf{T}^{-1} to both sides, and applying the above inequality for \mathbf{T}^{-1} , we obtain

$$f \vee g \geq \mathbf{T}^{-1}(\mathbf{T}f \vee \mathbf{T}g) \geq \mathbf{T}^{-1}\mathbf{T}f \vee \mathbf{T}^{-1}\mathbf{T}g = f \vee g.$$

Thus, the first above inequality is an equality, and, applying to it \mathbf{T} , we get

$$\mathbf{T}(f \vee g) = \mathbf{T}f \vee \mathbf{T}g.$$

By a symmetric argument, \mathbf{T} preserves the operation \wedge . In this manner this exercise has been reduced to a particular case of the next one.

Exercise 11.5.

It is obvious that pointwise generated operators preserve lattice operations. Notice that characteristic functions (i.e., assuming only the values 0 and 1) are precisely these functions f for which

$$f = 2f \wedge 1.$$

Thus a doubly stochastic operator which preserves the lattice operations (and it always preserves constants), sends characteristic functions to characteristic functions. It remains to show that then it is pointwise generated. Recall, that \mathbf{T} is a doubly stochastic operator on $L^1(\mu)$, where (X, \mathfrak{A}, μ) is a standard probability space.

If \mathbf{T} sends characteristic functions to characteristic functions, then it induces a map, say \mathbb{T} , from \mathfrak{A} into itself. Since \mathbf{T} preserves integrals with respect to μ , \mathbb{T} preserves the measure μ . By linearity and preservation of the constant $\mathbb{1}$, \mathbb{T} preserves set operations, and by the Lebesgue Dominated Theorem, also countable unions (up to measure μ). So, \mathbb{T} is a homomorphism from \mathfrak{A} to some sub-sigma-algebra $\mathfrak{B} = \mathbb{T}(\mathfrak{A}) \subset \mathfrak{A}$. It is well known that in standard spaces any such homomorphism is associated with a measure-preserving map $T : X \rightarrow X$ by the formula $\mathbb{T}(A) = T^{-1}(A)$. (Roughly, this map is defined (almost everywhere) as follows: a single point $\{x\}$ is almost surely sent by \mathbb{T} to an atom of \mathfrak{B} . This atom creates the preimage $T^{-1}(x)$.)

Perhaps we have a good opportunity to clarify an issue concerning doubly stochastic operators in general. Something that wasn't clearly said in the book. We have mentioned that a transition probability always determines a stochastic operator, and that not all stochastic operators are such. What we have not said, is this

Fact: every doubly stochastic operator is in fact determined by a transition probability.

Proof. First, each doubly stochastic operator T determines a shift-invariant measure μ on the countable product space $X^{\mathbb{N}_0}$ (interpreted as the space of trajectories), by the following formula (it suffices to define the measure on cylinders $A_0 \times A_1 \times \cdots \times A_n$):

$$\mu(A_0 \times A_1 \times \cdots \times A_n) = \int \mathbb{I}_{A_0} T(\cdots T(\mathbb{I}_{A_{n-2}} T(\mathbb{I}_{A_{n-1}} T \mathbb{I}_{A_n})) \cdots) d\mu.$$

In order to produce the transition probability $P(x, \cdot)$ it now suffices to take the disintegration measure μ_x of μ with respect to the sigma-algebra on the coordinate 0 at x and apply it to the sets depending on the coordinate 1. We skip the tedious but straightforward verification that μ is indeed a shift-invariant probability measure on the product sigma-algebra, and that the stochastic operator associated with so defined transition probability preserves μ and coincides on $L^1(\mu)$ with T . \square

The above fact opens yet another way to prove that doubly stochastic operators sending characteristic functions to characteristic functions are pointwise generated. We need to show that the transition probabilities $P(x, \cdot)$ are almost surely point-masses δ_y and then the associated map will be $x \mapsto y$.

By assumption, for every measurable set A , the function $(T \mathbb{I}_A)(x) = \int \mathbb{I}_A P(x, dy)$ takes on almost surely only the values 0 and 1. At almost every point this is true for a countable family of sets A that generates the sigma-algebra. This already implies that $P(x, \cdot)$ is concentrated at one point (otherwise at least one set from the generating family would have an intermediate measure value).

Exercise 11.6.

We have

$$\begin{aligned} \Theta_{m,t,s}(f)(x) = 1 &\iff m((f(x) - t) \wedge (s - f(x))) \geq 1 \\ &\iff f(x) \in [t + \frac{1}{m}, s - \frac{1}{m}] \\ \Theta_{m,t,s}(f)(x) = 0 &\iff (f(x) - t) \wedge (s - f(x)) \leq 0 \iff f(x) \notin (t, s). \end{aligned}$$

Thus $\mathbb{I}_{\{t + \frac{1}{m} \leq f \leq t - \frac{1}{m}\}} \leq \Theta_{m,t,s}(f) \leq \mathbb{I}_{\{t < f < s\}}$ (we have already proved (11.2.26)) and the extreme functions disagree only when $f \in (t, t + \frac{1}{m}) \cup (s - \frac{1}{m}, s)$. Thus the $L^1(\mu)$ distance between $\Theta_{m,t,s}(f)$ and $\mathbb{I}_{\{t < f < s\}}$ does not exceed the measure μ of the set of points for which f falls into this intervals. Including the internal endpoints leads to a not smaller value, hence the inequality (11.2.27) holds.

