7

Continuous time systems: an introduction

In previous chapters we concentrated on discrete time systems because they offer a convenient way of introducing many concepts and techniques of importance in the study of irregular behaviors in model systems. Now we turn to a study of continuous time systems.

Continuous and discrete systems differ in several important and interesting ways, which we will touch on throughout the remainder of this book. For example, in a continuous time system, complicated irregular behaviors are possible only if the dimension of the phase space of the system is three or greater. As we have seen, this is in sharp contrast to discrete time processes that can have extremely complicated dynamics in only one dimension. Further, continuous time processes in a finite dimensional phase space are in general invertible, which immediately implies that exactness is a property that will not occur for these systems (recall that noninvertibility is a necessary condition for exactness). However, systems in an infinite dimensional phase space, namely, time delay equations and some partial differential equations, are generally not invertible and, thus, may display exactness.

This chapter is devoted to an introduction of the concept of continuous time systems, an extension of many properties developed previously for discrete time systems, and the development of tools and techniques specifically designed for studying continuous time systems.

7.1 Two examples of continuous time systems

Here a continuous time process in a phase space X is given by a family of mappings

$$S_t : X \to X, \qquad t \geqslant 0 .$$

As illustrated in Figure 7.1.1, the value $S_t(x^0)$ is the position of the system at a time t that started from an initial point $x^0 \in X$ at time $t = 0$. We consider only those processes in which the dynamical law S does not explicitly depend on time

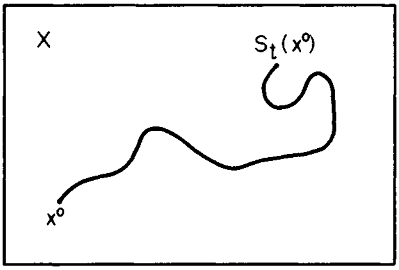

Figure 7.1.1. The trajectory of a continuous time process in the phase space X. At time $t = 0$ the system is at x^0, and at time t it is at $S_t(x^0)$.

so that the property

$$S_t(S_{t'}(x)) = S_{t+t'}(x) \tag{7.1.1}$$

holds. This simply means that the dynamics governing the evolution of the system are the same on the intervals $[0, t']$ and $[t, t + t']$.

Example 7.1.1. A well-known example of a continuous time process is given by an autonomous d-dimensional system of ordinary differential equations

$$\frac{dx}{dt} = F(x) \tag{7.1.2}$$

where $x = (x_1, \ldots, x_d)$ and $F: R^d \to R^d$ is sufficiently smooth to ensure the existence and uniqueness of solutions, such as F is C^1 and satisfies $|F(x)| \leq \alpha + \beta|x|$, with α and β finite. In this case, $S_t(x^0)$ is the solution of (7.1.2) with the initial condition

$$x(0) = x^0. \tag{7.1.3}$$

In this example time t need not be restricted to $t \geq 0$, and the system can also be studied for $t \leq 0$. As we will see in Section 7.8, this is a commonly encountered situation for problems in finite dimensional phase spaces. □

Example 7.1.2. Consider the delay-differential equation

$$\frac{dx(t)}{dt} = F(x(t), x(t - 1)) \tag{7.1.4}$$

with the initial condition

$$x(\tau) = x^0(\tau), \qquad \text{for } \tau \in [-1, 0]. \tag{7.1.5}$$

For rather simple restrictions on F (namely that F is C^1 and $|F(x,y)| \leqslant \alpha(y) + \beta(y)|x|$, where α and β are arbitrary continuous function of y), there is a unique solution to (7.1.4) with (7.1.5) [see Hale, 1977].

Let X be the space of all functions $[-1,0] \to R$ with the usual uniform convergence topology. Given $x^0 \in X$ and the solution x of (7.1.4) and (7.1.5), we may define

$$S_t x^0(\tau) = x(t + \tau), \qquad \text{for } \tau \in [-1,0]. \tag{7.1.6}$$

If $x(t)$ is the solution of (7.1.4)–(7.1.5), then, since F is not an explicit function of t, $x(t + a)$, $a \geqslant 0$, is also a solution to the problem. Using this fact it is easy to verify that transformation (7.1.6) satisfies property (7.1.1), although it is impossible to define $S_t x(\tau)$ for $t < 0$. □

A very important difference exists between these two examples with respect to their invertibility. Thus, although the solution to the system of ordinary differential equations in Example 7.1.1 may be studied for $t \leqslant 0$, in general no solution exists for the differential-delay equation of Example 7.1.2 when $t < 0$. This lack of invertibility is generally the case for delay-differential equations and, indeed, for many continuous time systems whose phase space X is not finite dimensional (e.g., some partial differential equations).

7.2 Dynamical and semidynamical systems

It is possible to establish many results for continuous time processes in a phase space X endowed with no other property than a measure μ, as was done in earlier chapters for discrete time processes. However it is simpler to consider continuous time processes in a measure space that is also equipped with a topology. Thus, from this point on, let X be a topological Hausdorff space and $\mathcal{A}$ the σ-algebra of Borel sets, that is, the smallest σ-algebra that contains all open, and thus closed, subsets of X.

Dynamical systems

Definition 7.2.1. A **dynamical system** $\{S_t\}_{t \in R}$ on X is a family of transformations $S_t: X \to X$, $t \in R$, satisfying

(a) $S_0(x) = x$ for all $x \in X$;
(b) $S_t(S_{t'}(x)) = S_{t+t'}(x)$ for all $x \in X$, with $t, t' \in R$; and
(c) The mapping $(t,x) \to S_t(x)$ from $X \times R$ into X is continuous.

Remark 7.2.1. System (7.1.2) of ordinary differential equations, introduced in the preceding section, is clearly an example of a dynamical system. □

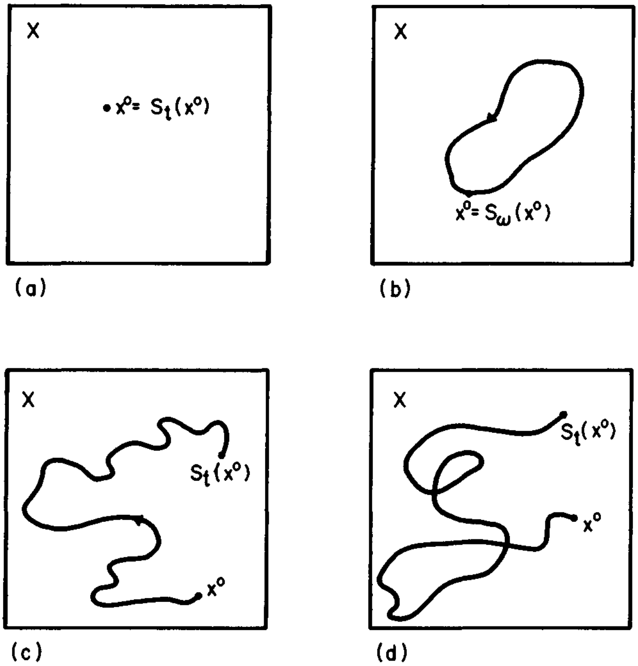

Figure 7.2.1. Trajectories of a dynamical system in its phase space X. In (a) the trajectory is a stationary point, whereas in (b) the trajectory is a periodic orbit. Trajectory (c) is of the nonintersecting type. The intersecting trajectory shown in (d) is not possible in a dynamical system.

Remark 7.2.2. It is clear from the group property of Definition 7.2.1 that

$$S_t(S_{-t}(x)) = x \quad \text{and} \quad S_{-t}(S_t(x)) = x \qquad \text{for all } t \in R.$$

Thus, for all $t_0 \in R$, any transformation S_{t_0} of a dynamical system $\{S_t\}_{t \in R}$ is invertible. □

In applied problems the space X is customarily called the **phase space** of the dynamical system $\{S_t\}_{t \in R}$, whereas, for every fixed $x^0 \in X$, the function $S_t(x^0)$, considered as a function of t, is called a **trajectory** of the system. The trajectories of a dynamical system $\{S_t\}_{t \in R}$ in its phase space X are of only three possible types, as shown in Figure 7.2.1a, b, c for $X = R^2$. First, (Figure 7.2.1a), the trajectory can be a stationary point x^0 such that

$$S_t(x^0) = x^0 \qquad \text{for all } t \in R.$$

Second, as shown in Figure 7.2.1b, the trajectory can be periodic with period $\omega > 0$, that is,

$$S_{t+\omega}(x^0) = S_t(x^0) \qquad \text{for all } t \in R.$$

Finally, the trajectory can be nonintersecting (see Figure 7.2.1c), by which we mean that

$$S_{t_1}(x^0) \neq S_{t_2}(x^0) \qquad \text{for all } t_1 \neq t_2, \qquad \text{with } t_1, t_2 \in R\,.$$

It is straightforward to show that the trajectory of a dynamical system cannot be of the intersecting nonperiodic form shown in Figure 7.2.1d. To demonstrate this, assume the contrary, that, for a given $x^0 \in X$, we have

$$S_{t_1}(x^0) = S_{t_2}(x^0) \qquad t_2 > t_1\,.$$

By applying S_{t-t_1} to both sides of this equation, we have

$$S_{t-t_1}(S_{t_1}(x^0)) = S_{t-t_1}(S_{t_2}(x^0))\,.$$

By the group property (b) of Definition 7.2.1, we also have

$$S_{t-t_1}(S_{t_1}(x^0)) = S_t(x^0)$$

and

$$S_{t-t_1}(S_{t_2}(x^0)) = S_{t+(t_2-t_1)}(x^0)\,.$$

Hence, with $\omega = (t_2 - t_1)$, our assumption leads to

$$S_t(x^0) = S_{t+\omega}(x^0)\,,$$

implying that the only possible intersecting trajectories of a dynamical system are periodic.

However, it is often the case that the evolution in time of data is observed to be of the intersecting nonperiodic type. For example, the two-dimensional projection of the trajectory of a three-dimensional system might easily be of this type. The projection of a trajectory of a dynamical system is called the trace of the system. The following is a more precise definition.

Definition 7.2.2. Let X and Y be two topological Hausdorff spaces, $\phi\colon Y \to X$ a given continuous function, and $S_t\colon Y \to Y$ a given dynamical system on Y. A function $g\colon R \to X$ is called the **trace** of the dynamical system $\{S_t\}_{t\in R}$ if there is a $y \in Y$ such that

$$g(t) = \phi(S_t(y)) \qquad \text{for all } t \in R\,.$$

From our precise definition of the trace of a dynamical system, the following obvious question arises: Given an observed continuous function in a space X that is intersecting and nonperiodic, when is this function the trace of a dynamical system $\{S_t\}_{t\in R}$ operating in some higher-dimensional phase space Y? The answer is as surprising as it is simple, for it turns out that *all continuous functions in X are traces of a single dynamical system!* This is stated more formally in the following theorem.

Theorem 7.2.1. Let X be an arbitrary topological Hausdorff space. Then there is another topological Hausdorff space Y, a dynamical system $(S_t)_{t\in R}$ operating in Y, and a continuous function $\phi: Y \to X$ such that every continuous function $g: R \to X$ is the trace of $\{S_t\}$ that is, for every g there is a $y \in Y$ such that

$$g(t) = \phi(S_t(y)) \qquad \text{for all } t \in R.$$

Proof: Let Y be the space of all continuous functions from R into X (note that the elements of space Y are functions, not points). Let a dynamical system $\{S_{t'}\}_{t'\in R}, S_{t'}: Y \to Y$, operating on Y, be a simple shift so that starting from a given $y \in Y$ we have, after the operation of $S_{t'}$, a new function $y(t + t')$. This may be represented by a diagram,

$$y(t) \xrightarrow{S_{t'}} y(t + t'),$$

or, more formally,

$$S_{t'}(y)(t) = y(t + t').$$

Define a projection $\phi: Y \to X$ by

$$\phi(y) = y(0),$$

then projection ϕ is just the evaluation of y at point $t = 0$. Let $g: R \to X$ be an arbitrary continuous function so that, by our definitions,

$$S_{t'}(g)(t) = g(t + t')$$

and

$$\phi(S_{t'}(g)) = S_{t'}g(0) = g(t'),$$

showing that g is the trace of a trajectory of the dynamical system $\{S_{t'}\}_{t'\in R}$ operating in Y, namely, a trajectory starting from the initial point g. Further, Y will be a topological Hausdorff space, and $(t', y) \to S_{t'}(y)$ a continuous mapping if we equip the function space Y with the topology of uniform convergence on compact intervals. This, coupled with the trivial observation that $S_{t'+t''}(y) = S_{t'}(S_{t''}(y))$ and $S_0(y) = y$, complete the proof of the theorem. ■

Remark 7.2.3. Note that the proof of this theorem rests on the identification of the functions on X as the objects on which the new dynamical system $\{S_{t'}\}_{t'\in R}$ operates. □

Semidynamical systems

Definition 7.2.3. A **semidynamical system** $\{S_t\}_{t\geq 0}$ on X is a family of transformations $S_t: X \to X$, $t \in R^+$, satisfying

(a) $S_0(x) = x$ for all $x \in X$;
(b) $S_t(S_{t'}(x)) = S_{t+t'}(x)$ for all $x \in X$, with $t, t' \in R^+$;
(c) The mapping $(t, x) \to S_t(x)$ from $X \times R^+$ into X is continuous.

Remark 7.2.4. The only difference between dynamical and semidynamical systems is contained in the group property [compare conditions (b) of Definitions 7.2.1 and 7.2.3]. The consequence of this difference is most important, however, because semidynamical systems, in contrast to dynamical systems, are not invertible. It is this property that makes the study of semidynamical systems so important for applications. Henceforth, we will confine our attention to semidynamical systems. □

Remark 7.2.5. An examination of the proof of Theorem 7.2.1 shows that it is also true for semidynamical systems. □

Remark 7.2.6. On occasion a family of transformations $\{S_t\}_{t\geq 0}$ satisfying properties (a) and (b) will be called a **semigroup of transformations.** This is because property (b) in Definition 7.2.3 ensures that transformations S_t form an Abelian semigroup in which the group operation is the composition of two functions. Thus a semidynamical system is a **continuous semigroup.** □

Remark 7.2.7. The area of topological dynamics examines the behavior of semidynamical systems from a topological perspective. Here, however, since we are primarily interested in highly irregular behaviors, our main tools will be measures on X. □

7.3 Invariance, ergodicity, mixing, and exactness in semidynamical systems

Invariance and the individual ergodic theorem

From the continuity property (c) of Definition 7.2.3, all our transformations S_t are measurable, that is, for all $A \in \mathcal{A}$,

$$S_t^{-1}(A) \in \mathcal{A}$$

where, as usual, $S_t^{-1}(A)$ denotes the counterimage of A, namely, the set of all points x such that $S_t(x) \in A$. Thus we can state the following definition.

Definition 7.3.1. A measure μ is called **invariant** under a family $\{S_t\}$ of measurable transformations $S_t: X \to X$ if

$$\mu(S_t^{-1}(A)) = \mu(A) \qquad \text{for all } A \in \mathcal{A}. \tag{7.3.1}$$

As for discrete time processes, we will say interchangeably either that a measure μ is invariant under $\{S_t\}$ or that transformations $\{S_t\}$ are **measure preserving** when equation (7.3.1) holds.

Given a finite invariant measure μ, we can formulate a continuous time analog of Theorem 4.2.3, which is also known as the Birkhoff individual ergodic theorem.

Theorem 7.3.1. Let μ be a finite invariant measure with respect to the semidynamical system $\{S_t\}_{t\geq 0}$, and let $f\colon X \to R$ be an arbitrary integrable function. Then the limit

$$f^*(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(S_t(x))\,dt \tag{7.3.2}$$

exists for all $x \in X$ except perhaps for a set of measure zero.

Proof: This theorem may be rather easily demonstrated using the corresponding discrete time result, Theorem 4.2.3, if we assume, in addition, that for almost all $x \in X$ the integrand $f(S_t(x))$ is a bounded measurable function of t.

Set

$$g(x) = \int_0^1 f(S_t(x))\,dt$$

and assume at first the T is an integer, $T = n$. Note also that the group property (b) of semidynamical systems implies that

$$f(S_t(x)) = f(S_{t-k}(S_k(x)))\,.$$

Then the integral on the right-hand side of (7.3.2) may be written as

$$\begin{aligned}\frac{1}{T}\int_0^T f(S_t(x))\,dt &= \frac{1}{n}\int_0^n f(S_t(x))\,dt \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\int_k^{k+1} f(S_t(x))\,dt \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\int_k^{k+1} f(S_{t-k}(S_k(x)))\,dt \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\int_0^1 f(S_{t'}(S_k(x)))\,dt' \\ &= \frac{1}{n}\sum_{k=0}^{n-1} g(S_k(x))\,.\end{aligned}$$

However, $S_k = S_1 \circ S_{k-1} = S_1 \overset{k}{\circ\circ\circ} S_1 = S_1^k$, so that

$$\lim_{n\to\infty}\frac{1}{n}\int_0^n f(S_t(x))\,dt = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} g(S_1^k(x)),$$

and the right-hand side exists by Theorem 4.2.3. Call this limit $f^*(x)$.

