4

Studying chaos with densities

Here we introduce the concept of measure-preserving transformations and then define and illustrate three levels of irregular behavior that such transformations can display. These three levels are known as ergodicity, mixing, and exactness. The central theme of the chapter is to show the utility of the Frobenius–Perron and Koopman operators in the study of these behaviors.

All these basic notions arise in ergodic theory. Roughly speaking, preservation of an initial measure μ by a transformation corresponds to the fact that the constant density $f(x) = 1$ is a stationary density of the Frobenius–Perron operator, $P1 = 1$. Ergodicity corresponds to the fact that $f(x) \equiv 1$ is the unique stationary density of the Frobenius–Perron operator. Finally, mixing and exactness correspond to two different kinds of stability of the stationary density $f(x) \equiv 1$.

In Section 4.5, we briefly introduce Kolmogorov automorphisms, which are closely related to exact transformations. This section is only of a reference nature, and, therefore, all proofs are omitted and the examples are treated superficially.

4.1 Invariant measures and measure-preserving transformations

We start with a definition.

Definition 4.1.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $S\colon X \to X$ a measurable transformation. Then S is said to be **measure preserving** if

$$\mu(S^{-1}(A)) = \mu(A) \qquad \text{for all } A \in \mathcal{A}.$$

Since the property of measure preservation is dependent on S as well as μ, we will alternately say that the measure μ is **invariant** under S if S is measure preserving. Note that every measure-preserving transformation is necessarily nonsingular.

 https://doi.org/10.1017/CBO9780511897474.005

Theorem 4.1.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a nonsingular transformation, and P the Frobenius–Perron operator associated with S. Consider a nonnegative $f \in L^1$. Then a measure μ_f given by

$$\mu_f(A) = \int_A f(x)\mu(dx)$$

is invariant if and only if f is a fixed point of P.

Proof: First we show the "only if" portion. Assume μ_f is invariant. Then, by the definition of an invariant measure,

$$\mu_f(A) = \mu_f(S^{-1}(A)) \qquad \text{for all } A \in \mathcal{A},$$

or

$$\int_A f(x)\mu(dx) = \int_{S^{-1}(A)} f(x)\mu(dx) \qquad \text{for } A \in \mathcal{A}. \tag{4.1.1}$$

However, by the very definition of the Frobenius–Perron operator, we have

$$\int_{S^{-1}(A)} f(x)\mu(dx) = \int_A Pf(x)\mu(dx), \qquad \text{for } A \in \mathcal{A}. \tag{4.1.2}$$

Comparing (4.1.1) with (4.1.2) we immediately have $Pf = f$.

Conversely, if $Pf = f$ for some $f \in L^1, f \geqslant 0$, then from the definition of the Frobenius–Perron operator equation (4.1.1) follows and thus μ_f is invariant. ■

Remark 4.1.1. Note that the original measure μ is invariant if and only if $P1 = 1$. □

Example 4.1.1. Consider the r-adic transformation originally introduced in Example 1.2.1,

$$S(x) = rx \qquad (\text{mod } 1),$$

where $r > 1$ is an integer, on the measure space $([0,1], \mathcal{B}, \mu)$ where $\mathcal{B}$ is the Borel σ-algebra and μ is the Borel measure (cf. Remark 2.1.3). As we have shown in Example 1.2.1, for any interval $[0,x] \subset [0,1]$

$$S^{-1}([0,x]) = \bigcup_{i=0}^{r-1} \left[\frac{i}{r}, \frac{i}{r} + \frac{x}{r}\right]$$

and the Frobenius–Perron operator P corresponding to S is given by equation (1.2.13):

$$Pf(x) = \frac{1}{r}\sum_{i=0}^{r-1} f\left(\frac{i}{r} + \frac{x}{r}\right).$$

Thus

$$P1 = \frac{1}{r}\sum_{i=0}^{r-1} 1 = 1$$

and by our previous remark the Borel measure is invariant under the r-adic transformation. □

Remark 4.1.2. It should be noted that, as defined, the r-adic transformation is not continuous at $\frac{1}{2}$. However, if instead of defining the r-adic transformation on the interval $[0, 1]$ we define it on the unit circle (circle with circumference of 1) obtained by identifying 0 with 1 on the interval $[0, 1]$, then it is continuous and differentiable throughout. □

Example 4.1.2. Again consider the measure space $([0, 1], \mathcal{B}, \mu)$, where μ is the Borel measure. Let $S: [0, 1] \to [0, 1]$ be the quadratic map $S(x) = 4x(1 - x)$ of Chapter 1. As was shown there, for $[0, x] \subset [0, 1]$,

$$S^{-1}([0, x]) = [0, \tfrac{1}{2} - \tfrac{1}{2}\sqrt{1 - x}] \cup [\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - x}, 1]$$

and the Frobenius–Perron operator is given by

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

Clearly,

$$P1 = \frac{1}{2\sqrt{1 - x}},$$

so that the Borel measure μ is not invariant under S by Remark 4.1.1. To find an invariant measure we must find a solution to the equation $Pf = f$ or

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

This problem was first solved by Ulam and von Neumann [1947] who showed that the solution is given by

$$f_*(x) = \frac{1}{\pi\sqrt{x(1 - x)}}, \tag{4.1.3}$$

which justifies our assertion in Section 1.2. It is straightforward to show that f_* as given by (4.1.3) does, indeed, constitute a solution to $Pf = f$. Hence the measure

$$\mu_{f_*}(A) = \int_A \frac{dx}{\pi\sqrt{x(1 - x)}}$$

is invariant under the quadratic transformation $S(x) = 4x(1-x)$. □

Remark 4.1.3. The factor of π appearing in equation (4.1.3) ensures that f_* is a density and thus that the measure μ_{f_*} is normalized. □

Example 4.1.3 (The baker transformation). Now let X be the unit square in a plane, which we denote by $X = [0,1] \times [0,1]$ (see Section 2.2). The Borel σ-algebra $\mathcal{B}$ is now generated by all possible rectangles of the form $[0,a] \times [0,b]$ and the Borel measure μ is the unique measure on $\mathcal{B}$ such that

$$\mu([0,a] \times [0,b]) = ab.$$

(Thus the Borel measure is a generalization of the concept of the area.) We define a transformation $S\colon X \to X$ by

$$S(x,y) = \begin{cases} (2x, \frac{1}{2}y) & 0 \leqslant x < \frac{1}{2},\ 0 \leqslant y \leqslant 1 \\ (2x-1, \frac{1}{2}y + \frac{1}{2}) & \frac{1}{2} \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1. \end{cases} \tag{4.1.4}$$

To understand the operation of this transformation, examine Figure 4.1.1, where X is shown in Figure 4.1.1a. The first operation of S involves a compression of X in the y direction by $\frac{1}{2}$ and a stretching of X in the x direction by a factor of 2 (Figure 4.1.1b). The transformation S is completed by vertically dividing the compressed and stretched rectangle, shown in Figure 4.1.1b, into two equal parts and then placing the right-hand part on top of the left-hand part (Figure 4.1.1c). This transformation has become known as the baker transformation because it mimics some aspects of kneading dough. From Figure 4.1.1 it is obvious that the counterimage of any rectangle is again a rectangle or a pair of rectangles with the same total area. Thus the baker transformation is measurable.

Now we calculate the Frobenius–Perron operator for the baker transformation. It will help to refer to Figure 4.1.2 and to note that two cases must be distinguished: $0 \leqslant y < \frac{1}{2}$ and $\frac{1}{2} \leqslant y \leqslant 1$. Thus, for the simpler case of $0 \leqslant y < \frac{1}{2}$ and $0 \leqslant x < 1$ we have

$$S^{-1}([0,x] \times [0,y]) = [0, \tfrac{1}{2}x] \times [0, 2y]$$

so from equation (3.2.9)

$$\begin{aligned} Pf(x,y) &= \frac{\partial^2}{\partial x\, \partial y} \int_0^{x/2} ds \int_0^{2y} f(s,t)\, dt \\ &= f(\tfrac{1}{2}x, 2y), \qquad 0 \leqslant y < \tfrac{1}{2}. \end{aligned}$$

In the second case, for $\frac{1}{2} \leqslant y \leqslant 1$, we find that

$$S^{-1}([0,x] \times [0,y]) = ([0,\tfrac{1}{2}x] \times [0,1]) \cup ([\tfrac{1}{2}, \tfrac{1}{2} + \tfrac{1}{2}x] \times [0, 2y-1])$$

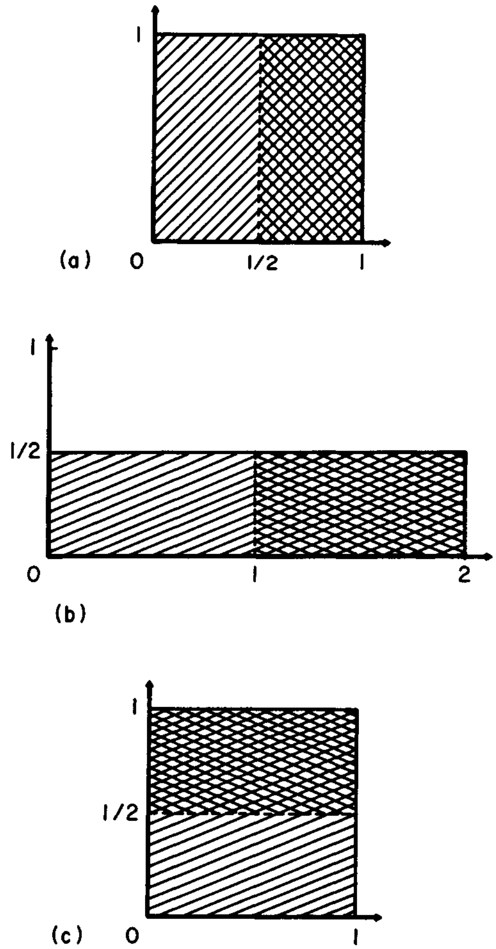


Figure 4.1.1. Steps showing the operation of the baker transformation given in equation (4.1.4).

