9

Entropy

The concept of entropy was first introduced by Clausius and later used in a different form by L. Boltzmann in his pioneering work on the kinetic theory of gases published in 1866. Since then, entropy has played a pivotal role in the development of many areas in physics and chemistry and has had important ramifications in ergodic theory. However, the Boltzmann entropy is different from the Kolmogorov–Sinai–Ornstein entropy [Walters, 1975; Parry 1981] that has been so successfully used in solving the problem of isomorphism of dynamical systems, and which is related to the work of Shannon [see Shannon and Weaver, 1949].

In this short chapter we consider the Boltzmann entropy of sequences of densities $\{P^n f\}$ and give conditions under which the entropy may be constant or increase to a maximum. We then consider the inverse problem of determining the behavior of $\{P^n f\}$ from the behavior of the entropy.

9.1 Basic definitions

If $(X, \mathcal{A}, \mu)$ is an arbitrary measure space and $P: L^1 \to L^1$ a Markov operator, then under certain circumstances valuable information concerning the behavior of $\{P^n f\}$ (or, in the continuous time case, $\{P^t f\}$) can be obtained from the behavior of the sequence

$$H(P^n f) = \int_X \eta(P^n f(x))\mu(dx), \tag{9.1.1}$$

where $\eta(u)$ is some function appropriately defined for $u \geqslant 0$.

The classical work of Boltzmann on the statistical properties of dilute gases suggested that the function η should be of the form

$$\eta(u) = -u \log u, \qquad \eta(0) = 0, \qquad (\log u = \log_e u) \tag{9.1.2}$$

and gives us our definition of entropy.

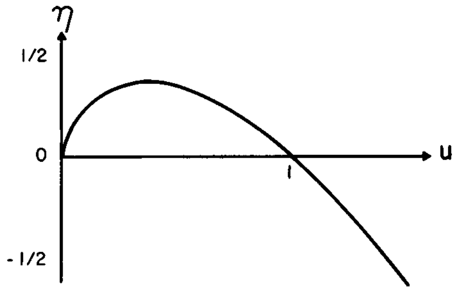

Figure 9.1.1. Plot of function $\eta(u) = -u \log u$.

Definition 9.1.1. If $f \geqslant 0$ and $\eta(f) \in L^1$ then the **entropy** of f is defined by

$$H(f) = \int_X \eta(f(x))\mu(dx). \tag{9.1.3}$$

Remark 9.1.1. If $\mu(X) < \infty$, then the integral (9.1.3) is always well defined for every $f \geqslant 0$. In fact, the integral over the positive parts of $\eta(f(x))$,

$$[\eta(f(x))]^+ = \max[0, \eta(f(x))],$$

is always finite. Thus $H(f)$ is either finite or equal to $-\infty$. □

Since we take $\eta(0) = 0$, the function $\eta(u)$ is continuous for all $u \geqslant 0$. The graph of η is shown in Figure 9.1.1. One of the most important properties of η is that it is convex. To see this, note that

$$\eta''(u) = -1/u$$

so $\eta''(u) < 0$ for all $u > 0$. From this it follows immediately that the graph of η always lies below the tangent line, or

$$\eta(u) \leqslant (u - v)\eta'(v) + \eta(v) \tag{9.1.4}$$

for every $u, v > 0$. Combining (9.1.4) with the definition of η given in equation (9.1.2) leads to the **Gibbs inequality**

$$u - u \log u \leqslant v - u \log v \qquad \text{for } u, v > 0, \tag{9.1.5}$$

which we shall have occasion to use frequently.

If f and g are two densities such that $\eta(f(x))$ and $f(x) \log g(x)$ are integrable, then from (9.1.5) we have the useful integral inequality

$$-\int_X f(x) \log f(x)\mu(dx) \leqslant -\int_X f(x) \log g(x)\mu(dx), \tag{9.1.6}$$

and the equality holds only for $f = g$. Inequality (9.1.6) is often of help in proving some extremal properties of $H(f)$ as shown in the following.

Proposition 9.1.1. Let $\mu(X) < \infty$, and consider all the possible densities f defined on X. Then, in the family of all such densities, the maximal entropy occurs for the constant density

$$f_0(x) = 1/\mu(X), \tag{9.1.7}$$

and for any other f the entropy is strictly smaller.