If T is not an integer, let n be the largest integer such that $n < T$. Then we may write

$$\frac{1}{T}\int_0^T f(S_t(x))\,dt = \frac{n}{T}\cdot\frac{1}{n}\int_0^n f(S_t(x))\,dt + \frac{1}{T}\int_n^T f(S_t(x))\,dt.$$

As $T \to \infty$, the first term on the right-hand side converges to $f^*(x)$, as we have shown previously, whereas the second term converges to zero since $f(S_t(x))$ is bounded. ∎

As in the discrete time case, the limit $f^*(x)$ satisfies two conditions:

(C1) $\quad f^*(S_t(x)) = f^*(x), \quad$ a.e. in x for every $t \geq 0$ (7.3.3)

and

(C2) $\quad \int_X f^*(x)\,dx = \int_X f(x)\,dx.$ (7.3.4)

Ergodicity and mixing

We now develop the notions of ergodicity and mixing for semidynamical systems. Exact semidynamical systems are considered in the next section.

Under the action of a semidynamical system $\{S_t\}_{t\geq 0}$, a set $A \in \mathcal{A}$ is called **invariant** if

$$S_t^{-1}(A) = A \qquad \text{for } t \geq 0. \tag{7.3.5}$$

Again we require that for every $t \geq 0$ the equality (7.3.5) is satisfied modulo zero (see Remark 3.1.3). By using this notion of invariant sets, we can define ergodicity for semidynamical systems.

Definition 7.3.2. A semidynamical system $\{S_t\}_{t\geq 0}$, consisting of nonsingular transformations $S_t\colon X \to X$ is **ergodic** if every invariant set $A \in \mathcal{A}$ is such that either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. (Recall that a set A for which $\mu(A) = 0$ or $\mu(X \setminus A) = 0$ is called **trivial.**)

Example 7.3.1. Again we consider the example of rotation on the unit circle, originally introduced in Example 4.2.2. Now $X = [0, 2\pi)$ and

$$S_t(x) = x + \omega t \qquad (\mathrm{mod}\ 2\pi). \tag{7.3.6}$$

S_t is measure preserving (with respect to the natural Borel measure on the circle) and, for $\omega \neq 0$, it is also ergodic. To see this, first pick $t = t_0$ such that $\omega t_0/2\pi$ is irrational. Then the transformation $S_{t_0}: X \to X$ is ergodic, as was shown in Example 4.4.1. Since S_{t_0} is ergodic for at least one t_0, every (invariant) set A that satisfies $S_{t_0}^{-1}(A) = A$ must be trivial by Definition 4.2.1. Thus, any set A that satisfies (7.3.5) must likewise be trivial, and the semidynamical system $\{S_t\}_{t\geq 0}$ with S_t given by (7.3.6) is ergodic. □

Remark 7.3.1. It is interesting to note that, for any t_0 commensurate with $2\pi/\omega$ (e.g., $t_0 = \pi/\omega$), the transformation S_{t_0} is not ergodic. This curious result illustrates a very general property of semidynamical systems: For a given ergodic semidynamical system $\{S_t\}_{t\geq 0}$, there might be a specific t_0 for which S_{t_0} is not ergodic. However, if at least one S_{t_0} is ergodic, then the entire semidynamical system $\{S_t\}_{t\geq 0}$ is ergodic. □

We now turn our attention to mixing in semidynamical systems, starting with the following definition.

Definition 7.3.3. A semidynamical system $\{S_t\}_{t\geq 0}$ on a measure space $(X, \mathcal{A}, \mu)$ with a normalized invariant measure μ is **mixing** if

$$\lim_{t\to\infty} \mu(A \cap S_t^{-1}(B)) = \mu(A)\mu(B) \qquad \text{for all } A, B \in \mathcal{A}. \tag{7.3.7}$$

Thus, in continuous time systems, the interpretation of mixing is the same as for discrete time systems. For example, consider all points x in the set $A \cap S_t^{-1}(B)$, that is, points x such that $x \in A$ and $S_t(x) \in B$. From (7.3.7), for large t the measure of these points is just $\mu(A)\mu(B)$, which means that the fraction of points starting in A that eventually are in B is given by the product of the measures of A and B in the phase space X.

By Definition 7.3.3 the semidynamical system $\{S_t\}_{t\geq 0}$, consisting of rotation on the unit circle given by (7.3.6), is evidently not mixing. This is because, given any two nontrivial disjoint sets $A, B \in \mathcal{A}$, the left-hand side of (7.3.7) is always zero for $\omega t = 2\pi n$ (n an integer), whereas $\mu(A)\mu(B) \neq 0$. A continuous time system that is mixing is illustrated in Example 7.7.2.

Remark 7.3.2. The concepts of ergodicity and mixing are also applicable to dynamical systems. In this case, condition (7.3.7) can be replaced by

$$\lim_{t\to\infty} \mu(A \cap S_t(B)) = \mu(A)\mu(B) \tag{7.3.8}$$

since

$$\mu(A \cap S_t(B)) = \mu(S_t^{-1}(A \cap S_t(B))) = \mu(S_t^{-1}(A) \cap B). \quad □$$

Exactness

Definition 7.3.4. Let $(X, \mathcal{A}, \mu)$ be a normalized measure space. A measure-preserving semidynamical system $\{S_t\}_{t\geq 0}$ such that $S_t(A) \in \mathcal{A}$ for $A \in \mathcal{A}$ is **exact** if

$$\lim_{t\to\infty} \mu(S_t(A)) = 1 \qquad \text{for all } A \in \mathcal{A},\ \mu(A) > 0\,. \tag{7.3.9}$$

Example 11.1.1 illustrates exactness for a continuous time semidynamical system.

Remark 7.3.3. As in discrete time systems, exactness of $\{S_t\}_{t\geq 0}$ implies that $\{S_t\}_{t\geq 0}$ is mixing. □

Remark 7.3.4. Due to their invertibility, dynamical systems cannot be exact. This is easily seen, since $\mu(S_{-t}(S_t(A))) = \mu(A)$ and, thus, the limit in (7.3.9) is $\mu(A)$ and not 1, for all $A \in \mathcal{A}$. If the system is nontrivial and contains a set A such that $0 < \mu(A) < 1$, then, of course, condition (7.3.9) is not satisfied. □

7.4 Semigroups of the Frobenius–Perron and Koopman operators

As we have seen in the discrete time case, many properties of dynamical systems are more easily studied by examining ensembles of trajectories rather than single trajectories. This is primarily because the ensemble approach leads to semigroups of linear operators, and, hence, the techniques of linear functional analysis may be applied to a study of their properties.

Since, for any fixed t in a semidynamical system $\{S_t\}_{t\geq 0}$, the transformation S_t is measurable, we can adopt the discrete time definitions of the Frobenius–Perron and Koopman operators directly for the continuous time case.

Frobenius–Perron operator

Assume that a measure μ on X is given and that all transformations S_t of a semidynamical system $\{S_t\}_{t\geq 0}$ are nonsingular, that is,

$$\mu(S_t^{-1}(A)) = 0 \qquad \text{for each } A \in \mathcal{A} \text{ such that } \mu(A) = 0\,.$$

Then, analogously to (3.2.2), the condition

$$\int_A P_t f(x)\mu(dx) = \int_{S_t^{-1}(A)} f(x)\mu(dx) \qquad \text{for } A \in \mathcal{A} \tag{7.4.1}$$

for each fixed $t \geq 0$ uniquely defines the Frobenius–Perron operator $P_t: L^1(X) \to L^1(X)$, corresponding to the transformation S_t.

It is easy to show, with the aid of (7.4.1), that P_t has the following properties:

(FP1) $P_t(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 P_t f_1 + \lambda_2 P_t f_2$ for all $f_1, f_2 \in L^1$, $\lambda_1, \lambda_2 \in R$; (7.4.2)

(FP2) $P_t f \geq 0$ if $f \geq 0$; (7.4.3)

(FP3) $\int_X P_t f(x)\mu(dx) = \int_X f(x)\mu(dx)$ for all $f \in L^1$. (7.4.4)

Thus, for every fixed t, the operator $P_t: L^1(X) \to L^1(X)$ is a **Markov operator.**

The entire family of Frobenius–Perron operators $P_t: L^1(X) \to L^1(X)$ satisfies some properties similar to (a) and (b) of Definition 7.2.3. To see this, first note that since $S_{t+t'} = S_t \circ S_{t'}$, then $S_{t+t'}^{-1} = S_{t'}^{-1}(S_t^{-1})$ and, thus,

$$\begin{aligned}\int_A P_{t+t'} f(x)\mu(dx) &= \int_{S_{t+t'}^{-1}(A)} f(x)\mu(dx) = \int_{S_{t'}^{-1}(S_t^{-1}(A))} f(x)\mu(dx)\\ &= \int_{S_t^{-1}(A)} P_{t'} f(x)\mu(dx)\\ &= \int_A P_t(P_{t'} f(x))\mu(dx).\end{aligned}$$

This implies that

$$P_{t+t'} f = P_t(P_{t'} f) \qquad \text{for all } f \in L^1(X),\ t, t' \geq 0 \tag{7.4.5}$$

and, thus, P_t satisfies a group property analogous to (b) of Definition 7.2.3.

Further, since $S_0(x) = x$, we have $S_0^{-1}(A) = A$ and, consequently,

$$\int_A P_0 f(x)\mu(dx) = \int_{S_0^{-1}(A)} f(x)\mu(dx) = \int_A f(x)\mu(dx)$$

implying that

$$P_0 f = f \qquad \text{for all } f \in L^1(X). \tag{7.4.6}$$

Hence P_t satisfies properties (a) and (b) of the definition of a semidynamical system.

The properties of P_t in (7.4.2)–(7.4.6) are important enough to warrant the following definition.

Definition 7.4.1. Let $(X, \mathscr{A}, \mu)$ be a measure space. A family of operators $P_t: L^1(X) \to L^1(X)$, $t \geq 0$, satisfying properties (7.4.2)–(7.4.6) is called a **stochastic semigroup.** Further, if, for every $f \in L^1$ and $t_0 \geq 0$,

$$\lim_{t\to t_0} \|P_tf - P_{t_0}f\| = 0,$$

then this semigroup is called **continuous.**

A very important and useful property of stochastic semigroups is that

$$\|P_tf_1 - P_tf_2\| \leqslant \|f_1 - f_2\| \qquad \text{for } f_1, f_2 \in L^1, \tag{7.4.7}$$

and, thus, from the group property (7.4.5), the function $t \to \|P_tf_1 - P_tf_2\|$ is a nonincreasing function of t. This is simply shown by

$$\|P_{t+t'}f_1 - P_{t+t'}f_2\| = \|P_{t'}(P_tf_1 - P_tf_2)\| \leqslant \|P_tf_1 - P_tf_2\|,$$

which follows from (7.4.7). By using this property, we may now proceed to prove a continuous time analog of Theorem 5.6.2.

Theorem 7.4.1. Let $\{P_t\}_{t\geqslant 0}$ be a semigroup of Markov operators, not necessarily continuous. Assume that there is an $h \in L^1$, $h(x) \geqslant 0$, $\|h\| > 0$ such that

$$\lim_{t\to\infty} \|(P_tf - h)^-\| = 0 \qquad \text{for every } f \in D. \tag{7.4.8}$$

Then there is a unique density f_* such that $P_tf_* = f_*$ for all $t \geqslant 0$. Furthermore,

$$\lim_{t\to\infty} P_tf = f_* \qquad \text{for every } f \in D. \tag{7.4.9}$$

Proof: Take any $t_0 > 0$ and define $P = P_{t_0}$ so that $P_{nt_0} = P^n$. Then, from (7.4.8)

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

Thus, by Theorem 5.6.2, there is a unique $f_* \in D$ such that $Pf_* = f_*$ and

$$\lim_{n\to\infty} P^nf = f_* \qquad \text{for every } f \in D.$$

Having shown that $P_tf_* = f_*$ for the set $\{t_0, 2t_0, \ldots\}$, we now turn to a demonstration that $P_tf_* = f_*$ for all t. Pick a particular time t', set $f_1 = P_{t'}f_*$, and note that $f_* = P^nf_* = P_{nt_0}f_*$. Therefore,

$$\begin{aligned}\|P_{t'}f_* - f_*\| &= \|P_{t'}(P_{nt_0}f_*) - f_*\| \\ &= \|P_{nt_0}(P_{t'}f_*) - f_*\| \\ &= \|P^n(P_{t'}f_*) - f_*\| \\ &= \|P^nf_1 - f_*\|. \end{aligned} \tag{7.4.10}$$

Thus, since,

$$\lim_{n\to\infty} \|P^nf_1 - f_*\| = 0$$

and the left-hand side of (7.4.10) is independent of n, we must have $\|P_{t'}f_* - f_*\| = 0$ so $P_{t'}f_* = f_*$. Since t' is arbitrary, we have $P_t f_* = f_*$ for all $t \geq 0$.

Finally, to show (7.4.9), pick a function $f \in D$ so that $\|P_t f - f_*\| = \|P_t f - P_t f_*\|$ is a nonincreasing function. Pick a subsequence $t_n = nt_0$. We know from before that $\lim_{n\to\infty} \|P_{t_n} f - f_*\| = 0$. Thus we have a nonincreasing function that converges to zero on a subsequence and, hence

$$\lim_{t\to\infty} \|P_t f - f_*\| = 0 . \quad \blacksquare$$

Remark 7.4.1. The proof of this theorem illustrates a very important property of stochastic semigroups: namely, a stochastic semigroup $\{P_t\}_{t\geq 0}$ is called **asymptotically stable** if there exists a unique $f_* \in D$ such that $Pf_* = f_*$ and if condition (7.4.9) holds for every $f \in D$. □

Stochastic semigroups that are not semigroups of Frobenius–Perron operators can arise, as illustrated by the following example.

Example 7.4.1. Let $X = R$, $f \in L^1(X)$, and define $P_t\colon L^1(X) \to L^1(X)$ by

$$P_t f(x) = \int_{-\infty}^{\infty} K(t,x,y) f(y)\, dy, \qquad P_0 f(x) = f(x) , \tag{7.4.11}$$

where

$$K(t,x,y) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{(x-y)^2}{2\sigma^2 t} \right] . \tag{7.4.12}$$

It may be easily shown that the kernel $K(t,x,y)$ satisfies:

(a) $K(t,x,y) \geq 0$;

(b) $\displaystyle\int_{-\infty}^{\infty} K(t,x,y)\, dx = 1;$ and

(c) $\displaystyle K(t + t',x,y) = \int_{-\infty}^{\infty} K(t,x,z)K(t',z,y)\, dz .$

From these properties it follows that P_t defined by (7.4.11) forms a continuous stochastic semigroup. The demonstration that $\{P_t\}_{t\geq 0}$ defined by (7.4.11) and (7.4.12) is not a semigroup of Frobenius–Perron operators is postponed to Remark 7.10.2.

That (7.4.11) and (7.4.12) look familiar should come as no surprise as the function $u(t,x) = P_t f(x)$ is the solution to the heat equation

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} \qquad \text{for } t > 0,\ x \in R \tag{7.4.13}$$

with the initial condition

$$u(0,x) = f(x) \qquad \text{for } x \in R. \quad \square \tag{7.4.14}$$

The Koopman operator

Again let $\{S_t\}_{t\geq 0}$ be a semigroup of nonsingular transformations S_t in our topological Hausdorff space X with Borel sets $\mathcal{A}$ and measure μ. Recall that the S_t are nonsingular if, and only if, for every $A \in \mathcal{A}$ such that $\mu(A) = 0$, $\mu(S_t^{-1}(A)) = 0$. Further, let $f \in L^\infty(X)$. Then the function $U_t f$, defined by

$$U_t f(x) = f(S_t(x)), \tag{7.4.15}$$

is again a function in $L^\infty(X)$. Equation (7.4.15) defines, for every $t \geq 0$, the Koopman operator associated with the transformation S_t. The family of operators $\{U_t\}_{t\geq 0}$, defined by (7.4.15), satisfies all the properties of the discrete time Koopman operator introduced in Section 3.3.

It is also straightforward to show that $\{U_t\}_{t\geq 0}$ is a semigroup. To check this, first note from the defining formula (7.4.15) that

$$\begin{aligned} U_{t+t'} f(x) &= f(S_{t+t'}(x)) = f(S_t(S_{t'}(x))) \\ &= U_t(U_{t'} f(x)), \end{aligned}$$

which implies

$$U_{t+t'} f \equiv U_t(U_{t'} f) \qquad \text{for all } f \in L^\infty.$$

Furthermore, $U_0 f(x) = f(S_0(x)) = f(x)$, or

$$U_0 f \equiv f \qquad \text{for all } f \in L^\infty,$$

so that $\{U_t\}_{t\geq 0}$ is a semigroup.