hence

$$Pf(x,y) = \frac{\partial^2}{\partial x\,\partial y}\left\{\int_0^{x/2} ds \int_0^1 f(s,t)\,dt + \int_{1/2}^{(1/2)+(x/2)} ds \int_0^{2y-1} f(s,t)\,dt\right\}$$

$$= f(\tfrac{1}{2} + \tfrac{1}{2}x, 2y - 1), \qquad \tfrac{1}{2} \leqslant y \leqslant 1 .$$

Thus, finally,

$$Pf(x,y) = \begin{cases} f(\tfrac{1}{2}x, 2y), & 0 \leqslant y < \tfrac{1}{2} \\ f(\tfrac{1}{2} + \tfrac{1}{2}x, 2y - 1), & \tfrac{1}{2} \leqslant y \leqslant 1 \end{cases} \tag{4.1.5}$$

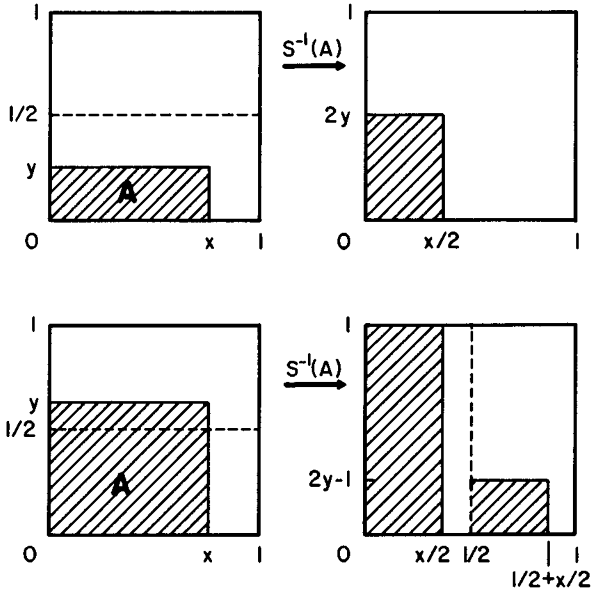


Figure 4.1.2. Two cases for calculating the counterimage of a set A by the baker transformation.

so that $P1 = 1$, and the Borel measure is, therefore, invariant under the baker transformation. □

Remark 4.1.4. Note that the transformation of the x-coordinate in the baker transformation is the dyadic transformation. However, the dyadic transformation is not 1–1, whereas the baker transformation is a.e. Given an $X \subset R$ and any (not necessarily invertible) one-dimensional transformation $S: X \to X$, we may construct a two-dimensional invertible transformation $T: X \times X \to X \times X$ with $0 < \beta$ and

$$T(x, y) = (S(x) + y, \beta x).$$

As an example let $S: [0,1] \to [0,1]$ be the quadratic map $S(x) = 4x(1 - x)$. Then T is given by

$$T(x, y) = (4x(1 - x) + y, \beta x),$$

which is equivalent to the **Henon map** first studied by Henon [1976]. □

Remark 4.1.5. Our derivation of the Frobenius–Perron operator (4.1.5) corresponding to the baker transformation is longer than it need be. Since the baker transformation is invertible (except on the line $y = \frac{1}{2}$), and indeed

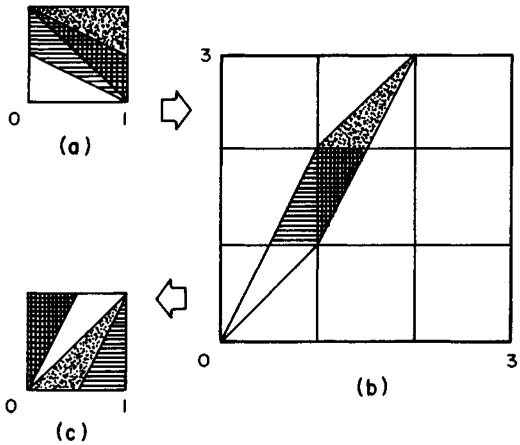


Figure 4.1.3. Operation of the Anosov diffeomorphism [equation (4.1.6)].

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

equation (4.1.5) may be immediately obtained from Corollary 3.2.1. □

Example 4.1.4 (Anosov diffeomorphisms). The baker transformation of the previous example may be considered to be a prototype of a very important class of transformations originally introduced by Anosov [1963]. One of the simplest of the Anosov diffeomorphisms is given by

$$S(x,y) = (x + y, x + 2y) \qquad (\text{mod } 1). \tag{4.1.6}$$

To see the effect of this transformation consult Figure 4.1.3. In the first part (a) of the figure we depict the unit square in the plane and divide it into four triangular areas. In Figure 4.1.3b we show how the unit square is transformed after one application of $(x,y) \to (x + y, x + 2y)$, whereas Figure 4.1.3c shows the result of the full Anosov diffeomorphism. It is clear that the effect of this transformation will be to very quickly scramble, or mix, various regions of the unit square. This property of mixing, also shared by the baker transformation, is most important and is dealt with in more detail in Section 4.3.

The determinant of the Jacobian of transformation (4.1.6) is given by

$$J = \det\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1,$$

so that the transformation is measure preserving; we already noted this result on geometric grounds in Figure 4.1.3. The two eigenvectors associated with S have eigenvalues

$$\lambda_1 = \frac{3}{2} - \frac{\sqrt{5}}{2} \quad \text{and} \quad \lambda_2 = \frac{3}{2} + \frac{\sqrt{5}}{2},$$

hence $0 < \lambda_1 < 1 < \lambda_2$. Thus, as for the baker transformation, the Anosov diffeomorphism involves a stretching in one direction and a corresponding compression in the orthogonal direction.

With some patience it is possible to derive an explicit formula for the Frobenius–Perron operator corresponding to the Anosov diffeomorphism (4.1.6) using a technique analogous to that employed for the baker transformation of the previous example. However, we can obtain this result immediately from Corollary 3.2.1 since the Anosov diffeomorphism is invertible. An easy calculation gives

$$S^{-1}(x,y) = (2x - y, y - x) \qquad (\mathrm{mod}\ 1)$$

and thus

$$Pf(x,y) = f(2x - y, y - x), \tag{4.1.7}$$

where, as in (4.1.6), the terms $2x - y$ and $y - x$ should be interpreted modulo 1. From (4.1.7) it is clear that $P1 = 1$, which corresponds to the fact that S preserves the Borel measure. □

Remark 4.1.6. Observe that, if we replace the unit square $[0,1] \times [0,1]$ with the torus, that is, if we identify points $(x,1)$ with $(x,0)$ and $(1,y)$ with $(0,y)$, then this example of an Anosov diffeomorphism becomes continuous and differentiable just as the r-adic transformation does when the unit interval is replaced by the unit circle. The word diffeomorphism comes from the fact that the Anosov transformation is invertible, and that both the transformation and its inverse are differentiable. □

4.2 Ergodic transformations

Because a transformation S has an invariant measure or because the Frobenius–Perron operator P associated with S has a stationary density does not imply that S has interesting statistical properties. For example, if S is the identity on X, that is, $S(x) = x$ for every $x \in X$, then

$$S^{-1}(A) = A \tag{4.2.1}$$

for every $A \subset X$, and, consequently, $Pf = f$ for every $f \in L^1$. This is, of course, not an interesting transformation. However, even if (4.2.1) holds for just one subset A of X, then the transformation S may be studied on the sets A and $X\backslash A$ separately. To see this, assume that A is fixed and condition (4.2.1) holds. Consider a trajectory

$$x^0, S(x^0), S^2(x^0), \ldots, .$$

Equality (4.2.1) implies that S maps A into itself and no element of $X\backslash A$ is mapped into A. Thus, if $x^0 \in A$ then $S^n(x^0) \in A$ for all n, and *if* $x^0 \notin A$ then $S^n(x^0) \notin A$ for all n.

Example 4.2.1. A simple example is

$$S(k) = \begin{cases} k + 2 & \text{for } k = 1, \ldots, 2(N-1) \\ 1 & \text{for } k = 2N - 1 \\ 2 & \text{for } k = 2N \end{cases}$$

operating on the space $X = \{1, \ldots, 2N\}$ with the counting measure. This transformation can be studied separately on the sets $A = \{1, 3, \ldots, 2N - 1\}$ and $X\backslash A = \{2, 4, \ldots, 2N\}$ of odd and even integers. □

Any set A satisfying (4.2.1) is called **invariant.** We require this equality to be satisfied modulo zero (see Remark 3.1.3). Then we can make the following definition.