Proof: Pick an arbitrary $f \in D$ so that the entropy of f is given by

$$H(f) = -\int_X f(x) \log f(x)\mu(dx)$$

and, by inequality (9.1.6),

$$\begin{aligned} H(f) &\leqslant -\int_X f(x) \log f_0(x)\mu(dx) \\ &= -\log\left[\frac{1}{\mu(X)}\right]\int_X f(x)\mu(dx) \end{aligned}$$

or

$$H(f) \leqslant -\log\left[\frac{1}{\mu(X)}\right].$$

and the equality is satisfied only for $f = f_0$. However, the entropy of f_0 is simply

$$H(f_0) = -\int_X \frac{1}{\mu(X)} \log\left[\frac{1}{\mu(X)}\right]\mu(dx) = -\log\left[\frac{1}{\mu(X)}\right],$$

so $H(f) \leqslant H(f_0)$ for all $f \in D$. ∎

If $\mu(X) = \infty$, then there are no constant densities and this proposition fails. However, if additional constraints are placed on the density, then we may obtain other results for maximal entropies as illustrated in the following two examples.

Example 9.1.1. Let $X = [0, \infty)$ and consider all possible densities f such that the first moment of f is given by

$$\int_0^\infty xf(x)\,dx = 1/\lambda. \tag{9.1.8}$$

Then the density

$$f_0(x) = \lambda e^{-\lambda x} \tag{9.1.9}$$

maximizes the entropy.

The proof proceeds as in Proposition 9.1.1. From inequality (9.1.6) we have, for arbitrary $f \in D$ satisfying (9.1.8),

$$H(f) \leq -\int_0^\infty f(x)\log(\lambda e^{-\lambda x})\,dx$$

$$= -\log\lambda\int_0^\infty f(x)\,dx + \int_0^\infty \lambda x f(x)\,dx$$

$$= -\log\lambda + 1\,.$$

Also, however, with f_0 given by (9.1.9),

$$H(f_0) = -\int_0^\infty \lambda e^{-\lambda x}\log(\lambda e^{-\lambda x})\,dx = -\log\lambda + 1$$

and thus $H(f) \leq H(f_0)$ for all $f \in D$ satisfying (9.1.8). □

Example 9.1.2. For our next example take $X = (-\infty, \infty)$ and consider all possible densities $f \in D$ such that the second moment of f is finite, that is,

$$\int_{-\infty}^{\infty} x^2 f(x)\,dx = \sigma^2. \tag{9.1.10}$$

Then the maximal entropy is achieved for the Gaussian density

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{9.1.11}$$

As before, we calculate that, for arbitrary $f \in D$ satisfying (9.1.10),

$$H(f) \leq -\int_{-\infty}^{\infty} f(x)\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right)\right]dx$$

$$= -\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]\int_{-\infty}^{\infty} f(x)\,dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \frac{1}{2} - \log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right].$$

Further

$$H(f_0) = -\int_{-\infty}^{\infty} f_0(x)\log f_0(x)\,dx = \frac{1}{2} - \log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]$$

so that the entropy is maximized with the Gaussian density (9.1.11). □

These two examples are simply special cases covered by the following simple statement.

Proposition 9.1.2. Let $(X, \mathcal{A}, \mu)$ be a measure space. Assume that a sequence $g_1, \ldots, g_m$ of nonnegative functions, $g_i \in L^\infty(X)$, is given as well as two sequences of positive real constants $\overline{g}_1, \ldots, \overline{g}_m$ and $\nu_1, \ldots, \nu_m$ that satisfy

$$\overline{g}_i = \int_X g_i(x) \exp[-\nu_i g_i(x)]\mu(dx) \bigg/ \int_X \prod_{i=1}^m \exp[-\nu_i g_i(x)]\mu(dx), \qquad i = 1, \ldots, m.$$

Then the maximum of the entropy $H(f)$ for all $f \in D$, subject to the conditions

$$\overline{g}_i = \int_X g_i(x) f(x)\mu(dx), \qquad i = 1, \ldots, m$$

occurs for

$$f_0(x) = \prod_{i=1}^m \exp[-\nu_i g_i(x)] \bigg/ \int_X \prod_{i=1}^m \exp[-\nu_i g_i(x)]\mu(dx).$$

Proof: For simplicity, set

$$Z = \int_X \prod_{i=1}^m \exp[-\nu_i g_i(x)]\mu(dx)$$

so

$$f_0(x) = Z^{-1} \prod_{i=1}^m \exp[-\nu_i g_i(x)].$$

From inequality (9.1.6), we have

$$\begin{aligned} H(f) &\leq -\int_X f(x) \log f_0(x)\mu(dx) \\ &= -\int_X f(x) \left[-\log Z - \sum_{i=1}^m \nu_i g_i(x)\right]\mu(dx) \\ &= \log Z + \sum_{i=1}^m \nu_i \int_X f(x) g_i(x)\mu(dx) \\ &= \log Z + \sum_{i=1}^m \nu_i \overline{g}_i . \end{aligned}$$

