5

The asymptotic properties of densities

The preceding chapter was devoted to an examination of the various degrees of "chaotic" behavior (ergodicity, mixing, and exactness) that measure-preserving transformations may display. In particular, we saw the usefulness of the Koopman and Frobenius–Perron operators in answering these questions.

Theorem 4.1.1 reduced the problem of finding an invariant measure to one of finding solutions to the equation $Pf = f$. Perhaps the most obvious, although not the simplest, way to find these solutions is to pick an arbitrary $f \in D$ and examine the sequence $\{P^n f\}$ of successive iterations of f by the Frobenius–Perron operator. If $\{P^n f\}$ converges to f_*, then clearly $\{P^{n+1} f\} = \{P(P^n f)\}$ converges simultaneously to f_* and Pf_* and we are done. However, to prove that $\{P^n f\}$ converges (weakly or strongly) to a function f_* is difficult.

In this chapter we first examine the convergence of the sequence $\{A_n f\}$ of averages defined by

$$A_n f = \frac{1}{n} \sum_{k=0}^{n-1} P^k f$$

and show how this may be used to demonstrate the existence of a stationary density of P. We then show that under certain conditions $\{P^n f\}$ can display a new property, namely, asymptotic periodicity. Finally, we introduce the concept of asymptotic stability for Markov operators, which is a generalization of exactness for Frobenius–Perron operators. We then show how the lower-bound function technique may be used to demonstrate asymptotic stability. This technique is used throughout the remainder of the book.

5.1 Weak and strong precompactness

In calculus one of the most important observations, originally due to Weierstrass, is that any bounded sequence of numbers contains a convergent subsequence. This observation can be extended to spaces of any finite dimension. Unfortunately, for more complicated objects, such as densities, this is not the case. One example is

$$f_n(x) = n1_{[0,1/n]}(x), \qquad 0 \leqslant x \leqslant 1$$

which is bounded in L^1 norm, that is, $\|f_n\| = 1$, but which does not converge weakly or strongly in $L^1([0,1])$ to any density. In fact, as $n \to \infty$, $f_n(x) \to \delta(x)$, the Dirac delta function that is supported on a single point, $x = 0$.

One of the great achievements in mathematical analysis was the discovery of sufficient conditions for the existence of convergent subsequences of functions, which subsequently found applications in the calculus of variations, optimal control theory, and proofs for the existence of solutions to ordinary and partial differential equations and integral equations.

To make these comments more precise we introduce the following definitions. Let $(X, \mathscr{A}, \mu)$ be a measure space and $\mathscr{F}$ a set of functions in L^p.

Definition 5.1.1. The set $\mathscr{F}$ will be called **strongly precompact** if every sequence of functions $\{f_n\}$, $f_n \in \mathscr{F}$, contains a strongly convergent subsequence $\{f_{\alpha_n}\}$ that converges to an $\bar{f} \in L^p$.

Remark 5.1.1. The prefix "pre-" is used because we take $\bar{f} \in L^p$ rather than $\bar{f} \in \mathscr{F}$. □

Definition 5.1.2. The set $\mathscr{F}$ will be called **weakly precompact** if every sequence of functions $\{f_n\}$, $f_n \in \mathscr{F}$, contains a weakly convergent subsequence $\{f_{\alpha_n}\}$ that converges to an $\bar{f} \in L^p$.

Remark 5.1.2. These two definitions are often applied to sets consisting simply of sequences of functions. In this case the precompactness of $\mathscr{F} = \{f_n\}$ simply means that every subsequence $\{f_{\alpha_n}\}$ contains a convergent subsequence. □

Remark 5.1.3. From the definitions it immediately follows that any subset of a weakly or strongly precompact set is itself weakly or strongly precompact. □

There are several simple and general criteria useful for demonstrating the weak precompactness of sets in L^p [see Dunford and Schwartz, 1957]. The three we will have occasion to use are as follows:

1 Let $g \in L^1$ be a nonnegative function. Then the set of all functions $f \in L^1$ such that

$$|f(x)| \leqslant g(x) \qquad \text{for } x \in X \text{ a.e.} \tag{5.1.1}$$

is weakly precompact in L^1.

2 Let $M > 0$ be a positive number and $p > 1$ be given. If $\mu(X) < \infty$, then the set of all functions $f \in L^1$ such that

$$\|f\|_{L^p} \leqslant M \tag{5.1.2}$$

is weakly precompact in L^1.

3 A set of functions $\mathcal{F} \subset L^1$, $\mu(X) < \infty$, is weakly precompact if and only if:

(a) There is an $M < \infty$ such that $\|f\| \leqslant M$ for all $f \in \mathcal{F}$; and

(b) For every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\int_A |f(x)|\mu(dx) < \varepsilon \qquad \text{if } \mu(A) < \delta \text{ and } f \in \mathcal{F}.$$

Remark 5.1.4. If the measure is not finite these two conditions must be supplemented by

(c) For every $\varepsilon > 0$ there is a set $B \in \mathcal{A}$, $\mu(B) < \infty$, such that

$$\int_{X \setminus B} |f(x)|\mu(dx) < \varepsilon. \quad \square$$

Strong precompactness is generally more difficult to demonstrate than weak precompactness. One of the simplest criteria, which we present only for one-dimensional spaces, is as follows:

4 Let $\mathcal{F}$ be a set of functions defined on a bounded interval Δ of the real line. $\mathcal{F}$ is strongly precompact in $L^1(\Delta)$ if and only if:

(a) There exists a constant $M > 0$ independent of f such that

$$\|f\| \leqslant M \qquad \text{for all } f \in \mathcal{F}; \tag{5.1.3a}$$

(b) For all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\int_\Delta |f(x+h) - f(x)|\, dx < \varepsilon \tag{5.1.3b}$$

for all $f \in \mathcal{F}$ and all h such that $|h| < \delta$. To ensure that this integral is well defined we assume $f(x+h) - f(x) = 0$ for $x + h \notin \Delta$.

Remark 5.1.5. This necessary and sufficient condition for strong precompactness is valid for unbounded intervals Δ if, in addition, for every $\varepsilon > 0$ there is an $r > 0$ such that

$$\int_{|x| \geqslant r} |f(x)|\, dx < \varepsilon \qquad \text{for all } f \in \mathcal{F}. \quad \square \tag{5.1.4}$$

Remark 5.1.6. In practical situations it is often difficult to verify inequality (5.1.3b). However, if the functions $f \in \mathcal{F}$ have uniformly bounded derivatives, that is, if there is a constant K such that $|f'(x)| \leqslant K$, then the condition is automatically satisfied. To see this, note that

$$|f(x+h)-f(x)| \leq Kh$$

implies

$$\int_\Delta |f(x+h)-f(x)|\,dx \leq Kh\mu(\Delta)$$

and thus if, for a given ε, we pick

$$\delta = \varepsilon / K\mu(\Delta)$$

the condition (5.1.3b) is satisfied. Clearly this will not work for unbounded intervals because for $\mu(\Delta) \to \infty$, $\delta \to 0$. □

To close this section we state the following corollary.

Corollary 5.1.1. For every $f \in L^1$, Δ bounded or not,

$$\lim_{h\to 0} \int_\Delta |f(x+h)-f(x)|\,dx = 0. \tag{5.1.5}$$

Proof: To see this note that the set $\{f\}$ consisting of only one function f is obviously strongly precompact since the sequence $\{f, f, \ldots\}$ is always convergent. Thus equation (5.1.5) follows from the foregoing condition (4b) for strong precompactness. ■

5.2 Properties of the averages $A_n f$

In this section we assume a measure space $(X, \mathcal{A}, \mu)$ and a Markov operator $P: L^1 \to L^1$. We are going to demonstrate some simple properties of the averages defined by

$$A_n f = \frac{1}{n}\sum_{k=0}^{n-1} P^k f, \qquad \text{for } f \in L^1. \tag{5.2.1}$$

We then state and prove a special case of the Kakutani–Yosida abstract ergodic theorem as well as two corollaries to that theorem.

Proposition 5.2.1. For all $f \in L^1$,

$$\lim_{n\to\infty} \|A_n f - A_n Pf\| = 0.$$

Proof: By the definition of $A_n f$ (5.2.1) we have

$$A_n f - A_n Pf = (1/n)(f - P^n f)$$

and thus

$$\|A_n f - A_n Pf\| \leqslant (1/n)(\|f\| + \|P^n f\|).$$

Since it is an elementary property of Markov operators that $\|P^n f\| \leqslant \|f\|$, we have

$$\|A_n f - A_n Pf\| \leqslant (2/n)\|f\| \to 0$$

as $n \to \infty$, which completes the proof. ■

Proposition 5.2.2. If, for $f \in L^1$, there is a subsequence $\{A_{\alpha_n} f\}$ of the sequence $\{A_n f\}$ that converges weakly to $f_* \in L^1$, then $Pf_* = f_*$.

Proof: First, since $PA_{\alpha_n} f = A_{\alpha_n} Pf$, then $\{A_{\alpha_n} Pf\}$ converges weakly to Pf_*. Since $\{A_{\alpha_n} Pf\}$ has the same limit as $\{A_{\alpha_n} f\}$, we have $Pf_* = f_*$. ■

The following theorem is a special case of an abstract ergodic theorem originally due to Kakutani and Yosida (see Dunford and Schwartz [1957]). The usefulness of the theorem lies in the establishment of a simple condition for the existence of a fixed point for a given Markov operator P.

Theorem 5.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a Markov operator. If for a given $f \in L^1$ the sequence $\{A_n f\}$ is weakly precompact, then it converges strongly to some $f_* \in L^1$ that is a fixed point of P, namely, $Pf_* = f_*$. Furthermore, if $f \in D$, then $f_* \in D$, so that f_* is a stationary density.

Proof: Because $\{A_n f\}$ is weakly precompact by assumption, there exists a subsequence $\{A_{\alpha_n} f\}$ that converges weakly to some $f_* \in L^1$. Further, by Proposition 5.2.2, we know $Pf_* = f_*$.

Write $f \in L^1$ in the form

$$f = (f - f_*) + f_* \tag{5.2.2}$$

and assume for the time being that for every $\varepsilon > 0$ the function $f - f_*$ can be written in the form

$$f - f_* = Pg - g + r, \tag{5.2.3}$$

where $g \in L^1$ and $\|r\| < \varepsilon$. Thus, from equations (5.2.2) and (5.2.3), we have

$$A_n f = A_n(Pg - g) + A_n r + A_n f_*.$$

Because $Pf_* = f_*$, $A_n f_* = f_*$, and we obtain

$$\|A_n f - f_*\| = \|A_n(f - f_*)\| \leqslant \|A_n(Pg - g)\| + \|A_n r\|.$$

By Proposition 5.2.1 we know that $\|A_n(Pg - g)\|$ is strongly convergent to zero as $n \to \infty$, and by our assumptions $\|A_n r\| \leqslant \|r\| < \varepsilon$. Thus, for sufficiently large

n, we must have

$$\|A_n f - f_*\| \leq \varepsilon .$$

Since ε is arbitrary, this proves that $\{A_n f\}$ is strongly convergent to f_*.

