

# **A MEASURE-THEORETIC APPROACH TO CHAOTIC DYNAMICAL SYSTEMS**

by

**Pranitha Singh**

This dissertation is submitted in fulfilment of the  
requirements for the degree of  
Master of Science  
in the  
School of Mathematical and Statistical Sciences  
University of Natal

**Durban**

**December 2001**

## Abstract

The past few years have witnessed a growth in the study of the long-time behaviour of physical, biological and economic systems using measure-theoretic and probabilistic methods. In this dissertation we present a study of the evolution of dynamical systems that display various types of irregular behaviour for large times.

Large systems, containing many elements, like e.g. bacteria populations or ensembles of gas particles, are very difficult to analyse and contain elements of uncertainty. Also, in general, it is not necessary to know the evolution of each bacteria or each gas particle. Therefore we replace the "pointwise" description of the evolution of the system with that of the evolution of suitable averages of the population like e.g. the gas or the bacteria spatial density. In particular cases, when the quantity in the evolution that we analyse has the probabilistic interpretation, say, the probability of finding the particle in certain state at certain time, we will be talking about the evolution of (probability) densities.

We begin with the establishment of results for discrete time systems and this is later followed with analogous results for continuous time systems. We observe that in many cases the system has two important properties: at each step it is determined by a non-negative function (for example the spatial density or the probability density) and the overall quantity of the elements remains preserved. Because of these properties the most suitable framework to investigate such systems is the theory of Markov operators.

We shall discuss three levels of "chaotic" behaviour that are known as ergodicity, mixing and exactness. They can be described as follows: ergodicity means that the only invariant sets are trivial, mixing means that for any set  $A$  the sequence of sets  $S^{-n}(A)$  becomes, asymptotically, independent of any other set  $B$ , and exactness

implies that if we start with any set of positive measure, then, after a long time the points of this set will spread and completely fill the state space.

In this dissertation we describe an application of two operators related to the generating Markov operator to study and characterize the abovementioned properties of the evolution system.

However, a system may also display regular behaviour. We refer to this as the asymptotic stability of the Markov operator generating this system and we provide some criteria characterizing this property.

Finally, we demonstrate the use of the above theory by applying it to a system that is modeled by the linear Boltzmann equation.

## Preface

The study described in this dissertation was carried out in the Department of Mathematics and Applied Mathematics, University of Natal, Durban, from February 2000 to December 2001, under the supervision of Professor Jacek Banasiak.

These studies represent original work by the authoress and have not been submitted in any form to another University. Where use was made of the work of others it has been duly acknowledged in the text.

---

Pranitha Singh

December 2001

*To*  
**Navitha**  
**and**  
**Nashil**

## Acknowledgements

I wish to express my sincere thanks and appreciation to the following people for their roles in the successful compilation of this dissertation:

- My supervisor Professor Jacek Banasiak for his dedication, guidance and expert knowledge in the conducting of this research.
- My husband Navin, daughter Navitha and son Nashil for sparing me their quality time.
- My parents for their love and support.
- Professor S D Maharaj and Dr Sudan Hansraj for helpful discussions.
- The Foundation for Research and Development for financial assistance.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Mathematical Preliminaries</b>	<b>7</b>
2.1	Measures and Measure Spaces . . . . .	7
2.2	Lebesgue Integration . . . . .	10
2.3	Convergence of Sequences of Functions . . . . .	25
2.4	Probability Theory . . . . .	28
<b>3</b>	<b>The Relationship between Densities and Dynamical Systems</b>	<b>32</b>
3.1	Introduction . . . . .	32
3.2	Defining a Dynamical System . . . . .	33
3.3	Defining a Density . . . . .	34
<b>4</b>	<b>Markov Operators</b>	<b>38</b>
4.1	Markov Operators . . . . .	38
4.2	The Frobenius-Perron Operator . . . . .	46
4.3	The Koopman Operator . . . . .	53