Furthermore, it is easy to show that

$$H(f_0) = \log Z + \sum_{i=1}^m \nu_i \overline{g}_i$$

and thus $H(f) \leq H(f_0)$. ∎

Remark 9.1.2. Note that if $m = 1$ and $g(x)$ is identified as the energy of a system, then the maximal entropy occurs for

$$f_0(x) = Z^{-1}e^{-\nu g(x)},$$

which is just the **Gibbs canonical distribution function,** with the **partition function** Z given by

$$Z = \int_X e^{-\nu g(x)}\mu(dx).$$

Further, the maximal entropy

$$H(f_0) = \log Z + \nu\overline{g}$$

is just the thermodynamic entropy. As is well known, all of the results of classical thermodynamics can be derived with the partition function Z and the preceding entropy $H(f_0)$. Indeed, the contents of Proposition 9.1.2 have been extensively used by Jaynes [1957] and Katz [1967] in an alternative formulation and development of classical and quantum statistical mechanics. □

Thus, the simple Gibbs inequality has far-reaching implications in pure mathematics as well as in more applied fields. Another inequality that we will have occasion to use often is the **Jensen inequality:** If $\eta(u)$, $u \geqslant 0$ is a function such that $\eta'' \leqslant 0$ (i.e., the area below the graph of η is convex), $P: L^p \to L^p$ $1 \leqslant p \leqslant \infty$, is a linear operator such that $P1 = 1$, and $Pf \geqslant 0$ for all $f \geqslant 0$, then for every $f \in L^p$, $f \geqslant 0$,

$$\eta(Pf) \geqslant P\eta(f) \qquad \text{whenever } P\eta(f) \text{ exists.} \tag{9.1.12}$$

The proof of this result is difficult and requires many specialized techniques. However, the following considerations provide some insight into why it is true. Let $\eta(y)$ be a convex function defined for $y \geqslant 0$. Pick u, v, and z such that $0 \leqslant u \leqslant z \leqslant v$. Since $z \in [u, v]$ there exist nonnegative constants, α and β, with $\alpha + \beta = 1$, such that

$$z = \alpha u + \beta v.$$

Further, from the convexity of η it is clear that $\eta(z) \geqslant r$, where

$$r = \alpha\eta(u) + \beta\eta(v).$$

Thus $\eta(z) \geqslant r$ gives

$$\eta(\alpha u + \beta v) \geqslant \alpha\eta(u) + \beta\eta(v).$$

Further, it is easy to verify by induction that for every sequence $0 \leqslant u_1 < u_2 \cdots$

$$\eta\left(\sum_i \alpha_i u_i\right) \geqslant \sum_i \alpha_i\eta(u_i), \tag{9.1.13}$$

where $\alpha_i \geqslant 0$ and $\Sigma_i \alpha_i = 1$. Now suppose we have a linear operator $P: R^n \to R^n$ satisfying $P1 = 1$. Since P is linear its coordinates must be of the form

$$(Pf)_i = \sum_{j=1}^{n} k_{ij} f_j \,,$$

where $f = (f_1, \dots, f_n)$ and $\Sigma_j k_{ij} = 1$, $k_{ij} \geqslant 0$. By applying inequality (9.1.13) to $(Pf)_i$, we have

$$\eta((Pf)_i) \geqslant \sum_{j=1}^{n} k_{ij} \eta(f_j) = P(\eta f)_i \,,$$

or, suppressing the coordinate index,

$$\eta(Pf) \geqslant P\eta(f) \,.$$

In an arbitrary (not necessarily finite dimensional) space the proof of the Jensen inequality is much more difficult, but still uses (9.1.13) as a starting point.

The final inequality we will have occasion to use is a direct consequence of integrating inequality (9.1.13) over the entire space X, namely

$$H\left(\alpha_i \sum_i f_i\right) \geqslant \sum_i \alpha_i H(f_i) \,, \tag{9.1.14}$$

where again $\alpha_i \geqslant 0$ and $\Sigma_i \alpha_i = 1$.

9.2 Entropy of $P^n f$ when P is a Markov operator

We are now in a position to examine the behavior of the entropy $H(P^n f)$ when P is a Markov operator. We begin with the following theorem.

Theorem 9.2.1. Let $(X, \mathcal{A}, \mu)$ be a finite measure space $[\mu(X) < \infty]$ and $P: L^1 \to L^1$ a Markov operator. If P has a constant stationary density $[P1 = 1]$, then

$$H(Pf) \geqslant H(f) \tag{9.2.1}$$

for all $f \geqslant 0, f \in L^1$.