To show that if $f \in D$, then $f_* \in D$, recall from Definition 3.1.3 that $f \in D$ means that

$$f \geq 0 \quad \text{and} \quad \|f\| = 1 .$$

Therefore $Pf \geq 0$ and $\|Pf\| = 1$ so that $P^n f \geq 0$ and $\|P^n f\| = 1$. As a consequence, $A_n f \geq 0$ and $\|A_n f\| = 1$ and, since $\{A_n f\}$ is strongly convergent to f_*, we must have $f_* \in D$. This completes the proof under the assumption that representation (5.2.3) is possible for every ε.

In proving this assumption, we will use a simplified version of the Hahn–Banach theorem (see Remark 5.2.1). Suppose that for some ε there does not exist an r such that equation (5.2.3) is true. If this were the case, then $f - f_* \notin$ closure$(P - I)L^1(X)$ and, thus, by the Hahn–Banach theorem, there must exist a $g_0 \in L^\infty$ such that

$$\langle f - f_*, g_0 \rangle \neq 0 \tag{5.2.4}$$

and

$$\langle h, g_0 \rangle = 0 \qquad \text{for all } h \in \text{closure}(P - I)L^1(X) .$$

In particular,

$$\langle (P - I)P^j f, g_0 \rangle = 0 .$$

Thus

$$\langle P^{j+1} f, g_0 \rangle = \langle P^j f, g_0 \rangle \qquad \text{for } j = 0, 1, \ldots ,$$

and by induction we must, therefore, have

$$\langle P^j f, g_0 \rangle = \langle f, g_0 \rangle . \tag{5.2.5}$$

As a consequence

$$\frac{1}{n} \sum_{j=0}^{n-1} \langle P^j f, g_0 \rangle = \frac{1}{n} \sum_{j=0}^{n-1} \langle f, g_0 \rangle = \langle f, g_0 \rangle$$

or

$$\langle A_n f, g_0 \rangle = \langle f, g_0 \rangle . \tag{5.2.6}$$

Since $\{A_{\alpha_n} f\}$ was assumed to converge weakly to f_*, we have

$$\lim_{n \to \infty} \langle A_{\alpha_n} f, g_0 \rangle = \langle f_*, g_0 \rangle$$

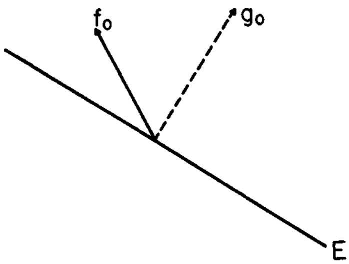

Figure 5.2.1. Diagram showing that, for $f_0 \notin E$, we can find a g_0 such that g_0 is not orthogonal to f_0 but it is orthogonal to all $f \in E$. Since g_0 belongs to $L^{p'}$, but not necessarily to L^p, it is drawn as a dashed line.

and, by (5.2.6),

$$\langle f, g_0 \rangle = \langle f_*, g_0 \rangle ,$$

which gives

$$\langle f - f_*, g_0 \rangle = 0 .$$

However, this result contradicts (5.2.4), and therefore we conclude that the representation (5.2.3) is, indeed, always possible. ■

Remark 5.2.1. The Hahn–Banach theorem is one of the classical results of functional analysis. Although it is customarily stated as a general property of some linear topological spaces (e.g., Banach spaces and locally convex spaces), here we state it for L^p spaces. We need only two concepts. A set $E \subset L^p$ is a **linear subspace** of L^p if $\lambda_1 f_1 + \lambda_2 f_2 \in E$ for all $f_1, f_2 \in E$ and all scalars λ_1, λ_2. A linear subspace is **closed** if $\lim f_n \in E$ for every strongly convergent sequence $\{f_n\} \subset E$. □

Next we state a simple consequence of the Hahn–Banach theorem in the language of L^p spaces [see Dunford and Schwartz, 1957].

Proposition 5.2.3. Let $1 \leq p < \infty$ and let p' be adjoint to p, that is, $(1/p) + (1/p') = 1$ for $p > 1$ and $p' = \infty$ for p = 1. Further, let $E \subset L^p$ be a linear closed subspace. If $f_0 \in L^p$ and $f_0 \notin E$, then there is a $g_0 \in L^{p'}$ such that $\langle f_0, g_0 \rangle \neq 0$ and $\langle f, g_0 \rangle = 0$ for $f \in E$.

Geometrically, this proposition means that, if we have a closed subspace E and a vector $f_0 \notin E$, then we can find another vector g_0 orthogonal to E but not orthogonal to f_0 (see Figure 5.2.1).

Remark 5.2.2. By proving Theorem 5.2.1 we have reduced the problem of demonstrating the existence of a stationary density f_* for the operator P, that is, $Pf_* = f_*$, to the simpler problem of demonstrating the weak precompactness of the sequence $\{A_n f\}$. In the special case that P is a Frobenius–Perron operator this also suffices to demonstrate the existence of an invariant measure. □

There are two simple and useful corollaries to Theorem 5.2.1.

Corollary 5.2.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a Markov operator. If, for some $f \in D$ there is a $g \in L^1$ such that

$$P^n f \leqslant g \tag{5.2.7}$$

for all n, then there is an $f_* \in D$ such that $Pf_* = f_*$, that is, f_* is a stationary density.

Proof: By assumption, $P^n f \leqslant g$ so that

$$0 \leqslant A_n f = \frac{1}{n} \sum_{k=0}^{n-1} P^k f \leqslant g$$

and, thus, $|A_n f| \leqslant g$. By applying our first criterion for weak precompactness (Section 5.1), we know that $\{A_n f\}$ is weakly precompact. Then Theorem 5.2.1 completes the argument. ■

Corollary 5.2.2. Again let $(X, \mathcal{A}, \mu)$ be a finite measure space and $P: L^1 \to L^1$ a Markov operator. If for some $f \in D$ there exists $M > 0$ and $p > 1$ such that

$$\|P^n f\|_{L^p} \leqslant M \tag{5.2.8}$$

for all n, then there is an $f_* \in D$ such that $Pf_* = f_*$.

Proof: We have

$$\|A_n f\|_{L^p} = \left\| \frac{1}{n} \sum_{k=0}^{n-1} P^k f \right\|_{L^p} \leqslant \frac{1}{n} \sum_{k=0}^{n-1} \|P^k f\|_{L^p} \leqslant \frac{1}{n}(nM) = M.$$

Hence, by our second criterion for weak precompactness, $\{A_n f\}$ is weakly precompact, and again Theorem 5.2.1 completes the proof. ■

Remark 5.2.3. The conditions $P^n f \leqslant g$ or $\|P^n f\|_{L^p} \leqslant M$ of these two corollaries guaranteeing the existence of a stationary density f_* rely on the properties of $\{P^n f\}$ for large n. To make this clearer suppose $P^n f \leqslant g$ only for $n > n_0$. Then, of course, $P^{n+n_0} f \leqslant g$ for all n, but this can be rewritten in the alternate form $P^n P^{n_0} f = P^n \tilde{f} \leqslant g$, where $\tilde{f} = P^{n_0} f$. The same argument holds for $\|P^n f\|_{L^p}$, thus demonstrating that it is sufficient for some n_0 to exist such that for all $n > n_0$ either (5.2.7) or (5.2.8) holds. □

We have proved that either convergence or precompactness of $\{A_n f\}$ implies the existence of a stationary density. We may reverse the question to ask whether the existence of a stationary density gives any clues to the asymptotic properties of sequences $\{A_n f\}$. The following theorem gives a partial answer to this question.

Theorem 5.2.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a Markov operator with a unique stationary density f_*. If $f_*(x) > 0$ for all $x \in X$, then

$$\lim_{n\to\infty} A_n f = f_* \qquad \text{for all } f \in D.$$

Proof: First assume f/f_* is bounded. By setting $c = \sup(f/f_*)$, we have

$$P^n f \leqslant P^n cf_* = cP^n f_* = cf_* \quad \text{and} \quad A_n f \leqslant cA_n f_* = cf_*.$$

Thus the sequence $\{A_n f\}$ is weakly precompact and, by Theorem 5.2.1, is convergent to a stationary density. Since f_* is the unique stationary density, $\{A_n f\}$ must converge to f_*. Thus the theorem is proved when f/f_* is bounded.

In the general case, write $f_c = \min(f, cf_*)$. We then have

$$f = \frac{1}{\|f_c\|} f_c + r_c\,, \tag{5.2.9}$$

where

$$r_c = \left(1 - \frac{1}{\|f_c\|}\right) f_c + f - f_c\,.$$

Since $f_*(x) > 0$ we also have

$$\lim_{c\to\infty} f_c(x) = f(x) \qquad \text{for all } x$$

and, evidently, $f_c(x) \leqslant f(x)$. Thus, by the Lebesgue dominated convergence theorem, $\|f_c - f\| \to 0$ and $\|f_c\| \to \|f\| = 1$ as $c \to \infty$. Thus the remainder r_c is strongly convergent to zero as $c \to \infty$. By choosing $\varepsilon > 0$ we can find a value c such that $\|r_c\| < \varepsilon/2$. Then

$$\|A_n r_c\| \leqslant \|r_c\| < \frac{\varepsilon}{2}. \tag{5.2.10}$$

However, since $f_c/\|f_c\|$ is a density bounded by $c\|f_c\|^{-1} f_*$, according to the first part of the proof,

$$\left\| A_n\left(\frac{1}{\|f_c\|} f_c\right) - f_* \right\| \leqslant \frac{\varepsilon}{2} \tag{5.2.11}$$

for sufficiently large n. Combining inequalities (5.2.10) and (5.2.11) with the decomposition (5.2.9), we immediately obtain

$$\|A_n f - f_*\| \leq \varepsilon$$

for sufficiently large n. ■

In the case that P is the Frobenius–Perron operator corresponding to a nonsingular transformation S, Theorem 5.2.2 offers a convenient criterion for ergodicity. As we have seen in Theorem 4.2.2, the ergodicity of S is equivalent to the uniqueness of the solution to $Pf = f$. Using this relationship, we can prove the following corollary.