Definition 4.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and let a nonsingular transformation $S: X \to X$ be given. Then S is called **ergodic** if every invariant set $A \in \mathcal{A}$ is such that either $\mu(A) = 0$ or $\mu(X\backslash A) = 0$; that is, S is ergodic if all invariant sets are **trivial** subsets of X.

From this definition it follows that any ergodic transformation S must be studied on the entire space X. Determining ergodicity on the basis of Definition 4.2.1 is, in general, difficult except for simple examples on finite spaces. Thus, for example, the transformation in Example 4.2.1 is not ergodic on the space X of integers, but it is ergodic on the sets of even and odd integers.

In studying more interesting examples the following theorem may be of use.

Theorem 4.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $S: X \to X$ a nonsingular transformation. S is ergodic if and only if, for every measurable function $f: X \to R$,

$$f(S(x)) = f(x) \qquad \text{for almost all } x \in X \tag{4.2.2}$$

implies that f is constant almost everywhere.

Proof: We first show that ergodicity implies f is constant. Assume that, as in Figure 4.2.1, we have a function f satisfying (4.2.2), which is not constant almost

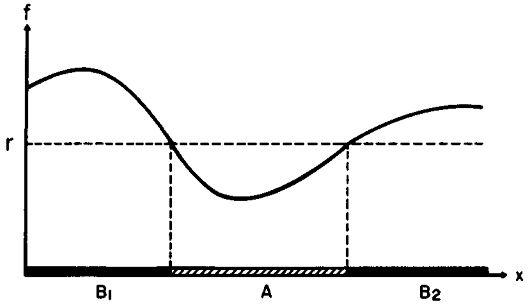


Figure 4.2.1. Definition of the sets A and $B = B_1 \cup B_2$.

everywhere, and that S is ergodic. Then there is some r such that the sets

$$A = \{x: f(x) \leqslant r\} \quad \text{and} \quad B = \{x: f(x) > r\}$$

have positive measure. These sets are also invariant because

$$\begin{aligned} S^{-1}(A) &= \{x: S(x) \in A\} = \{x: f(S(x)) \leqslant r\} \\ &= \{x: f(x) \leqslant r\} = A \end{aligned}$$

and similarly for B. Because sets A and B are invariant, S is not ergodic, which is a contradiction. Thus every f satisfying (4.2.2) must be constant.

To prove the converse assume that S is not ergodic. Then, by Definition 4.2.1, there is a nontrivial set $A \in \mathcal{A}$ that is invariant. Set $f = 1_A$, and, since A is nontrivial, f is not a constant function. Moreover, since $A = S^{-1}(A)$ we have

$$f(S(x)) = 1_A(S(x)) = 1_{S^{-1}(A)}(x) = 1_A(x) = f(x) \text{ a.e.}$$

and (4.2.2) is satisfied by a nonconstant function. ■

Remark 4.2.1. It is clear from the proof that it is sufficient to verify only (4.2.2) for bounded measurable functions since in the last part of the proof we used characteristic functions that are bounded. □

An immediate consequence of Theorem 4.2.1 in combination with the definition of the Koopman operator is the following corollary.

Corollary 4.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a nonsingular transformation, and U the Koopman operator with respect to S. Then S is ergodic if and only if all the fixed points of U are constant functions.

In addition to Theorem 4.2.1 and the preceding corollary, another result of use in checking the ergodicity of S using the Frobenius–Perron operator is contained in the following theorem.

Theorem 4.2.2. Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a nonsingular transformation, and P the Frobenius–Perron operator associated with S. If S is ergodic, then there is at most one stationary density f_* of P. Further, if there is a unique stationary density f_* of P and $f_*(x) > 0$ a.e., then S is ergodic.

Proof: To prove the first part of the theorem assume that S is ergodic and that f_1 and f_2 are different stationary densities of P. Set $g = f_1 - f_2$ so that $Pg = g$ by the linearity of P. Thus, by Proposition 3.1.3, g^+ and g^- are both stationary densities of P:

$$Pg^+ = g^+ \quad \text{and} \quad Pg^- = g^-. \tag{4.2.3}$$

Since, by assumption, f_1 and f_2 are not only different but are also densities, we have

$$g^+ \not\equiv 0 \quad \text{and} \quad g^- \not\equiv 0\,.$$

Set

$$A = \operatorname{supp} g^+ = \{x: g^+(x) > 0\}$$

and

$$B = \operatorname{supp} g^- = \{x: g^-(x) > 0\}\,.$$

It is evident that A and B are disjoint sets and both have positive (nonzero) measure. By equality (4.2.3) and Proposition 3.2.1, we have

$$A \subset S^{-1}(A) \quad \text{and} \quad B \subset S^{-1}(B)\,.$$

Since A and B are disjoint sets, $S^{-1}(A)$ and $S^{-1}(B)$ are also disjoint. By induction we, therefore, have

$$A \subset S^{-1}(A) \subset S^{-2}(A) \cdots \subset S^{-n}(A)$$

and

$$B \subset S^{-1}(B) \subset S^{-2}(B) \cdots \subset S^{-n}(B)\,,$$

where $S^{-n}(A)$ and $S^{-n}(B)$ are also disjoint for all n. Now define two sets by

$$\overline{A} = \bigcup_{n=0}^{\infty} S^{-n}(A) \quad \text{and} \quad \overline{B} = \bigcup_{n=0}^{\infty} S^{-n}(B)\,.$$

These two sets $\overline{A}$ and $\overline{B}$ are also disjoint and, furthermore, they are invariant because

$$S^{-1}(\overline{A}) = \bigcup_{n=1}^{\infty} S^{-n}(A) = \bigcup_{n=0}^{\infty} S^{-n}(A) = \overline{A}$$

and

$$S^{-1}(\overline{B}) = \bigcup_{n=1}^{\infty} S^{-n}(B) = \bigcup_{n=0}^{\infty} S^{-n}(B) = \overline{B}\,.$$

Neither $\overline{A}$ nor $\overline{B}$ are of measure zero since A and B are not of measure zero. Thus, $\overline{A}$ and $\overline{B}$ are nontrivial invariant sets, which contradicts the ergodicity of S. Thus the first portion of the theorem is proved.

To prove the second portion of the theorem, assume that $f_* > 0$ is the unique density satisfying $Pf_* = f_*$ but that S is not ergodic. If S is not ergodic, then there exists a nontrivial set A such that

$$S^{-1}(A) = A$$

and with $B = X\backslash A$

$$S^{-1}(B) = B\,.$$

With these two sets A and B, we may write $f_* = 1_A f_* + 1_B f_*$, so that

$$1_A f_* + 1_B f_* = P(1_A f_*) + P(1_B f_*)\,. \tag{4.2.4}$$

The function $1_B f_*$ is equal to zero in the set $X\backslash B = A = S^{-1}(A)$. Thus, by Proposition 3.2.1, $P(1_B f_*)$ is equal to zero in $A = X\backslash B$ and, likewise, $P(1_A f_*)$ is equal to zero in $B = X\backslash A$. Thus, equality (4.2.4) implies that

$$1_A f_* = P(1_A f_*) \quad \text{and} \quad 1_B f_* = P(1_B f_*)\,.$$

Since f_* is positive on A and on B, we may replace $1_A f_*$ by $f_A = 1_A f_*/\|1_A f_*\|$, and $1_B f_*$ by $f_B = 1_B f_*/\|1_B f_*\|$ in the last pair of equalities to obtain

$$f_A = Pf_A \quad \text{and} \quad f_B = Pf_B\,.$$

This implies that there exist two stationary densities of P, which is in contradiction to our assumption. Thus, if there is a unique positive stationary density f_* of P, then S is ergodic. ■

Example 4.2.2. Consider a circle of radius 1, and let S be a rotation through an angle ϕ. This transformation is equivalent to the map $S\colon [0, 2\pi) \to [0, 2\pi)$ defined by

$$S(x) = x + \phi \qquad (\text{mod } 2\pi)\,.$$

If ϕ is commensurate with 2π (that is $\phi/2\pi$ is rational), then S is evidently nonergodic. For example, if $\phi = \pi/3$ then the sets A and B of Figure 4.2.2 are invariant. For any $\phi = 2\pi(k/n)$, where k and n are integers, we will still find two invariant sets A and B, each containing n parts. As n becomes large the intermingling of the two sets A and B becomes more complicated and suggests

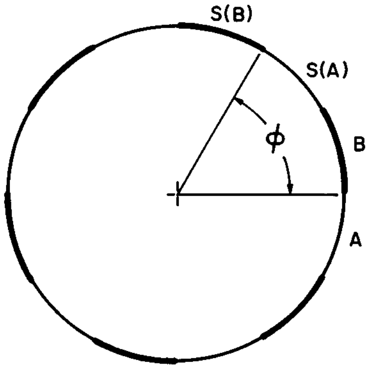


Figure 4.2.2. The two disjoint sets A (containing all the arcs denoted by thin lines) and B (containing arcs marked by heavy lines) are invariant under the rotational transformation when $\phi/2\pi$ is rational.

that the rotational transformation S may be ergodic for $(\phi/2\pi)$ irrational. This does in fact hold, but it will be proved later when we have more techniques at our disposal.