4.4	Frobenius-Perron and Koopman operators in the Space $L^p$ . . . . .	59
<b>5</b>	<b>Various Types of Chaotic Behaviour</b>	<b>62</b>
5.1	Measure-Preserving Transformations . . . . .	62
5.2	Ergodicity . . . . .	67
5.3	Mixing and Exactness . . . . .	78
5.4	Characterizing Chaotic Behaviour . . . . .	80
5.5	Other Types of Chaotic Behaviour . . . . .	94
<b>6</b>	<b>Asymptotic Stability of Markov Operators</b>	<b>97</b>
6.1	Introduction . . . . .	97
6.2	Asymptotic Properties of the Averages $\{A_n f\}$ . . . . .	98
6.3	The Existence of a Stationary Density . . . . .	108
6.4	Asymptotic Stability of $\{P^n\}$ . . . . .	117
6.5	Asymptotic Stability of the Markov Operator defined by a Stochastic Kernel . . . . .	126
6.6	Examples of Asymptotic Stability . . . . .	135
<b>7</b>	<b>An Introduction to Continuous Time Systems</b>	<b>144</b>
7.1	Introduction . . . . .	144
7.2	Continuous Time Dynamical Systems . . . . .	144
7.3	Semigroups of the Frobenius-Perron and Koopman Operators . . . .	150



7.4	The Hille-Yosida Theorem . . . . .	162
<b>8</b>	<b>An Application</b>	<b>171</b>
8.1	Introduction . . . . .	171
8.2	The Abstract linear Boltzmann Equation . . . . .	171
8.3	Solutions of the Linear Boltzmann Equation . . . . .	183
8.4	The Effect of the Properties of the Markov Operator on Solutions of the Linear Boltzmann Equation . . . . .	191
<b>9</b>	<b>Conclusion</b>	<b>197</b>
<b>10</b>	<b>References</b>	<b>199</b>

# 1 Introduction

"Traditionally scientists have looked for the simplest world around us. Now, mathematics and computer power have produced a theory that helps researchers to understand the complexities of nature. The theory of chaos touches all disciplines." - Ian Percival (Hall 1992)

The field of dynamical systems and especially the study of chaotic systems has emerged as one of the important breakthroughs in recent science. Chaotic dynamics has been hailed as the third great scientific revolution in physics this century, comparable to relativity and quantum mechanics (Smith 1998).

A *dynamical system* is a system which changes over time: for example the planets in motion, a convecting fluid, a continuous chemical reaction etc. There are regular changes and chaotic ones. The entity that changes is some variable, or a set of variables, which determines the state of the system. The basic goal of the theory of dynamical systems is to understand the eventual or asymptotic behaviour of a discrete or continuous process. If the process is a discrete process such as the iteration of a function, the theory hopes to understand the eventual behaviour of the points  $x, f(x), f^2(x), \dots, f^n(x)$  as  $n$  becomes large. If the process is a continuous process the dynamics are usually described by a differential equation whose independent variable is time, then the theory attempts to predict the ultimate behaviour of solutions of the equation in either the distant future ( $t \rightarrow \infty$ ) (or the distant past

( $t \rightarrow -\infty$ ), which is not considered in our analysis). That is, dynamical systems raise the question of where do points go and what do they do when they get there.

During the 19th century there were 2 kinds of theories for changing systems, namely deterministic theories and theories of probability (Hall 1992). The two approaches appeared to be incompatible. In the first, the future is determined from the past, with no apparent need for probability. In the second, the future depends in some random way on the past, and cannot be determined from it. The first challenge to this picture came with quantum theory in the 1920's and the 1930's in which theorists describe the behavior of an electron in terms of a "probability wave". The second challenge came from chaos. Analysis showed that even in simple systems which obey Newtons' laws of motion you cannot always predict what is going to happen next. The reason for this is a persistent instability. This often arises when an object feels the effect of more than one force. A well-known example is a pendulum with a bob that is attracted equally to 2 magnets below it. Its future motion becomes extremely sensitive to small changes in the present position and velocity, so the motion can become chaotic.