Corollary 5.2.3. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space, $S: X \to X$ a measure-preserving transformation, and P the corresponding Frobenius–Perron operator. Then S is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k f = 1 \qquad \text{for every } f \in D. \tag{5.2.12}$$

Proof: The proof is immediate. Since S is measure preserving, we have $P1 = 1$. If S is ergodic, then by Theorem 4.2.2, $f_*(x) \equiv 1$ is the unique stationary density of P and, by Theorem 5.2.2, the convergence of (5.2.12) follows. Conversely, if the convergence of (5.2.12) holds, applying (5.2.12) to a stationary density f gives $f = 1$. Thus $f_*(x) = 1$ is the unique stationary density of P and again, by Theorem 4.2.2, the transformation S is ergodic. ■

5.3 Asymptotic periodicity of $\{P^n f\}$

In the preceding section we reduced the problem of examining the asymptotic properties of the averages $A_n f$ to one of determining the precompactness of $\{A_n f\}$. This, in turn, was reduced by Corollaries 5.2.1 and 5.2.2 to the problem of finding an upper-bound function for $P^n f$ or an upper bound for $\|P^n f\|_{L^p}$. In this section we show that if conditions similar to those in Corollaries 5.2.1 and 5.2.2 are satisfied for Frobenius–Perron operators, then the surprising result is that $\{P^n f\}$ is asymptotically periodic. Even more generally, we will show that almost any kind of upper bound on the iterates $P^n f$ of a Markov operator P suffices to establish that $\{P^n f\}$ will also have very regular (asymptotically periodic) behavior.

A general way to develop this upper-bound concept is via a definition of the convergence to a set by a sequence of functions.

Definition 5.3.1. Let $\mathcal{F}$ be a nonempty set in L^1 and let $g \in L^1$. Then the **distance** d between g and $\mathcal{F}$ is the "shortest of the distances" between g and

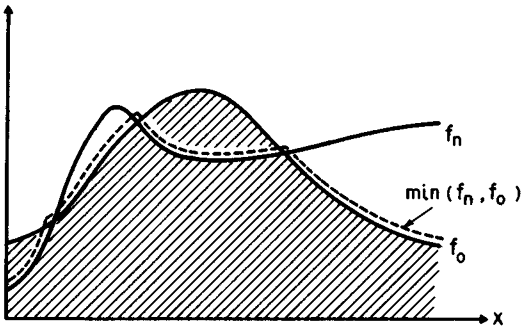

Figure 5.3.1. Graph showing convergence of a sequence of functions $\{f_n\}$ to a set $\mathcal{F}$, where the hatched region contains all possible functions drawn from $\mathcal{F}$. (See Example 5.3.1 for details.)

the elements $f \in \mathcal{F}$ or, more precisely,

$$d(g, \mathcal{F}) = \inf_{f \in \mathcal{F}} \|f - g\|.$$

Definition 5.3.2. A sequence of functions $\{f_n\}, f_n \in L^1$, is **convergent to a set** $\mathcal{F} \subset L^1$ if

$$\lim_{n \to \infty} d(f_n, \mathcal{F}) = 0.$$

Example 5.3.1. As a simple example of this type of convergence consider a sequence of functions $\{f_n\}, f_n \in L^1$, with the property that

$$\lim_{n \to \infty} \|(f_n - f_0)^+\| = 0,$$

where $f_0 \in L^1$ is given. Then, from our definitions, we may easily show that

$$\lim_{n \to \infty} d(f_n, \mathcal{F}) = 0,$$

where $\mathcal{F}$ is the set of all functions f such that

$$f \leqslant f_0.$$

To see this, examine Figure 5.3.1, illustrating that

$$d(f_n, \mathcal{F}) = \|f_n - \min(f_n, f_0)\| = \|(f_n - f_0)^+\|. \quad \square$$

Definition 5.3.3. A Markov operator P will be called **strongly (weakly) constrictive** if there exists a strongly (weakly) precompact set such that

$$\lim_{n \to \infty} d(P^n f, \mathcal{F}) = 0 \qquad \text{for all } f \in D.$$

Strongly and weakly constrictive Markov operators have a number of interesting properties, which we develop here. We defer to Chapter 6, however, the discussion of specific examples.

The first property that holds for Markov operators is contained in the following theorem.

Theorem 5.3.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, and $P: L^1 \to L^1$ a Markov operator. If P is weakly constrictive, then it is strongly constrictive.

Remark 5.3.1. This theorem is proved in the paper of Komornik [In press]. It is useful because, from our considerations in Section 5.1, it is much easier to check for weak precompactness than for strong precompactness. For example, if a Markov operator P has the property that there is an $f_0 \in L^1$ such that

$$\lim_{n\to\infty}\|(P^n f - f_0)^+\| = 0 \qquad \text{for all } f \in D,$$

then P is weakly constrictive and thus strongly constrictive. In what follows, the adjectives strongly and weakly will not be used for constrictive operators. □

A second property, and the main result of this section, which is proved in Lasota, Li, and Yorke ([1984]; see also Schaefer [1980] and Keller [1982]) is as follows.

***Theorem 5.3.2.* (spectral decomposition theorem).** Let P be a constrictive Markov operator. Then there is an integer r, two sequences of nonnegative functions $g_i \in D$ and $k_i \in L^\infty$, $i = 1, \ldots, r$, and an operator $Q: L^1 \to L^1$ such that for all $f \in L^1$, Pf may be written in the form

$$Pf(x) = \sum_{i=1}^{r} \lambda_i(f) g_i(x) + Qf(x), \tag{5.3.1}$$

where

$$\lambda_i(f) = \int_X f(x) k_i(x) \mu(dx). \tag{5.3.2}$$

The functions g_i and operator Q have the following properties:

(1) $g_i(x) g_j(x) = 0$ for all $i \neq j$, so that functions g_i have disjoint supports;
(2) For each integer i there exists a unique integer $\alpha(i)$ such that $Pg_i = g_{\alpha(i)}$. Further $\alpha(i) \neq \alpha(j)$ for $i \neq j$ and thus operator P just serves to permute the functions g_i.
(3) $\|P^n Qf\| \to 0$ as $n \to \infty$ for every $f \in L^1$.

Remark 5.3.2. Note from representation (5.3.1) that operator Q is automatically determined if we know the functions g_i and k_i, that is,

$$Qf(x) = Pf(x) - \sum_{i=1}^{r} \lambda_i(f) g_i(x) . \quad \square$$

From representation (5.3.1) of Theorem 5.3.2 for Pf, it immediately follows that the structure of $P^n f$ is given by

$$P^n f(x) = \sum_{i=1}^{r} \lambda_i(f) g_{\alpha^n(i)}(x) + Q_n f(x) , \tag{5.3.3}$$

where $Q_n = P^{n-1}Q$, and $\alpha^n(i) = \alpha(\alpha^{n-1}(i)) = \cdots$, and $\|Q_n f\| \to 0$ as $n \to \infty$. The terms under the summation in (5.3.3) are just permuted with each application of P, and since r is finite the sequence

$$\sum_{i=1}^{r} \lambda_i(f) g_{\alpha^n(i)}(x) \tag{5.3.4}$$

must be periodic with a period $\tau \leqslant r!$. Since $\{\alpha^n(1), \dots, \alpha^n(r)\}$ is simply a permutation of $\{1, \dots, r\}$, there is a unique i corresponding to each $\alpha^n(i)$. Thus it is clear that summation (5.3.4) may be rewritten as

$$\sum_{i=1}^{r} \lambda_{\alpha^{-n}(i)}(f) g_i(x) ,$$

where $\{\alpha^{-n}(i)\}$ denotes the inverse permutation of $\{\alpha^n(i)\}$.

Rewriting the summation in this form clarifies how successive applications of operator P really work. Since the functions g_i are supported on disjoint sets, each successive application of operator P leads to a new set of scaling coefficients $\lambda_{\alpha^{-n}}(f)$ associated with each function $g_i(x)$.

It is actually rather easy to obtain an upper bound on the integer r appearing in equation (5.3.1) if we can find an upper bound function for $P^n f$. Assume that P is a Frobenius–Perron operator and there exists a function $h \in L^1$ such that

$$\lim_{n\to\infty} \|(P^n f - h)^+\| = 0 \qquad \text{for } f \in D . \tag{5.3.5}$$

Thus, by Theorem 5.3.1, P is constrictive and representation (5.3.1) is valid. Let τ be the period of sequence (5.3.4), so that, from (5.3.1) and (5.3.5), we have

$$Lf(x) \doteq \lim_{n\to\infty} P^{n\tau} f(x) = \sum_{i=1}^{r} \lambda_i(f) g_i(x) \leqslant h(x), \qquad f \in D .$$

Set $f = g_k$ so that $Lf(x) = g_k(x) \leqslant h(x)$. By integrating over the support of g_k, bearing in mind that the supports of the g_k are disjoint, and summing from $k = 1$ to $k = r$, we have

$$\sum_{k=1}^{r} \int_{\operatorname{supp} g_k} g_k(x) \mu(dx) \leqslant \sum_{k=1}^{r} \int_{\operatorname{supp} g_k} h(x) \mu(dx) \leqslant \|h\| .$$

Since $g_k \in D$ this reduces to

$$r \leqslant \|h\|, \tag{5.3.6}$$

which is the desired result.

If the explicit representation (5.3.1) for Pf for a given Markov operator P is known, then it is especially easy to check for the existence of invariant measures and to determine ergodicity, mixing, or exactness, as shown in the following sections. Unfortunately, we seldom have an explicit representation for a given Markov operator, but in the remainder of this chapter we show that the mere existence of representation (5.3.1) allows us to deduce some interesting properties.

5.4 The existence of stationary densities

In this section we first show that every constrictive Markov operator has a stationary density and then give an explicit representation for $P^n f$ when that stationary density is a constant. We start with a proposition.

Proposition 5.4.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a constrictive Markov operator. Then P has a stationary density.

Proof: Let a density f be defined by

$$f(x) = \frac{1}{r} \sum_{i=1}^{r} g_i(x), \tag{5.4.1}$$

where r and g_i were defined in Theorem 5.3.2. Because of property (2), Theorem 5.3.2,

$$Pf(x) = \frac{1}{r} \sum_{i=1}^{r} g_{\alpha(i)}(x)$$

and thus $Pf = f$, which completes the proof. ■

Now assume that the measure μ is normalized $[\mu(X) = 1]$ and examine the consequences for the representation of $P^n f$ when we have a constant stationary density $f = 1_X$. Remember that, if P is a Frobenius–Perron operator, this is equivalent to μ being invariant.