In this example the behavior of the trajectories is moderately regular and insensitive to changes in the initial value. Thus, independent of whether or not $(\phi/2\pi)$ is rational, if the value of ϕ is known precisely but the initial condition is located between α and β, $x_0 \in (\alpha, \beta)$, then

$$S^n(x_0) \in (\alpha + n\phi, \beta + n\phi) \qquad (\text{mod } 2\pi)$$

and all of the following points of the trajectory are known with the same accuracy, $(\beta - \alpha)$. □

Before closing this section we state, without proof, the **Birkhoff individual ergodic theorem** [Birkhoff, 1931a, b].

Theorem 4.2.3. Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a measurable transformation, and $f: X \to R$ an integrable function. If the measure μ is invariant, then there exists an integrable function f^* such that

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) \qquad \text{for almost all } x \in X. \tag{4.2.5}$$

Without additional assumptions the limit $f^*(x)$ is generally difficult to determine. However, it can be shown that $f^*(x)$ satisfies

$$f^*(x) = f^*(S(x)) \qquad \text{for almost all } x \in X, \tag{4.2.6}$$

and when $\mu(X) < \infty$

$$\int_X f^*(x)\mu(dx) = \int_X f(x)\mu(dx). \tag{4.2.7}$$

Equation (4.2.6) follows directly from (4.2.5) if x is replaced by $S(x)$. The second property, (4.2.7), follows from the invariance of μ and equation (4.2.5). Thus, by Theorem 3.2.1,

$$\int_X f(x)\mu(dx) = \int_X f(S(x))\mu(dx) = \cdots$$

so that integrating equation (4.2.5) over X and passing to the limit yields (4.2.7) by the Lebesgue-dominated convergence theorem when f is bounded. When f is not bounded the argument is more difficult.

Remark 4.2.2. Theorem 4.2.3 is known as the individual ergodic theorem because it may be used to give information concerning the asymptotic behavior of trajectories starting from a given point $x \in X$. As our emphasis is on densities and not on individual trajectories, we will seldom use this theorem. □

With the notion of ergodicity we may derive an important and often quoted extension of the Birkhoff individual ergodic theorem.

Theorem 4.2.4. Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $S: X \to X$ be measure preserving and ergodic. Then, for any integrable f, the average of f along the trajectory of S is equal almost everywhere to the average of f over the space X; that is,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) = \frac{1}{\mu(X)} \int_X f(x)\mu(dx) \quad \text{a.e.} \tag{4.2.8}$$

Proof: From (4.2.6) and Theorem 4.2.1 it follows that f^* is constant almost everywhere. Hence, from (4.2.7), we have

$$\int_X f^*(x)\mu(dx) = f^* \int_X \mu(dx) = f^*\mu(X) = \int_X f(x)\mu(dx),$$

so that

$$f^*(x) = \frac{1}{\mu(X)} \int_X f(x)\mu(dx) \quad \text{a.e.}$$

Thus equation (4.2.5) of the Birkhoff theorem and the preceding formula imply (4.2.8), and the theorem is proved. ■

One of the most quoted consequences of this theorem is the following.

Corollary 4.2.2. Let $(X, \mathcal{A}, \mu)$ be a finite measure space and $S: X \to X$ be measure preserving and ergodic. Then for any set $A \in \mathcal{A}$, $\mu(A) > 0$, and almost all $x \in X$, the fraction of the points $\{S^k(x)\}$ in A as $k \to \infty$ is given by $\mu(A)/\mu(X)$.

Proof: Using the characteristic function 1_A of A, the fraction of points from $\{S^k(x)\}$ in A is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_A(S^k(x)).$$

However from (4.2.8) this is simply $\mu(A)/\mu(X)$. ■

Remark 4.2.3. Corollary 4.2.2 says that every set of nonzero measure is visited infinitely often by the iterates of almost every $x \in X$. This result is a special case of the **Poincaré recurrence Theorem.** □

4.3 Mixing and exactness

Mixing transformations

The examples of the previous section show that ergodic behavior per se need not be very complicated and suggests the necessity of introducing another concept, that of mixing.

Definition 4.3.1. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space, and $S: X \to X$ a measure-preserving transformation. S is called **mixing** if

$$\lim_{n\to\infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B) \qquad \text{for all } A, B \in \mathcal{A}. \tag{4.3.1}$$

Condition (4.3.1) for mixing has a very simple interpretation. Consider points x belonging to the set $A \cap S^{-n}(B)$. These are the points such that $x \in A$ and $S^n(x) \in B$. Thus, from (4.3.1), as $n \to \infty$ the measure of the set of such points is just $\mu(A)\mu(B)$. This can be interpreted as meaning that the fraction of points starting in A that ended up in B after n iterations (n must be a large number) is just given by the product of the measures of A and B and is independent of the position of A and B in X.

It is easy to see that any mixing transformation must be ergodic. Assume that $B \in \mathcal{A}$ is an invariant set, so that $B = S^{-1}(B)$ and, even further, $B = S^{-n}(B)$ by induction. Take $A = X \backslash B$ so that $\mu(A \cap B) = \mu(A \cap S^{-n}(B)) = 0$. However, from (4.3.1), we must have

$$\lim_{n\to\infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B) = (1 - \mu(B))\mu(B),$$

and thus $\mu(B)$ is either 0 or 1, which proves ergodicity.

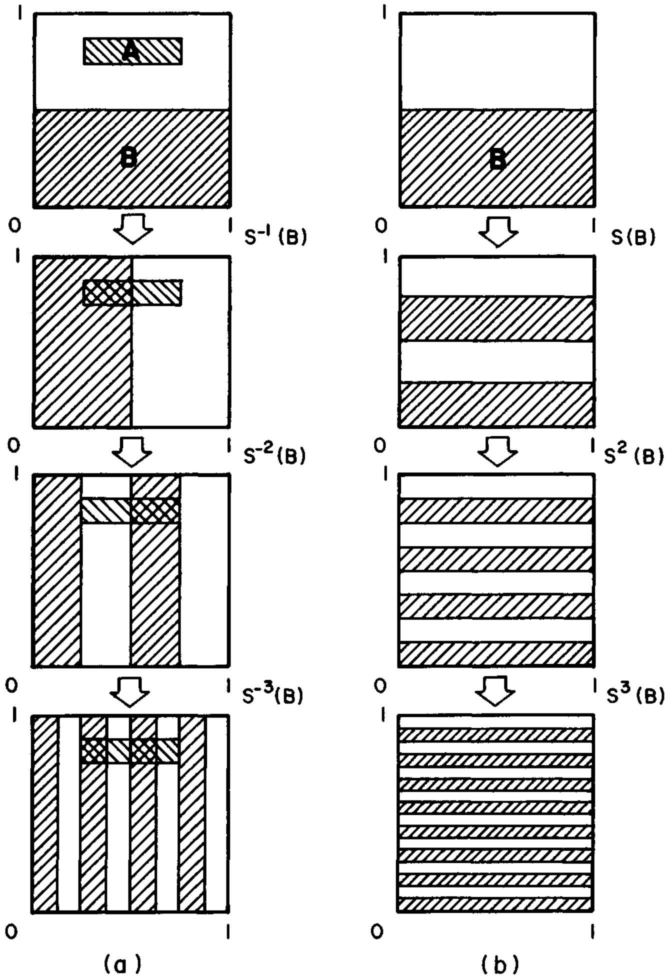


Figure 4.3.1. Mixing illustrated by the behavior of counterimages and images of a set B by the baker transformation. (a) The nth counterimage of the set B consists of 2^{n-1} vertical rectangles, each of equal area. (b) Successive iterates of the same set B results in 2^n horizontal rectangles after n iterations.

Many of the transformations considered in our examples to this point are mixing, for example, the baker, quadratic, Anasov, and r-adic. (The rotation transformation is not mixing according to our foregoing discussion.) To illustrate the mixing property we consider the baker and r-adic transformations in more detail.

Example 4.3.1. (See also Example 4.1.3.) In considering the baker transformation, it is relatively easy to check the mixing condition (4.3.1) for genera-

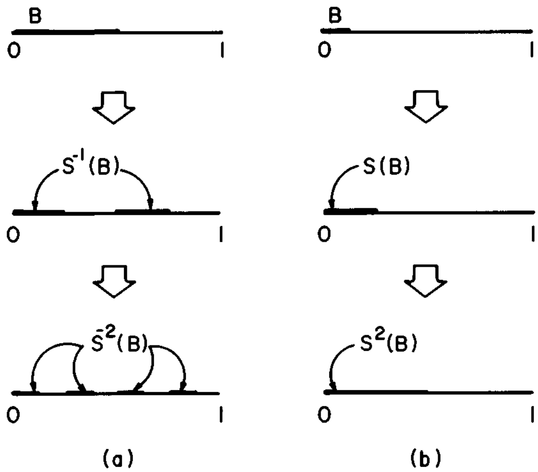


Figure 4.3.2. The behavior of counterimages and images of a set B by the dyadic transformation. (a) Successive counterimages of a set B that result after n such counterimages, in 2^n disjoint sets on $[0, 1]$. (b) The behavior of images of a set B generated by the dyadic transformation, which is quite different than that for the baker transformation. (See the text for further details.)

tors of the σ-algebra $\mathcal{B}$, namely, for rectangles. Although the transformation is simple, writing the algebraic expressions for the counterimages is tedious, and the property of mixing is easier to see pictorially. Consider Figure 4.3.1a, where two sets A and B are represented with $\mu(B) = \frac{1}{2}$. We take repeated counterimages of the set B by the baker transformation and find that after n such steps, $S^{-n}(B)$ consists of 2^{n-1} vertical rectangles of equal area. Eventually the measure of $A \cap S^{-n}(B)$ approaches $\mu(A)/2$, and condition (4.3.1) is evidently satisfied. The behavior of any pair of sets A and B is similar.