Exercise 11.7.

It is easy to see that $T^n(f)(x) = \frac{1}{2^n} f(\sigma^n x) + \frac{2^n - 1}{2^n} \int f d\mu$ for every $f \in L^1(\mu)$. Thus $T^n(f)$ converges to the constant $\int f d\mu$, implying that the entropy $h_\mu(T)$ is zero. On the other hand, if $f_0(x) = x_0$ then $T^n(f_0)(x) = \frac{1}{2^n} f(x_n) + \frac{2^n - 1}{2^n} \frac{1}{2}$ and $(T^n(f_0))^{-1}(\mathcal{Z})$ equals the n th coordinate partition. Thus, the partition generated jointly by $(T^i(f_0))^{-1}(\mathcal{Z})$ for $i = 0, 2, \dots, n - 1$ equals the partition into the blocks of length n , which has static entropy $n \log 2$.

12 Exercises in Chapter 12

Exercise 12.1.

(i) is clear using Definition 12.1.2, since a family \mathcal{F} of continuous functions on the factor lifts to a family \mathcal{F}' of continuous functions on X , and for each cover \mathcal{V} the preimage $\mathcal{F}^{-1}(\mathcal{V})$ lifts to $\mathcal{F}'^{-1}(\mathcal{V})$. Recall that the cardinality of a minimal subcover is preserved under preimage of a continuous surjection.

(ii) is now obvious, as conjugate systems are factors of each-other.

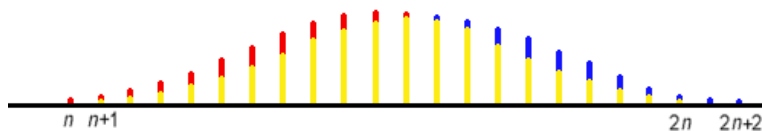
(iii) is best seen using Definition 12.1.3. Each family \mathcal{F} of continuous functions on Y prolongs to a family \mathcal{F}' on X and then every $(d_{\mathcal{F}}, \varepsilon)$ -separated set in Y remains $(d_{\mathcal{F}'}, \varepsilon)$ -separated in X . Also note that, by invariance of Y , for each n we can use $T^n(\mathcal{F}')$ as a prolongation of $T^n(\mathcal{F})$.

(iv) The proof is analogous to that in Exercise 11.3 (without subadditivity we must cope with \limsup , hence the simple way as in Fact 6.2.3 cannot be applied). We only outline the steps. Let us use Definition 12.1.1 for a change. We begin by proving an analog of Exercise 11.2: $\mathbf{h}_1(T, T^l(\mathcal{F}), \varepsilon) = \mathbf{h}_1(T, \mathcal{F}, \varepsilon)$. This is done the same way as that exercise with \sqcup replaced by the ordinary union of families and $H_\mu(\mathcal{F})$ replaced by $\mathbf{H}_1(\mathcal{F}, \varepsilon)$. Monotonicity (the analogue of (11.2.2)) and subadditivity (the analogue of (11.2.11)) are now obvious properties of joining the covers $\mathcal{U}_{\mathcal{F}}^\varepsilon$.

Next we follow Exercise 11.3. with the same substitutions.

Exercise 12.2.

We have $T^n f = \frac{1}{2^n} \left(\sum_{k=0}^n \binom{n}{k} f \circ \sigma^{n+k} \right)$, which is the convex combination of the functions $f \circ \sigma^j$ with coefficients as in the binomial $(1/2, 1/2)$ -distribution on $[n, 2n]$. The difference $T^n f - T^{n+1} f$ is thus a combination of the same functions with coefficients as in the difference of binomial distributions on $[n, 2n]$ and on $[n + 1, 2n + 2]$. Skipping the precise calculations, we agree that this difference is a signed distribution



on $[n, 2n + 2]$ whose absolute value has small total mass, if n is large (see the figure: the

red “tops” represent the positive atoms of the difference distribution, the blue “tops” are negative). Thus, since f is bounded, the differences $\mathbf{T}^n f - \mathbf{T}^{n+1} f$ converge to zero uniformly as $n \rightarrow \infty$. To complete this exercise, we will prove a more general fact.

Fact: If \mathcal{F} consist of functions f such that the differences $\mathbf{T}^n f - \mathbf{T}^{n+1} f$ converge to zero uniformly as $n \rightarrow \infty$, then $\mathbf{h}_2(\mathbf{T}, \mathcal{F}, \mathcal{V}) = 0$ for any cover \mathcal{V} of the interval.