Proposition 5.4.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a constrictive Markov operator. If P has a constant stationary density, then the representation for $P^n f$ takes the simple form

$$P^n f(x) = \sum_{i=1}^{r} \lambda_{\alpha^{-n}(i)}(f) \bar{1}_{A_i}(x) + Q_n f(x) \qquad \text{for all } f \in L^1, \tag{5.4.2}$$

where

$$\bar{1}_{A_i}(x) = [1/\mu(A_i)]1_{A_i}(x).$$

The sets A_i form a partition of X, that is,

$$\bigcup_i A_i = X \quad \text{and} \quad A_i \cap A_j = \emptyset \qquad \text{for } i \neq j.$$

Furthermore, $\mu(A_i) = \mu(A_j)$ whenever $j = \alpha^n(i)$ for some n.

Proof: First observe that with $f = 1_X$ and stationary, $P1_X = 1_X$ so that $P^n 1_X = 1_X$. However, if P is constrictive, then, from Theorem 5.3.2,

$$P^n 1_X(x) = \sum_{i=1}^{r} \lambda_{\alpha^{-n}(i)}(1_X) g_i(x) + Q_n 1_X(x). \tag{5.4.3}$$

From our considerations in the preceding section, we know that the summation in equation (5.4.3) is periodic. Let τ be the period of the summation portion of P^n (remember that $\tau \leqslant r!$) so that

$$\alpha^{-n\tau}(i) = i$$

and

$$P^{n\tau} 1_X(x) = \sum_{i=1}^{r} \lambda_i(1_X) g_i(x) + Q_{n\tau} 1_X(x).$$

Passing to the limit as $n \to \infty$ and using the stationarity of 1_X, we have

$$1_X(x) = \sum_{i=1}^{r} \lambda_i(1_X) g_i(x). \tag{5.4.4}$$

However, since functions g_i are supported on disjoint sets, therefore, from (5.4.4), we must have each g_i constant or, more specifically,

$$g_i(x) \equiv [1/\lambda_i(1_X)]1_{A_i}(x),$$

where $A_i \subset X$ denotes the support of g_i, that is, the set of all x such that $g_i(x) \neq 0$. From (5.4.4) it also follows that $\bigcup_i A_i = X$.

Apply operator P^n to equation (5.4.4) to give

$$P^n 1_X(x) \equiv 1_X(x) = \sum_{i=1}^{r} \lambda_i(1_X) g_{\alpha^n(i)}(x),$$

and, by the same reasoning employed earlier, we have

$$g_{\alpha^n(i)}(x) \equiv 1/\lambda_i(1_X) \qquad \text{for all } x \in A_{\alpha_i}.$$

Thus, the functions $g_i(x)$ and $g_{\alpha^n(i)}$ must be equal to the same constant. And, since the functions $g_i(x)$ are densities, we must have

$$\int_{A_i} g_i(x)\mu(dx) = 1 = \mu(A_i)/\lambda_i(1_X)\,.$$

Thus $\mu(A_i) = \lambda_i(1_X)$ and

$$g_i(x) = [1/\mu(A_i)]1_{A_i}(x)\,. \tag{5.4.5}$$

Moreover, $\mu(A_{\alpha^n(i)}) = \mu(A_i)$ for all n. ■

5.5 Ergodicity, mixing, and exactness

We now turn our attention to the determination of ergodicity, mixing, and exactness for operators P that can be written in the form of equation (5.3.1). We assume throughout that $\mu(X) = 1$ and that $P1_X = 1_X$. We further note that a permutation $\{\alpha(1), \dots, \alpha(r)\}$ of the set $\{1, \dots, r\}$ (see Theorem 5.3.2) for which there is no invariant subset is called a **cycle** or **cyclical permutation.**

Ergodicity

Theorem 5.5.1. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space and $P: L^1 \to L^1$ a constrictive Markov operator. Then P is ergodic if and only if the permutation $\{\alpha(1), \dots, \alpha(r)\}$ of the sequence $\{1, \dots, r\}$ is cyclical.

Proof: We start the proof with the "if" portion. Recall from equation (5.2.1) that the average A_nf is defined by

$$A_nf(x) = \frac{1}{n}\sum_{j=0}^{n-1} P^jf(x)\,.$$

Thus, with representation (5.4.2), A_nf can be written as

$$A_nf(x) = \sum_{i=1}^{r}\left[\frac{1}{n}\sum_{j=0}^{n-1}\lambda_{\alpha^{-j}(i)}(f)\right]\overline{1}_{A_i}(x) + \tilde{Q}_nf(x)\,,$$

where the remainder $\tilde{Q}_nf$ is given by

$$\tilde{Q}_nf = \frac{1}{n}\sum_{j=0}^{n-1} Q_jf \qquad Q_0f = -\sum_{i=1}^{r}\lambda_i(f)\overline{1}_{A_i} + f.$$

Now consider the coefficients

$$\frac{1}{n}\sum_{j=0}^{n-1}\lambda_{\alpha^{-j}(i)}(f)\,. \tag{5.5.1}$$

Since, as we showed in the Section 5.4, the sequence $\{\lambda_{\alpha^{-j}(i)}\}$ is periodic in j, the summation (5.5.1) must always have a limit as $n \to \infty$. Let this limit be $\overline{\lambda}_i(f)$.

Assume there are no invariant subsets of $\{1, \ldots, r\}$ under the permutation α. Then the limits $\overline{\lambda}_i(f)$ must be independent of i since every piece of the summation (5.5.1) of length r for different i consists of the same numbers but in a different order. Thus

$$\lim_{n\to\infty} A_n f = \sum_{i=1}^{r} \overline{\lambda}(f)\overline{1}_{A_i}.$$

Further, since α is cyclical, Proposition 5.4.2 implies that $\mu(A_i) = \mu(A_j) = 1/r$ for all i, j and $\overline{1}_{A_i} = r1_{A_i}$, so that

$$\lim_{n\to\infty} A_n f = r\overline{\lambda}(f).$$

Hence, for $f \in D$, $\overline{\lambda}(f) = 1/r$, and we have proved that if the permutation $\{\alpha(1), \ldots, \alpha(r)\}$ of $\{1, \ldots, r\}$ is cyclical, then $\{P^n f\}$ is Cesaro convergent to 1 and, therefore, ergodic.

The converse is also easy to prove. Suppose P is ergodic and that $\{\alpha(i)\}$ is not a cyclical permutation. Thus $\{\alpha(i)\}$ has an invariant subset I. As an initial f take

$$f(x) = \sum_{i=1}^{r} c_i \overline{1}_{A_i}(x)$$

wherein

$$c_i = \begin{cases} c \neq 0 & \text{if } i \text{ belongs to the invariant subset } I \text{ of the} \\ & \text{permutation of } \{1, \ldots, r\} \\ 0 & \text{otherwise}. \end{cases}$$

Then

$$\lim_{n\to\infty} A_n f = \frac{1}{r}\sum_{i=1}^{r} \overline{\lambda}_i(f)\overline{1}_{A_i},$$

where $\overline{\lambda}_i(f) \neq 0$ if i is contained in the invariant subset I, and $\overline{\lambda}_i(f) = 0$ otherwise. Thus the limit of $A_n f$ as $n \to \infty$ is not a constant function with respect to x, so that P cannot be ergodic. This is a contradiction; hence, if P is ergodic, $\{\alpha(i)\}$ must be a cyclical permutation. ■

Mixing and exactness

Theorem 5.5.2. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space and $P: L^1 \to L^1$ a constrictive Markov operator. If $r = 1$ in representation (5.3.1) for P, then P is exact.

Proof: The proof is simple. Assume $r = 1$, so by (5.4.2) we have

$$P^n f(x) = \lambda(f) 1_X(x) + Q_n f(x)$$

and, thus,

$$\lim_{n\to\infty} P^n f = \lambda(f) 1_X .$$

In particular, when $f \in D$ then $\lambda(f) \equiv 1$ since P preserves the norm. Hence, for all $f \in D$, $\{P^n f\}$ converges strongly to 1, and P is therefore exact (and, of course, also mixing). ∎

The converse is surprising, for we can prove that P mixing implies that $r = 1$.

Theorem 5.5.3. Again let $(X, \mathscr{A}, \mu)$ be a normalized measure space and $P: L^1 \to L^1$ a constrictive Markov operator. If P is mixing, then $r = 1$ in representation (5.3.1).

Proof: To see this, assume P is mixing but that $r > 1$ and take an initial $f \in D$ given by

$$f(x) = c_1 1_{A_1}(x), \qquad \text{where } c_1 = 1/\mu(A_1) .$$

Therefore

$$P^n f(x) = c_1 1_{A(n)}(x) ,$$

where $A(n) = A_{\alpha^n(1)}$. Since P was assumed to be mixing, $\{P^n f\}$ converges weakly to 1. However, note that

$$\langle P^n f, 1_{A_1} \rangle = \begin{cases} c_1 & \text{if } \alpha^n(1) = 1 \\ 0 & \text{if } \alpha^n(1) \neq 1 . \end{cases}$$

Hence $\{P^n f\}$ will converge weakly to 1 only if $\alpha^n(1) = 1$ for all sufficiently large n. Since α is a cyclical permutation, r cannot be greater than 1, thus demonstrating that $r = 1$. ∎

Remark 5.5.1. It is somewhat surprising that in this case P mixing implies P exact. □

Remark 5.5.2. Observe that, except for the remainder $Q_n f$, $P^n f$ behaves like permutations for which the notions of ergodicity, mixing, and exactness are quite simple. □

5.6 Asymptotic stability of $\{P^n\}$

Our considerations of ergodicity, mixing, and exactness for Markov operators in the previous section were based on the assumption that we are working with a normalized measure space $(X, \mathcal{A}, \mu)$. We now turn to a more general situation and take $(X, \mathcal{A}, \mu)$ to be an arbitrary measure space. We show how Theorem 5.3.2 allows us to obtain a most interesting result concerning the asymptotic stability of $\{P^n f\}$.

We first present a generalization for Markov operators of the concept of exactness for Frobenius–Perron operators associated with a transformation.

Definition 5.6.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a Markov operator. Then $\{P^n\}$ is said to be **asymptotically stable** if there exists a unique $f_* \in D$ such that $Pf_* = f_*$ and

$$\lim_{n\to\infty} \|P^n f - f_*\| = 0 \qquad \text{for every } f \in D. \tag{5.6.1}$$

When P is a Frobenius–Perron operator the following definition holds.

Definition 5.6.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ the Frobenius–Perron operator corresponding to a nonsingular transformation $S: X \to X$. If $\{P^n\}$ is asymptotically stable, then the transformation S is said to be **statistically stable.**

The following theorem is a direct consequence of Theorem 5.3.2.