Proof: Integrating Jensen's inequality (9.1.12) over the entire space X gives

$$\int_X \eta(Pf(x))\mu(dx) \geqslant \int_X P\eta(f(x))\mu(dx)$$
$$= \int_X \eta(f(x))\mu(dx)$$

since P preserves the integral. However, the left-most integral is $H(Pf)$, and the last integral is $H(f)$, so that (9.2.1) is proved. ■

Remark 9.2.1. For a finite measure space, we know that the maximal entropy $H_{\max}$ is $-\log[1/\mu(X)]$, so that

$$-\log[1/\mu(X)] \geq H(P^n f) \geq H(f).$$

This, in conjunction with Theorem 9.2.1, tells us that in a finite measure space when P has a constant stationary density, the entropy never decreases and is bounded above by $-\log[1/\mu(X)]$. Thus, in this case the entropy $H(P^n f)$ always converges as $n \to \infty$, although not necessarily to the maximum. Note further that, if we have a normalized measure space, then $\mu(X) = 1$ and $H_{\max} = 0$. □

Remark 9.2.2. In the case of a Markov operator without a constant stationary density, it may happen that the sequence $H(P^n f)$ is not increasing as n increases. As a simple example consider the parabolic transformation $S(x) = 4x(1 - x)$. The Frobenius–Perron operator for S, derived in Section 1.2, is

$$Pf(x) = \frac{1}{4\sqrt{1-x}}\{f(\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1-x}) + f(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1-x})\}$$

and it is easy to verify that

$$f_*(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

is a stationary density for P. Take as an initial density $f = 1$, so $H(f) = 0$ and

$$Pf(x) = \frac{1}{2\sqrt{1-x}}.$$

Then

$$H(Pf) = -\int_0^1 \frac{1}{2\sqrt{1-x}} \log\left(\frac{1}{2\sqrt{1-x}}\right) dx = (\log 2) - 1.$$

Clearly $H(Pf) < H(f) = 0$. □

It is for this reason that it is necessary to introduce the concept of conditional entropy for Markov operators with nonconstant stationary densities.

Definition 9.2.1. For $f \geq 0$ and $g \in D$, the **conditional entropy** of f with respect to g is defined by

$$H(f|g) = \int_X g(x)\eta\left[\frac{f(x)}{g(x)}\right]\mu(dx) = -\int_X f(x)\log\left[\frac{f(x)}{g(x)}\right]\mu(dx). \tag{9.2.2}$$

Remark 9.2.3. Since g is a density, the integral $H(f|g)$ is always defined, that is, it is either finite or equal to $-\infty$. In some sense, which is suggested by the equation (9.2.2), the value $H(f|g)$ measures the deviation of f from the density g. □

The conditional entropy $H(f|g)$ has two properties, which we will use later. They are

1. If $f, g \in D$, then, by inequality (9.1.6), $H(f|g) \leqslant 0$. The equality holds if and only if $f = g$.
2. If g is the constant density, $g = 1$, then $H(f|1) = H(f)$. Thus the conditional entropy $H(f|g)$ is a generalization of the entropy $H(f)$.

Theorem 9.2.2. Let $(X, \mathcal{A}, \mu)$ be an arbitrary measure space and $P: L^1 \to L^1$ a Markov operator. Then

$$H(Pf|Pg) \geqslant H(f|g) \qquad \text{for } f \geqslant 0 \text{ and } g \in D. \tag{9.2.3}$$

Remark 9.2.4. Note from this theorem that if g is a stationary density of P, then $H(Pf|Pg) = H(Pf|g)$ and thus

$$H(Pf|g) \geqslant H(f|g).$$

Thus the conditional entropy with respect to a stationary density is always increasing and bounded above by zero. It follows that $H(P^n f|g)$ always converges, but not necessarily to zero, as $n \to \infty$. □

Proof of Theorem 9.2.2: Here we give the proof of Theorem 9.2.2 only in the case when $Pg > 0$, $g > 0$, and the function f/g is bounded. [Consult Voigt (1981) for the full proof.] Take $g \in L^1$ with $g > 0$. Define an operator $R: L^\infty \to L^\infty$ by

$$Rh = P(hg)/Pg \qquad \text{for } h \in L^\infty,$$

where hg denotes multiplication, not composition. R has the following properties:

1. $Rh \geqslant 0$ for $h \geqslant 0$; and
2. $R1 = Pg/Pg = 1$.