Chaos in the ordinary sense, is precisely the absence of order. It is a persistent instability. Chaos is a dynamic phenomenon. It occurs when the state of a system changes. Chaos introduces an interface between determinism and randomness (Lasota and Mackey 1985). It presents a universe that is deterministic, obeying the fundamental physical laws, but with a predisposition for disorder, complexity and unpredictability. It reveals many systems that are changing are extremely sensitive to their initial state - position, velocity, and so on. As the system evolves in time, minute changes amplify rapidly through feedback.

Chaos is usually viewed as a nonlinear phenomenon. The original example of chaotic behaviour given in the paper by Lorentz, and all the subsequent results on chaotic

behaviour of solutions to ordinary differential equations, refer to nonlinear systems, that is, in the equation

$$\dot{u} = f(u), \tag{1.1}$$

the function  $f$  is nonlinear (possibly acting in a multi dimensional space  $R^n$  and  $u = (u_1, \dots, u_n)$ ).

However, the idea of chaotic behaviour also arises in discrete and continuous linear dynamical systems. But, for this they must be infinite dimensional. This means that the function  $f$  on the right-hand side of Equation (1.1) must be for instance an infinite matrix, or a differentiation operator (Banasiak 2000).

Modern dynamical systems theory has a relatively short history as noted in Devaney (1989). It began in the 19th century with the French mathematician Henri Poincaré. He revolutionized the study of nonlinear differential equations by applying topology to discuss the global properties of solutions to systems. This approach turned out to be a powerful tool for describing types of behaviour. An understanding of the global behaviour of all solutions of the system was more important than the local behaviour of particular, analytically precise solutions. This point of view was adopted and furthered by Birkhoff in the first part of the twentieth century. Birkhoff realized the importance of the study of mappings and emphasized discrete dynamics as a means of understanding the more difficult dynamics arising from differential equations.

During the last twenty years or so, several definitions of chaos have been proposed (Devaney 1989, Kirchgraber and Stoffer 1989, and Wiggins 1990). One of the most used among these definitions is that as presented by (Devaney 1989). This approach uses topological notions.

In this dissertation, chaotic behaviour of continuous linear semigroups is analysed by the use of measure-theoretic ergodic notions. Using measure theory, we develop concepts of 'measure' that enable us to talk about the measure not just of simple regions but, more generally of Borel sets (sets formed by repeatedly taking unions and intersections by starting from interval-like regions and their complements), and then we talk not just of Borel sets but of arbitrary subsets of  $R^n$ . In parallel to these concepts of measure we also develop notions of integration, which allow us to integrate functions over sets which are not simple regions. These notions of measure are used to present a spectrum of theorems, although the basic results (e.g. 'the empty set has measure zero', 'the measure of  $[0,2]$  is the sum of the measures of  $[0,1]$  and  $[1,2]$ ', etc.) stay fixed. Thus, we observe the useful notions of Borel measure, Lebesgue measure, Riemann integral, the Lebesgue integral and so on. Using ergodic theory we associate chaotic linear semigroups with the idea of exactness applied to first-order partial differential equations. Exact linear semigroups are applied to nonlinear continuous semidynamical systems via the Frobenius-Perron operators and the Koopman operators.

Suppose a real-world phenomenon has an acceptable mathematical model. We consider possible definitions of chaos for (the dynamics in) a mathematical model. If we consider cases of mathematical models with complex behaviour for example the logistic map (Hall 1992 and Lasota and Mackey 1985), the tent map (Smith 1998) or the baker transformation (Lasota 1985) we observe that their features and structure present a problem in explaining natural phenomena. We overcome the problem of characterizing and providing a simpler physical description of large systems containing inherent elements of uncertainty by abandoning the pointwise description of the evolution of our system in favour of a system of suitable averages.

In view of the above discussion of chaotic dynamical systems, our main objectives

in this dissertation are:

- To construct a suitable system of averages that would replace the "pointwise" description of the evolution of a large system.
- To provide descriptions based on measure theoretic properties of a system of the types of "chaotic" behaviour that a system may display and to demonstrate techniques with which we can identify these irregular behaviours.
- To examine conditions under which our system may display irregular behaviour.
- To utilize the above results and the theory of semigroups to determine the unique solution to the linear Boltzmann equation.