Theorem 5.6.1. Let P be a constrictive Markov operator. Assume there is a set $A \subset X$ of nonzero measure, $\mu(A) > 0$, with the property that for every $f \in D$ there is an integer $n_0(f)$ such that

$$P^n f(x) > 0 \tag{5.6.2}$$

for almost all $x \in A$ and all $n > n_0(f)$. Then $\{P^n\}$ is asymptotically stable.

Proof: Since, by assumption, P is constrictive, representation (5.3.1) is valid. We will first show that $r = 1$.

Assume $r > 1$, and choose an integer i_0 such that A is not contained in the support of g_{i_0}. Take a density $f \in D$ of the form $f(x) = g_{i_0}(x)$ and let τ be the period of the permutation α. Then we have

$$P^{n\tau} f(x) = g_{i_0}(x).$$

Clearly, $P^{n\tau} f(x)$ is not positive on the set A since A is not contained in the support of g_{i_0}. This result contradicts (5.6.2) of the theorem and, thus, we must have $r = 1$.

Since $r = 1$, equation (5.3.3) reduces to

$$P^n f(x) = \lambda(f) g(x) + Q_n f(x)$$

so

$$\lim_{n\to\infty} P^n f = \lambda(f)g.$$

If $f \in D$ then $\lim_{n\to\infty} P^n f \in D$ also; therefore, by integrating over X we have

$$1 = \lambda(f).$$

Thus $\lim_{n\to\infty} P^n f = g$ for all $f \in D$ and $\{P^n\}$ is asymptotically stable; this finishes the proof. ∎

The disadvantage with this theorem is that it requires checking for two different criteria: that P is constrictive as well as the existence of the set A. It is interesting that, by a slight modification of the assumption that $P^n f$ is positive on a set A, we can completely eliminate the necessity of assuming P to be constrictive. To do this, we first introduce the notion of a lower-bound function.

Definition 5.6.3. A function $h \in L^1$ is a **lower-bound function** for a Markov operator $P: L^1 \to L^1$ if

$$\lim_{n\to\infty} \|(P^n f - h)^-\| = 0 \qquad \text{for every } f \in D. \tag{5.6.3}$$

Condition (5.6.3) may be rewritten as

$$(P^n f - h)^- = \varepsilon_n,$$

where $\|\varepsilon_n\| \to 0$ as $n \to \infty$ or, even more explicitly, as

$$P^n f \geqslant h - \varepsilon_n.$$

Thus, figuratively speaking, a lower-bound function h is one such that, for every density f, successive iterates of that density by P are eventually almost above h.

It is, of course, clear that any nonpositive function is a lower-bound function, but, since $f \in D$ and thus $P^n f \in D$ and all densities are positive, a negative lower bound function is of no interest. Thus we give a second definition.

Definition 5.6.4. A lower-bound function h is called **nontrivial** if $h \geqslant 0$ and $\|h\| > 0$.

Having introduced the concept of nontrivial lower-bound functions, we can now state the following theorem.

Theorem 5.6.2. Let $P: L^1 \to L^1$ be a Markov operator. $\{P^n\}$ is asymptotically stable if and only if there is a nontrivial lower bound function for P.

Proof: The "only if" part is obvious since (5.6.1) implies (5.6.3) with $h = f_*$.

The proof of the "if" part is not so direct, and will be done in two steps. We first show that

$$\lim_{n\to\infty}\|P^n(f_1 - f_2)\| = 0 \tag{5.6.4}$$

for every $f_1, f_2 \in D$ and then proceed to construct the function f_*.

Step I. For every pair of densities $f_1, f_2 \in D$, the $\|P^n(f_1 - f_2)\|$ is a decreasing function of n. To see this, note that, since every Markov operator is contractive,

$$\|Pf\| \leqslant \|f\|$$

and, as a consequence,

$$\|P^{n+m}(f_1 - f_2)\| = \|P^m P^n(f_1 - f_2)\| \leqslant \|P^n(f_1 - f_2)\|.$$

Now set $g = f_1 - f_2$ and note that, since $f_1, f_2 \in D$,

$$c = \|g^+\| = \|g^-\| = \tfrac{1}{2}\|g\|.$$

Assume c > 0. We have $g = g^+ - g^-$ and

$$\|P^n g\| = c\|(P^n(g^+/c) - h) - (P^n(g^-/c) - h)\|. \tag{5.6.5}$$

Since g^+/c and g^-/c belong to D, by equation (5.6.3), there must exist an integer n_1 such that for all $n \geqslant n_1$

$$\|(P^n(g^+/c) - h)^-\| \leqslant \tfrac{1}{4}\|h\|$$

and

$$\|(P^n(g^-/c) - h)^-\| \leqslant \tfrac{1}{4}\|h\|.$$

Now we wish to establish upper bounds for $\|P^n(g^+/c) - h\|$ and $\|P^n(g^-/c) - h\|$. To do this, first note that, for any pair of nonnegative real numbers a and b,

$$|a - b| = a - b + 2(a - b)^-.$$

Next write

$$\begin{aligned}
\|P^n(g^+/c) - h\| &= \int_X |P^n(g^+/c)(x) - h(x)|\mu(dx) \\
&= \int_X P^n(g^+/c)(x)\mu(dx) - \int_X h(x)\mu(dx) \\
&\quad + 2\int_X (P^n(g^+/c)(x) - h(x))^-\mu(dx) \\
&= \|P^n(g^+/c)\| - \|h\| + 2\|(P^n(g^+/c) - h)^-\| \\
&\leqslant 1 - \|h\| + 2\cdot\tfrac{1}{4}\|h\| = 1 - \tfrac{1}{2}\|h\| \qquad \text{for} \quad n \geqslant n_1.
\end{aligned}$$

Analogously,

$$\|P^n(g^-/c) - h\| \leq 1 - \tfrac{1}{2}\|h\| \qquad \text{for } n \geq n_1 .$$

Thus equation (5.6.5) gives

$$\begin{aligned}\|P^n g\| &\leq c\|P^n(g^+/c) - h\| + c\|P^n(g^-/c) - h\| \\ &\leq c(2 - \|h\|) = \|g\|(1 - \tfrac{1}{2}\|h\|) \qquad \text{for } n \geq n_1 .\end{aligned} \tag{5.6.6}$$

From (5.6.6), for any $f_1, f_2 \in D$, we can find an integer n_1 such that

$$\|P^{n_1}(f_1 - f_2)\| \leq \|f_1 - f_2\|(1 - \tfrac{1}{2}\|h\|) .$$

By applying the same argument to the pair $P^{n_1}f_1$, $P^{n_1}f_2$, we may find a second integer n_2 such that

$$\begin{aligned}\|P^{n_1+n_2}(f_1 - f_2)\| &\leq \|P^{n_1}(f_1 - f_2)\|(1 - \tfrac{1}{2}\|h\|) \\ &\leq \|f_1 - f_2\|(1 - \tfrac{1}{2}\|h\|)^2 .\end{aligned}$$

After k repetitions of this procedure, we have

$$\|P^{n_1+\cdots+n_k}(f_1 - f_2)\| \leq \|f_1 - f_2\|(1 - \tfrac{1}{2}\|h\|)^k ,$$

and since $\|P^n(f_1 - f_2)\|$ is a decreasing function of n, this implies (5.6.4).

Step II. To complete the proof, we construct a maximal lower-bound function for P. Thus, let

$$\rho = \sup\{\|h\|: h \text{ is a lower-bound function for } P\} .$$

Since by assumption there is a nontrivial h, we must have $0 < \rho \leq 1$. Observe that for any two lower-bound functions h_1 and h_2, the function $h = \max(h_1, h_2)$ is also a lower-bound function. To see this, note that

$$\|(P^n f - h)^-\| \leq \|(P^n f - h_1)^-\| + \|(P^n f - h_2)^-\| .$$

Choose a sequence $\{h_j\}$ of lower-bound functions such that $\|h_j\| \to \rho$. Replacing, if necessary, h_j by $\max(h_1, \ldots, h_j)$, we can construct an increasing sequence $\{h_j\}$ of lower functions, which will always have a limit (finite or infinite). This limiting function

$$h_* = \lim_{j\to\infty} h_j$$

is also a lower-bound function since

$$\|(P^n f - h_*)^-\| \leq \|(P^n f - h_j)^-\| + \|h_j - h_*\|$$

and, by the Lebesgue monotone convergence theorem,

$$\|h_j - h_*\| = \int_X h_*(x)\mu(dx) - \int_X h_j(x)\mu(dx) \to 0 \qquad \text{as } j \to \infty .$$

Now the limiting function h_* is also the maximal lower function. To see this, note that for any other lower function h, the function $\max(h, h_*)$ is also a lower function and that

$$\|\max(h, h_*)\| \leqslant \rho = \|h_*\|,$$

which implies $h \leqslant h_*$.

Observe that, since $(Pf)^- \leqslant Pf^-$, for every m and n $(n > m)$,

$$\|(P^n f - P^m h_*)^-\| \leqslant \|P^m(P^{n-m}f - h_*)^-\| \leqslant \|(P^{n-m}f - h_*)^-\|,$$

which implies that, for every m, the function $P^m h_*$ is a lower function. Thus, since h_* is the maximal lower function, $P^m h_* \leqslant h_*$ and, since P^m preserves the integral, $P^m h_* = h_*$. Thus the function $f_* = h_*/\|h_*\|$ is a density satisfying $Pf_* = f_*$.

Finally, by equation (5.6.4), we have

$$\lim_{n\to\infty}\|P^n f - f_*\| = \lim_{n\to\infty}\|P^n f - P^n f_*\| = 0 \qquad \text{for } f \in D,$$

which automatically gives equation (5.6.1). ∎

In checking for the conditions of Theorem 5.6.2, it is necessary to show that (5.6.3) is satisfied for all $f \in D$; clearly, this is difficult to do. In fact, it is sufficient to check that this is true only for an arbitrary class of functions $f \in D_0 \subset D$, where the set D_0 is dense in D. To be more precise, we give the following definition.

Definition 5.6.5. A set $D_0 \subset D(X)$ is called **dense** in $D(X)$ if, for every $h \in D$ and $\varepsilon > 0$, there is a $g \in D_0$ such that $\|h - g\| < \varepsilon$.