Finally, the Koopman operator is adjoint to the Frobenius–Perron operator, or

$$\langle P_t f, g\rangle = \langle f, U_t g\rangle \qquad \text{for all } f \in L^1(X),\ g \in L^\infty(X) \text{ and } t \geq 0. \tag{7.4.16}$$

The family of Koopman operators is, in general, not a stochastic semigroup because U_t does not map L^1 into itself (though it does map L^∞ into itself) and satisfies the inequality

$$\text{ess sup}|U_t f| \leq \text{ess sup}|f|$$

instead of preserving the norm. In order to have a common notion for families of operators such as $\{P_t\}$ and $\{U_t\}$, we introduce the following definition.

Definition 7.4.2. Let $L = L^p$, $1 \leq p \leq \infty$. A family $\{T_t\}_{t\geq 0}$ of operators, $T_t\colon L \to L$, defined for $t \geq 0$, is called a **semigroup of contracting linear**

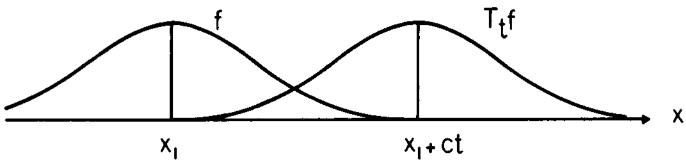

Figure 7.4.1. Plots of $f(x)$ and $T_t f(x) = f(x - ct)$, for $c > 0$.

operators (or a **semigroup of contractions**) if T_t satisfies the following conditions:

(a) $T_t(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 T_t f_1 + \lambda_2 T_t f_2$ for $f_1, f_2 \in L$, $\lambda_1, \lambda_2 \in R$;
(b) $\|T_t f\|_L \leqslant \|f\|_L$ for $f \in L$;
(c) $T_0 f = f$ for $f \in L$; and
(d) $T_{t+t'} f = T_t(T_{t'} f)$ for $f \in L$.

Moreover, if

$$\lim_{t \to t_0} \|T_t f - T_{t_0} f\|_L = 0 \qquad \text{for } f \in L,\ t_0 \geqslant 0,$$

then this semigroup is called **continuous.**

Example 7.4.2. Consider the family of operators $\{T_t\}_{t \geqslant 0}$ defined by (cf. Figure 7.4.1)

$$T_t f = f(x - ct) \qquad \text{for } x \in R,\ t \geqslant 0. \tag{7.4.17}$$

These operators map $L = L^p(R)$, $1 \leqslant p \leqslant \infty$, into itself, satisfy properties (a)–(d) of Definition 7.4.2, and form a semigroup of contractions.

To see that property (b) holds for T_t, use the "change of variables" formula,

$$\|T_t f\|_{L^p}^p = \int_{-\infty}^{\infty} |f(x - ct)|^p\, dx = \int_{-\infty}^{\infty} |f(y)|^p\, dy = \|f\|_{L^p}^p$$

when $p < \infty$, and the obvious equality,

$$\|T_t f\|_{L^\infty} = \operatorname*{ess\,sup}_x |f(x - ct)| = \operatorname*{ess\,sup}_x |f(x)| = \|f\|_{L^\infty}$$

when $p = \infty$. The remaining properties (a), (c), and (d) follow immediately from the definition of T_t in equation (7.4.17).

Finally, we note that if $p = 1$ then this semigroup of contractions is continuous. To see this, first use

$$\begin{aligned}\|T_t f - T_{t_0} f\|_{L^1} &= \int_{-\infty}^{\infty} |f(x - ct) - f(x - ct_0)|\, dx \\ &= \int_{-\infty}^{\infty} |f(y) - f(y - c(t_0 - t))|\, dy\end{aligned}$$

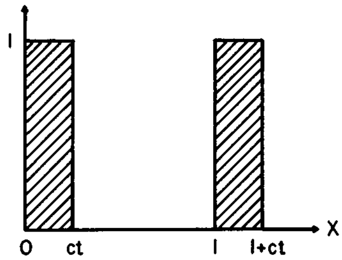

Figure 7.4.2. Function $|1_{(1,1+ct]}(x) - 1_{[0,ct)}(x)|$ versus x.

and note that the right-hand side converges to zero by Corollary 5.1.1. A slightly more complicated calculation shows that $\{T_t\}_{t\geq 0}$ is a continuous semigroup of contractions for every $1 \leq p < \infty$. However, in L^∞ the semigroup $\{T_t\}_{t\geq 0}$ given by (7.4.17) is not continuous except in the trivial case when $c = 0$. This may be easily shown by setting $f = 1_{[0,1]}$. We then have

$$T_t f(x) = 1_{[0,1]}(x - ct) = 1_{[ct,ct+1]}(x)$$

and, as a consequence,

$$\|T_t f - f\|_{L^\infty} = \operatorname*{ess\,sup}_x |1_{(1,1+ct]}(x) - 1_{[0,ct)}(x)| = 1$$

for $0 < ct < 1$. Thus $\|T_t f - f\|_{L^\infty}$ does not converge to zero as $t \to 0$. This may be simply interpreted as shown in Figure 7.4.2 where the hatched areas corresponding to the function $|1_{(1,1+ct]} - 1_{[0,ct)}|$ disappear as $t \to 0$ but the heights do not. □

7.5 Infinitesimal operators

The problems associated with the study of continuous time processes are more difficult than those encountered in discrete time systems. This is partially due to concerns over continuity of processes with respect to time. Also, equivalent formulations of discrete and continuous time properties may appear more complicated in the continuous case because of the use of integrals rather than summations, for example, in the Birkhoff ergodic theorem. However, there is one great advantage in the study of continuous time problems over discrete time dynamics, and this is the existence of a new tool – the infinitesimal operator.

In the case of a semidynamical system $\{S_t\}_{t\geq 0}$ arising from a system of ordinary differential equations (7.1.2), the infinitesimal operator is simply the function $F(x)$. This connection between the infinitesimal operator and $F(x)$ stems from the formula

$$\lim_{n\to\infty} \frac{x(t) - x(0)}{t} = F(x^0),$$

where $x(t)$ is the solution of (7.1.2) with the initial condition (7.1.3). This can be rewritten in terms of the transformations S_t as

$$\lim_{t\to 0} \frac{S_t(x^0) - x^0}{t} = F(x^0).$$

This relation offers some insight into how the infinitesimal operator may be defined for semigroups of contractions in general, and for semigroups of the Frobenius–Perron and Koopman operators in particular.

Definition 7.5.1. Let $L = L^p$, $1 \leqslant p \leqslant \infty$, and $\{T_t\}_{t\geqslant 0}$ be a semigroup of contractions. We define by $\mathfrak{D}(A)$ the set of all $f \in L$ such that the limit

$$Af = \lim_{t\to 0} \frac{T_t f - f}{t} \tag{7.5.1}$$

exists, where the limit is considered in the sense of strong convergence (cf. Definition 2.3.3). Thus (7.5.1) is equivalent to

$$\lim_{t\to 0}\left\| Af - \frac{T_t f - f}{t}\right\|_L = 0.$$

Operator $A: \mathfrak{D}(A) \to L$ is called the **infinitesimal operator.** It is evident that the subspace $\mathfrak{D}(A)$ is linear or that

$$\lambda_1 f_1 + \lambda_2 f_2 \in \mathfrak{D}(A) \qquad \text{for all } f_1, f_2 \in \mathfrak{D}(A), \text{ and } \lambda_1, \lambda_2 \in R.$$

Furthermore, operator $A: \mathfrak{D}(A) \to L$ is linear or

$$A(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 A f_1 + \lambda_2 A f_2 \qquad \text{for all } f_1, f_2 \in \mathfrak{D}(A) \text{ and } \lambda_1, \lambda_2 \in R.$$

In general, the domain $\mathfrak{D}(A)$ of operator A is not the entire space L.

Before deriving the infinitesimal operators for the Frobenius–Perron and Koopman semigroups, we consider the following example.

Example 7.5.1. Let $X = R$ and $L = L^p(R)$, $1 \leqslant p < \infty$. Consider a semigroup $\{T_t\}_{t\geqslant 0}$ on L defined, as in Example 7.4.2, by

$$T_t f(x) = f(x - ct)$$

(cf. Figure 7.4.1). By the mean value theorem, if f is C^1 on R, then

$$\frac{f(x - ct) - f(x)}{t} = -cf'(x - \theta ct),$$

where $|\theta| \leqslant 1$ and $f' \equiv df/dx$. Thus, if f' is bounded and uniformly continuous on R, then

$$Af = \lim_{t\to 0} \frac{T_t f - f}{t} = -cf',$$

and the limit is uniform on R and consequently strong in L^∞. Further, if f (and thus f') has **compact support** (zero outside a bounded interval), then the limit is strong in every L^p, $1 \leqslant p \leqslant \infty$. Thus, all such f belong to $\mathfrak{D}(A)$ and for them A is just differentiation with respect to x and multiplication by $-c$. □

Having introduced the notion of infinitesimal operators, and illustrated their calculation in Example 7.5.1, we now wish to state a theorem that makes explicit the relation among semigroups of contractions, infinitesimal operators, and differential equations.

First, however, we must define the strong derivative of a function with values in $L = L^p$. Given a function $u: \Delta \to L$, where $\Delta \subset R$, and a point $t_0 \in \Delta$, we define the **strong derivative** $u'(t_0)$ by

$$u'(t_0) = \lim_{t \to t_0} \frac{u(t) - u(t_0)}{t - t_0},$$

where the limit is considered in the sense of strong convergence. This definition is equivalent to

$$\lim_{t \to t_0} \left\| \frac{u(t) - u(t_0)}{t - t_0} - u'(t_0) \right\|_L = 0. \tag{7.5.2}$$

By using this concept, we can see that the value of the infinitesimal operator for $f \in \mathfrak{D}(A)$, Af, is simply the derivative of the function $u(t) = T_t f$ at $t = 0$. The following theorem gives a more sophisticated relation between the strong derivative and the infinitesimal operator.

Theorem 7.5.1. Let $\{T_t\}_{t \geqslant 0}$ be a continuous semigroup of contractions acting on L, and $A: \mathfrak{D}(A) \to L$ the corresponding infinitesimal operator. Further, let $u(t) = T_t f$ for fixed $f \in \mathfrak{D}(A)$. Then $u(t)$ satisfies the following properties:

(1) $u(t) \in \mathfrak{D}(A)$ for $t \geqslant 0$;

(2) $u'(t)$ exists for $t \geqslant 0$; and

(3) $u(t)$ satisfies the differential equation

$$u'(t) = Au(t) \qquad \text{for } t \geqslant 0 \tag{7.5.3}$$

and the initial condition

$$u(0) = f. \tag{7.5.4}$$

Proof: For $t = 0$, properties (1)–(3) are satisfied by assumption. Thus we may concentrate on $t > 0$. Let $t_0 > 0$ be fixed. By the definition of $u(t)$, we have

$$\frac{u(t) - u(t_0)}{t - t_0} = \frac{T_t f - T_{t_0} f}{t - t_0}.$$

Noting that $T_t = T_{t-t_0} T_{t_0}$ for $t > t_0$ this differential quotient may be rewritten as

$$\frac{u(t) - u(t_0)}{t - t_0} = T_{t_0}\left(\frac{T_{t-t_0}f - f}{t - t_0}\right) \qquad \text{for } t > t_0. \tag{7.5.5}$$

Because $f \in \mathfrak{D}(A)$, the limit of

$$\frac{T_{t-t_0}f - f}{t - t_0}$$

exists as $t \to t_0$ and gives Af. Thus the limit of (7.5.5) also exists and is equal to $T_{t_0}Af$. In an analogous fashion, if $t < t_0$, we have $T_{t_0} = T_t T_{t_0-t}$ and, as a consequence,

$$\frac{u(t) - u(t_0)}{t - t_0} = T_t\left(\frac{T_{t_0-t}f - f}{t_0 - t}\right) \qquad \text{for } t < t_0 \tag{7.5.6}$$

and

$$\begin{aligned}\left\|\frac{u(t) - u(t_0)}{t - t_0} - T_{t_0}Af\right\|_L &\leqslant \left\|T_t\left(\frac{T_{t_0-t}f - f}{t_0 - t} - Af\right)\right\|_L \\ &\quad + \|T_tAf - T_{t_0}Af\|_L \leqslant \left\|\frac{T_{t_0-t}f - f}{t_0 - t} - Af\right\|_L \\ &\quad + \|T_tAf - T_{t_0}Af\|_L .\end{aligned}$$

Again, since T_tAf converges to $T_{t_0}Af$ as $t \to t_0$, the limit of (7.5.6) exists as $t \to 0$ and is equal to $T_{t_0}Af$. Thus the existence of the derivative $u'(t_0)$ is proved.

Now we can rewrite equation (7.5.5) in the form

$$\frac{u(t) - u(t_0)}{t - t_0} = \frac{T_{t-t_0}(T_{t_0}f) - (T_{t_0}f)}{t - t_0} \qquad \text{for } t > t_0.$$

Since the limit of the differential quotient on the left-hand side exists as $t \to t_0$, the limit on the right-hand side also exists as $t \to t_0$, and we obtain

$$u'(t_0) = AT_{t_0}f,$$

which proves that $T_{t_0}f \in \mathfrak{D}(A)$ and that $u'(t_0) = Au(t_0)$. ■

Remark 7.5.1. The main property of the set $\mathfrak{D}(A)$ that follows directly from Theorem 7.5.1 is that, for $f \in \mathfrak{D}(A)$, the function $u(t) = T_tf$ is a solution of equations (7.5.3) and (7.5.4). Moreover, the solution can be proved to be unique. Unfortunately, in general $\mathfrak{D}(A)$ is not the entire space L, although it can be proved that, for continuous semigroups of contractions, $\mathfrak{D}(A)$ is dense in L. □

In Theorem 7.5.1, the notion of a function $u: [0, \infty) \to L$, where L is again a space of functions, may seem strange. In fact, u actually represents a function of two variables, t and x, since, for each $t \geqslant 0$, $u(t) \in L^p$. Thus we frequently write $u(t)(x) = u(t,x)$, and equation (7.5.3) is to be interpreted as an equation in two variables.

Applying this theorem to the semigroup considered in Examples 7.4.2 and 7.5.1 with $L = L^p$, $1 \leqslant p < \infty$, it is clear that this semigroup satisfies equation (7.5.3), where

$$u(t,x) = T_t f(x) = f(x - ct)$$

and

$$Af = -c\frac{df}{dx}, \qquad f \in \mathfrak{D}(A).$$

These relations can, in turn, be interpreted as meaning that $u(t,x)$ satisfies the first-order partial differentiation equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \tag{7.5.7}$$

with the initial condition

$$u(0,x) = f(x).$$

Remark 7.5.2. It is important to stress the large difference in the two interpretations of this problem as embodied in equations (7.5.3) and (7.5.7). From the point of view of (7.5.7), $u(t,x)$ is thought of as a *function of isolated coordinates* t and x that evolve independently and whose derivatives $\partial u/\partial t$ and $\partial u/\partial x$ are evaluated at specific points in the (t,x)-plane. However, in the semigroup approach that leads to (7.5.3), we are considering the evolution in time of a *family of functions*, and the derivative $du(t)/dt$ is to be thought of as taken over an entire ensemble of points. This is made somewhat clearer when we take into account that $u(t) = T_t f$ has a time derivative $u'(t_0)$ at a point t_0 if (7.5.2) is satisfied, that is,

$$\lim_{t\to t_0} \int_{-\infty}^{\infty} \left| \frac{u(t)(x) - u(t_0)(x)}{t - t_0} - u'(t_0)(x) \right|^p dx = 0.$$

Moreover, $u(t)(x)$ and $u'(t)(x)$ with fixed t are defined as functions of x up to a set of measure zero. □

7.6 Infinitesimal operators for semigroups generated by systems of ordinary differential equations

We now turn to an explicit calculation of the infinitesimal operators for the semigroups $\{P_t\}_{t\geqslant 0}$ and $\{U_t\}_{t\geqslant 0}$ generated by a d-dimensional system of ordinary differential equations

$$\frac{dx}{dt} = F(x) \tag{7.6.1a}$$

or

$$\frac{dx_i}{dt} = F_i(x), \qquad i = 1, \ldots, d, \tag{7.6.1b}$$

where $x = (x_1, \ldots, x_d)$.

The semigroup of transformations $\{S_t\}_{t\geq 0}$ corresponding to equations (7.6.1) is defined by the formula

$$S_t(x^0) = x(t), \tag{7.6.2}$$

where $x(t)$ is the solution of (7.6.1) corresponding to the initial condition

$$x(0) = x^0. \tag{7.6.3}$$

We will assume that the F_i have continuous derivatives $\partial F_i/\partial x_j$, $i, j = 1, \ldots, d$, and that for every $x^0 \in R^d$ the solution $x(t)$ exists for all $t \in R$. This guarantees that (7.6.2) actually defines a group of transformations. Because of a well-known theorem on the continuous dependence of solutions of differential equations on the initial condition, $\{S_t\}_{t\geq 0}$ is a dynamical system (see Example 7.1.1).