It is interesting that the baker transformation behaves in a similar fashion if, instead of examining $S^{-n}(B)$, we look at $S^n(B)$ as shown in Figure 4.3.1b. Now we have 2^n horizontal rectangles after n steps and all of our previous comments apply. So, for the baker transformation the behavior of images and counterimages is very similar and illustrates the property of mixing. This is not true for our next example, the dyadic transformation. □

In general, proving that a given transformation is mixing via the Definition 4.3.1 is difficult. In the next section, Theorem 4.4.1 and Proposition 4.4.1, we introduce easier and more powerful techniques for this purpose.

Example 4.3.2. (Cf. Examples 1.2.1 and 4.1.1.) To examine the mixing property (4.3.1) for the dyadic transformation, consider Figure 4.3.2a. Now we take the set $B = [0, b]$ and find that the nth counterimage of B consists of intervals on $[0,1]$ each of the same length. Eventually, as before, $\mu(A \cap S^{-n}(B)) \to \mu(A)\mu(B)$.

As for the baker transformation let us consider the behavior of images of a set B under the dyadic transformation (cf. Figure 4.3.2b). In this case, if $B = [0, b]$, then $S(B) = [0, 2b]$ and after a finite number of iterations $S^n(B) = [0, 1)$. The same procedure with any arbitrary set $B \subset [0,1]$ of positive measure will show that $\mu(S^n(B)) \to 1$ and thus the behavior of images of the dyadic transformation is different from the baker transformation. □

Exact transformations

The behavior illustrated by images of the dyadic transformation is called exactness, and is made precise by the following definition due to Rochlin [1964].

Definition 4.3.2. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space and $S: X \to X$ a measure-preserving transformation such that $S(A) \in \mathcal{A}$ for each $A \in \mathcal{A}$. If

$$\lim_{n\to\infty} \mu(S^n(A)) = 1 \qquad \text{for every } A \in \mathcal{A},\ \mu(A) > 0\,, \tag{4.3.2}$$

then S is called **exact.**

It can be proved, although it is not easy to do so from the definition, that exactness of S implies that S is mixing. As we have seen from the baker transformation the converse is not true. We defer the proof until the next section when we have other tools at our disposal.

Condition (4.3.2) has a very simple interpretation. If we start with a set A of initial conditions of nonzero measure, then after a large number of iterations of an exact transformation S the points will have spread and completely filled the space X.

Remark 4.3.1. It cannot be emphasized too strongly that invertible transformations cannot be exact. In fact, for any invertible measure-preserving transformation S, we have $\mu(S(A)) = \mu(S^{-1}(S(A))) = \mu(A)$ and by induction $\mu(S^n(A)) = \mu(A)$, which violates (4.3.2). □

In this and the previous section we have defined and examined a hierarchy of "chaotic" behaviors. However, by themselves the definitions are a bit sterile and may not convey the full distinction between the behaviors of ergodic, mixing, and exact transformations. To remedy this we present the first six successive iterates of a random distribution of 1000 points in the set $X = [0,1] \times [0,1]$ by the ergodic transformation

$$S(x,y) = (\sqrt{2} + x, \sqrt{3} + y) \qquad (\mathrm{mod}\ 1) \tag{4.3.3}$$

in Figure 4.3.3; by the mixing transformation

$$S(x,y) = (x + y, x + 2y) \qquad (\mathrm{mod}\ 1) \tag{4.3.4}$$

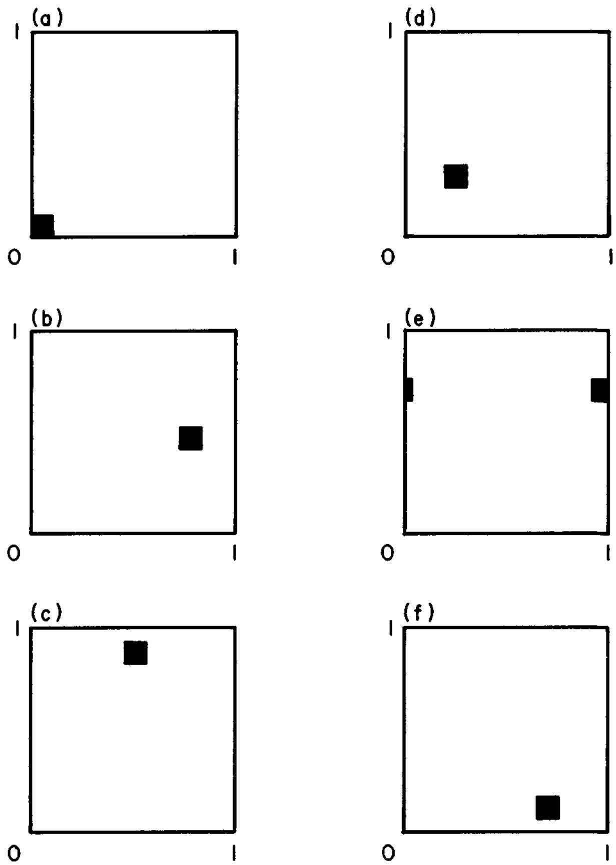


Figure 4.3.3. Successive iterates of a random distribution of 1000 points in $[0, 0.1] \times [0, 0.1]$ by the ergodic transformation (4.3.3). Note how the distribution moves about in the space $[0, 1] \times [0, 1]$.

in Figure 4.3.4; and by the exact transformation

$$S(x, y) = (3x + y, x + 3y) \qquad (\text{mod } 1) \tag{4.3.5}$$

in Figure 4.3.5. Techniques to prove these assertions will be developed in the next two chapters.

4.4 Using the Frobenius–Perron and Koopman operators for classifying transformations

The concepts developed in the previous two sections for classifying various degrees of irregular behaviors (ergodicity, mixing, and exactness) were stated in terms of the behavior of sequences of sets. The proof of ergodicity, mixing, or

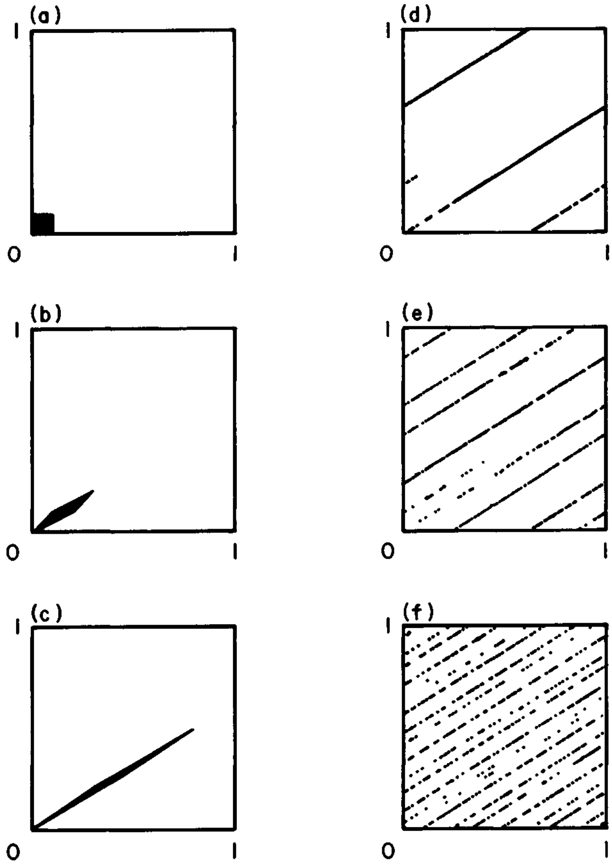


Figure 4.3.4. The effect of the mixing transformation (4.3.4) on the same initial distribution of points used in Figure 4.3.3.

exactness using these definitions is difficult. Indeed, in all the examples we gave to illustrate these concepts, no rigorous proofs were ever given, although it is possible to do so.

In this section we reformulate the concepts of ergodicity, mixing, and exactness in terms of the behavior of sequences of iterates of Frobenius–Perron and Koopman operators and show how they can be used to determine whether a given transformation S with an invariant measure is ergodic, mixing, or exact. The techniques of this chapter rely heavily on the notions of Cesaro, weak and strong convergences, which were developed in Section 2.3.

We will first state and prove the main theorem of this section and then show its utility.