Thus R satisfies the assumptions of Jensen's inequality, giving

$$\eta(Rh) \geqslant R\eta(h). \tag{9.2.4}$$

Setting $h = f/g$ the left-hand side of (9.2.4) may be written in the form

$$\eta(Rh) = -(Pf/Pg)\log(Pf/Pg)$$

and the right-hand side is given by

$$R\eta(h) = (1/Pg)P[(\eta \circ h)g] = -(1/Pg)P[f \log(f/g)].$$

Hence inequality (9.2.4) becomes

$$-Pf \log(Pf/Pg) \geq -P[f \log(f/g)].$$

Integrating this last inequality over the space X, and remembering that P preserves the integral, we have

$$\begin{aligned} H(Pf|Pg) &\geq -\int_X P\left\{f(x) \log\left[\frac{f(x)}{g(x)}\right]\right\}\mu(dx) \\ &= -\int_X f(x) \log\left[\frac{f(x)}{g(x)}\right]\mu(dx) = H(f|g), \end{aligned}$$

which finishes the proof. ∎

9.3 Entropy $H(P^n f)$ when P is a Frobenius–Perron operator

Inequalities (9.2.1) and (9.2.3) of Theorems 9.2.1 and 9.2.2 are not strong. In fact, the entropy may not increase at all during successive iterations of f. This is always the case when P is the Frobenius–Perron operator corresponding to an invertible transformation, which leads to the following theorem.

Theorem 9.3.1. Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $S: X \to X$ be an invertible measure-preserving transformation. If P is the Frobenius–Perron operator corresponding to S, then $H(P^n f) = H(f)$ for all n.

Proof: If S is invertible and measure preserving, then by equation (3.2.10) we have $Pf(x) = f(S^{-1}(x))$ since $J^{-1} \equiv 1$. If P_1 is the Frobenius–Perron operator corresponding to S^{-1}, we also have $P_1 f(x) = f(S(x))$. Thus $P_1 Pf = PP_1 f = f$, so $P_1 = P^{-1}$. From Theorem 9.2.1 we also have

$$H(P_1 Pf) \geq H(Pf) \geq H(f),$$

but, since $P_1 Pf = P^{-1}Pf = f$, we conclude that $H(Pf) = H(f)$, so $H(P^n f) = H(f)$ for all n. ∎

Remark 9.3.1. For any discrete or continuous time system that is invertible and measure preserving the entropy is always constant. In particular, for a continuous time system evolving according to the set of differential equations $x' = F(x)$, the entropy is constant if div $F = 0$ [see equation (7.8.18)]. Every Hamiltonian system satisfies this condition. □

However, for noninvertible (irreversible) systems this is not the case, and we have the following theorem.

Theorem 9.3.2. Let $(X, \mathcal{A}, \mu)$ be a measure space, $\mu(X) = 1$, $S: X \to X$ a measure-preserving transformation, and P the Frobenius–Perron operator corresponding to S. If S is exact then

$$\lim_{n\to\infty} H(P^n f) = 0$$

for all $f \in D$ such that $H(f) > -\infty$.

Proof: Assume initially that f is bounded, that is, $0 \leqslant f \leqslant c$. Then

$$0 \leqslant P^n f \leqslant P^n c = cP^n 1 = c.$$

Without any loss of generality, we can assume that $c > 1$. Further, since $\eta(u) \leqslant 0$ for $u \geqslant 1$, we have [note $\mu(X) = 1$ and $H_{\max} = 0$]

$$0 \geqslant H(P^n f) \geqslant \int_{A_n} \eta(P^n f(x))\mu(dx), \tag{9.3.1}$$

where

$$A_n = \{x: 1 \leqslant P^n f(x) \leqslant c\}.$$

Now, by the mean value theorem [using $\eta(1) = 0$], we obtain

$$\begin{aligned}\left|\int_{A_n} \eta(P^n f(x))\mu(dx)\right| &= \int_{A_n} |\eta(P^n f(x)) - \eta(1)|\mu(dx) \\ &\leqslant k\int_{A_n} |P^n f(x) - 1|\mu(dx) \\ &\leqslant k\int_X |P^n f(x) - 1|\mu(dx) = \|P^n f - 1\|,\end{aligned}$$

where

$$k = \sup_{1\leqslant u\leqslant c} |\eta'(u)|.$$

Since S is exact, from Theorem 4.4.1, we have $\|P^n f - 1\| \to 0$ as $n \to \infty$ for all $f \in D$ and thus

$$\lim_{n\to\infty} \int_{A_n} \eta(P^n f(x))\mu(dx) = 0.$$

From inequality (9.3.1), it follows that $H(P^n f)$ converges to zero.