As the derivative of the infinitesimal operator A_K for the Koopman operator is simpler, we start from there. By definition we have

$$U_t f(x^0) = f(S_t(x^0)).$$

Therefore

$$\frac{U_t f(x^0) - f(x^0)}{t} = \frac{f(S_t(x^0)) - f(x^0)}{t} = \frac{f(x(t)) - f(x^0)}{t},$$

so that, if f is continuously differentiable with compact support, then by the mean value theorem

$$\frac{U_t f(x^0) - f(x^0)}{t} = \sum_{i=1}^{d} f_{x_i}(x(\theta t))x_i'(\theta t) = \sum_{i=1}^{d} f_{x_i}(x(\theta t))F_i(x(\theta t)),$$

where $0 < \theta < 1$. Now by using equation (7.6.2), we obtain

$$\frac{U_t f(x^0) - f(x^0)}{t} = \sum_{i=1}^{d} f_{x_i}(S_{\theta t}(x^0))F_i(S_{\theta t}(x^0)). \tag{7.6.4}$$

Since the derivatives f_{x_i} have compact support

$$\lim_{t\to 0} f_{x_i}(S_{\theta t}(x^0))F_i(S_{\theta t}(x^0)) = f_{x_i}(x^0)F_i(x^0)$$

uniformly for all x^0. Thus (7.6.4) has a strong limit in L^∞, and the infinitesimal operator A_K is given by

$$A_K f(x) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i} F_i(x). \tag{7.6.5}$$

Observe that equation (7.6.5) was derived only for functions f with some special properties, namely, continuously differentiable f with compact support. These functions do not form a dense set in L^∞, which is not surprising since it can be proved that the semigroup $\{U_t\}_{t\geq 0}$ is not, in general, continuous in L^∞. It does become continuous in a subspace of L^∞ consisting of all continuous functions with compact support (see Remark 7.6.2).

Hence, if f is continuously differentiable with compact support, then by Theorem 7.5.1 for such f the function

$$u(t,x) = U_t f(x)$$

satisfies the first-order partial differential equation (7.5.3). From (7.6.5) it may be written as

$$\frac{\partial u}{\partial t} - \sum_{i=1}^{d} F_i(x) \frac{\partial u}{\partial x_i} = 0. \tag{7.6.6}$$

Remark 7.6.1. It should be noted that the same equation can be immediately derived for $u(t,x) = f(S_t(x))$ by differentiating the equality $u(t, S_{-t}(x)) = f(x)$ with respect to t. In this case f may be an arbitrary continuously differentiable function, not necessarily having compact support. However, in this case (7.6.6) is satisfied locally at every point (t,x) and is not an evolution equation in L^∞ (cf. Remark 7.5.2). □

We now turn to a derivation of the infinitesimal operator for the semigroup of Frobenius–Perron operators generated by the semigroup of (7.6.1a). This is difficult to do if we start from the formal definition of the Frobenius–Perron operator, that is,

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

However, the derivation is straightforward if we start from the fact that the Frobenius–Perron and Koopman operators are adjoint, that is,

$$\langle P_t f, g\rangle = \langle f, U_t g\rangle, \qquad \text{for } f \in L^1,\ g \in L^\infty. \tag{7.6.7}$$

Subtract $\langle f, g\rangle$ from both sides of (7.6.7) to give

$$\langle P_t f - f, g\rangle = \langle f, U_t g - g\rangle$$

or, after division on both sides by t,

$$\langle (P_t f - f)/t, g\rangle = \langle f, (U_t g - g)/t\rangle. \tag{7.6.8}$$

Now let $f \in \mathfrak{D}(A_{FP})$ and $g \in \mathfrak{D}(A_K)$, where A_{FP} and A_K denote, respectively, the infinitesimal operators for the semigroups of Frobenius–Perron and Koopman operators. Take the limit as $t \to 0$ in (7.6.8) to obtain

$$\langle A_{FP}f, g\rangle = \langle f, A_K g\rangle . \tag{7.6.9}$$

However, from equation (7.6.5) the right-hand side of (7.6.9) can be written as

$$\left\langle f, \sum_{i=1}^{d} F_i \frac{\partial g}{\partial x_i}\right\rangle .$$

provided g is a continuously differentiable function with compact support. If we write out this scalar product explicitly and note that $X = R^d$ and $d\mu = dx_1 \cdots dx_d = dx$, we obtain

$$\begin{aligned}\left\langle f, \sum_{i=1}^{d} F_i \frac{\partial g}{\partial x_i}\right\rangle &= \int_{R^d} f \sum_{i=1}^{d} F_i \frac{\partial g}{\partial x_i}\, dx \\ &= \sum_{i=1}^{d} \int_{R^d} \left\{\frac{\partial (fF_i g)}{\partial x_i} - g\frac{\partial (fF_i)}{\partial x_i}\right\} dx\end{aligned}$$

for $f \in \mathfrak{D}(A_{FP})$, which is also continuously differentiable. Since g has compact support,

$$\sum_{i=1}^{d} \int_{R^d} \frac{\partial (fF_i g)}{\partial x_i}\, dx = 0$$

and thus

$$\begin{aligned}\left\langle f, \sum_{i=1}^{d} F_i \frac{\partial g}{\partial x_i}\right\rangle &= -\sum_{i=1}^{d} \int_{R^d} g\frac{\partial (fF_i)}{\partial x_i}\, dx \\ &= \left\langle -\sum_{i=1}^{d} \frac{\partial (fF_i)}{\partial x_i}, g\right\rangle ,\end{aligned}$$

which is a d-dimensional version of the "integration by parts" formula. From this and equation (7.6.9), we finally obtain

$$\langle A_{FP}f, g\rangle = \left\langle -\sum_{i=1}^{d} \frac{\partial (fF_i)}{\partial x_i}, g\right\rangle .$$

This formula holds for every continuously differentiable $f \in \mathfrak{D}(A_{FP})$ and for every continuously differentiable g with compact support. Such a function g is automatically contained in $\mathfrak{D}(A_K)$.

Therefore

$$A_{FP}f = -\sum_{i=1}^{d} \frac{\partial (fF_i)}{\partial x_i} \tag{7.6.10}$$

for continuously differentiable $f \in \mathfrak{D}(A_{FP})$. Again, by using Theorem 7.5.1, we conclude that the function

$$u(t,x) = P_t f(x)$$

satisfies the partial differential equation (continuity equation)

$$\frac{\partial u}{\partial t} + \sum_{i=1}^{d} \frac{\partial (uF_i)}{\partial x_i} = 0 . \tag{7.6.11}$$

Example 7.6.1. As a special case of the system (7.6.1) of ordinary differential equations, let $d = 2n$ and consider a **Hamiltonian system** whose dynamics are governed by the canonical equations of motion (Hamilton's equations)

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \qquad i = 1, \ldots, n, \tag{7.6.12}$$

where $H(p,q)$ is the system **Hamiltonian.** In systems of this type, q and p are referred to as the generalized position and momenta, respectively, whereas H is called the energy. Equation (7.6.11) for Hamiltonian systems takes the form

$$\frac{\partial u}{\partial t} + \sum_{i=1}^{n} \frac{\partial u}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial u}{\partial p_i}\frac{\partial H}{\partial q_i} = 0 ,$$

which is often written as

$$\frac{\partial u}{\partial t} + [u,H] = 0 ,$$

where $[u,H]$ is the **Poisson bracket** of u with H. For Hamiltonian systems, the change with time of an arbitrary function g of the variables $q_1, \ldots, q_n, p_1, \ldots, p_n$ is given by

$$\frac{dg}{dt} = \sum_{i=1}^{n} \frac{\partial g}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial g}{\partial p_i}\frac{\partial H}{\partial q_i} = [g,H] .$$

In particular, if we take g to be a function of the energy H, then

$$\frac{dg}{dt} = \frac{dg}{dH}\frac{dH}{dt} = \frac{dg}{dH}[H,H] \equiv 0$$

since $[H,H] \equiv 0$. Thus any function of the generalized energy H is a **constant of the motion.** □

Remark 7.6.2. The semigroup of Frobenius–Perron operators $\{P_t\}_{t\geq 0}$ corresponding to the system $\{S_t\}_{t\geq 0}$ generated by equation (7.6.1) is continuous.

To show this note that, since S_t is invertible ($S_t^{-1} = S_{-t}$), by Corollary 3.2.1 we have

$$P_t f(x) = f(S_{-t}(x))J_{-t}(x) , \tag{7.6.13}$$

where J_{-t} is the Jacobian of the transformation S_{-t}. Thus, for every continuous f with compact support,

$$\lim_{t\to t_0} f(S_{-t}(x))J_{-t}(x) = f(S_{-t_0}(x))J_{-t_0}(x)$$

uniformly with respect to x. This implies that

$$\lim_{t\to t_0}\|P_t f - P_{t_0} f\| = \lim_{t\to t_0}\int_{R^d} |P_t f(x) - P_{t_0} f(x)|\,dx = 0$$

since the integrals are, in actuality, over a bounded set. Because continuous functions with compact support form a dense subset of L^1, this completes the proof that $\{P_t\}_{t\geq 0}$ is continuous.

Much the same argument holds for the semigroup $\{U_t\}_{t\geq 0}$ if we restrict ourselves to continuous functions with compact support. In this case, from the relation

$$U_t f(x) = f(S_t(x)),$$

it immediately follows that $U_t f$ is uniformly convergent to $U_{t_0} f$ as $t \to t_0$. For this class of functions the proof of Theorem 7.5.1 can be repeated, thus showing that equation (7.5.3) is true for $f \in \mathfrak{D}(A_k)$. □

In the whole space L^∞, it may certainly be the case that $\{U_t\}_{t\geq 0}$ is not a continuous semigroup. As an example, consider the differential equation

$$\frac{dx}{dt} = -c$$

whose corresponding dynamical system is $S_t x = x - ct$. Thus the semigroup $\{U_t\}_{t\geq 0}$ is given by $U_t f(x) = f(x - ct)$. As we know from Example 7.4.2, when $c \neq 0$, this semigroup is certainly not continuous in L^∞.

The continuity of $\{P_t\}_{t\geq 0}$ is very important since it proves that the set $\mathfrak{D}(A_{FP})$ is dense in L^1. Using equation (7.6.13) it may also be shown that $\mathfrak{D}(A_{FP})$ contains all f with compact support, that have continuous first- and second-order derivatives.

7.7 Applications of the semigroups of the Frobenius–Perron and Koopman operators

After developing the concept of the semigroups of the Frobenius–Perron operators in Section 7.4 and introducing the general notion of an infinitesimal operator in Section 7.5 and of infinitesimal operators for semigroups generated by a system of ordinary differential equations in Section 7.6, we are now in a position to examine the utility and applications of these semigroups to questions concerning the existence of invariant measures and ergodicity. This material forms the core of this and the following section.

Theorem 7.7.1. Let $(X, \mathcal{A}, \mu)$ be a measure space, and $S_t\colon X \to X$ a family of nonsingular transformations. Also let $P_t\colon L^1 \to L^1$ be the Frobenius–Perron operator corresponding to $\{S_t\}_{t\geq 0}$. Then the measure

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

is invariant with respect to $\{S_t\}$ if and only if $P_t f = f$ for all $t \geq 0$.

Proof: The proof is trivial, since μ_f invariant implies

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

which, with the definition of P_t, implies $P_t f = f$. The converse is equally easy to prove. ■

Now assume that μ_f is invariant. Since by the preceding theorem we know that $P_t f = f$, and

$$A_{FP} f = \lim_{t\to 0} \frac{P_t f - f}{t},$$

then $A_{FP} f = 0$. Thus the condition $A_{FP} f = 0$ is necessary for μ_f to be invariant. To demonstrate that $A_{FP} f = 0$ is also sufficient for μ_f to be invariant is not so easy, since we must pass from the infinitesimal operator to the semigroup. To deal with this very general and difficult problem, we must examine the way in which semigroups are constructed from their infinitesimal operators. This construction is very elegantly demonstrated by the Hille–Yosida theorem, which is described in Section 7.8.

Analogously to the way in which the semigroup of the Frobenius–Perron operator is employed in studying invariant measures of a semidynamical system $\{S_t\}_{t\geq 0}$, the semigroup of the Koopman operator can be used to study the ergodicity of $\{S_t\}_{t\geq 0}$.

We start by stating the following theorem.

Theorem 7.7.2. A semidynamical system $\{S_t\}_{t\geq 0}$, with nonsingular transformations $S_t\colon X \to X$, is ergodic if and only if the fixed points of $\{U_t\}_{t\geq 0}$ are constant functions.

Proof: The proof is quite similar to that of Theorem 4.2.1. First note that if $\{S_t\}_{t\geq 0}$ is not ergodic then there is an invariant nontrivial subset $C \subset X$, that is,

$$S_t^{-1}(C) = C \qquad \text{for } t \geq 0.$$

By setting $f = 1_C$, we have

$$U_t f = 1_C \circ S_t = 1_{S_t^{-1}(C)} = 1_C = f.$$

Since C is not a trivial set, therefore f is not a constant function (cf. Theorem 4.2.1). Thus, if $\{S_t\}_{t\geq 0}$ is not ergodic, then there is a nonconstant fixed point of $\{U_t\}_{t\geq 0}$.

Conversely, assume there exists a nonconstant fixed point f of $\{U_t\}_{t\geq 0}$. Then it is possible to find a number r such that the set

$$C = \{x: f(x) < r\}$$

is nontrivial (cf. Figure 4.2.1). Since, for each $t \geq 0$,

$$\begin{aligned} S_t^{-1}(C) &= \{x: S_t(x) \in C\} = \{x: f(S_t(x)) < r\} \\ &= \{x: U_t f < r\} = \{x: f(x) < r\} = C, \end{aligned}$$

subset C is invariant, implying that $\{S_t\}_{t\geq 0}$ is not ergodic. ■

Proceeding further with an examination of the infinitesimal operator generated by the Koopman operator, note that the condition $U_t f = f$, $t \geq 0$, implies that

$$A_K f = \lim_{t\to 0} \frac{U_t f - f}{t} = 0\,.$$

Thus, if the only solutions of $A_K f = 0$ are constant, then the semidynamical system $\{S_t\}_{t\geq 0}$ must be ergodic.

Example 7.7.1. In this example we consider the ergodic motion of a point on a d-dimensional torus, which is a generalization of the rotation of the circle treated in Example 7.3.1. We first note that the unit circle S^1 is a circle of radius 1, or

$$S^1 = \{m: m = e^{ix}, x \in R\}\,.$$

Formally, the **d-dimensional torus** T^d is defined as the Cartesian product of d unit circles S^1, that is,

$$\begin{aligned} T^d &= S^1 \times \overset{d}{\cdots} \times S^1 \\ &= \{(m_1, \ldots, m_d): m_k = e^{ix_k}, x_k \in R, k = 1, \ldots, d\} \end{aligned}$$

(cf. Example 6.8.1 where we introduced the two-dimensional torus). T^d is clearly a d-dimensional Riemannian manifold, and the functions $m_k = e^{ix_k}$, $k = 1, \ldots, d$, give a one to one correspondence between points on the torus T^d and points on the Cartesian product

$$[0, 2\pi) \times \overset{d}{\cdots} \times [0, 2\pi)\,. \tag{7.7.1}$$

The x_k have an important geometrical interpretation since they are lengths on S^1. The natural Borel measure on S^1 is generated by these arc lengths and, by Fubini's theorem, these measures, in turn, generate a Borel measure on T^d. Thus, from a measure theoretic point of view, we identify T^d with the Cartesian product

(7.7.1), and the measure μ on T^d with the Borel measure on R^d. We have, in fact, used exactly this identification in the intuitively simpler cases $d = 1$ (r-adic transformation; see Example 4.1.1 and Remark 4.1.2) and $d = 2$ (Anosov diffeomorphism; see Example 4.1.4 and Remark 4.1.6). The disadvantage of this identification is that curves that are continuous on the torus may not be continuous on the Cartesian product (7.7.1).

Thus we consider a dynamical system $\{S_t\}_{t\in R}$ that, in the coordinate system $\{x_k\}$, is defined by

$$S_t(x_1, \ldots, x_d) = (x_1 + \omega_1 t, \ldots, x_d + \omega_d t) \qquad (\mathrm{mod}\ 2\pi).$$

We call this system **rotation on the torus** with angular velocities $\omega_1, \ldots, \omega_d$. Since $\det(dS_t(x)/dx) = 1$, the transformation S_t preserves the measure. We will prove that $\{S_t\}_{t\in R}$ is ergodic if and only if the angular velocities $\omega_1, \ldots, \omega_d$ are linearly independent over the ring of integers. This linear independence means that the only integers $k_1, \ldots, k_d$ satisfying

$$k_1\omega_1 + \cdots + k_d\omega_d = 0 \tag{7.7.2}$$

are $k_1 = \cdots = k_d = 0$.