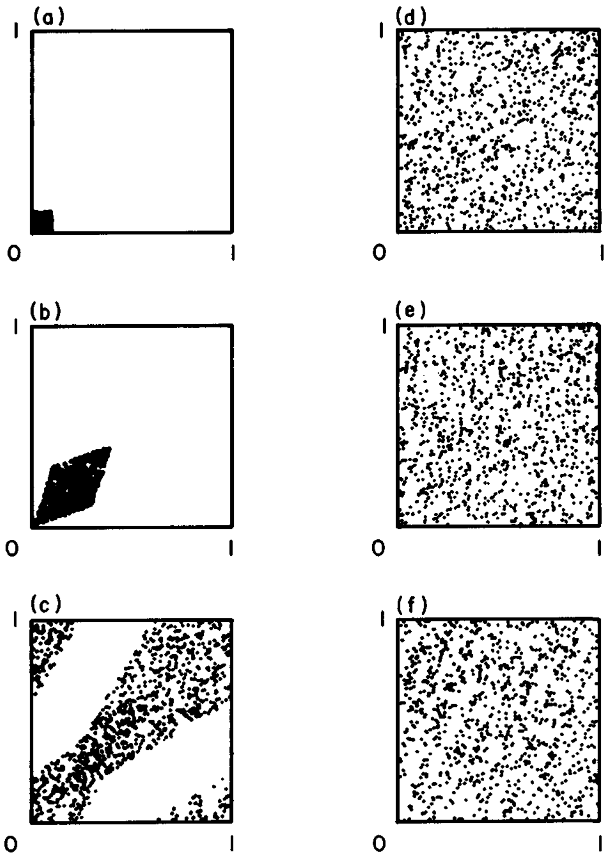


Figure 4.3.5. Successive applications of the exact transformation [equation (4.3.5)]. Note the rapid spread of the initial distribution of points throughout the phase space.

Theorem 4.4.1. 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. Then

(a) S is ergodic if and only if the sequence $\{P^n f\}$ is Cesaro convergent to 1 for all $f \in D$;
(b) S is mixing if and only if $\{P^n f\}$ is weakly convergent to 1 for all $f \in D$;
(c) S is exact if and only if $\{P^n f\}$ is strongly convergent to 1 for all $f \in D$.

Before giving the proof of Theorem 4.4.1, we note that, since P is linear, convergence of $\{P^n f\}$ to 1 for $f \in D$ is equivalent to the convergence of $\{P^n f\}$ to $\langle f, 1\rangle$ for every $f \in L^1$. This observation is, of course, valid for all types of convergence: weak, strong, and weak Cesaro. Thus we may restate Theorem 4.4.1 in the equivalent form.

Corollary 4.4.1. Under the assumptions of Theorem 4.4.1, the following equivalences hold:

(a) S is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle P^k f, g\rangle = \langle f, 1\rangle\langle 1, g\rangle \qquad \text{for } f \in L^1,\ g \in L^\infty;$$

(b) S is mixing if and only if

$$\lim_{n\to\infty} \langle P^n f, g\rangle = \langle f, 1\rangle\langle 1, g\rangle \qquad \text{for } f \in L^1,\ g \in L^\infty;$$

(c) S is exact if and only if

$$\lim_{n\to\infty} \|P^n f - \langle f, 1\rangle\| = 0 \qquad \text{for } f \in L^1.$$

Proof of Theorem 4.4.1: The proof of part (a) follows easily from Corollary 5.2.3.

Next consider the mixing portion of the theorem. Assume S is mixing, which, by definition, means

$$\lim_{n\to\infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B) \qquad \text{for all } A, B \in \mathcal{A}.$$

The mixing condition can be rewritten in integral form as

$$\lim_{n\to\infty} \int_X 1_A(x) 1_B(S^n(x))\mu(dx) = \int_X 1_A(x)\mu(dx) \int_X 1_B(x)\mu(dx).$$

By applying the definitions of the Koopman operator and the scalar product to this equation, we obtain

$$\lim_{n\to\infty} \langle 1_A, U^n 1_B\rangle = \langle 1_A, 1\rangle\langle 1, 1_B\rangle. \tag{4.4.1}$$

Since the Koopman operator is adjoint to the Frobenius–Perron operator, equation (4.4.1) may be rewritten as

$$\lim_{n\to\infty} \langle P^n 1_A, 1_B\rangle = \langle 1_A, 1\rangle\langle 1, 1_B\rangle$$

or

$$\lim_{n\to\infty} \langle P^n f, g\rangle = \langle f, 1\rangle\langle 1, g\rangle$$

for $f = 1_A$ and $g = 1_B$. Since this relation holds for characteristic functions it must also hold for the simple functions

$$f = \sum_i \lambda_i 1_{A_i} \quad \text{and} \quad g = \sum_i \sigma_i 1_{B_i}.$$

Further, every function $g \in L^\infty$ is the uniform limit of simple functions $g_k \in L^\infty$, and every function $f \in L^1$ is the strong (in L^1 norm) limit of a sequence of simple functions $f_k \in L^1$. Obviously,

$$\begin{aligned} |\langle P^n f, g\rangle - \langle f, 1\rangle\langle 1, g\rangle| \leq\ & |\langle P^n f, g\rangle - \langle P^n f_k, g_k\rangle| \\ & + |\langle P^n f_k, g_k\rangle - \langle f_k, 1\rangle\langle 1, g_k\rangle| \\ & + |\langle f_k, 1\rangle\langle 1, g_k\rangle - \langle f, 1\rangle\langle 1, g\rangle| . \end{aligned} \tag{4.4.2}$$

If $\|f_k - f\| \leq \varepsilon$ and $\|g_k - g\|_{L^\infty} \leq \varepsilon$, then the first and last terms on the right-hand side of (4.4.2) satisfy

$$\begin{aligned} &|\langle P^n f, g\rangle - \langle P^n f_k, g_k\rangle| \\ &\qquad \leq |\langle P^n f, g\rangle - \langle P^n f_k, g\rangle| + |\langle P^n f_k, g\rangle - \langle P^n f_k, g_k\rangle| \\ &\qquad \leq \varepsilon\|g\|_{L^\infty} + \varepsilon\|f_k\| \leq \varepsilon(\|g\|_{L^\infty} + \|f\| + \varepsilon) \end{aligned}$$

and analogously

$$|\langle f_k, 1\rangle\langle 1, g_k\rangle - \langle f, 1\rangle\langle 1, g\rangle| \leq \varepsilon(\|g\|_{L^\infty} + \|f\| + \varepsilon) .$$

Thus these terms are arbitrarily small for small ε. Finally, for fixed k the middle term of (4.4.2),

$$|\langle P^n f_k, g_k\rangle - \langle f_k, 1\rangle\langle 1, g_k\rangle|$$

converges to zero as $n \to \infty$, which shows that the right-hand side of inequality (4.4.2) can be as small as we wish it to be for large n. This completes the proof that mixing implies the convergence of $\langle P^n f, g\rangle$ to $\langle f, 1\rangle\langle 1, g\rangle$ for all $f \in L^1$ and $g \in L^\infty$. Conversely, this convergence implies the mixing condition (4.4.1) if we set $f = 1_A$ and $g = 1_B$.

Lastly, we show that the strong convergence of $\{P^n f\}$ to $\langle f, 1\rangle$ implies exactness. Assume $\mu(A) > 0$ and define

$$f_A(x) = (1/\mu(A))1_A(x) .$$

Clearly, f_A is a density. If the sequence $\{r_n\}$ is defined by

$$r_n = \|P^n f_A - 1\| ,$$

then it is also clear that the sequence is convergent to zero. By the definition of r_n, we have

$$\begin{aligned} \mu(S^n(A)) &= \int_{S^n(A)} \mu(dx) \\ &= \int_{S^n(A)} P^n f_A(x)\mu(dx) - \int_{S^n(A)} (P^n f_A(x) - 1)\mu(dx) \\ &\geq \int_{S^n(A)} P^n f_A(x)\mu(dx) - r_n . \end{aligned} \tag{4.4.3}$$

From the definition of the Frobenius–Perron operator, we have

$$\int_{S^n(A)} P^n f_A(x)\mu(dx) = \int_{S^{-n}(S^n(A))} f_A(x)\mu(dx)$$

and, since $S^{-n}(S^n(A))$ contains A, the last integral is equal to 1. Thus inequality (4.4.3) gives

$$\mu(S^n(A)) \geqslant 1 - r_n,$$

which completes the proof that the strong convergence of $\{P^n f\}$ to $\langle f, 1\rangle$ implies exactness. ■

We omit the proof of the converse (that exactness implies the strong convergence of $\{P^n f\}$ to $\langle f, 1\rangle$) since we will never use this fact and its proof is based on quite different techniques (see Lin [1971]).

Because the Koopman and Frobenius–Perron operators are adjoint, it is possible to rewrite conditions (a) and (b) of Corollary 4.4.1 in terms of the Koopman operator. The advantage of such a reformulation lies in the fact that the Koopman operator is much easier to calculate than the Frobenius–Perron operator. Unfortunately, this reformulation cannot be extended to condition (c) for exactness of Corollary 4.4.1 since it is not expressed in terms of a scalar product.

Thus, from Corollary 4.4.1, the following proposition can easily be stated.