This dissertation is organised as follows:

- Chapter 1: Introduction
- Chapter 2: We discuss the mathematical concepts: measures and measure spaces, Lebesgue integration, convergence of sequences of functions and probability theory which are necessary for our study.
- Chapter 3: In this chapter our aim is to derive a suitable method of describing the evolution of a dynamical system containing many elements.
- Chapter 4: We motivate the use of the Markov operator as a suitable generator of our system.
- Chapter 5: This chapter is the focal point of our study. We describe 3 types of irregular behaviour by using an approach based on measure theory.

- Chapter 6: We discuss a property of Markov operators known as asymptotic stability.
- Chapter 7: In this chapter we introduce continuous dynamical systems and present a continuous time analogue of definitions and results for the discrete time case. We also describe a theory which enables us to determine solutions to a system.
- Chapter 8: We conclude our analysis by applying our results to a system modeled by the linear Boltzmann equation.
- Chapter 9: Conclusion

## 2 Mathematical Preliminaries

In this chapter we present the background knowledge that we shall use in the development of our analysis. We briefly outline essential concepts from measure theory, the theory of Lebesgue integration, the theory of the convergence of sequences of functions and probability theory.

### 2.1 Measures and Measure Spaces

**Definition 2.1** *Let  $X$  be an arbitrary set. A collection  $\mathcal{A}$  of subsets of  $X$  is called a  $\sigma$ -algebra in  $X$  if*

(a)  $X \in \mathcal{A}$

(b) For each set  $A \in \mathcal{A}$ , the set  $A^c$  (or  $X \setminus A$ )  $\in \mathcal{A}$

(c) For each finite or infinite sequence  $A_i$  of subsets of  $X$ ,  $A_i \in \mathcal{A}$ , the union  $\bigcup_i A_i \in \mathcal{A}$ .

From this definition we note the following:

(1) From properties (a) and (b)  $\emptyset \in \mathcal{A}$  since  $\emptyset = X^c$  (or  $X \setminus X$ ) and;

(2) From (b) and (c)  $\bigcap_i A_i \in \mathcal{A}$ , since,  $\bigcap_i A_i = (\bigcup_i A_i^c)^c$  (From de Morgan's laws).



**Definition 2.2** A real-valued function  $\mu : \mathcal{A} \rightarrow [0, \infty)$ , defined on a  $\sigma$ -algebra  $\mathcal{A}$  is called a **measure** if

(a)  $\mu(\emptyset) = 0$ ;

(b)  $\mu(A) \geq 0$  for all  $A \in \mathcal{A}$ ; and

(c)  $\mu(\cup_i A_i) = \sum_i \mu(A_i)$  if  $\{A_i\}$  is a finite or infinite sequence of pairwise disjoint subsets of  $\mathcal{A}$ , that is,  $A_i \cap A_j = \emptyset$  for  $i \neq j$ .

In addition  $\mu$  is called a countably additive measure if

$\mu(\emptyset) = 0$  and  $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ ; and is called a finitely additive measure if  $\mu(\emptyset) = 0$  and  $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$ .

**Remark 2.1** Since  $\mu(A_i)$  is non-negative, the sum  $\sum_i \mu(A_i)$  always exists either as a real number or as  $+\infty$ . ■

**Definition 2.3** If  $X$  is a set,  $\mathcal{A}$  is a  $\sigma$ -algebra on  $X$ , and  $\mu$  is a measure on  $\mathcal{A}$ , then the triple  $(X, \mathcal{A}, \mu)$  is called a **measure space**. The sets belonging to  $\mathcal{A}$  are called **measurable sets**, since, for them, the measure is defined.

The following are two examples of measure spaces.