To prove this, we will use Theorem 7.7.2. Choose $f \in L^2(T^d)$ and assume $U_t f = f$ for $t \in R$, where $U_t f = f \circ S_t$ is the group of Koopman operators corresponding to S_t. Write f as a Fourier series

$$f(x_1, \ldots, x_d) = \sum a_{k_1\ldots k_d} \exp[i(k_1x_1 + \cdots + k_dx_d)],$$

where the summation is taken over all possible integers $k_1, \ldots, k_d$. Substitution of this series into the identity $f(x) = f(S_t(x))$ yields

$$\sum a_{k_1\ldots k_d} \exp[it(k_1x_1 + \cdots + k_dx_d)] = \sum a_{k_1\ldots k_d} \exp[it(\omega_1k_1 + \cdots + \omega_dk_d)]$$
$$\exp[i(k_1x_1 + \cdots + k_dx_d)].$$

As a consequence we must have

$$a_{k_1\ldots k_d} = a_{k_1\ldots k_d} \exp[i(\omega_1k_1 + \cdots + \omega_dk_d)] \qquad \text{for } t \in R \tag{7.7.3}$$

and all sequences $k_1, \ldots, k_d$. Equation (7.7.3) will be satisfied either when $a_{k_1\ldots k_d} = 0$ or when (7.7.2) holds. If $\omega_1, \ldots, \omega_d$ are linearly independent, then the only Fourier coefficient that can be different from zero is $a_{0\ldots0}$. In this case, then, $f(x) = a_{0\ldots0}$ is constant and the ergodicity of $\{S_t\}_{t\in R}$ is proved.

Conversely, if the $\omega_1, \ldots, \omega_d$ are not linearly independent, and condition (7.7.2) is thus satisfied for a nontrivial sequence $k_1, \ldots, k_d$, then (7.7.3) holds for $a_{k_1\ldots k_d} = 1$. In this case the nonconstant function

$$f(x) = \exp[i(k_1x_1 + \cdots + k_dx_d)]$$

satisfies $f(x) = f(S_t(x))$ and $\{S_t\}_{t\in R}$ is not ergodic. □

Remark 7.7.1. The reason why rotation on the torus is so important stems from its frequent occurrence in applied problems. As a simple example, consider a system of d independent and autonomous oscillators

$$\frac{dp_k}{dt} + \omega_k^2 q_k = 0, \quad \frac{dq_k}{dt} = p_k, \qquad k = 1, \ldots, d, \tag{7.7.4}$$

where $q_1, \ldots, q_d$ are the positions of the oscillators and $p_1, \ldots, p_d$ are their corresponding velocities. For this system the total energy of each oscillator is given by

$$E_k = \tfrac{1}{2}p_k^2 + \tfrac{1}{2}\omega_k^2 q_k^2, \qquad k = 1, \ldots, d,$$

and it is clear that the E_k are constants of the motion. Assuming that $E_1, \ldots, E_d$ are given and positive, equations (7.7.4) may be solved to give

$$p_k(t) = A_k\omega_k \cos(\omega_k t + \alpha_k), \quad q_k(t) = A_k \sin(\omega_k t + \alpha_k),$$

where $A_k = \sqrt{2E_k}/\omega_k$ and the α_k are determined, modulo 2π, by the initial conditions of the system. Set $\tilde{p}_k = p_k/A_k\omega_k$ and $\tilde{q}_k = q_k/A_k$ so that the vector $(\tilde{p}(t), \tilde{q}(t))$ describes the position of a point on a d-dimensional torus moving with the angular velocities $\omega_1, \ldots, \omega_d$. Thus, for fixed and positive $E_1, \ldots, E_d$, all possible trajectories of the system (7.7.4) are described by the group $\{S_t\}_{t\in R}$ of the rotation on the torus.

At first it might appear that the set of oscillators described by (7.7.4) is a very special mechanical system. Such is not the case, as equations (7.7.4) are approximations to a very general situation. We present an argument below that supports this claim.

Consider a Hamiltonian system

$$\frac{dq_k}{dt} = \frac{\partial H}{\partial p_k}, \quad \frac{dp_k}{dt} = -\frac{\partial H}{\partial q_k} \qquad k = 1, \ldots, d. \tag{7.7.5}$$

Typically the energy H has the form

$$H(p, q) = \tfrac{1}{2}\sum_{j,k} a_{jk}(q)p_j p_k + V(q),$$

where the first term represents the kinetic energy and V is a potential function. Because the first term in H is associated with the kinetic energy, the quadratic form $\Sigma_{j,k}\, a_{jk}(q)$ is symmetric and positive definite. Further, if q^0 is a stable equilibrium point, then

$$\left.\frac{\partial V}{\partial q_k}\right|_{q=q^0} = 0 \qquad k = 1, \ldots, d$$

and the quadratic form,

$$\sum_{j,k} \frac{\partial^2 V}{\partial q_j \partial q_k},$$

is also positive definite (we neglect some special cases in which it might be semidefinite). Further, we assume that $H(0, q^0) = V(q^0) = 0$ since the potential is only defined up to an additive constant. Thus, developing H in a Taylor series in the neighborhood of $(0, q^0)$, and neglecting terms of order three and higher, we obtain

$$H(p, q) = \tfrac{1}{2}\sum_{j,k} a_{jk} p_j p_k + \tfrac{1}{2}\sum_{j,k} b_{jk}(q_j - q_j^0)(q_k - q_k^0) \tag{7.7.6}$$

where $a_{jk} = a_{jk}(q^0)$ and $b_{jk} = (\partial^2 V/\partial q_j \partial q_k)|_{q^0}$. Both the quadratic forms $\Sigma_{j,k} a_{jk}$ and $\Sigma_{j,k} b_{jk}$ are symmetric and positive definite. With approximation (7.7.6), the original Hamiltonian equations (7.7.5) may be rewritten as

$$\frac{d(q_k - q_k^0)}{dt} = \sum_j a_{jk} p_j, \quad \frac{dp_k}{dt} = -\sum_j b_{jk}(q_j - q_j^0), \tag{7.7.7}$$

where the variables p_k and $q_k - q_k^0$ denote, respectively, the deviation of the system from the equilibrium point $(0, q^0)$.

Since matrices $A = (a_{jk})$ and $B = (b_{jk})$ are symmetric and positive definite, there exists a nonsingular matrix C such that (Gantmacher, 1959)

$$CBC^T = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_d \end{pmatrix} \quad \text{and} \quad CA^{-1}C^T = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

with positive elements λ_i on the diagonal. By introducing new variables $q - q^0 = C^T\bar{q}$ and $p = C^{-1}\bar{p}$ into equations (7.7.7), we obtain

$$\frac{d\bar{q}_k}{dt} = \bar{p}_k, \qquad \frac{d\bar{p}_k}{dt} = -\lambda_k \bar{q}_k. \tag{7.7.8}$$

This new system is completely equivalent to our system (7.7.4) of independent oscillators with angular velocities $\omega_k^2 = \lambda_k$.

Finally we note that, although our approximation shows the correspondence between rotation on the torus and Hamiltonian systems, the terms we neglected in our expansion of H might play a very important role in modifying the eventual asymptotic behavior of a Hamiltonian system. □

Remark 7.7.2. Note that the statement and proof of Theorem 7.7.2 are virtually identical with the corresponding discrete time result given in Theorem 4.2.1. Indeed, necessary and sufficient conditions for ergodicity, mixing, and exactness using the Frobenius–Perron operator, identical to those in Theorem 4.4.1, can be stated by replacing n by t. Analogously, conditions for ergodicity and mixing in continuous time systems using the Koopman operator can be obtained from Proposition 4.4.1 by setting $n = t$. Since all of these conditions are completely equivalent we will not rewrite them for continuous time systems. □

Example 7.7.2. To illustrate the property of mixing in a continuous time system we consider a model for an ideal gas in R^3 adapted from Cornfeld, Fomin, and Sinai [1982]. However, our proof of the mixing property is based on a different technique. At any given moment of time the state of this system is described by the set of pairs

$$y = \{(x_i, v_i)\}, \qquad x_i \in R^3,\ v_i \in R^3,$$

where x_i denotes the position, and v_i the velocity of a particle. We emphasize that y is a set of pairs and not a sequence of pairs, which means that the coordinate pairs (x_i, v_i) are not taken in any specific order. Physically this means that the particles are not distinguishable. It is further assumed that the gas is sufficiently dilute, both in spatial position and in velocity, so that the only states that must be considered are such that in every bounded set $B \subset R^6$ there is, at most, a finite number of pairs (x_i, v_i).

The collection of all possible states of this gas will be denoted by Y, and we assume that the motion of each particle at the gas is governed by a group of transformations $S_t\colon Y \to Y$ given by

$$S_t(y) = \{(x_i + v_i t, v_i)\} \qquad \text{for } y = \{(x_i, v_i)\},$$

or, more compactly, by $S_t(y) = \{s_t(x_i, v_i)\}$, where $\{s_t\}_{t\in R}$ is the family of transformations in R^6 such that

$$s_t(x, v) = (x + vt, v).$$

Thus particles move with a constant speed and do not interact. The surprising result, proved below, is that this system is mixing.

To study the asymptotic properties of $\{S_t\}_{t\in R}$, we must define a σ-algebra and a measure on Y. We do this by first introducing a special measure on R^6, which is the phase space for the motion of a single particle. Let g be a density on R^3. As usual, the measure associated with g is

$$m_g(A) = \int_A g(v)\, dv$$

for every Borel set $A \subset R^3$, and the measure m in $R^6 = R^3 \times R^3$ is defined as the product of the usual Borel measure and m_g, that is,

$$m(A_1 \times A_2) = \int_{A_1} dx \int_{A_2} g(v)\,dv, \qquad A_1, A_2 \subset R^3.$$

From a physical point of view this definition of the measure simply reflects the fact that the particle positions are uniformly distributed in R^3, whereas the velocities are distributed with a given density g, for instance, the Maxwellian $g(v) = c\,\exp(-|v|^2)$.

With these comments we now proceed to define a σ-algebra and a measure on Y. Let $B_1, \ldots, B_n$ be a given sequence of bounded Borel subsets of R^6 for an arbitrary n, and $k_1, \ldots, k_n$ be a given sequence of integers. We use $C(B_1, \ldots, B_n; k_1, \ldots, k_n)$ to denote the set of all $y = \{(x_i, v_i)\}$ such that the number of elements (x_i, v_i) that belong to B_j is equal to k_j, that is,

$$C(B_1, \ldots, B_n; k_1, \ldots, k_n) = \{y \in Y: {}^{\#}(y \cap B_1) = k_1, \ldots, {}^{\#}(y \cap B_n) = k_n\}, \tag{7.7.9}$$

where ${}^{\#}Z$ denotes the number of elements of the set Z. Sets of the form (7.7.9) are called **cylinders.** If the sets $B_1, \ldots, B_n$ are disjoint, then the cylinder is said to be **proper.** For every proper cylinder, we define

$$\mu(C(B_1, \ldots, B_n;\, k_1, \ldots, k_n)) = \frac{[m(B_1)]^{k_1} \cdots [m(B_n)]^{k_n}}{k_1! \cdots k_n!} \exp\left[-\sum_{i=1}^{n} m(B_i)\right]. \tag{7.7.10}$$

From (7.7.10) it follows immediately that

$$\mu(C(B_1, \ldots, B_n; k_1, \ldots, k_n) = \mu(C(B_1; k_1)) \ldots \mu(C(B_n; k_n)) \tag{7.7.11}$$

whenever the sets $B_1, \ldots, B_n$ are mutually disjoint.

It is also easy to calculate the measure of $C(B_1, B_2; k_1, k_2)$ when B_1 and B_2 are not disjoint by writing C as the union of proper cylinders. Thus, y belongs to $C(B_1, B_2; k_1, k_2)$ if, for some $r \leqslant \min(k_1, k_2)$, the set $B_1^0 = B_1 \backslash B_2$ contains $k_1 - r$ particles, $B^0 = B_1 \cap B_2$ contains r particles, and $B_2^0 = B_2 \backslash B_1$ has $k_2 - r$ particles. As a consequence,

$$C(B_1, B_2; k_1, k_2) = \bigcup_{r=0}^{k} [C(B_1^0; k_1 - r) \cap C(B^0; r) \cap C(B_2^0; k_2 - r)],$$

where $k = \min(k_1, k_2)$, and, thus,

$$\mu(C(B_1, B_2; k_1, k_2)) = \sum_{r=0}^{k} \frac{[m(B_1^0)]^{k_1 - r}[m(B^0)]^r[m(B_2^0)]^{k_2 - r}}{(k_1 - r)!\,r!\,(k_2 - r)!} \cdot \exp[-m(B_1^0) - m(B^0) - m(B_2^0)] \tag{7.7.12}$$

By employing arguments of this type we can calculate the measure μ of any cylinder. However, the formulas for arbitrary cylinders are much more complicated as it is necessary to sum these various contributions first with respect to $q = \binom{n}{2}$ parameters $r_1, \ldots, r_q$, corresponding to all possible intersections $B_i \cap B_j$, $i \neq j$, then with respect to $\binom{n}{3}$ parameters corresponding to all possible intersections $B_i \cap B_j \cap B_l$, $i \neq j \neq l$, and so forth.

With respect to the σ-algebra, we define $\mathscr{A}$ to be the smallest σ-algebra that contains all the cylinders or, equivalently, all proper cylinders. Using standard results from measure theory, it is possible to prove that μ given by (7.7.10) for proper cylinders can be uniquely extended to a measure on $\mathscr{A}$ and that the characteristic functions of proper cylinders

$$1_{C(B_1, \ldots, B_n; k_1, \ldots, k_n)}$$

form a linearly dense subset of $L^2(Y, \mathscr{A}, \mu)$. We omit the proof of these facts as they are quite technical in nature and, instead, turn to consider the asymptotic properties of system $\{S_t\}_{t \in R}$ on the phase space Y.

First we note that the measure μ is normalized. To show this, take an arbitrary bounded Borel set B. Then

$$Y = \bigcup_{k=0}^{\infty} C(B;k)$$

since every y belongs to one of the cylinders $C(B;k)$, namely, the one for which $^{\#}(y \cap B) = k$. As the cylinders $C(B;k)$, $k = 0, 1, \ldots$, are mutually disjoint, we have

$$\mu(Y) = \sum_{k=0}^{\infty} \mu(C(B;k)) = \sum_{k=0}^{\infty} \frac{[m(B)]^k}{k!} e^{-m(B)} = 1 .$$

Second, the measure μ is invariant with respect to $\{S_t\}_{t \in R}$. To show this, note that for every cylinder

$$S_t(C(B_1, \ldots, B_n; k_1, \ldots, k_n)) = C(s_t(B_1), \ldots, s_t(B_n); k_1, \ldots, k_n) .$$

It is clear that $(x, v) \in s_t(B_j)$ if and only if $(\bar{x}, \bar{v}) \in B_j$, where $\bar{x} = x - vt$, $\bar{v} = v$, and, as a consequence,

$$m(s_t(B_j)) = \iint_{s_t(B_j)} g(v)\, dx\, dv = \iint_{B_j} g(\bar{v})\, d\bar{x}\, d\bar{v} = m(B_j) .$$

From this equality, $m(s_t(B_j)) = m(B_j)$ and, from equation (7.7.10), we, therefore, have

$$\mu(S_t(C(B_1, \ldots, B_n; k_1, \ldots, k_n))) = \mu(C(B_1, \ldots, B_n; k_1, \ldots, k_n))$$

for every proper cylinder. Writing $\mu_t(E) = \mu(S_t(E))$ for $E \in \mathscr{A}$, we define for every fixed t a measure μ_t on $\mathscr{A}$ that is identical with μ for proper cylinders. Since μ is uniquely determined by its values on cylinders, we must have $\mu_t(E) = \mu(E)$ for all $E \in \mathscr{A}$, and thus the invariance of μ with respect to S_t is proved.