Proposition 4.4.1. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space, $S: X \to X$ a measure-preserving transformation, and U the Koopman operator corresponding to S. Then

(a) S is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle f, U^k g\rangle = \langle f, 1\rangle\langle 1, g\rangle \qquad \text{for } f \in L^1,\ g \in L^\infty;$$

(b) S is mixing if and only if

$$\lim_{n\to\infty} \langle f, U^n g\rangle = \langle f, 1\rangle\langle 1, g\rangle \qquad \text{for } f \in L^1,\ g \in L^\infty.$$

Proof: The proof of this proposition is trivial since, according to equation (3.3.4), we have

$$\langle f, U^n g\rangle = \langle P^n f, g\rangle \qquad \text{for } f \in L^1,\ g \in L^\infty,\ n = 1, 2, \ldots,$$

which shows that conditions (a) and (b) of Corollary 4.4.1 and Proposition 4.4.1 are identical. ■

Remark 4.4.1. We stated Theorem 4.4.1 and Corollary 4.4.1 in terms of L^1 and

L^∞ spaces to underline the role of the Frobenius–Perron operator as a transformation of densities. The same results can be proved using adjoint spaces L^p and $L^{p'}$ instead of L^1 and L^∞, respectively. Moreover, when verifying conditions (a) through (c) of Theorem 4.4.1 and Corollary 4.4.1, or conditions (a) and (b) of Proposition 4.4.1, it is not necessary to check for their validity for all $f \in L^p$ and $g \in L^{p'}$. Due to special properties of the operators P and U, which are linear contractions, it is sufficient to check these conditions for f and g belonging to linearly dense subsets of L^p and $L^{p'}$, respectively (see Section 2.3). □

Example 4.4.1. In Example 4.2.2 we showed that the rotational transformation

$$S(x) = x + \phi \qquad (\text{mod } 2\pi)$$

is not ergodic when $\phi/2\pi$ is rational. Here we prove that it is ergodic when $\phi/2\pi$ is irrational.

It is straightforward to show that S preserves the Borel measure μ and the normalized measure $\mu/2\pi$. We take as our linearly dense set in $L^{p'}([0, 2\pi])$ that consisting of the functions $\{\sin kx, \cos lx\colon k, l = 0, 1, \ldots\}$. We will show that, for each function g belonging to this set,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} U^k g(x) = \langle 1, g\rangle \tag{4.4.4}$$

uniformly for all x, thus implying that condition (a) of Proposition 4.4.1 is satisfied for all f. To simplify the calculations, note that

$$\sin kx = \frac{e^{ikx} - e^{-ikx}}{2i}, \qquad \cos kx = \frac{e^{ikx} + e^{-ikx}}{2}$$

where $i = \sqrt{-1}$. Consequently, it is sufficient to verify (4.4.4) only for $g(x) = \exp(ikx)$ with k an arbitrary (not necessarily positive) integer.

We have, for $k \neq 0$,

$$U^l g(x) = g(S^l(x)) = e^{ik(x+l\phi)},$$

so that

$$u_n(x) = \frac{1}{n} \sum_{l=0}^{n-1} U^l g(x)$$

obeys

$$\begin{aligned} u_n(x) &= \frac{1}{n} \sum_{l=0}^{n-1} e^{ik(x+l\phi)} \\ &= \frac{1}{n} e^{ikx} \frac{e^{ink\phi} - 1}{e^{ik\phi} - 1}. \end{aligned}$$

and

$$\|u_n(x)\|_{L^2} \leqslant \frac{1}{n|e^{ik\phi} - 1|}\left\{\int_0^{2\pi} |e^{ikx}[e^{ink\phi} - 1]|^2 \frac{dx}{2\pi}\right\}^{1/2}$$

$$\leqslant \frac{2}{n|e^{ik\phi} - 1|}.$$

Thus $u_n(x)$ converges in L^2 to zero. Also, however, with our choice of $g(x)$,

$$\langle 1, g\rangle = \int_0^{2\pi} e^{ikx}\frac{dx}{2\pi} = \frac{1}{ik}[e^{2\pi ik} - 1] = 0$$

and condition (a) of Proposition 4.4.1 for ergodicity is satisfied when $k \neq 0$.

When $k = 0$ the calculation is even simpler, since $g(x) \equiv 1$ and thus

$$u_n(x) \equiv 1 .$$

Noting also that

$$\langle 1, g\rangle = \int_0^{2\pi} \frac{dx}{2\pi} \equiv 1 ,$$

we have again that $u_n(x)$ converges to $\langle 1, g\rangle$. □

Example 4.4.2. In this example we demonstrate the exactness of the r-adic transformation

$$S(x) = rx \qquad (\mathrm{mod}\ 1) .$$

From Corollary 4.4.1 it is sufficient to demonstrate that $\{P^n f\}$ converges strongly to $\langle f, 1\rangle$ for f in a linearly dense set in $L^p([0,1])$. We take that linearly dense set to be the set of continuous functions.

From equation (1.2.13) we have

$$Pf(x) = \frac{1}{r}\sum_{i=0}^{r-1} f\left(\frac{i}{r} + \frac{x}{r}\right),$$

and thus by induction

$$P^n f(x) = \frac{1}{r^n}\sum_{k=0}^{r^n-1} f\left(\frac{i}{r^n} + \frac{x}{r^n}\right).$$

However, in the limit as $n \to \infty$, the right-hand side of this equation approaches the Riemann integral of f over $[0,1]$, that is,

$$\lim_{n\to\infty} P^n f(x) = \int_0^1 f(s)\,ds, \qquad \text{uniformly in } x,$$

which, by definition, is just $\langle f, 1\rangle$. Thus the condition for exactness is fulfilled. □

Example 4.4.3. Here we show that the Anosov diffeomorphism

$$S(x,y) = (x + y, x + 2y) \qquad (\text{mod } 1)$$

is mixing. For this, from Proposition 4.4.1, it is sufficient to show that $U^n g(x,y) \equiv g(S^n(x,y))$ converges weakly to $\langle 1, g\rangle$ for g in a linearly dense set in $L^{p'}([0,1] \times [0,1])$.

Observe that for $g(x,y)$ periodic in x and y with period 1, $g(S(x,y)) = g(x + y, x + 2y)$, $g(S^2(x,y)) = g(2x + 3y, 3x + 5y)$, and so on. By induction we easily find that

$$U^n g(x,y) = g(a_{2n-2}x + a_{2n-1}y, a_{2n-1}x + a_{2n}y),$$

where the a_n are the Fibonacci numbers given by $a_0 = a_1 = 1$, $a_{n+1} = a_n + a_{n-1}$. Thus, if we take $g(x,y) = \exp[2\pi i(kx + ly)]$ and $f(x,y) = \exp[-2\pi i(px + qy)]$, then we have

$$\langle f, U^n g\rangle = \int_0^1 \int_0^1 \exp\{2\pi i[ka_{2n-2} + la_{2n-1} - p)x + (ka_{2n-1} + la_{2n} - q)y]\}\,dx\,dy,$$

and it is straightforward to show that

$$\langle f, U^n g\rangle = \begin{cases} 1 & \text{if } (ka_{2n-2} + la_{2n-1} - p) = (ka_{2n-1} + la_{2n} - q) = 0 \\ 0 & \text{otherwise}. \end{cases}$$

Now we show that for large n either

$$ka_{2n-2} + la_{2n-1} - p \quad \text{or} \quad ka_{2n-1} + la_{2n} - q$$

is different from zero if at least one of k, l, p, q is different from zero. If $k = l = 0$ but $p \neq 0$ or $q \neq 0$ this is obvious. We may suppose that either k or l is not zero. Assume $k \neq 0$ and that $ka_{2n-2} + la_{2n-1} - p = 0$ for infinitely many n. Thus,

$$k\frac{a_{2n-2}}{a_{2n-1}} + l - \frac{p}{a_{2n-1}} = 0.$$

It is well known [Hardy and Wright, 1959] that

$$\lim_{n\to\infty} \frac{a_{2n-2}}{a_{2n-1}} = \frac{2}{1 + \sqrt{5}} \quad \text{and} \quad \lim_{n\to\infty} a_n = \infty,$$

hence

$$\lim_{n\to\infty}\left[k\left(\frac{a_{2n-2}}{a_{2n-1}}\right) + l - \frac{p}{a_{2n-1}}\right] = \frac{2k}{1 + \sqrt{5}} + l.$$

However, this limit can never be zero because k and l are integers. Thus $ka_{2n-2} + la_{2n-1} - p \neq 0$ for large n. Therefore, for large n,

$$\langle f, U^n g\rangle = \begin{cases} 1 & \text{if } k = l = p = q = 0 \\ 0 & \text{otherwise}. \end{cases}$$

But

$$\langle 1, g\rangle = \int_0^1 \int_0^1 \exp[2\pi i(kx + ly)]\,dx\,dy$$
$$= \begin{cases} 1 & k = l = 0 \\ 0 & k \neq 0 \quad \text{or} \quad l \neq 0, \end{cases}$$

so that

$$\langle f, 1\rangle\langle 1, g\rangle = \int_0^1 \int_0^1 \langle 1, g\rangle \exp[-2\pi i(px + qy)]\,dx\,dy$$
$$= \begin{cases} \langle 1, g\rangle & \text{if } p = q = 0 \\ 0 & \text{if } p \neq 0 \quad \text{or} \quad q \neq 0 \end{cases}$$
$$= \begin{cases} 1 & \text{if } k = l = p = q = 0 \\ 0 & \text{otherwise}. \end{cases}$$

Thus

$$\langle f, U^n g\rangle = \langle f, 1\rangle\langle 1, g\rangle$$

for large n and, as a consequence, $\{U^n g\}$ converges weakly to $\langle 1, g\rangle$. Therefore, mixing of the Anosov diffeomorphism is demonstrated. □

In this chapter we have shown how the study of ergodicity, mixing, and exactness for transformations S can be greatly facilitated by the use of the Frobenius–Perron operator P corresponding to S (cf. Theorem 4.4.1 and Corollary 4.4.1). Since the Frobenius–Perron operator is a special type of Markov operator, there is a certain logic to extending the notions of ergodicity, mixing, and exactness for transformations to Markov operators in general. Thus, we close this section with the following definition.