If X is an interval of the real line R or, more generally, an open set in R^d, then, for example, the following subsets of $D(X)$ are dense:

$$D_1 = \{\text{nonnegative continuous functions on } X\} \cap D(X)$$

$$D_2 = \{\text{nonnegative continuous functions with compact support in } X\} \cap D(X)$$

$$D_3 = \{\text{nonnegative differentiable functions on } X\} \cap D(X)$$

$$D_4 = \{\text{positive differentiable functions on } X\} \cap D(X).$$

If a set $D_0 \subset D(X)$ is dense in $D(X)$, it is sufficient to check the convergence of (5.6.3) for any $f \in D_0$. Then, for any other $f \in D(X)$, the convergence will be automatically ensured because condition (5.6.3) is framed in terms of a norm. The fact that we need check condition (5.6.3) only for a subset D shows the usefulness and convenience of Theorem 5.6.2 relative to Theorem 5.6.1, which must be checked for all $f \in D(X)$.

The results of Theorem 5.6.2 with respect to the uniqueness of stationary densities for asymptotically stable Markov operators may be generalized by the following observation.

Proposition 5.6.1. Let $(X, \mathcal{A}, \mu)$ be a measure space and $P: L^1 \to L^1$ a Markov operator. If $\{P^n\}$ is asymptotically stable and f_* is the unique stationary density of P, then for every normalized $f \in L^1$ ($\|f\| = 1$) the condition

$$Pf = f \tag{5.6.7}$$

implies that either $f = f_*$ or $f = -f_*$.

Proof: From Proposition 3.1.3, equation (5.6.7) implies that both f^+ and f^- are fixed points of P. Assume $\|f^+\| > 0$, so that $\tilde{f} = f^+/\|f^+\|$ is a density and $P\tilde{f} = \tilde{f}$. Uniqueness of f_* implies $\tilde{f} = f_*$, hence

$$f^+ = f_*\|f^+\|,$$

which must also hold for $\|f^+\| = 0$. In an analogous fashion,

$$f^- = f_*\|f^-\|$$

so that

$$f = f^+ - f^- = (\|f^+\| - \|f^-\|)f_* = \alpha f_*.$$

Since $\|f\| = \|f_*\|$, we have $|\alpha| = 1$, and the proof is complete. ■

Before closing this section we state and prove a result that draws the connection between statistical stability and exactness when P is a Frobenius–Perron operator.

Proposition 5.6.2. Let $(X, \mathcal{A}, \mu)$ be a measure space, $S: X \to X$ a nonsingular transformation such that $S(A) \in \mathcal{A}$ for $A \in \mathcal{A}$, and P the Frobenius–Perron operator corresponding to S. If S is statistically stable and f_* is the density of the unique invariant measure, then the transformation S with the measure

$$\mu_{f_*}(A) = \int_A f_*(x)\mu(dx) \qquad \text{for } A \in \mathcal{A}$$

is exact.

Proof: From Theorem 4.1.1 it follows immediately that μ_{f_*} is invariant. Thus, it only remains to prove the exactness.

Assume $\mu_{f_*}(A) > 0$ and define

$$f_A(x) = [1/\mu_{f_*}(A)]f_*(x)1_A(x) \qquad \text{for } x \in X.$$

Clearly, $f_A \in D(X, \mathcal{A}, \mu)$ and

$$\lim_{n\to\infty} r_n = \lim_{n\to\infty}\|P^n f_A - f_*\| = 0\,.$$

From the definition of μ_{f_*}, we have

$$\mu_{f_*}(S^n(A)) = \int_{S^n(A)} f_*(x)\mu(dx) \geqslant \int_{S^n(A)} P^n f_A(x)\mu(dx) - r_n\,. \qquad (5.6.8)$$

By Proposition 3.2.1, we know that $P^n f_A$ is supported on $S^n(A)$, so that

$$\int_{S^n(A)} P^n f_A(x)\mu(dx) = \int_X P^n f_A(x)\mu(dx) = 1\,.$$

Substituting this result into (5.6.8) and taking the limit as $n \to \infty$ gives

$$\lim_{n\to\infty} \mu_{f_*}(S^n(A)) = 1\,,$$

hence $S\colon X \to X$ is exact by definition. ■

Remark 5.6.1. In the most general case, Proposition 5.6.2 is not invertible, that is, statistical stability of S implies the existence of a unique invariant measure and exactness, but not vice versa. Lin [1971] has shown that the inverse implication is true when the initial measure μ is invariant. □

5.7 Markov operators defined by a stochastic kernel

As a sequel to Section 5.6, we wish to develop some important consequences of Theorems 5.6.1 and 5.6.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and $K\colon X \times X \to R$ be a measurable function that satisfies

$$0 \leqslant K(x,y) \qquad (5.7.1)$$

and

$$\int_X K(x,y)\,dx = 1, \qquad [dx = \mu(dx)]\,. \qquad (5.7.2)$$

Any function K satisfying (5.7.1) and (5.7.2) is called a **stochastic kernel.** Further, we define an integral operator P by

$$Pf(x) = \int_X K(x,y)f(y)\,dy \qquad \text{for } f \in L^1\,. \qquad (5.7.3)$$

The operator P is clearly linear and nonnegative. Since we also have

$$\begin{aligned}\int_X Pf(x)\,dx &= \int_X dx \int_X K(x,y)f(y)\,dy \\ &= \int_X f(y)\,dy \int_X K(x,y)\,dx = \int_X f(y)\,dy,\end{aligned}$$

therefore P is a Markov operator. In the special case that X is a finite set and μ is a counting measure, we have a Markov chain and P is a stochastic matrix.

Now consider two Markov operators P_a and P_b and their corresponding stochastic kernels, K_a and K_b. Clearly, $P_a P_b$ is also a Markov operator, and we wish to know how its kernel is related to K_a and K_b. Thus, write

$$\begin{aligned}(P_a P_b)f(x) = P_a(P_b f)(x) &= \int_X K_a(x,z)\,(P_b f(z))\,dz \\ &= \int_X K_a(x,z)\left\{\int_X K_b(z,y)f(y)\,dy\right\}dz \\ &= \int_X\left\{\int_X K_a(x,z)K_b(z,y)\,dz\right\}f(y)\,dy.\end{aligned}$$

Then $P_a P_b$ is also an integral operator with the kernel

$$K(x,y) = \int_X K_a(x,z)K_b(z,y)\,dz. \tag{5.7.4}$$

We denote this composed kernel K by

$$K = K_a * K_b \tag{5.7.5}$$

and note that the composition has the properties:

(i) $K_a * (K_b * K_c) = (K_a * K_b) * K_c$ (associative law); and
(ii) Any kernel formed by the composition of stochastic kernels is stochastic.

However, in general kernels K_a and K_b do not commute, that is, $K_a * K_b \neq K_b * K_a$. Note that the foregoing operation of composition definition is just a generalization of matrix multiplication.

Now we are in a position to show that Theorem 5.6.2 can be applied to operators P defined by stochastic kernels and, in fact, gives a simple sufficient condition for the asymptotic stability of $\{P^n\}$.

Corollary 5.7.1. Let $(X, \mathcal{A}, \mu)$ be a measure space $K\colon X \times X \to R$ a stochastic kernel, i.e. K satisfies (5.7.1) and (5.7.2), and P the corresponding Markov operator defined by (5.7.3). Denote by K_n the kernel corresponding to P^n. If, for some m,

$$\int_X \inf_y K_m(x,y)\,dx > 0, \tag{5.7.6}$$

then $\{P^n\}$ is asymptotically stable.

Proof: By the definition of K_n, for every $f \in D(X)$ we have

$$P^n f(x) = \int_X K_n(x,y) f(y)\, dy .$$

Furthermore, from the associative property of the composition of kernels,

$$K_{n+m}(x,y) = \int_X K_m(x,z) K_n(z,y)\, dz ,$$

so that

$$\begin{aligned} P^{n+m} f(x) &= \int_X K_{n+m}(x,y) f(y)\, dy \\ &= \int_X \left\{ \int_X K_m(x,z) K_n(z,y)\, dz \right\} f(y)\, dy . \end{aligned}$$

If we set

$$h(x) = \inf_y K_m(x,y) ,$$

then

$$\begin{aligned} P^{n+m} f(x) &\geqslant h(x) \int_X \left\{ \int_X K_n(z,y)\, dz \right\} f(y)\, dy \\ &= h(x) \int_X f(y)\, dy \end{aligned}$$

since K_n is a stochastic kernel. Furthermore, since $f \in D(X)$,

$$\int_X f(y)\, dy = 1 ,$$

and, therefore,

$$P^{n+m} f(x) \geqslant h(x) \qquad \text{for } n \geqslant 1, f \in D(X) .$$

Thus

$$(P^n f - h)^- = 0 \qquad \text{for } n \geqslant m + 1 ,$$

which implies that (5.6.3) holds, and we have finished the proof. ■

In the case that X is a finite set and K is a stochastic matrix, this result is equivalent to one originally obtained by Markov.

Although condition (5.7.6) on the kernel is quite simple, it is seldom satisfied when $K(x,y)$ is defined on an unbounded space. For example, in Section 8.9 we discuss the evolution of densities under the operation of a Markov operator defined by the kernel [cf. equation (8.9.6)]

$$K(x,y) = \begin{cases} -e^y \mathrm{Ei}(-y), & 0 < x \leqslant y \\ -e^y \mathrm{Ei}(-x), & 0 < y < x, \end{cases} \tag{5.7.7}$$

where

$$-\mathrm{Ei}(-x) \equiv \int_x^\infty (e^{-y}/y)\,dy, \qquad x > 0,$$

is the exponential integral. In this case

$$\inf_y K(x,y) = 0 \qquad \text{for all } x > 0,$$

and the same holds for all of its iterates $K_m(x,y)$. A similar problem occurs with the kernel

$$K(x,y) = g(ax + by),$$

where $b \neq 0$ and g is an integrable function defined on R or even on R^+ (cf. Example 5.7.2).

In these and other cases where condition (5.7.6) is not satisfied, an alternative approach, reminiscent of the stability methods developed by Liapunov, offers a way to examine the asymptotic properties of iterates of densities by Markov operators.

Let G be an unbounded measurable subset of a d-dimensional Euclidian space $R^d, G \subset R^d$, and $K: G \times G \to R$ a measurable stochastic kernel. We will call any continuous nonnegative function $V: G \to R$ satisfying

$$\lim_{|x|\to\infty} V(x) = \infty \tag{5.7.8}$$

a **Liapunov function.**

Next, we introduce the **Chebyshev inequality** through the following proposition.

Proposition 5.7.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, $V: X \to R$ an arbitrary nonnegative measurable function, and for all $f \in D$ set

$$E(V|f) = \int_X V(x)f(x)\mu(dx).$$

If

$$G_a = \{x: V(x) < a\},$$

then

$$\int_{G_a} f(x)\mu(dx) \geqslant 1 - E(V|f)/a \tag{5.7.9}$$

(the Chebyshev inequality).