With these results in hand, we now prove that the dynamical system $\{S_t\}_{t\in R}$ is mixing. Since the characteristic functions of proper cylinders are linearly dense in $L^2(Y, \mathscr{A}, \mu)$, by Remark 7.7.2 it is sufficient to verify the condition

$$\lim_{t\to\infty}\langle U_t 1_{C_1}, 1_{C_2}\rangle = \langle 1_{C_1}, 1\rangle\langle 1, 1_{C_2}\rangle \tag{7.7.13}$$

for every two proper cylinders C_1 and C_2. Since

$$U_t 1_{C_1}(y) = 1_{C_1}(S_t(y)) = 1_{S_{-t}(C_1)}(y)$$

and $\langle 1_{C_j}, 1\rangle = \mu(C_j)$, condition (7.7.13) is equivalent to

$$\lim_{t\to\infty} \mu(S_{-t}(C_1) \cap C_2) = \mu(C_1)\mu(C_2)\,. \tag{7.7.14}$$

We will verify that (7.7.14) holds only in the simplest case when each of the cylinders C_j is determined by only one bounded Borel set. Thus we assume

$$C_j = C(B_j; k_j), \qquad j = 1, 2\,. \tag{7.7.15}$$

(This is not an essential simplification, since the argument proceeds in exactly the same way for arbitrary proper cylinders. However, in the general case the formulas are so complicated that the simple geometrical ideas behind the calculations are obscured.) When the C_j are given by (7.7.15), the right-hand side of equation (7.7.14) may be easily calculated by (7.7.10). Thus

$$\mu(C_1)\mu(C_2) = \frac{[m(B_1)]^{k_1}[m(B_2)]^{k_2}}{k_1!\,k_2!}\exp[-m(B_1) - m(B_2)]\,. \tag{7.7.16}$$

To compute the left-hand side of equation (7.7.14), observe that

$$S_{-t}(C_1) = C(s_{-t}(B_1); k_1)$$

so

$$\begin{aligned}\mu(S_{-t}(C_1) \cap C_2) &= \mu(C(s_{-t}(B_1); k_1) \cap C(B_2; k_2))\\ &= \mu(C(s_{-t}(B_1), B_2; k_1, k_2))\,.\end{aligned} \tag{7.7.17}$$

With (7.7.12) we have

$$\begin{aligned}\mu(S_{-t}(C_1) \cap C_2) = \sum_{r=0}^{k} &\frac{[m(B_1^t)]^{k_1-r}[m(B^t)]^r[m(B_2^t)]^{k_2-r}}{(k_1 - r)!\,r!\,(k_2 - r)!}\\ &\cdot \exp[-m(B_1^t) - m(B^t) - m(B_2^t)]\,,\end{aligned} \tag{7.7.18}$$

where $B_1^t = s_{-t}(B_1)\backslash B_2$, $B^t = s_{-t}(B_1) \cap B_2$, and $B_2^t = B_2\backslash s_{-t}(B_1)$. From our definition of m, we have

$$m(B^t) = \iint\limits_{s_{-t}(B_1)\cap B_2} g(v)\,dx\,dv = \iint\limits_{B_2} 1_{s_{-t}(B_1)}(x,v)g(v)\,dx\,dv$$
$$= \iint\limits_{B_2} 1_{B_1}(x+vt,v)g(v)\,dx\,dv\,.$$

Since B_1 and B_2 are bounded, $1_{B_1}(x + vt, v) = 0$ for almost every point $(x, v) \in B_2$ if t is sufficiently large (except for some points at which $v = 0$). Thus, by the Lebesgue dominated convergence theorem,

$$\lim_{t\to\infty} m(B^t) = 0\,. \tag{7.7.19}$$

Furthermore, since $B_2^t = B_2 \backslash B^t$, it follows that

$$\lim_{t\to\infty} m(B_2^t) = \lim_{t\to\infty}[(m(B_2) - m(B^t))] = m(B_2)\,. \tag{7.7.20}$$

Finally, since $B_1^t = s_{-t}(B_1)\backslash B^t$ and s_t is measure preserving,

$$m(B_1^t) = m(s_{-t}(B_1)) - m(B^t) = m(B_1) - m(B^t)\,,$$

and

$$\lim_{t\to\infty} m(B_1^t) = m(B_1)\,. \tag{7.7.21}$$

Passing to the limit in equation (7.7.18) and using (7.7.19) through (7.7.21) gives

$$\lim_{t\to\infty} \mu(S_{-t}(C_1) \cap C_2) = \left\{\frac{[m(B_1)]^{k_1}[m(B_2)]^{k_2}}{k_1!k_2!}\right\} \exp[-m(B_1) - m(B_2)]\,,$$

which, together with (7.7.16), proves (7.7.14).

From this proof, it should be clear that mixing in this model is a consequence of the following two facts. The first is that, for disjoint B_1 and B_2 and given k_1 and k_2, the events consisting of B_1 containing k_1 particles and B_2 containing k_2 particles are independent [this follows from equation (7.7.11)]. Second, for every two bounded Borel sets B_1 and B_2, the sets $s_{-t}(B_1)$ and B_2 are "almost" disjoint for large t. Taken together these produce the surprising result that mixing can appear in a system without particle interaction. □

Example 7.7.3. The preceding example gave a continuous time, dynamical system that was mixing. The phase space of this system was infinite dimensional. This fact is not essential. There is a large class of finite dimensional, mixing, dynamical systems that play an important role in classical mechanics. In this example we briefly describe these systems. An exhaustive treatment requires highly specialized techniques from differential geometry and cannot be given within the measure-theoretic framework that we have adopted. All necessary information can be found in the books by Arnold and Avez [1968], by Abraham and Marsden [1978], and articles by Anosov [1967] and by Smale [1967].

Let M be a compact connected smooth Riemannian manifold. Having M, we define the sphere bundle Σ as the set of all pairs (m, ξ), where m is an arbitrary point of M and ξ is a unit tangent vector starting at m. This definition can be written as

$$\Sigma = \{(m, \xi): m \in M, \xi \in T_m, \|\xi\| = 1\}.$$

It can be proved that Σ, with an appropriately defined metric, is also a Riemannian manifold. Thus a measure μ_Σ is automatically given on Σ. In a physical interpretation, M is the configuration space of a system that moves with constant speed and Σ is its phase space. To describe precisely the dynamical system that corresponds to this interpretation we need only the concept of geodesics. Let $\gamma: R \to M$ be a C^1 curve. This curve is called a **geodesic** if for every point $m_0 = \gamma(t_0)$ there is an $\varepsilon > 0$ such that for every $m_1 = \gamma(t_1)$, with $|t_1 - t_0| \leqslant \varepsilon$, the length of the arc γ between the points m_0 and m_1 is equal to the distance between m_0 and m_1. It can be proved that, for every $(m, \xi) \in \Sigma$, there exists exactly one geodesic satisfying

$$\gamma(0) = m, \quad \gamma'(0) = \xi, \quad \|\gamma'(t)\| = 1 \quad \text{for } t \in R. \tag{7.7.22}$$

We define a dynamical system $\{S_t\}_{t\in R}$ on Σ by setting

$$S_t(m, \xi) = (\gamma(t), \gamma'(t)) \qquad \text{for } t \in R,$$

where the geodesic γ satisfies (7.7.22). This system is called a **geodesic flow.**

In the case $\dim M = 2$, the geodesic flow has an especially simple interpretation: It describes the motion of a point that moves on the surface M in the absence of external forces and without friction. The motion described by the geodesic flow looks quite specific but, in fact, it represents a rather general situation. If M is the configuration space of a mechanical system with the typical Hamiltonian function (see Remark 7.7.1),

$$H(q, p) = \tfrac{1}{2}\sum_{j,k} a_{jk}(q)p_j p_k + V(q),$$

then it is possible to change the Riemannian metric on M in such a way that trajectories of the system become geodesics.

The behavior of the geodesic flow depends on the geometrical properties of the manifold M and most of all on its curvature. In the simplest case, $\dim M = 2$, the curvature K is a scalar function and has a clear geometrical interpretation. In order to define K at a point $m \in M$, we consider, in a neighborhood W of m, a triangle made by three geodesics. We denote the angles of that triangle by α_1, α_2,

α_3, and its area by σ. Then

$$K(m) = \lim[(\alpha_1 + \alpha_2 + \alpha_3 - \pi)/\sigma],$$

where the limit is taken over a sequence of neighborhoods that shrinks to the point m. In the general case, $\dim M > 2$, the curvature must be defined separately for every two-dimensional section of a neighborhood of the point m. (Thus, in this case, the curvature becomes a tensor.) When the curvature of M is negative, the behavior of the geodesic flow is quite specific and highly chaotic. Such flows have been studied since the beginning of the century, starting with Hadamard [1898]. Results were first obtained for manifolds with constant negative curvature and then finally completed by Anosov [1967]. It follows that the geodesic flow on a compact, connected, smooth Riemannian manifold with negative curvature is mixing and even a K-flow (a continuous time analog of K-automorphism). This fact has some profound consequences for the foundations of classical statistical mechanics. A heuristic geometrical argument of Arnold [1963] shows that the Boltzmann–Gibbs model of a dilute gas (ideal balls with elastic collisions) may be considered as a geodesic flow on a manifold with negative curvature. Thus, such a system is not only ergodic but also mixing. A sophisticated proof of the ergodicity and mixing of the Boltzmann–Gibbs model has been given by Sinai [1963, 1970]. □

7.8 The Hille–Yosida theorem and its consequences

Theorem 7.8.1 (Hille–Yosida). Let $A:\mathfrak{D}(A) \to L$ be a linear operator, where $\mathfrak{D}(A) \subset L$ is a linear subspace of L. In order for A to be an infinitesimal operator for a continuous semigroup of contractions, it is necessary and sufficient that the following three conditions are satisfied:

(a) $\mathfrak{D}(A)$ is dense in L, that is, every point in L is a strong limit of a sequence of points from $\mathfrak{D}(A)$;

(b) For each $f \in L$ there exists a unique solution $g \in \mathfrak{D}(A)$ of the **resolvent equation**

$$\lambda g - Ag = f; \tag{7.8.1}$$

(c) For every $g \in \mathfrak{D}(A)$ and $\lambda > 0$,

$$\|\lambda g - Ag\|_L \geq \lambda \|g\|_L. \tag{7.8.2}$$

Further, if A satisfies (a)–(c), then the semigroup corresponding to A is unique and is given by

$$T_t f = \lim_{\lambda\to\infty} e^{tA_\lambda} f, \qquad f \in L, \tag{7.8.3}$$

where $A_\lambda = \lambda A R_\lambda$ and $R_\lambda f = g$ (the **resolvent operator**) is the unique solution of $\lambda g - Ag = f$.

Consult Dynkin [1965] or Dunford and Schwartz [1957] for the proof.

Operator $A_\lambda = \lambda A R_\lambda$ can be written in several alternative forms, each of which is useful in different situations. Thus, after substitution of $g = R_\lambda f$ into (7.8.1), we have

$$\lambda R_\lambda f - A R_\lambda f = f \qquad \text{for } f \in L. \tag{7.8.4}$$

By applying the operator R_λ to both sides of (7.8.1) and using $g = R_\lambda f$, we also obtain

$$\lambda R_\lambda g - R_\lambda A g = g \qquad \text{for } g \in \mathcal{D}(A). \tag{7.8.5}$$

Equations (7.8.4) and (7.8.5) immediately give

$$R_\lambda A f = A R_\lambda f \qquad \text{for } f \in \mathcal{D}(A). \tag{7.8.6}$$

Equation (7.8.4) also gives

$$A R_\lambda f = (\lambda R_\lambda - I) f \qquad \text{for } f \in L, \tag{7.8.7}$$

where I is the identity operator ($If \equiv f$ for all f). Thus we have three possible representations for A_λ: the original definition,

$$A_\lambda = \lambda A R_\lambda; \tag{7.8.8}$$

or, from (7.8.7),

$$A_\lambda = \lambda(\lambda R_\lambda - I); \tag{7.8.9}$$

and, finally, from (7.8.6),

$$A_\lambda = \lambda R_\lambda A. \tag{7.8.10}$$

The representations in (7.8.8) and (7.8.9) hold in the entire space L, whereas (7.8.10) holds in $\mathcal{D}(A)$.

From conditions (b) and (c) of the Hille–Yosida theorem, using $g = R_\lambda f$, it follows that

$$\|f\|_L \geqslant \lambda \|R_\lambda f\|_L. \tag{7.8.11}$$

Consequently, using (7.8.9),

$$\|A_\lambda f\|_L = \|\lambda^2 R_\lambda f - \lambda f\|_L \leqslant \|\lambda^2 R_\lambda f\|_L + \|\lambda f\|_L \leqslant 2\lambda \|f\|_L,$$

so that the operator $\exp(tA_\lambda)$ can be interpreted as the series

$$e^{tA_\lambda} f = \sum_{n=0}^{\infty} \frac{t^n}{n!} A_\lambda^n f, \tag{7.8.12}$$

which is strongly convergent.

In addition to demonstrating the existence of a semigroup $\{T_t\}_{t\geq 0}$ corresponding to a given operator A, the Hille–Yosida theorem also allows us to determine some properties of $\{T_t\}_{t\geq 0}$.

One very interesting corollary is the following. Suppose we have an operator $A: \mathfrak{D}(A) \to L$ (remembering that $L = L^p$) that satisfies conditions (a)–(c) of the Hille–Yosida theorem, and such that the solution $g = R_\lambda f$ of equation (7.8.1) has the property that $R_\lambda f \geq 0$ for $f \geq 0$. Then, as we will show next, $T_t f \geq 0$ for every $f \geq 0$.

To see this, note that from (7.8.9) we have

$$e^{tA_\lambda}f = e^{-t\lambda}(e^{t\lambda^2 R_\lambda}f)\,, \tag{7.8.13}$$

where

$$e^{t\lambda^2 R_\lambda}f = \sum_{n=0}^{\infty} \frac{t^n\lambda^n}{n!}(\lambda R_\lambda)^n f. \tag{7.8.14}$$

Further, for any $f \geq 0$, $R_\lambda f \geq 0$ and, by induction, $R_\lambda^n f \geq 0$. Thus, from (7.8.14), since $\lambda > 0$ and $t \geq 0$, $\exp(t\lambda^2 R_\lambda)f \geq 0$ and so, from (7.8.13), $\exp(tA_\lambda)f \geq 0$. Finally, from (7.8.3), we have $T_t f \geq 0$ since it is the limit of nonnegative functions.

Now suppose that $L = L^1$ and that the operator λR_λ preserves the integral, that is,

$$\lambda \int_X R_\lambda f(x)\mu(dx) = \int_X f(x)\mu(dx) \qquad \text{for all } f \in L^1,\ \lambda > 0\,. \tag{7.8.15}$$

We will show that these properties imply that

$$\int_X T_t f(x)\mu(dx) = \int_X f(x)\mu(dx), \qquad f \geq 0,\ t \geq 0\,.$$

This is straightforward. Since (7.8.14) is strongly convergent, and using equation (7.8.15), we obtain

$$\begin{aligned}\int_X e^{t\lambda^2 R_\lambda}f(x)\mu(dx) &= \sum_{n=0}^{\infty}\frac{t^n\lambda^n}{n!}\int_X (\lambda R_\lambda)^n f(x)\mu(dx)\\ &= \sum_{n=0}^{\infty}\frac{t^n\lambda^n}{n!}\int_X f(x)\mu(dx) = e^{t\lambda}\int_X f(x)\mu(dx)\,.\end{aligned} \tag{7.8.16}$$

Now,

$$\begin{aligned}\int_X T_t f(x)\mu(dx) &= \lim_{\lambda\to\infty}\int_X e^{tA_\lambda}f(x)\mu(dx)\\ &= \lim_{\lambda\to\infty}\int_X e^{-t\lambda}(e^{t\lambda^2 R_\lambda}f(x))\mu(dx) = \int_X f(x)\mu(dx)\end{aligned}$$

by the use of equation (7.8.16), and the claim is demonstrated.

These two results may be summarized in the following corollary.

Corollary 7.8.1. Let $A: \mathfrak{D}(A) \to L^1$ be an operator satisfying conditions (a)–(c) of the Hille–Yosida theorem. If the solution $g = R_\lambda f$ of (7.8.1) is such that λR_λ is a Markov operator, then $\{T_t\}_{t\geq 0}$ generated by A is a continuous semigroup of Markov operators.

In fact, in this corollary only conditions (a) and (b) of the Hille–Yosida theorem need be checked, as condition (c) is automatically satisfied for any Markov operator.

To see this, set $f = \lambda g - Ag$ and write inequality (7.8.2) in the form

$$\|f\| \geq \|\lambda R_\lambda f\|.$$

This is always satisfied if λR_λ is a Markov operator, as we have shown in Section 3.1 [cf. inequality (3.1.6)].

The Hille–Yosida theorem has several other important applications. The first is that it provides an immediate and simple way to demonstrate that $A_{FP}f = 0$ is a sufficient condition that μ_f is an invariant measure.

Thus, $Af = 0$ implies, from (7.8.10), that $A_\lambda f = 0$ and from (7.8.12)

$$e^{tA_\lambda}f = f.$$

This, combined with (7.8.3), gives

$$T_t f = f \qquad \text{for all } t \geq 0.$$

Thus, in the special case $A_{FP}f = 0$ this implies that $P_t f = f$ and thus μ_f is invariant.

By combining this result with that of Section 7.7, we obtain the following theorem.

Theorem 7.8.2. Let $\{S_t\}_{t\geq 0}$ be a semidynamical system such that the corresponding semigroup of Frobenius–Perron operators is continuous. Under this condition, an absolutely continuous measure μ_f is invariant if and only if $A_{FP}f = 0$.