Definition 4.4.1. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space and P: $L^1(X, \mathcal{A}, \mu) \to L^1(X, \mathcal{A}, \mu)$ a Markov operator with stationary density 1, that is, $P1 = 1$. Then we say:

(a) The operator P is **ergodic** if $\{P^n f\}$ is Cesaro convergent to 1 for all $f \in D$;
(b) The operator P is **mixing** if $\{P^n f\}$ is weakly convergent to 1 for all $f \in D$; and
(c) The operator P is **exact** if $\{P^n f\}$ is strongly convergent to 1 for all $f \in D$.

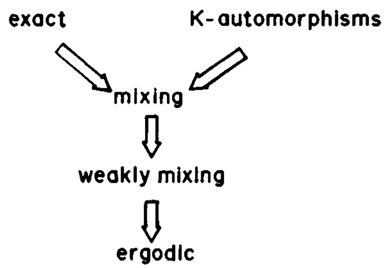
4.5 Kolmogorov automorphisms

Until now we have considered three types of transformations exhibiting gradually stronger chaotic properties: ergodicity, mixing, and exactness. This is not a complete list of possible behaviors. These three types are probably the most important, but it is possible to find some intermediate types and some new unexpected connections between them. For example, between ergodic and mixing transformations, there is a class of **weakly mixing** transformations that, by definition, are measure preserving [on a normalized measure space $(X, \mathcal{A}, \mu)$] and satisfy the condition

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(A \cap S^{-k}(B)) - \mu(A)\mu(B)| = 0 \qquad \text{for } A, B \in \mathcal{A}.$$

It is not easy to construct an example of a weakly mixing transformation that is not mixing. Interesting comments on this problem can be found in Brown [1976].

However, Kolmogorov automorphisms, which are invertible and therefore cannot be exact, are stronger than mixing. As we will see later, to some extent they are parallel to exact transformations. Schematically this situation can be visualized as follows:



where K-automorphism is the usual abbreviation for a Kolmogorov automorphism and the arrows indicate that the property above implies the one below. Before giving the precise definition of K-automorphisms, we introduce two simple notations.

If $S: X \to X$ is a given transformation and $\mathcal{A}$ is a collection of subsets of X (e.g., a σ-algebra), then $S(\mathcal{A})$ denotes the collection of sets of the form $S(A)$ for $A \in \mathcal{A}$, and $S^{-1}(\mathcal{A})$ the collection of $S^{-1}(A)$ for $A \in \mathcal{A}$. More generally,

$$S^n(\mathcal{A}) = \{S^n(A): A \in \mathcal{A}\}, \qquad n = 0, \pm 1, \pm 2, \ldots$$

Definition 4.5.1. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space and let $S: X \to X$ be an invertible transformation such that S and S^{-1} are measurable and measure preserving. The transformation S is called a **K-automorphism** if there exists a σ-algebra $\mathcal{A}_0 \subset \mathcal{A}$ such that the following three conditions are satisfied:

(i) $S^{-1}(\mathcal{A}_0) \subset \mathcal{A}_0$;

(ii) the σ-algebra

$$\bigcap_{n=0}^{\infty} S^{-n}(\mathcal{A}_0) \tag{4.5.1}$$

is **trivial**, that is, it contains only sets of measure 1 or 0; and

(iii) the smallest σ-algebra containing

$$\bigcup_{n=0}^{\infty} S^{n}(\mathcal{A}_0) \tag{4.5.2}$$

is identical to $\mathcal{A}$.

The word automorphism comes from algebra and in this case it means that the transformation S is invertible and measure preserving (analogously the word endomorphism is used for measure preserving but not necessarily invertible transformations).

Example 4.5.1. The baker transformation is a K-automorphism. As $\mathcal{A}_0$ we can take all the sets of the form

$$\mathcal{A}_0 = \{A \times [0,1]: A \subset [0,1],\ A \text{ is a Borel set}\}.$$

It is easy to verify condition (i) of Definition 4.5.1. Thus, if $B = A \times [0,1]$, then $B_1 = S^{-1}(B)$ has the form $B_1 = A_1 \times [0,1]$, where

$$A_1 = \tfrac{1}{2}A \cup (\tfrac{1}{2} + \tfrac{1}{2}A) \tag{4.5.3}$$

and thus condition (i) is satisfied. From this follows a hint of how to prove condition (ii). Namely, from (4.5.3) it follows that the basis A_1 of the set $B_1 = S^{-1}(B)$ is the union of two sets of equal measure that are contained in the intervals $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$, respectively. Furthermore, set $B_2 = S^{-2}(B)$ has the form $A_1 \times [0,1]$ and its basis A_2 is the union of four sets of equal measure contained in the intervals $[0,\frac{1}{4}], \dots, [\frac{3}{4},1]$. Finally, every set B_∞ belonging to the σ-algebra (4.5.1) is of the form $A_\infty \times [0,1]$ and A_∞ has the property that for each integer n the measure of the intersection of A_∞ with $[k/2^n, (k+1)/2^n]$ does not depend on k. From this it is easy to show that the measure of the intersection of A_∞ with $[0,x]$ is a linear function of x or

$$\int_0^x 1_{A_\infty}(y)\,dy = cx,$$

where c is a constant. Differentiation gives

$$1_{A_\infty}(x) = c \quad \text{a.e.} \qquad \text{for } 0 \leqslant x \leqslant 1 .$$

Since 1_{A_∞} is a characteristic function, either $c = 1$ or $c = 0$. In the first case, A_∞ as well as B_∞ have measure 1, and if $c = 0$, then A_∞ and B_∞ have measure 0. Thus condition (ii) is verified.

To verify (iii), observe that $\mathscr{A}_0 \cup S(\mathscr{A}_0)$ contains not only sets of the form $A \times [0,1]$ but also the sets of the form $A \times [0,\frac{1}{2}]$ and $A \times [\frac{1}{2},1]$. Further, $\mathscr{A}_0 \cup S(\mathscr{A}_0) \cup S^2(\mathscr{A}_0)$ also contains the sets $A \times [0,\frac{1}{4}], \ldots, A \times [\frac{3}{4},1]$ and so on. Thus, by using the sets from the family (4.5.2), we can approximate every rectangle contained in $[0,1] \times [0,1]$. Consequently, the smallest σ-algebra containing (4.5.2) is the σ-algebra of Borel sets. □

Example 4.5.2. The baker transformation considered in the previous example has an important geometrical property. At every point it is contracting in one direction and expanding in the orthogonal one. The transformation

$$S(x,y) = (x + y, x + 2y) \qquad (\mathrm{mod}\ 1)$$

considered in Example 4.1.4 has the same property. As we have observed the Jacobian of S has two eigenvalues λ_1, λ_2 such that $0 < \lambda_1 < 1 < \lambda_2$. To these eigenvalues correspond the eigenvectors

$$\xi_1 = (1, \tfrac{1}{2} - \tfrac{1}{2}\sqrt{5}), \qquad \xi_2 = (1, \tfrac{1}{2} + \tfrac{1}{2}\sqrt{5}) .$$

Thus S contracts in the direction ξ_1 and expands in the direction ξ_2. With this fact it can be verified that S is also a K-automorphism. The construction of $\mathscr{A}_0$ is related with vectors ξ_1 and ξ_2; that is, $\mathscr{A}_0$ may be defined as a σ-algebra generated by a class of rectangles with sides parallel to vectors ξ_1 and ξ_2. The precise definition of $\mathscr{A}_0$ requires some technical details, which can be found in an article by Arnold and Avez [1968]. □

As we observed in Remark 4.1.4, the first coordinate in the baker transformation is transformed independently of the second, which is the dyadic transformation. The baker transformation is a K-automorphism and the dyadic is exact. This fact is not a coincidence. It may be shown that every exact transformation is, in some sense, a restriction of a K-automorphism. To make this statement precise we need the following definition.

Definition 4.5.2. Let $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ be two normalized measure spaces and let $S: X \to X$ and $T: Y \to Y$ be two measure-preserving transformations. If there exists a transformation $F: Y \to X$ that is also measure preserving, namely

$$\nu(F^{-1}(A)) = \mu(A) \qquad \text{for } A \in \mathscr{A}$$

and such that $S \circ F = F \circ T$, then S is called a **factor** of T.

The situation described by Definition 4.5.2 can be visualized by the diagram

$$
\begin{tikzcd}
Y \arrow[r, "T"] \arrow[d, "F"] & Y \arrow[d, "F"] \\
X \arrow[r, "S"] & X
\end{tikzcd}
\tag{4.5.4}
$$

and the condition $S \circ F = F \circ T$ may be expressed by saying that the diagram (4.5.4) **commutes**. We have the following theorem due to Rochlin [1961].

Theorem 4.5.1. Every exact transformation is a factor of a K-automorphism.

The relationship between K-automorphisms and mixing transformations is much simpler; it is given by the following theorem.

Theorem 4.5.2. Every K-automorphism is mixing.

The proofs and more information concerning K-automorphisms can be found in the books by Walters [1982] and by Parry [1981].