Proof: The proof is easy. Clearly,

$$E(V|f) \geq \int_{X \setminus G_a} V(x)f(x)\mu(dx) \geq a\int_{X \setminus G_a} f(x)\mu(dx)$$
$$\geq a\left\{1 - \int_{G_a} f(x)\mu(dx)\right\}.$$

thus the Chebyshev inequality is proved. ■

With the lower-bound Theorem 5.6.2 and the Chebyshev inequality, it is possible to prove the following theorem.

Theorem 5.7.1. If the kernel $K(x,y)$ satisfies

$$\int_G \inf_{|y|\leq r} K(x,y)\,dx > 0 \qquad \text{for every } r > 0, \tag{5.7.10}$$

and has a Liapunov function $V\colon G \to R$ such that

$$\int_G K(x,y)V(x)\,dx \leq \alpha V(y) + \beta, \qquad 0 \leq \alpha < 1, \beta \geq 0, \tag{5.7.11}$$

then for the Markov operator $P\colon L^1(G) \to L^1(G)$, defined by equation (5.7.3), $\{P^n\}$ is asymptotically stable.

Proof: First define the function

$$E_n(V|f) = \int_G V(x)P^n f(x)\,dx \tag{5.7.12}$$

that can be thought of as the expected value of $V(x)$ with respect to the density $P^n f(x)$. By the definition of $P^n f$, we have directly

$$E_n(V|f) = \int_G V(x)\,dx \int_G K(x,y)P^{n-1}f(y)\,dy$$
$$= \int_G P^{n-1}f(y)\,dy \int_G K(x,y)V(x)\,dx. \tag{5.7.13}$$

Substituting equation (5.7.11) into (5.7.13) yields

$$E_n(V|f) \leq \int_G P^{n-1}f(y)[\beta + \alpha V(y)]\,dy$$
$$= \beta + \alpha\int_G P^{n-1}f(y)V(y)\,dy = \beta + \alpha E_{n-1}(V|f).$$

By an induction argument, it is easy to show that from this equation we obtain

$$E_n(V|f) \leqslant [\beta/(1-\alpha)] + \alpha^n E_0(V|f).$$

Even though $E_0(V|f)$ is clearly dependent on our initial choice of f, it is equally clear that, for every f such that

$$E_0(V|f) < \infty, \tag{5.7.14}$$

there is some integer $n_0 = n_0(f)$ such that

$$E_n(V|f) \leqslant [\beta/(1-\alpha)] + 1 \qquad \text{for all } n \geqslant n_0. \tag{5.7.15}$$

Now let

$$G_a = \{x \in G: V(x) < a\}$$

so that from the Chebyshev inequality we have

$$\int_{G_a} P^n f(x)\,dx \geqslant 1 - \frac{E_n(V|f)}{a}. \tag{5.7.16}$$

Further, set

$$a > 1 + [\beta/(1-\alpha)],$$

then

$$\frac{E_n(V|f)}{a} \leqslant \frac{1}{a}\left(1 + \frac{\beta}{1-\alpha}\right) < 1 \qquad \text{for } n \geqslant n_0$$

and thus (5.7.16) becomes

$$\int_{G_a} P^n f(x)\,dx \geqslant 1 - \frac{1}{a}\left(1 + \frac{\beta}{1-\alpha}\right) \doteqdot \varepsilon > 0 \qquad \text{for } n \geqslant n_0. \tag{5.7.17}$$

Since $V(x) \to \infty$ as $|x| \to \infty$ there is an $r > 0$ such that $V(x) > a$ for $|x| > r$. Thus the set G_a is entirely contained in the ball $|x| \leqslant r$, and we may write

$$\begin{aligned} P^{n+1}f(x) &= \int_G K(x,y)P^n f(y)\,dy \geqslant \int_{G_a} K(x,y)P^n f(y)\,dy \\ &\geqslant \inf_{y\in G_a} K(x,y) \int_{G_a} P^n f(y)\,dy \\ &\geqslant \inf_{|y|\leqslant r} K(x,y) \int_{G_a} P^n f(y)\,dy \\ &\geqslant \varepsilon \inf_{|y|\leqslant r} K(x,y) \end{aligned} \tag{5.7.18}$$

for all $n \geqslant n_0$.

By setting

$$h(x) = \varepsilon \inf_{|y|\leqslant r} K(x,y)$$

in inequality (5.7.18) we have, by assumption (5.7.10), that

$$\|h\| > 0 .$$

Finally, because of the continuity of V, the set $D_0 \subset D$ of all f such that (5.7.14) is satisfied is dense in D. Thus all the conditions of Theorem 5.6.2 are satisfied. ■

Another important property of Markov operators defined by a stochastic kernel is that they transform every weakly precompact set into a strongly precompact set. In this case, instead of using the Komornik Theorem 5.3.1, we may use the following result of Krasnosel'skii [1976].

Theorem 5.7.2. Let X be a bounded Borel subset of R^d, P be a Markov operator defined by (5.7.3), and $\mathcal{F}$ be a weakly precompact set of functions. Then the set $P(\mathcal{F})$ of all functions of the form

$$\int_X K(x,y)f(y)\,dy, \qquad f \in \mathcal{F},$$

is strongly precompact.

We will apply this theorem to integral operators with a stochastic kernel satisfying

$$K(x,y) \leqslant g(x), \qquad g \in L^1.$$

For $f \in D$, we have

$$Pf(x) = \int_X K(x,y)f(y)\,dy \leqslant g(x)\int_X f(y)\,dy = g(x).$$

Thus, for this kernel, the set $\mathcal{F}$ of all possible Pf with $f \in D$ is weakly precompact. Further, since $P^k f \in D$ for $f \in D$, we have

$$P^n f = PP^{n-1}f \in P(\mathcal{F}) \qquad \text{for } n > 1 .$$

By combining this result with Theorem 5.7.2, we can see that the set of all functions $\{P^n f\}$ with $f \in D$, $n > 1$, is strongly precompact.

These results show that for every bounded kernel $K(x,y)$, or for any kernel with a bounded iterate $K_m(x,y)$, representation (5.3.1) for $Pf(x)$ given in the spectral decomposition theorem is valid. To be more precise, we state the following corollary.

Corollary 5.7.2. Every sequence of functions $\{P^n f\}$, defined by

$$Pf(x) = \int_X K(x,y)f(y)\,dy, \qquad \text{with } K(x,y) \leqslant g(x), f, g \in L^1,$$

is asymptotically periodic.

Another consequence of Theorem 5.7.2 in conjunction with Theorem 5.6.1 is the following statement.

Corollary 5.7.3. If there exists an integer m and a $g \in L^1$ such that

$$K_m(x, y) \leqslant g(x),$$

where $K_m(x, y)$ is the mth iterate of a stochastic kernel, and there is a set $S \subset X$, $\mu(S) > 0$, such that

$$0 < K_m(x, y) \qquad \text{for } x \in S,\ y \in X,$$

then the sequence $\{P^n\}$ is asymptotically stable.

Example 5.7.1. To see the power of Theorem 5.7.1, we first consider the case where the kernel $K(x, y)$ is given by the exponential integrals in equation (5.7.7). It is easy to show that $-e^y(-\mathrm{Ei}(y))$ is decreasing and consequently

$$\inf_{0 \leqslant y \leqslant r} K(x, y) \geqslant \min\{-\mathrm{Ei}(-x), -e^r\mathrm{Ei}(-r)\} > 0.$$

Furthermore, taking $V(x) = x$, we have, after integration,

$$\int_0^\infty xK(x, y)\, dx = \tfrac{1}{2}(1 + y).$$

Therefore it is clear that $V(x) = x$ is a Liapunov function for this system when $\alpha = \beta = \frac{1}{2}$. Also, observe that with $f(x) = \exp(-x)$, we have

$$Pf(x) = \int_0^\infty K(x, y)e^{-y}\, dy = e^{-x}.$$

Thus the limiting density attained by repeated application of the Markov operator P is $f_*(x) = \exp(-x)$. □

Example 5.7.2. As a second example, let $g: R \to R$ be a continuous positive function satisfying

$$\int_{-\infty}^\infty g(x)\, dx = 1 \quad \text{and} \quad m_1 = \int_{-\infty}^\infty |x| g(x)\, dx < \infty.$$

Further, let a stochastic kernel be defined by

$$K(x, y) = |a| g(ax + by), \qquad |a| > |b|,\ b \neq 0$$

and consider the corresponding Markov operator

$$Pf(x) = \int_{-\infty}^\infty K(x, y) f(y)\, dy.$$

Let $V(x) = |x|$, so that we have

$$\int_{-\infty}^{\infty} K(x,y)V(x)\,dx = |a|\int_{-\infty}^{\infty} |x|g(ax+by)\,dx = \int_{-\infty}^{\infty} g(s)\left|\frac{s-by}{a}\right| ds$$
$$\leqslant \int_{-\infty}^{\infty} g(s)\left|\frac{s}{a}\right| ds + \int_{-\infty}^{\infty} g(s)\left|\frac{by}{a}\right| ds = \frac{m_1}{|a|} + \left|\frac{by}{a}\right| .$$

Thus, when $\alpha = |b/a|$ and $\beta = m_1/|a|$, it is clear that $V(x)$ satisfies condition (5.7.11) and hence Theorem 5.7.1 is satisfied.

As will become evident in Section 10.5, in this example Pf has the following interesting probabilistic interpretation. If ξ and η are two independent random variables with densities $f(x)$ and $g(x)$, respectively, then

$$Pf(x) = |a|\int_{-\infty}^{\infty} g(ax+by)f(y)\,dy, \qquad \text{with } a = \frac{1}{c_2} \text{ and } b = -\frac{c_1}{c_2},$$

is the density of the random variable $(c_1\xi + c_2\eta)$ [cf. equation (10.1.8)]. □

Example 5.7.3. As a final example of the applicability of the results of this section, we consider a simple model for the cell cycle [Lasota and Mackey, 1984]. First, it is assumed that there exists an intracellular substance (mitogen), necessary for mitosis and that the rate of change of mitogen is governed by

$$\frac{dm}{dt} = g(m), \qquad m(0) = r$$

with solution $m(r,t)$. The rate g is a C^1 function on $[0,\infty)$ and $g(x) > 0$ for $x > 0$. Second, it is assumed that the probability of mitosis in the interval $[t, t+\Delta t]$ is given by $\phi(m(t))\,\Delta t + o(\Delta t)$, where ϕ is a C^1 function on $[0,\infty)$ such that

$$\phi(0) = 0 \quad \text{and} \quad \liminf_{x\to\infty} q(x) > 0\,,$$

where $q(x) = \phi(x)/g(x)$. Finally, it is assumed that at mitosis each daughter cell receives exactly one-half of the mitogen present in the mother cell.