Consider the special case where A_{FP} is the infinitesimal operator for a d-dimensional system of ordinary differential equations (cf. equation 7.6.10). Then the necessary and sufficient condition that μ_f be invariant, that is $A_{FP}f = 0$, reduces to

$$\sum_{i=1}^{d} \frac{\partial(fF_i)}{\partial x_i} = 0 \tag{7.8.17}$$

for continuously differentiable $f \in L^1$. This result was originally obtained by Liouville using quite different techniques and is known as **Liouville's theorem.**

Remark 7.8.1. Equation (7.8.17) is also a necessary and sufficient condition for the invariance of the measure

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

even if f is an arbitrary continuously differentiable function that is not necessarily integrable on R^d. This is related to the fact that operators $P_t f$ as given by (7.6.13) can also be considered for nonintegrable functions. Thus, if one wishes to determine when the Lebesgue measure

$$\mu(A) = \int_A dx_1 \ldots dx_d = \int_A dx$$

is invariant, it is necessary to substitute its density $f(x) \equiv 1$ into (7.8.17). This gives

$$\sum_{i=1}^{d} \frac{\partial F_i}{\partial x_i} = 0 \tag{7.8.18}$$

as a necessary and sufficient condition for the invariance of the Lebesgue measure. [In many sources, equation (7.8.18) is called Liouville's equation, even though it is a special case of equation (7.8.17).] □

Remark 7.8.2. It is quite straightforward to show that Hamiltonian systems (see Example 7.6.1) satisfy (7.8.18) since

$$\sum_{i=1}^{n} \left[\frac{\partial}{\partial q_i}\left(\frac{\partial H}{\partial p_i}\right) + \frac{\partial}{\partial p_i}\left(-\frac{\partial H}{\partial q_i}\right)\right] = 0$$

automatically, and thus they preserve the Lebesgue measure. □

Returning now to the problem of determining the ergodicity of a semidynamical system $\{S_t\}_{t\geq 0}$, recall that $U_t g = g$ implies $A_K g = 0$. Using this relation and Theorem 7.7.2 we are going to prove the following theorem.

Theorem 7.8.3. Let $\{S_t\}_{t\geq 0}$ be a semidynamical system such that the corresponding semigroup $\{P_t\}$ of Frobenius–Perron operators is continuous. Then $\{S_t\}_{t\geq 0}$ is ergodic if and only if $A_K g = 0$ has only constant solutions in L^∞.

Proof: The "if" part follows from Theorem 7.7.2. The proof of the "only if" part is more difficult since, in general, the semigroup $\{U_t\}$ is not continuous and we cannot use the Hille–Yosida theorem. Thus, assume that $A_K g = 0$ for some nonconstant g. Choose an arbitrary $f \in L^1$ and define the real-valued function ϕ by the formula

$$\phi(t) = \langle f, U_t g\rangle = \langle P_t f, g\rangle .$$

Due to the continuity of $\{P_t\}$, function ϕ is also continuous. Further, we have

$$\frac{\phi(t+h)-\phi(t)}{h} = \left\langle f, \frac{U_{t+h}g - U_t g}{h}\right\rangle$$
$$= \left\langle P_t f, \frac{U_h g - g}{h}\right\rangle \qquad \text{for } h > 0,\ t \geqslant 0$$

Since $A_K g = 0$, passing to the limit as $h \to 0$, we obtain

$$\phi'(t) = \langle P_t f, A_K g\rangle = 0\,.$$

Function ϕ is continuous with the right-hand derivative identically equal to zero, implying that $\phi(t) = \phi(0)$ for all $t \geqslant 0$. Consequently,

$$\langle f, U_t g - g\rangle = \phi(t) - \phi(0) = 0 \qquad \text{for } t \geqslant 0\,.$$

Since f is arbitrary this, in turn, implies that $U_t g = g$ for $t \geqslant 0$, which, by Theorem 7.7.2, completes the proof. ■

In particular, if $\{S_t\}_{t\geqslant 0}$ is a semigroup generated by a system of ordinary differential equations then, from equation (7.6.5), $A_K f = 0$ is equivalent to

$$\sum_{i=1}^{d} F_i(x)\frac{\partial f}{\partial x_i} = 0 \tag{7.8.19}$$

for continuously differentiable f with compact support. However, it must be pointed out that (7.8.19) is of negligible usefulness in checking ergodicity, because the property "$A_K f = 0$ implies f constant for *all* functions in L^∞" must be checked and not just the continuously differentiable functions. This is quite different from the situation where one is using the Liouville theorem (7.8.17) to check for invariant measures. In the latter case, it is necessary to find only a *single* solution of $A_{FP} f = 0$.

Example 7.8.1. Theorem 7.8.3 allows us easily to prove that Hamiltonian systems (see Example 7.6.1) are not ergodic. To show this, note that for a Hamiltonian system defined by equation (7.6.12), equation (7.6.5) becomes

$$A_K f = \sum_{i=1}^{n}\left[\frac{\partial f}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q_i}\right] = [f, H]\,.$$

Take $f \in L^\infty$ to be any nonconstant function of the energy H. By Example 7.6.1, we know that $A_K f \equiv 0$ since

$$[f(H), H] = \frac{\partial f}{\partial H}[H, H] = 0$$

and therefore Hamiltonian systems are not ergodic on the whole space. However, if we fix the total energy, or the energy for each degree of freedom as in Remark 7.7.1, then the system *may* become ergodic. □

7.9 Further applications of the Hille–Yosida theorem

Thus far we have used the Hille–Yosida theorem to demonstrate some simple properties of semigroups that followed directly from properties of the infinitesimal operator A and the resolvent equation (7.8.1). In these cases the semigroups were given. Now we are going to show a simple application of the theorem to the problem of determining a semigroup corresponding to a given infinitesimal operator A.

Let $X = R$ and $L = L^1(R)$, and consider the infinitesimal operator

$$Af = \frac{d^2f}{dx^2} \tag{7.9.1}$$

that can, of course, only be defined for some $f \in L^1$. Let $\mathfrak{D}(A)$ be the set of all $f \in L^1$ such that $f''(x)$ exists almost everywhere, is integrable on R, and

$$f'(x) = f(0) + \int_0^x f''(s)\,ds.$$

In other words, $\mathfrak{D}(A)$ is the set of all f such that f' is absolutely continuous and f'' is integrable on R. We will show that there is a unique semigroup corresponding to the infinitesimal operator A.

The set $\mathfrak{D}(A)$ is evidently dense in L^1 (even the set of C^∞ functions is dense in L^1), therefore we may concentrate on verifying properties (b) and (c) of the Hille–Yosida theorem.

The resolvent equation (7.8.1) has the form

$$\lambda g - \frac{d^2g}{dx^2} = f, \tag{7.9.2}$$

which is a second-order ordinary differential equation in the unknown function g. Using standard arguments, the general solution of (7.9.2) may be written as

$$g(x) = C_1e^{-\alpha x} + C_2e^{\alpha x} + \frac{1}{2\alpha}\int_{x_0}^{x} e^{-\alpha(x-y)}f(y)\,dy - \frac{1}{2\alpha}\int_{x_1}^{x} e^{\alpha(x-y)}f(y)\,dy$$

where $\alpha = \sqrt{\lambda}$, and C_1, C_2, x_0, and x_1 are arbitrary constants. To be specific, pick $x_0 = -\infty$, $x_1 = +\infty$, and set

$$K(x - y) = (1/2\alpha)e^{-\alpha|(x-y)|}. \tag{7.9.3}$$

Then the solution of (7.9.2) can be written in the more compact form

$$g(x) = C_1e^{-\alpha x} + C_2e^{\alpha x} + \int_{-\infty}^{\infty} K(x - y)f(y)\,dy. \tag{7.9.4}$$

The last term on the right-hand side of (7.9.4) is an integrable function on R, since

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} K(x-y)f(y)\,dy = \int_{-\infty}^{\infty} K(x-y)\,dx \int_{-\infty}^{\infty} f(y)\,dy$$
$$= \frac{1}{\lambda}\int_{-\infty}^{\infty} f(y)\,dy. \tag{7.9.5}$$

Thus, since neither $\exp(-\alpha x)$ nor $\exp(\alpha x)$ are integrable over R, a necessary and sufficient condition for f to be integrable over R is that $C_1 = C_2 = 0$. In this case we have shown that the resolvent equation (7.9.1) has a unique solution $g \in L^1$ given by

$$g(x) = R_\lambda f(x) = \int_{-\infty}^{\infty} K(x-y)f(y)\,dy, \tag{7.9.6}$$

and thus condition (b) of the Hille–Yosida theorem is satisfied.

Combining equations (7.9.5) and (7.9.6) it follows immediately that the operator λR_λ preserves the integral. Moreover, $\lambda R_\lambda f \geqslant 0$ if $f \geqslant 0$, so that λR_λ is a Markov operator. Thus condition (c) of the Hille–Yosida theorem is automatically satisfied, and we have shown that the operator d^2/dx^2 is an infinitesimal operator of a continuous semigroup $\{T_t\}_{t\geqslant 0}$ of Markov operators, where

$$T_t f = \lim_{\lambda\to\infty} e^{-t\lambda} \sum_{n=0}^{\infty} \frac{t^n\lambda^n}{n!}(\lambda R_\lambda)^n f \tag{7.9.7}$$

and R_λ is defined by (7.9.3) and (7.9.6).

It is interesting that the limit (7.9.7) can be calculated explicitly. To do this, denote by ϕ_f the Fourier transformation of f, that is,

$$\phi_f(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x)\,dx.$$

The Fourier transformation of $K(x)$ given by equation (7.9.3) is

$$1/(\lambda + \omega^2),$$

where $\lambda = \alpha^2$. Since, by (7.9.6), $R_\lambda f$ is the convolution of the functions K and f, and it is well known that

$$\phi_{f*g}(\omega) = \phi_f(\omega)\phi_g(\omega), \tag{7.9.8}$$

where $f * g$ denotes the convolution of f with g, the Fourier transformation of $R_\lambda^n f$ is

$$[1/(\lambda + \omega^2)^n]\phi_f(\omega).$$

As a consequence, the Fourier transformation of the series in (7.9.7) is

$$\sum_{n=0}^{\infty} \frac{t^n \lambda^{2n}}{(\lambda + \omega^2)^n n!} \phi_f(\omega) = \exp[\lambda^2 t/(\lambda + \omega^2)]\phi_f(\omega).$$

Thus the Fourier transformation of $T_t f$ is

$$\lim_{\lambda \to \infty} \exp(-\lambda t) \exp[\lambda^2 t/(\lambda + \omega^2)]\phi_f(\omega) = \exp(-\omega^2 t)\phi_f(\omega).$$

Using the fact that $\exp(-\omega^2 t)$ is the Fourier transformation of

$$\frac{1}{\sqrt{4\pi t}} \exp(-x^2/4t)$$

and (7.9.8), we then have

$$T_t f(x) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} \exp[-(x - y)^2/4t] f(y)\, dy. \tag{7.9.9}$$

Hence, using the semigroup method we have shown that $\mu(t, x) = T_t f(x)$ is the solution of the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

with the initial condition

$$u(0, x) = f(x).$$

Remark 7.9.1. It is a direct consequence of the elementary properties of the differential quotient (see Definition 7.5.1) that if A is the infinitesimal operator corresponding to a semigroup $\{T_t\}_{t\geq 0}$, then cA is the infinitesimal operator corresponding to $\{T_{ct}\}_{t\geq 0}$. Thus, since we have proved that $A = d^2/dx^2$ is the infinitesimal operator corresponding to the semigroup $\{T_t\}_{t\geq 0}$ given by (7.9.9), we know immediately that

$$T_{\sigma^2 t/2} f(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{\infty} \exp[-(x - y)^2/2\sigma^2 t] f(y)\, dy$$

has a corresponding infinitesimal operator equal to $(\sigma^2/2)(d^2/dx^2)$. (This is in perfect agreement with our observations in Example 7.4.1.) For simplicity, we have omitted the coefficient $(\sigma^2/2)$ in the foregoing calculations. □

The proof that d^2/dx^2 is an infinitesimal operator for a stochastic semigroup on R may be extended to R^d. Thus, for example, the operator

$$Af \equiv \Delta f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} \tag{7.9.10}$$

on R^d may be shown to be an infinitesimal operator for a stochastic semigroup, as can

$$Af = \sum_{i,j=1}^{n} a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}, \tag{7.9.11}$$

where the a_{ij} are constant, or sufficiently regular functions of x, and $\sum_{i,j} a_{ij}\xi_i\xi_j$ is positive definite. The procedure for proving these assertions is similar to that for operator d^2/dx^2 on R, but requires some special results from the theory of partial differential equations and functional analysis, allowing us to extend the definitions of the differential operators (7.9.10) and (7.9.11).

Operators such as d^2/dx^2, (7.9.10), or (7.9.11) may be considered not on the whole space (R or R^d), but also on bounded subspaces. However, in this case other boundary conditions must be specified, for example,

$$Af = \frac{d^2f}{dx^2} \qquad \text{on } L^1([a,b])$$

with

$$\left.\frac{df}{dx}\right|_a = 0 \quad \text{and} \quad \left.\frac{df}{dx}\right|_b = 0$$

is an infinitesimal operator for a stochastic semigroup. More details concerning such general elliptic operators may be found in Dynkin [1965].

Finally, we note that all semigroups that are generated by second-order differential operators are not semigroups of Frobenius–Perron operators for a semidynamical system and, thus, cannot arise from deterministic processes. This is quite contrary to the situation for first-order differential operators, as already discussed in Section 7.8.

Remark 7.9.2. Equation (7.8.3) of the Hille–Yosida theorem allows the construction of the semigroup $\{T_t\}_{t\geqslant 0}$ if the resolvent operator R_λ is known. As it turns out, the construction of the resolvent operator when the continuous semigroup of contractions is given is even simpler. Thus it can be shown that (Dynkin, 1965)

$$R_\lambda f = \int_0^\infty e^{-\lambda t} T_t f\,dt \qquad \text{for } f \in L,\ \lambda > 0. \tag{7.9.12}$$

In (7.9.12) the integral on the half-line $[0,\infty)$ is considered as the limit of Riemann integrals on $[0,a]$ as $a \to \infty$. This limit exists since

$$\left\| \int_0^\infty e^{-\lambda t} T_t f\,dt \right\| \leqslant \int_0^\infty e^{-\lambda t} \|T_t f\|\,dt \leqslant \frac{1}{\lambda}\|f\|.$$

It is an immediate consequence of (7.9.12) that for every stochastic semigroup

T_t: $L^1 \to L^1$, the operator λR_λ is a Markov operator. To show this note first that, for $f \geq 0$, equation (7.9.12) implies $\lambda R_\lambda f \geq 0$. Furthermore, for $f \geq 0$,

$$\|R_\lambda f\| = \int_X R_\lambda f(x)\,dx = \int_0^\infty e^{-\lambda t}\left\{\int_X T_t f(x)\,dx\right\}dt$$
$$= \int_0^\infty e^{-\lambda t}\|f\|\,dt = \frac{1}{\lambda}\|f\|.$$

In addition to demonstrating that λR_λ is a Markov operator, (7.9.12) also demonstrates that the semigroup corresponding to a given resolvent R_λ is unique. To see this, choose $g \in L^\infty$ and take the scalar product of both sides of equation (7.9.12) with g. We obtain

$$\langle g, R_\lambda f\rangle = \int_0^\infty e^{-\lambda t}\langle g, T_t f\rangle\,dt \qquad \text{for } \lambda > 0,$$

which shows that $\langle g, R_\lambda f\rangle$, as a function of λ, is the Laplace transformation of $\langle g, T_t f\rangle$ with respect to t. Since the Laplace transformation is one to one, this implies that $\langle g, T_t f\rangle$ is uniquely determined by $\langle g, R_\lambda f\rangle$. Further, since $g \in L^\infty$ is arbitrary, $\{T_t f\}$ is uniquely determined by $\{R_\lambda f\}$. The same argument also shows that for a bounded continuous function $u(t)$, with values in L^1, the equality

$$R_\lambda f = \int_0^\infty e^{-\lambda t}u(t)\,dt \quad \text{implies} \quad u(t) = T_t f. \quad \square$$

Some of the most sophisticated applications of semigroup theory occur in treating integro-differential equations. Thus we may not only prove the existence and uniqueness of solutions to such equations, but also determine the asymptotic properties of the solutions. One of the main tools in this area is the following extension of the Hille–Yosida theorem, generally known as the **Phillips perturbation theorem.**

Theorem 7.9.1. Let a continuous stochastic semigroup $\{T_t\}_{t\geq 0}$ and a Markov operator P be given. Further, let A be the infinitesimal operator of $\{T_t\}_{t\geq 0}$. Then there exists a unique continuous stochastic semigroup $\{P_t\}_{t\geq 0}$ for which

$$A_0 = A + P - I$$

(I is the identity operator on L^1) is the infinitesimal operator. Furthermore, the semigroup $\{P_t\}_{t\geq 0}$ is defined by

$$P_t f = e^{-t}\sum_{n=0}^{\infty} T_n(t) f \qquad f \in L^1, \tag{7.9.13}$$

where $T_0(t) = T_t$ and

$$T_n(t)f = \int_0^t T_0(t - \tau)PT_{n-1}(\tau)f\,d\tau. \tag{7.9.14}$$

Proof: Denote by $R_\lambda(A)$ the resolvent corresponding to operator A, that is, $g = R_\lambda(A)f$ is the solution of

$$\lambda g - Ag = f \qquad \text{for } f \in L^1.$$

Since $\{T_t\}_{t\geq 0}$ is a stochastic semigroup, $\lambda R_\lambda(A)$ is a Markov operator (see Remark 7.9.2). Now we observe that the resolvent equation for operator A_0,

$$\lambda g - A_0 g = f, \tag{7.9.15}$$

may be rewritten as

$$(\lambda + 1)g - Ag = f + Pg.$$

Thus (7.9.15) is equivalent to

$$g = R_{\lambda+1}(A)f + R_{\lambda+1}(A)Pg. \tag{7.9.16}$$

From inequality (7.8.11) we have $\|R_{\lambda+1}(A)Pg\| \leq \|Pg\|/(\lambda + 1)$. Since P is a Markov operator, this becomes

$$\|R_{\lambda+1}(A)Pg\| \leq \|g\|/(\lambda + 1).$$

Thus, equation (7.9.16) has a unique solution that can be constructed by the method of successive approximations. The result is given by

$$g = R_\lambda(A_0)f = \sum_{n=0}^{\infty} [R_{\lambda+1}(A)P]^n R_{\lambda+1}(A)f, \tag{7.9.17}$$

and the existence of a solution g to (7.9.15) is proved. Further, from (7.9.17) it follows that $R_\lambda(A_0)f \geq 0$ for $f \geq 0$ and that

$$\|R_\lambda(A_0)f\| = \sum_{n=0}^{\infty} \left(\frac{1}{\lambda + 1}\right)^{n+1} \|f\| = \frac{1}{\lambda}\|f\| \qquad \text{for } f \geq 0.$$

Thus $\lambda R_\lambda(A_0)$ is a Markov operator and A_0 satisfies all of the assumptions of the Hille–Yosida theorem. Hence the infinitesimal operator A_0 generates a unique stochastic semigroup and the first part of the theorem is proved.