**Example 2.1** Let  $X$  be an arbitrary set and  $\mathcal{A}$  a  $\sigma$ -algebra on  $X$ . Define a function  $\mu : \mathcal{A} \rightarrow [0, +\infty)$  by

$$\mu(A) = \begin{cases} n & \text{if } A \subset \mathcal{A} \text{ is a finite set with } n \text{ elements,} \\ +\infty & \text{if } A \subset \mathcal{A} \text{ is infinite.} \end{cases}$$

Then the measure  $\mu$  is called the **counting measure** on  $X$ .

**Example 2.2** If  $X = [0, 1]$  or  $\mathbb{R}$ , the real line, then the  $\sigma$ -algebra is the  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$  of all the Borel sets of  $\mathbb{R}$ . (That is the  $\sigma$ -algebra generated by the collection

of all the open sets of  $R$ .) Then there exists a unique measure  $\mu$ , called the Borel measure that assigns to each sub-interval of  $R$  its length i.e.  $\mu([a,b]) = b-a$ .

In almost all applications we use a more specific measure space defined as follows:

**Definition 2.4** A measure  $\mu$  in a measure space  $(X, \mathcal{A}, \mu)$  is called a  $\sigma$  - finite measure if there is a sequence  $A_i$ ,  $A_i \in \mathcal{A}$  satisfying  $X = \cup_{i=1}^{\infty} A_i$  and  $\mu(A_i) < \infty$  for all  $i$ .

**Definition 2.5** A measure space  $(X, \mathcal{A}, \mu)$  is called **finite** if  $\mu(X) < \infty$ . In particular, if  $\mu(X) = 1$ , then the measure space is called **normalized** or **probabilistic**.

Note that throughout our study, unless otherwise stated, a measure space will always be assumed to be  $\sigma$  - finite.

**Definition 2.6** Let  $(X, \mathcal{A}, \mu)$  be a measure space. A property of the points in a measure space  $X$  that holds everywhere except for a subset of that space having measure zero is said to be true **almost everywhere** (abbreviated as **a.e.**).

**Definition 2.7** Let  $f : X \rightarrow R$  be a real valued function and  $\Delta \subset R$  an interval. The set of all points that will be in  $\Delta$  after one application of  $f$  is called the **counterimage** of  $\Delta$  i.e

$$f^{-1}(\Delta) = \{x | f(x) \in \Delta\}$$

**Definition 2.8** Let  $(X, \mathcal{A}, \mu)$  be a measure space. A real-valued function  $f : X \rightarrow R$  is called **measurable** if  $f^{-1}(\Delta) \in \mathcal{A}$  for every interval  $\Delta \subset R$ .

More generally we define the following for a transformation  $S : X \rightarrow X$ :

**Definition 2.9** Let  $(X, \mathcal{A}, \mu)$  be a measure space. A transformation  $S : X \rightarrow X$  is called **measurable** if  $S^{-1}(A) \in \mathcal{A}$ , for all  $A \in \mathcal{A}$ .

A related concept to that of a measurable transformation is the following definition:

**Definition 2.10** A measurable transformation  $S : X \rightarrow X$  on a measure space  $(X, \mathcal{A}, \mu)$  is called **non-singular** if  $\mu(S^{-1}(A)) = 0$ , for all  $A \in \mathcal{A}$  such that  $\mu(A) = 0$ .

## 2.2 Lebesgue Integration

In this section we introduce the Lebesgue integral, which is defined for abstract measure spaces  $(X, \mathcal{A}, \mu)$ . Firstly we define

$$f^+(x) = \max(0, f(x)) \text{ and } f^-(x) = \max(0, -f(x));$$

and we hence observe that

$$f(x) = f^+(x) - f^-(x) \text{ and } |f(x)| = f^+(x) + f^-(x).$$

Next we define the characteristic/indicator function.

**Definition 2.11** *The characteristic/indicator function for any set  $A$  is defined as follows*

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Let  $(X, \mathcal{A}, \mu)$  be a measure space. We next illustrate an important property for a bounded, nonnegative measurable function.