Under these assumptions it can be shown that for a distribution $f_{n-1}(x)$ of mitogen in the $(n-1)$st generation of a large population of cells, the mitogen distribution in the following generation is given by

$$f_n(x) = \int_0^{\infty} K(x,r)f_{n-1}(r)\,dr\,,$$

where

$$K(x,r) = \begin{cases} 0 & x \in [0, \tfrac{1}{2}r) \\ 2q(2x)\exp\left[-\int_r^{2x} q(y)\,dy\right] & x \in [\tfrac{1}{2}r, \infty)\,. \end{cases}$$

It is straightforward to show that $K(x,r)$ satisfies (5.7.1) and (5.7.2) and is, thus, a stochastic kernel. Hence the operator $P: L^1 \to L^1$ defined by

$$Pf(x) = \int_0^\infty K(x,r)f(r)\,dr$$

is a Markov operator. To show that there is a unique stationary density $f_* \in D$ to which $\{P^n f\}$ converges strongly, we use Theorem 5.7.1.

First we examine the integral

$$\int_0^\infty xK(x,r)\,dx = \int_{r/2}^\infty 2xq(2x)\exp\left[-\int_r^{2x} q(y)\,dy\right]dx.$$

Integrating by parts, we have

$$\int_0^\infty xK(x,r)\,dx = -x\exp\left[-\int_r^{2x} q(y)\,dy\right]\Bigg|_{x=r/2}^{x=\infty} + \int_{r/2}^\infty \exp\left[-\int_r^{2x} q(y)\,dy\right]dx.$$

Since $\lim_{x\to\infty}\inf q(x) > 0$, it follows that there is an $\varepsilon > 0$ and $d \geqslant 0$ such that

$$q(x) \geqslant \varepsilon \qquad \text{for } x \geqslant d,$$

and, as a consequence,

$$\lim_{x\to\infty} x\exp\left[-\int_r^{2x} q(y)\,dy\right] = 0.$$

Furthermore,

$$\int_{r/2}^\infty \exp\left[-\int_r^{2x} q(y)\,dy\right]dx \leqslant \int_{r/2}^\infty \exp\{-\varepsilon[2x - \max(r,d)]\}\,dx$$

$$= \frac{1}{2\varepsilon}\exp\{-\varepsilon[r - \max(r,d)]\} \leqslant \frac{1}{2\varepsilon}\exp(\varepsilon d).$$

Consequently we obtain

$$\int_0^\infty xK(x,r)\,dx \leqslant \frac{r}{2} + \frac{1}{2\varepsilon}\exp(\varepsilon d)$$

so that the kernel satisfies inequality (5.7.11) of Theorem 5.7.1. It only remains to be shown that K satisfies (5.7.10).

Let $r_0 \geqslant 0$ be an arbitrary finite real number. Consider $K(x,r)$ for $0 \leqslant r \leqslant r_0$ and $x \geqslant \frac{1}{2}r$. Then

$$K(x,r) = 2q(2x)\exp\left[-\int_r^{2x} q(y)\,dy\right]$$

$$\geqslant 2q(2x)\exp\left[-\int_0^{2x} q(y)\,dy\right] \qquad \text{for } 0 \leqslant r \leqslant r_0,\ x \geqslant \tfrac{1}{2}r$$

and, as a consequence,

$$\inf_{0 \leqslant r \leqslant r_0} K(x,r) \geqslant h(x) = \begin{cases} 0 & \text{for } x < \frac{1}{2}r_0 \\ 2q(2x)\exp\left[-\int_0^{2x} q(y)\,dy\right] & \text{for } x \geqslant \frac{1}{2}r_0 . \end{cases}$$

Further,

$$\int_0^\infty h(x)\,dx = \int_{r_0/2}^\infty 2q(2x)\exp\left[-\int_0^{2x} q(y)\,dy\right]dx$$
$$= \exp\left[-\int_0^{r_0} q(y)\,dy\right] > 0 ;$$

hence $K(x,r)$ satisfies (5.7.10). Thus, in this simple model for cell division, we know that there is a globally asymptotically stable distribution of mitogen. □

5.8 Conditions for the existence of lower-bound functions

The consequences of the theorems of this chapter for the Frobenius–Perron operator are so far-reaching that an entire theory of invariant measures for a large class of transformations on the interval $[0,1]$, and even on manifolds, may be constructed. This forms the subject of Chapter 6. In this last section, we develop some simple criteria for the existence of lower-bound functions that will be of use in our specific examples of the next chapter.

Our first criteria for the existence of a lower function will be formulated in the special case when $X = (a,b)$ is an interval on the real line [(a,b) bounded or not] with the usual Borel measure. We will use some standard notions from the theory of differential inequalities [Szarski, 1967]. A function f: $(a,b) \to R$ is called **lower semicontinuous** if

$$\liminf_{\delta \to 0} f(x+\delta) \geqslant f(x) \qquad \text{for } x \in (a,b) .$$

It is **left lower semicontinuous** if

$$\liminf_{\substack{\delta \to 0 \\ \delta > 0}} f(x-\delta) \geqslant f(x) \qquad \text{for } x \in (a,b) .$$

For any function f: $(a,b) \to R$, we define its **right lower derivative** by setting

$$\frac{d_+ f(x)}{dx} = \liminf_{\substack{\delta \to 0 \\ \delta > 0}} \frac{1}{\delta}[f(x+\delta) - f(x)] \qquad \text{for } x \in (a,b) .$$

It is well known that every left lower semicontinuous function f: $(a,b) \to R$, satisfying

$$\frac{d_+ f(x)}{dx} \leqslant 0 \qquad \text{for } x \in (a,b) ,$$

is nonincreasing on (a, b). (The same is true for functions defined on a half-closed interval $[a, b)$.)

For every $f \in D_0$ that is a dense subset of D (Definition 5.6.5) write the trajectory $P^n f$ as

$$P^n f = f_n \qquad \text{for } n \geqslant n_0(f). \tag{5.8.1}$$

Then we have the following proposition.

Proposition 5.8.1. Let $P: L^1((a, b)) \to L^1((a, b))$ be a Markov operator. Assume that there exists a nonnegative function $g \in L^1((a, b))$ and a constant $k \geqslant 0$ such that for every $f \in D_0$ the functions f_n in (5.8.1) are left lower semicontinuous and satisfy the following conditions:

$$f_n(x) \leqslant g(x) \qquad \text{a.e. in } (a, b) \tag{5.8.2}$$

$$\frac{d_+ f_n(x)}{dx} \leqslant k f_n(x) \qquad \text{for all } x \in (a, b). \tag{5.8.3}$$

Then there exists an interval $\Delta \subset (a, b)$ and an $\varepsilon > 0$ such that $h = \varepsilon 1_\Delta$ is a lower function for P^n.

Proof: Let $x_0 < x_1 < x_2$ be chosen in (a, b) such that

$$\int_a^{x_1} g(x)\, dx < \tfrac{1}{4} \quad \text{and} \quad \int_{x_2}^b g(x)\, dx < \tfrac{1}{4}. \tag{5.8.4}$$

Set

$$\varepsilon = \min\{x_1 - x_0, M(x_2 - x_0)^{-1}\}, \qquad M = \tfrac{1}{4}\exp[-k(x_2 - x_0)].$$

Since $\|P^n f\| = 1$, condition (5.8.1) implies

$$\int_a^b f_n(x)\, dx = 1. \tag{5.8.5}$$

Now we are going to show that $h = \varepsilon 1_{(x_0, x_1)}$ is a lower function. Suppose it is not. Then there is $n' \geqslant n_0$ and $y \in (x_0, x_1)$ such that $f_{n'}(y) < h(y) = \varepsilon$. By integrating inequality (5.8.3), we obtain

$$f_{n'}(x) \leqslant f_{n'}(y) e^{k(x-y)} \leqslant \varepsilon/4M \qquad \text{for } x \in [y, x_2]. \tag{5.8.6}$$

Furthermore, since $f_{n'} \leqslant g$, we have

$$\int_a^b f_{n'}(x)\, dx \leqslant \int_a^{x_1} g(x)\, dx + \int_y^{x_2} f_{n'}(x)\, dx + \int_{x_2}^b g(x)\, dx.$$

Finally, by applying inequalities (5.8.4) and (5.8.6), we obtain

$$\int_a^b f_{n'}(x)\,dx \leqslant \tfrac{1}{4} + (x_2 - y)(\varepsilon/4M) + \tfrac{1}{4} \leqslant \tfrac{3}{4},$$

which contradicts equation (5.8.5). ■

Remark 5.8.1. In the proof of Proposition 5.8.1, the left lower semicontinuity of f_n and inequality (5.8.3) were only used to obtain the evaluation

$$f_n(x) \leqslant f_n(y)e^{k(x-y)} \qquad \text{for } x \geqslant y.$$

Therefore Proposition 5.8.1 remains true under this condition; for example, it is true if all f_n are nonincreasing. □

It is obvious that in Proposition 5.8.1 we can replace (5.8.3) by $d_-f_n/dx \geqslant -kf_n$ and assume f_n right lower continuous (or assume f_n nondecreasing; cf. Remark 5.8.1). In the case of a bounded interval, we may omit condition (5.8.2) and replace (5.8.3) by a two-sided inequality. This observation is summarized as follows.

Proposition 5.8.2. Let (a, b) denote a bounded interval and let $P\colon L^1((a,b)) \to L^1((a,b))$ be a Markov operator. Assume that for each $f \in D_0$ the functions f_n in (5.8.1) are differentiable and satisfy the inequality

$$\left|\frac{df_n(x)}{dx}\right| \leqslant kf_n(x) \qquad \text{for all } x \in (a,b), \tag{5.8.7}$$

where $k \geqslant 0$ is a constant independent of f. Then there exists an $\varepsilon > 0$ such that $h = \varepsilon 1_{(a,b)}$ is a lower-bound function.

Proof: As in the preceding proof, we have equation (5.8.5). Set

$$\varepsilon = [1/2(b-a)]e^{-k(b-a)}.$$

Now it is easy to show that $f_n \geqslant h$ for $n \geqslant n_0$. If not, then $f_{n'}(y) < \varepsilon$ for some $y \in (a,b)$ and $n' \geqslant n_0$. Consequently, by (5.8.7),

$$f_{n'}(x) \leqslant f_{n'}(y)e^{k|x-y|} \leqslant [1/2(b-a)].$$

This evidently contradicts (5.8.5). The inequality $f_n \geqslant h$ completes the proof. ■