Now relax the assumption that f is bounded and write f in the form

$$f = f_1 + f_2,$$

where

$$f_1(x) = \begin{cases} 0 & \text{if } f(x) > c \\ f(x) & \text{if } 0 \leqslant f(x) \leqslant c \end{cases}$$

and $f_2 = f - f_1$. Fixing $\varepsilon > 0$, we may choose c sufficiently large so that

$$\|f_2\| < \varepsilon \quad \text{and} \quad H(f_2) > -\varepsilon.$$

Write $P^n f$ in the form

$$P^n f = (1 - \delta)P^n\left(\frac{1}{1-\delta} f_1\right) + \delta P^n\left(\frac{1}{\delta} f_2\right),$$

where $\delta = \|f_2\|$. Now $f_1/(1 - \delta)$ is a bounded density, and so from the first part of our proof we know that for n sufficiently large

$$H\left(P^n\left(\frac{1}{1-\delta} f_1\right)\right) > -\varepsilon.$$

Furthermore,

$$\begin{aligned} \delta H\left(P^n\left(\frac{1}{\delta} f_2\right)\right) &= H(P^n f_2) - \log\left(\frac{1}{\delta}\right) \int_X P^n f_2(x)\mu(dx) \\ &= H(P^n f_2) - \|f_2\| \log\left(\frac{1}{\delta}\right) \\ &= H(P^n f_2) + \delta \log \delta. \end{aligned}$$

Since $H(P^n f_2) \geqslant H(f_2) > -\varepsilon$, this last expression becomes

$$\delta H\left(P^n\left(\frac{1}{\delta} f_2\right)\right) \geqslant -\varepsilon + \delta \log \delta.$$

Combining these results and inequality (9.1.14), we have

$$\begin{aligned} H(P^n f) &\geqslant (1 - \delta)H\left(P^n\left(\frac{1}{1-\delta} f_1\right)\right) + \delta H\left(P^n\left(\frac{1}{\delta} f_2\right)\right) \\ &\geqslant -\varepsilon(1 - \delta) - \varepsilon + \delta \log \delta \\ &= -2\varepsilon + \delta\varepsilon + \delta \log \delta. \end{aligned} \tag{9.3.2}$$

Since $\mu(X) = 1$, we have $H(P^n f) \leqslant 0$. Further since $\delta < \varepsilon$ and ε is arbitrary, the right-hand side of (9.3.2) is also arbitrarily small, and the theorem is proved. ∎

Example 9.3.1. We wish to compare the entropy of the baker transformation

$$S(x,y) = \begin{cases} (2x, \frac{1}{2}y), & 0 \le x < \frac{1}{2},\ 0 \le y \le 1 \\ (2x - 1, \frac{1}{2}y + \frac{1}{2}), & \frac{1}{2} \le x \le 1,\ 0 \le y \le 1, \end{cases}$$

originally introduced in Example 4.1.3, with that of the dyadic transformation. Observe that the x-coordinate of the baker transformation is transformed by the dyadic transformation

$$S_1(x) = 2x \qquad (\mathrm{mod}\ 1).$$

From our considerations of Chapter 4, we know that the baker transformation is invertible and measure preserving. Thus by Theorem 9.3.1 it follows that the entropy of the sequence $\{P^n f\}$, where P is the Frobenius–Perron operator corresponding to the baker transformation, is constant for every density f.

Conversely, the dyadic transformation S_1 is exact. Hence, from Theorem 9.3.2, the entropy of $\{P_1^n f\}$, where P_1 is the Frobenius–Perron operator corresponding to S_1, increases to zero for all bounded initial densities f. □

Remark 9.3.2. Observe that in going from the baker to the dyadic transformation, we are going from an invertible (reversible) to a noninvertible (irreversible) system through the loss of information about the y-coordinate. This loss of information is accompanied by an alteration of the behavior of the entropy. An analogous situation occurs in statistical mechanics where, in going from the Liouville equation to the Boltzmann equation, we also lose coordinate information and go from a situation where entropy is constant (Liouville equation) to one in which the entropy increases to its maximal value (Boltzmann H theorem). □

9.4 Behavior of $P^n f$ from $H(P^n f)$

In this section we wish to see what aspects of the eventual behavior of $P^n f$ can be deduced from $H(P^n f)$. This is a somewhat difficult problem, and the major stumbling block arises from the fact that η changes its sign. Thus, because of the integration in the definition of the entropy, it is difficult to determine f or its properties from $H(f)$. However, by use of the spectral representation Theorem 5.3.2 for Markov operators, we are able to circumvent this problem.

In our first theorem we wish to show that, if $H(P^n f)$ is bounded below, then P is constrictive. This is presented more precisely in the following theorem.