Now we show that this semigroup is given by equations (7.9.13) and (7.9.14). Using (7.9.14) it is easy to show by induction that

$$\|T_n(t)f\| \leq (t^n/n!)\|f\|. \tag{7.9.18}$$

Thus, the series (7.9.13) is uniformly convergent, with respect to t, on bounded intervals and $P_t f$ is a continuous function of t. Now set

$$Q_{\lambda,n} f = \int_0^\infty e^{-\lambda t} T_n(t) f \, dt \qquad n = 0, 1, \ldots$$

so

$$Q_{\lambda,0} f = \int_0^\infty e^{-\lambda t} T_t f \, dt = R_\lambda(A) f$$

and

$$\begin{aligned} Q_{\lambda,n} f &= \int_0^\infty e^{-\lambda t} \left\{ \int_0^t T_0(t - \tau) P T_{n-1}(\tau) f \, d\tau \right\} dt \\ &= \int_0^\infty \left\{ \int_\tau^\infty e^{-\lambda t} T_0(t - \tau) P T_{n-1}(\tau) f \, dt \right\} d\tau \\ &= \int_0^\infty \left\{ e^{-\lambda \tau} \int_0^\infty e^{-\lambda t} T_0(t) P T_{n-1}(\tau) f \, dt \right\} d\tau \\ &= \int_0^\infty \left\{ e^{-\lambda t} T_0(t) P \int_0^\infty e^{-\lambda \tau} T_{n-1}(\tau) f \, d\tau \right\} dt \\ &= R_\lambda(A) P Q_{\lambda,n-1} f . \end{aligned}$$

Hence, by induction, we have

$$Q_{\lambda,n} = [R_\lambda(A) P]^n R_\lambda(A) .$$

Define

$$Q_\lambda f = \int_0^\infty e^{-\lambda t} P_t f \, dt$$

and substitute equation (7.9.13) to give

$$\begin{aligned} Q_\lambda f &= \sum_{n=0}^\infty \int_0^\infty e^{-(\lambda+1)t} T_n(t) f \, dt = \sum_{n=0}^\infty Q_{\lambda+1,n} f \\ &= \sum_{n=0}^\infty [R_{\lambda+1}(A) P]^n R_{\lambda+1}(A) f . \end{aligned}$$

By comparing this result with (7.9.17), we see that $Q_\lambda = R_\lambda(A_0)$ or

$$R_\lambda(A_0) f = \int_0^\infty e^{-\lambda t} P_t f \, dt . \tag{7.9.19}$$

From (7.9.19) (see also the end of Remark 7.9.2), it follows that $\{P_t f\}_{t\geq 0}$ is the semigroup corresponding to A_0. ■

Example 7.9.1. Consider the integro-differential equation

$$\frac{\partial u(t,x)}{\partial t} + u(t,x) = \frac{\sigma^2}{2}\frac{\partial^2 u(t,x)}{\partial x^2} + \int_{-\infty}^{\infty} K(x,y)u(t,y)\,dy,$$
$$t > 0,\ x \in R \qquad (7.9.20)$$

with the initial condition

$$u(0,x) = \phi(x) \qquad x \in R. \qquad (7.9.21)$$

We assume that the kernel is measurable and stochastic, that is,

$$K(x,y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} K(x,y)\,dx = 1\,.$$

To treat the initial value problem, equations (7.9.20) and (7.9.21), using semigroup theory, we rewrite it in the form

$$\frac{du}{dt} = (A + P - I)u, \qquad u(0) = \phi, \qquad (7.9.22)$$

where $A = \frac{1}{2}\sigma^2(d^2/dx^2)$ is the infinitesimal operator for the semigroup

$$T_t f(x) = \frac{1}{\sqrt{2\sigma^2 \pi t}}\int_{-\infty}^{\infty} \exp[-(x - y)^2/2\sigma^2 t] f(y)\,dy \qquad (7.9.23)$$

(see Remark 7.9.1) and

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

From Theorem 7.9.1 it follows that there is a unique continuous semigroup $\{P_t\}_{t\geq 0}$ corresponding to operator $A_0 = A + P - I$, and, by Theorem 7.5.1, the function $u(t) = P_t\phi$ is the solution of (7.9.22) for every $\phi \in \mathfrak{D}(A_0) = \mathfrak{D}(A)$. Thus $u(t,x) = P_t\phi(x)$ can be interpreted as the generalized solution to equations (7.9.20) and (7.9.21) for every $\phi \in L^1(R)$.

This method of treating equation (7.9.20) is convenient from several points of view. First, it demonstrates the existence and uniqueness of the solution $u(t,x)$ for every $\phi \in L^1(R)$, and stochastic kernel K. Second, it shows that $P_t\phi$ is a density for $t \geq 0$ whenever ϕ is a density. Furthermore, some additional properties of the solution can be demonstrated by using the explicit representation for P_t

given in Theorem 7.9.1. For this example, it follows directly from (7.9.13) and (7.9.14) that

$$P_t\phi = e^{-t}\int_0^t T_0(t-\tau)g_\tau\,d\tau + e^{-t}T_0(t)\phi,$$

where

$$g_t = \sum_{n=1}^{\infty} PT_{n-1}(t)\phi.$$

Thus, using (7.9.23) (with $T_0(t) = T_t$), we have the explicit representation

$$P_t\phi(x) = e^{-t}\int_0^t \left\{\frac{1}{\sqrt{2\sigma^2\pi(t-\tau)}}\int_{-\infty}^{\infty}\exp[-(x-y)^2/2\sigma^2(t-\tau)]g_\tau(y)\,dy\right\}d\tau$$
$$+ e^{-t}\frac{1}{\sqrt{2\sigma^2\pi t}}\int_{-\infty}^{\infty}\exp[-(x-y)^2/2\sigma^2 t]\phi(y)\,dy.$$

This shows directly that the function $u(t,x) = P_t\phi(x)$ is continuous and strictly positive for $t > 0$ and every $\phi \in L^1(R)$, even if ϕ and the stochastic kernel K are not continuous! Finally, we will come back to this semigroup approach in Section 11.10 and use it to demonstrate some asymptotic properties of the solution $u(t,x)$. □

Example 7.9.2. As a second example of the applicability of the Phillips perturbation theorem, we consider the first-order integro-differential equation

$$\frac{\partial u(t,x)}{\partial t} + \frac{\partial u(t,x)}{\partial x} + u(t,x) = \int_x^{\infty} K(x,y)u(t,y)\,dy,$$
$$t > 0,\ x \geqslant 0 \qquad (7.9.24)$$

with

$$u(t,0) = 0 \quad \text{and} \quad u(0,x) = \phi(x) \qquad (7.9.25)$$

Again the kernel K is assumed to be measurable and stochastic, that is,

$$K(x,y) \geqslant 0 \quad \text{and} \quad \int_0^y K(x,y)\,dx = 1\,. \qquad (7.9.26)$$

Equation (7.9.24) occurs in queuing theory and astrophysics [Bharucha-Reid, 1960]. In its astrophysical form,

$$K(x,y) = (1/y)\psi(x/y)\,, \qquad (7.9.27)$$

and, with this specific expression for K, equation (7.9.24) is called the

Chandrasekhar–Münch equation. As developed by Chandrasekhar and Münch [1952], equation (7.9.24) with K is given by (7.9.27) describes fluctuations in the brightess x of the Milky Way as a function of the extent of the system t along the line of sight. The unknown function $u(t, x)$ is the probability density of the fluctuations, and the given function ψ in (7.9.27) is related to the probability density of light transmission through interstellar gas clouds. This function satisfies

$$\psi(z) \geqslant 0 \quad \text{and} \quad \int_0^1 \psi(z)\, dz = 1 \tag{7.9.28}$$

and, thus, K as given by (7.9.27) automatically satisfies (7.9.26).

To rewrite (7.9.24) as a differential equation in L^1, recall (see Example 7.5.1) that $-d/dx$ is the infinitesimal operator for the semigroup $T_t f(x) = f(x - t)$ defined on $L^1(R)$. On $L^1([0, \infty))$,

$$T_t f(x) = 1_{[0,\infty)}(x - t) f(x - t) \tag{7.9.29}$$

plays an analogous role. Proceeding much as in Example 7.5.1, a simple calculation shows that for continuously differentiable f with compact support in $(0, \infty)$ the infinitesimal operator corresponding to the semigroup in (7.9.29) is given by $Af = -df/dx$. Further, it is clear that $u(t, x) = T_t f(x)$ satisfies $u(t, 0) = 0$ for $t > 0$. Hence we may rewrite equations (7.9.24)–(7.9.25) in the form

$$\frac{du}{dt} = (A + P - I)u, \qquad u(0) = \phi, \tag{7.9.30}$$

where

$$Pf(x) = \int_x^\infty K(x, y) f(y)\, dy.$$

By Theorem 7.9.1 there is a continuous unique semigroup $\{P_t\}_{t \geqslant 0}$ corresponding to the infinitesimal operator $A + P - I$. For every $\phi \in \mathfrak{D}(A)$, the function $u(t) = P_t \phi$ is a solution of (7.9.30). □

7.10 The relation between the Frobenius–Perron and Koopman operators

The semigroup of Frobenius–Perron operators $\{P_t\}$ and the semigroup $\{U_t\}$ of Koopman operators, both generated by the same semidynamical system $\{S_t\}_{t \geqslant 0}$, are closely related because they are adjoint. However, each describes the behavior of the system $\{S_t\}_{t \geqslant 0}$ in a different fashion, and in this section we show the connection between the two.

Equation (7.4.16), $\langle P_t f, g\rangle = \langle f, U_t g\rangle$, which says that P_t and U_t are adjoint, may be written explicitly as

$$\int_X g(x)P_t f(x)\mu(dx) = \int_X f(x)g(S_t(x))\mu(dx) \qquad \text{for } f \in L^1,\ g \in L^\infty.$$

For some $A \subset X$ such that A and $S_t(A)$ are in $\mathcal{A}$, take $f(x) = 0$ for all $x \notin A$ and $g = 1_{X\setminus S_t(A)}$, then the preceding formula becomes

$$\int_X 1_{X\setminus S_t(A)}(x)P_t f(x)\mu(dx) = \int_X f(x)1_{X\setminus S_t(A)}(S_t(x))\mu(dx)$$
$$= \int_A f(x)1_{X\setminus S_t(A)}(S_t(x))\mu(dx)\,.$$

The right-hand side of this equation is obviously equal to zero since $S_t(x) \notin X\setminus S_t(A)$ for $x \in A$. The left-hand side is, however, just the L^1 norm of the integrand, so that

$$\|1_{X\setminus S_t(A)} P_t f\| = 0\,.$$

This, in turn, implies

$$1_{X\setminus S_t(A)}(x)P_t f(x) = 0$$

or

$$P_t f(x) = 0 \quad \text{for } x \notin S_t(A)\,. \tag{7.10.1}$$

Thus the operator P_t "carries" the function f, supported on A, forward in time to a function supported on a subset of $S_t(A)$ (see Example 3.2.1 and Proposition 3.2.1). Figuratively speaking, we may say that the density is transformed by P_t analogously to the way in which initial points x are transformed into $S_t(x)$.

Now consider the definition of the Koopman operator,

$$U_t f(x) = f(S_t(x))\,.$$

Assume $f \in L^\infty$ is zero outside a set A, so we have

$$f(S_t(x)) = 0 \qquad \text{if } S_t(x) \notin A\,. \tag{7.10.2}$$

This, in turn, implies that

$$U_t f(x) = 0 \qquad \text{for } x \notin S_t^{-1}(A)\,. \tag{7.10.3}$$

In contrast to P_t, therefore, U_t may be thought of as transporting the function supported on A, backward in time to a function supported on $S_t^{-1}(A)$.

These observations become even clearer when $\{S_t\}$ is a group of transformations, that is, when the group property holds for both positive and negative time,

$$S_{t+t'}(x) = S_t(S_{t'}(x)) \qquad \text{for all } t, t' \in R,\ x \in X,$$

and all the S_t are at least nonsingular. In this case, $S_t^{-1}(x) = S_{-t}(x)$ and (7.10.3) becomes

$$U_t f(x) = 0 \quad \text{for } x \notin S_{-t}(A).$$

If, in addition, the group $\{S_t\}$ preserves the measure μ, we have

$$\int_A P_t f(x)\mu(dx) = \int_{S_{-t}(A)} f(x)\mu(dx) = \int_A f(S_{-t}(x))\mu(dx),$$

which gives

$$P_t f(x) = f(S_{-t}(x))$$

or, finally,

$$P_t f(x) = U_{-t} f(x). \tag{7.10.4}$$

Equation (7.10.4) makes totally explicit our earlier comments on the forward and backward transport of densities in time by the Frobenius–Perron and Koopman operators.

Furthermore, from (7.10.4) we have directly that

$$\lim_{t\to 0}[(P_t f - f)/t] = \lim_{t\to 0}[(U_{-t} f - f)/t]$$

and, thus, for f in a dense subset of L^1,

$$A_{FP} f = -A_K f. \tag{7.10.5}$$

This relation was previously derived, although not explicitly stated, for dynamical systems generated by a system of ordinary differential equations [cf. equations (7.6.5) and (7.6.10)].

Remark 7.10.1. Equation (7.10.4) may, in addition be interpreted as saying that the operator adjoint to P_t is also its inverse. In the terminology of Hilbert spaces [and thus in $L^2(X)$] this means simply that $\{P_t\}$ is a semigroup of unitary operators. The original discovery that $\{U_t\}$, generated by a group $\{S_t\}$ of measure-preserving transformations, forms a group of unitary operators is due to Koopman [1931]. It was later used by von Neumann [1932] in his proof of the statistical ergodic theorem. □

Remark 7.10.2. Equation (7.10.1) can sometimes be used to show that a semigroup of Markov operators cannot arise from a deterministic dynamical system, which means that it is not a semigroup of Frobenius–Perron operators for any semidynamical system $\{S_t\}_{t\geq 0}$.

For example, consider the semigroup $\{P_t\}$ given by equations (7.4.11) and (7.4.12):

$$P_t f(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{\infty} f(y) \exp\left[-\frac{(x-y)^2}{2\sigma^2 t}\right] dy. \tag{7.10.6}$$

Setting $f(y) = 1_{[0,1]}(y)$, it is evident that we obtain

$$P_t f(x) > 0 \qquad \text{for all } x \text{ and } t > 0.$$

However, according to (7.10.1), if $P_t f(x)$ was the Frobenius–Perron operator generated by a semidynamical system $\{St\}_{t\geqslant 0}$, then it should be zero outside a bounded interval $S_t([0,1])$. [The interval $S_t([0,1])$ is a bounded interval since a continuous function maps bounded intervals into bounded intervals.] Thus $\{P_t\}$, where $P_t f(x)$ is given by (7.10.6), does not correspond to any semidynamical system. □