Proof. Fix some $\delta > 0$. For each $V \in \mathcal{V}$ define $V_\delta = \{t : d(t, V^c) < \delta\} \subset V$ and let $\mathcal{V}_\delta = \{V_\delta : V \in \mathcal{V}\}$. Notice that if $\delta < \text{Leb}(\mathcal{V})/2$ then \mathcal{V}_δ still covers the entire interval and it is inscribed in \mathcal{V} . Thus, for every $f : X \rightarrow [0, 1]$, $f^{-1}(\mathcal{V}_\delta) \supseteq f^{-1}(\mathcal{V})$. Moreover, if $\|g - f\| < \delta$ then $g^{-1}(\mathcal{V}_\delta) \supseteq f^{-1}(\mathcal{V})$. So, if the assumption on \mathcal{F} is satisfied, then for each k , we have

$$(\mathbf{T}^n(\mathcal{F}))^{-1}(\mathcal{V}_\delta) \supseteq \bigvee_{i=0}^{k-1} (\mathbf{T}^{n+i}(\mathcal{F}))^{-1}(\mathcal{V}),$$

if n is larger than some n_k . Now, in the expression $\bigvee_{i=0}^{m-1} (\mathbf{T}^i(\mathcal{F}))^{-1}(\mathcal{V})$ defining $(\mathcal{F}^m)^{-1}(\mathcal{V})$, we can group the terms as follows (assuming $m = n_k + rk + s$, where $s \leq k$):

$$\begin{aligned} (\mathcal{F}^m)^{-1}(\mathcal{V}) = & \bigvee_{i=0}^{n_k-1} (\mathbf{T}^i(\mathcal{F}))^{-1}(\mathcal{V}) \vee \bigvee_{j=0}^{r-1} \left(\bigvee_{i=0}^{k-1} (\mathbf{T}^{n_k+jk+i}(\mathcal{F}))^{-1}(\mathcal{V}) \right) \vee \bigvee_{i=0}^{s-1} (\mathbf{T}^{n_k+rk+i}(\mathcal{F}))^{-1}(\mathcal{V}) \preceq \\ & \bigvee_{i=0}^{n_k-1} (\mathbf{T}^i(\mathcal{F}))^{-1}(\mathcal{V}) \vee \bigvee_{j=0}^r (\mathbf{T}^{n_k+jk}(\mathcal{F}))^{-1}(\mathcal{V}_\delta) \end{aligned}$$

The number of covers involved above is $n_k + r$. Since each of them has at most $\#\mathcal{F}\#\mathcal{V}$ elements, the static entropy $\log N((\mathcal{F}^m)^{-1}(\mathcal{V}))$ is at most

$$(n_k + r)(\log \#\mathcal{F} + \log(\#\mathcal{V})).$$

Dividing by m , letting m grow to infinity and remembering that $r < m/k$ while n_k does not grow with m , we obtain

$$\mathbf{h}_2(\mathbf{T}, \mathcal{F}, \mathcal{V}) \leq \frac{1}{k}(\log \#\mathcal{F} + \log(\#\mathcal{V})).$$

Since k in this argument is arbitrary, we conclude that this entropy is zero. \square

Now, in our exercise, the above holds for every \mathcal{F} hence the topological entropy $\mathbf{h}_2(\mathbf{T})$ is zero. The last question is answered using “half” of the variational principle (Theorem 12.3.1): for every invariant measure of \mathbf{T} , the measure-theoretic entropy of the corresponding doubly stochastic operator is zero, as well.

Exercise 12.3.

For $p \in [0, 1]$ let \mathbf{p}_p denote the probability measure on $\{0, 1\}$ assigning p to $\{0\}$ and $1-p$ to $\{1\}$. For $y = (y_n) \in [0, 1]^{\mathbb{N}_0}$ let $\mu^{(y)} = \mathbf{p}_{y_1} \times \mathbf{p}_{y_2} \times \dots$. It is easy to see that the map $y \mapsto \mu^{(y)}$ is a homeomorphism between $[0, 1]^{\mathbb{N}_0}$ and its image, which is a subset of $\mathcal{M}(\{0, 1\}^{\mathbb{N}_0})$. It is also immediately seen how the operator \mathbf{T}^* dual to the operator \mathbf{T} induced on $C(\{0, 1\}^{\mathbb{N}_0})$ by the shift map acts on this image: $\mathbf{T}^* \mu^{(y)} = \mu^{(\sigma y)}$. We have shown, that the “hipersystem” $(\mathcal{M}(\{0, 1\}^{\mathbb{N}_0}), \mathbf{T}^*)$ of the full unilateral shift on two symbols contains a subsystem conjugate to the full shift on $[0, 1]^{\mathbb{N}_0}$. Since the latter obviously has infinite topological entropy, so does the hipersystem, although the full shift on two symbols has finite entropy.

Exercise 12.4.

The proof should roughly follow the standard way, however, there might be some technical issues. I decided to leave this exercise open.