Theorem 9.4.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, $\mu(X) < \infty$, and $P: L^1 \to L^1$ a Markov operator such that $P1 = 1$. If there exists a constant $c > 0$ such that for every bounded $f \in D$

$$H(P^nf) \geqslant -c \quad \text{for} \quad n \text{ sufficiently large},$$

then P is constrictive.

Proof: Observe that $P1 = 1$ implies that Pf is bounded for bounded f. Thus, to prove our theorem, it is sufficient to show that the set $\mathcal{F}$ of all bounded $f \in D$ that satisfy

$$H(f) \geqslant -c$$

is weakly precompact.

We will use criterion 3 of Section 5.1 to demonstrate the weak precompactness of $\mathcal{F}$. Since $\|f\| = 1$ for all $f \in D$, the first part of the criterion is satisfied. To check the second part take $\varepsilon > 0$. Pick $l = e^{-1}\mu(X), N = \exp[2(c + l)/\varepsilon]$ and $\delta = \varepsilon/2N$, and take a set $A \subset X$ such that $\mu(A) < \delta$. Then

$$\int_A f(x)\mu(dx) = \int_{A_1} f(x)\mu(dx) + \int_{A_2} f(x)\mu(dx), \tag{9.4.1}$$

where

$$A_1 = \{x \in A: f(x) \leqslant N\}$$

and

$$A_2 = \{x \in A: f(x) > N\}.$$

The first integral on the right-hand side of (9.4.1) clearly satisfies

$$\int_{A_1} f(x)\mu(dx) \leqslant N\delta = \varepsilon/2.$$

In evaluating the second integral, note that from $H(f) \geqslant -c$, it follows that

$$\begin{aligned}\int_{A_2} f(x) \log f(x)\mu(dx) &\leqslant c - \int_{X\setminus A_2} f(x) \log f(x)\mu(dx) \\ &\leqslant c + \int_{X\setminus A_2} \eta_{\max}\mu(dx) \\ &\leqslant c + (1/e)\mu(X) = c + l.\end{aligned}$$

Therefore

$$\int_{A_2} f(x) \log N\mu(dx) < c + l$$

or

$$\int_{A_2} f(x)\mu(dx) < \frac{c + l}{\log N} = \frac{\varepsilon}{2}$$

Thus

$$\int_A f(x)\mu(dx) < \varepsilon$$

and $\mathcal{F}$ is weakly precompact. Thus, by Definition 5.3.3, the operator P is constrictive. ■

Before stating our next theorem, consider the following. Let $(X, \mathcal{A}, \mu)$ be a finite measure space, $S: X \to X$ a nonsingular transformation, and P the Frobenius–Perron operator corresponding to S. Assume that for some $c > 0$ the condition

$$H(P^n f) \geqslant -c$$

holds for every bounded $f \in D$ and n sufficiently large. By Theorem 5.3.1, since P is a Markov operator and is weakly constrictive, it is strongly constrictive. Hence in this case we may write Pf in the form given by the spectral decomposition Theorem 5.3.2, and, for every initial f, the sequence $\{P^n f\}$ will be asymptotically periodic.

Theorem 9.4.2. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space, $S: X \to X$ a measure-preserving transformation, and P the Frobenius–Perron operator corresponding to S. If

$$\lim_{n\to\infty} H(P^n f) = 0$$

for all bounded $f \in D$, then S is exact.

Proof: It follows from Theorems 9.4.1 and 5.3.1 that P is constrictive. Furthermore, since S is measure preserving, we know that P has a constant stationary density. From Proposition 5.4.2 we, therefore, have

$$P^n f(x) = \sum_{i=1}^{r} \lambda_{\alpha^{-n}(i)}(f)\bar{1}_{A_i}(x) + Q_n f(x) \qquad \text{for } f \in L^1.$$

If we can demonstrate that $r = 1$, then from Theorem 5.5.2 we will have shown S to be exact.

Pick

$$f(x) = [1/\mu(A_1)]1_{A_1}(x)$$

as an initial f. If τ is the asymptotic period of $P^n f$, then we must have

$$P^{n\tau} f(x) = [1/\mu(A_1)] 1_{A_1}(x).$$

However, by assumption,

$$\lim_{n\to\infty} H(P^n f) = 0,$$

and, since the sequence $\{H(P^{n\tau} f)\}$ is a constant sequence, we must have

$$H([1/\mu(A_1)] 1_{A_1}) = 0.$$

Note that, by Proposition 9.1.1, $H(f) = 0$ only if

$$f(x) = 1_X(x).$$

So, clearly, we must have

$$[1/\mu(A_1)] 1_{A_1}(x) = 1_X(x).$$

This is possible if and only if A_1 is the entire space X, and thus $r = 1$. Hence S is exact. ■

This theorem in conjunction with Theorem 9.3.2 tells us that the convergence of $H(P^n f)$ to zero as $n \to \infty$ is both necessary and sufficient for the exactness of measure-preserving transformations. If the transformation is not measure preserving then an analogous result using the conditional entropy may be proved.

To see this, suppose we have an arbitrary measure space $(X, \mathcal{A}, \mu)$ and a nonsingular transformation $S: X \to X$. Let P be the Frobenius–Perron operator corresponding to S and $g \in D$ $(g > 0)$ the stationary density of P so $Pg = g$. Since S is not measure preserving, our previous results cannot be used directly in examining the exactness of S.

However, consider the new measure space $(X, \mathcal{A}, \tilde{\mu})$, where

$$\tilde{\mu}(A) = \int_A g(x)\mu(dx) \qquad \text{for } A \in \mathcal{A}.$$

Since $Pg = g$, therefore $\tilde{\mu}$ is an invariant measure. Thus, in this new space the corresponding Frobenius–Perron operator $\tilde{P}$ is defined by

$$\int_A \tilde{P}h(x)\tilde{\mu}(dx) = \int_{S^{-1}(A)} h(x)\tilde{\mu}(dx) \qquad \text{for } A \in \mathcal{A}$$

and satisfies $\tilde{P}1 = 1$. This may be rewritten as

$$\int_A [\tilde{P}h(x)] g(x)\mu(dx) = \int_{S^{-1}(A)} h(x) g(x)\mu(dx).$$

However, we also have

$$\int_{S^{-1}(A)} h(x)g(x)\mu(dx) = \int_A P(h(x)g(x))\mu(dx)$$

so that $(\tilde{P}h)g = P(hg)$ or

$$\tilde{P}h = (1/g)P(hg)\,.$$

Furthermore, by induction,

$$\tilde{P}^n h = (1/g)P^n(hg)\,.$$

In this new space $(X, \mathscr{A}, \tilde{\mu})$, we may also calculate the entropy $\tilde{H}(\tilde{P}^n h)$ as

$$\begin{aligned}\tilde{H}(\tilde{P}^n h) &= -\int_X \tilde{P}^n h(x) \log[\tilde{P}^n h(x)]\tilde{\mu}(dx) \\ &= -\int_X \frac{1}{g(x)} P^n(h(x)g(x)) \log\left[\frac{P^n(h(x)g(x))}{g(x)}\right] g(x)\mu(dx) \\ &= H(P^n(hg)|g)\,.\end{aligned}$$

Observe that $h \in D(X, \mathscr{A}, \tilde{\mu})$ is equivalent to

$$h \geqslant 0 \quad \text{and} \quad \int_X h(x)g(x)\mu(dx) = 1\,,$$

which is equivalent to $hg \in D(X, \mathscr{A}, \mu)$. Set $f = hg$, so

$$\tilde{H}(\tilde{P}^n h) = H(P^n f|g)\,.$$

We may, therefore, use our previous theorems to examine the exactness of S in the new space $(X, \mathscr{A}, \tilde{\mu})$ or its asymptotic stability in the original space $(X, \mathscr{A}, \mu)$, that is, S is asymptotically stable in $(X, \mathscr{A}, \mu)$ if and only if

$$\lim_{n\to\infty} H(P^n f|g) = 0 \qquad (9.4.2)$$

for all $f \in D$ such that f/g is bounded.

Example 9.4.1. Consider the linear Boltzmann equation [equation (8.3.8)]

$$\frac{\partial u(t,x)}{\partial t} + u(t,x) = Pu(t,x)\,,$$

with the initial condition $u(0,x) = f(x)$, which we examined in Chapter 8. There we showed that the solution of this equation was given by

$$u(t,x) = e^{t(P-I)}f(x) = \hat{P}_t f(x)\,,$$

and $e^{t(P-I)}$ is a semigroup of Markov operators. From Theorem 9.2.2 we know immediately that the conditional entropy $H(\hat{P}_t f|f_*)$ is continually increasing for every f_* that is a stationary density of P. Furthermore, by (9.4.2) and Corollary 8.7.3, if $f_*(x) > 0$ and f_* is the unique stationary density of P, then

$$\lim_{t\to\infty} H(\hat{P}_t f|f_*) = H(f_*|f_*) = 0\,.$$

Thus, in the case in which f_* is positive and unique, the conditional entropy for the solutions of the linear Boltzmann equation always achieves its maximal value.