Let  $f : X \rightarrow \mathbb{R}$  be a bounded, nonnegative measurable function,  $0 \leq f(x) < M < \infty$ . Take the partition of the interval  $[0, M]$ ,  $0 = a_0 < a_1 < \dots < a_n = M$ ,  $a_i = M \frac{i}{n}$ ,  $i = 0, \dots, n$ , and define the sets  $A_i$  by

$$A_i = \{x : f(x) \in [a_i, a_{i+1})\}, i = 0, \dots, n-1,$$

where  $a_i, \dots, a_n$  are nonnegative real numbers and  $A_1, \dots, A_n$  are disjoint subsets of  $X$  that belong to  $\mathcal{A}$ .

We then observe that the  $\{A_i\}$  are measurable and

$$\left| f(x) - \sum_{i=0}^{n-1} a_i I_{A_i}(x) \right| \leq \frac{M}{n}.$$

Therefore, every bounded nonnegative measurable function can be approximated by finite linear combinations of characteristic functions.

We next define a simple function.

**Definition 2.12** *A function  $g : X \rightarrow \mathbb{R}$  is called a simple function if it can be written in the form*

$$g(x) = \sum_{i=1}^n \lambda_i I_{A_i}(x)$$

where  $\lambda_i \in \mathbb{R}$  and  $A_i \in \mathcal{A}, i = 1, \dots, n$  are disjoint sets.

We now define the Lebesgue integral in following four steps.

**Definition 2.13** Let  $(X, \mathcal{A}, \mu)$  be a measure space, and  $g$  a simple function as defined in Definition 2.12. Then the **Lebesgue integral** of the function  $g$  is defined as

$$\int_X g(x) d\mu_x = \sum_i \lambda_i \mu(A_i).$$

We note that  $\int_X g(x) d\mu_x$  is the integral of  $g$  over the set  $X$  with respect to the measure  $\mu$ , and which in this case  $g$  is defined as a function of  $x$ . One sometimes writes  $\int_X g d\mu$  or  $\int_X g(x) \mu(dx)$  or  $\int_X g(x) d\mu(x)$  in place of  $\int_X g(x) d\mu_x$ .

**Definition 2.14** Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $f : X \rightarrow \mathbb{R}$  an arbitrary non-negative bounded measurable function, and  $\{g_n\}$  a sequence of simple functions converging uniformly to  $f$ . Then the **Lebesgue integral** of  $f$  is defined as

$$\int_X f(x) d\mu_x = \lim_{n \rightarrow \infty} \int_X g_n(x) d\mu_x.$$

**Definition 2.15** Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $f : X \rightarrow \mathbb{R}$  a nonnegative unbounded measurable function, and define

$$f_M(x) = \begin{cases} f(x) & \text{if } 0 \leq f(x) \leq M, \\ M & \text{if } M < f(x). \end{cases}$$

Then the **Lebesgue integral** of  $f$  is defined by

$$\int_X f(x) d\mu_x = \lim_{M \rightarrow \infty} \int_X f_M(x) d\mu_x.$$

**Definition 2.16** Let  $(X, \mathcal{A}, \mu)$  be measure space and  $f : X \rightarrow \mathbb{R}$  be a measurable function. Then the **Lebesgue integral** of  $f$  is defined by

$$\int_X f(x) d\mu_x = \int_X f^+(x) d\mu_x - \int_X f^-(x) d\mu_x$$

if at least one of the terms

$$\int_X f^+(x) d\mu_x, \text{ or } \int_X f^-(x) d\mu_x$$

is finite. If both these terms are finite then the function  $f$  is called **integrable**.

**Remark 2.2** We note that Definition 2.13 to Definition 2.16 are for the Lebesgue integral of  $f$  over the entire space  $X$ . For  $A \in \mathcal{A}$  we have, by definition,

$$\int_A f(x) d\mu_x = \int_X f(x) I_A(x) d\mu_x \blacksquare$$

**Remark 2.3** We also observe that  $f$  is integrable if and only if  $|f|$  is integrable.

Indeed, since  $f(x) = f^+(x) - f^-(x)$  and  $|f| = f^+ + f^-$ , then

if  $f$  is integrable, then so are  $f^+$  and  $f^-$  and therefore

$$\int_X |f(x)| d\mu_x = \int_X f^+(x) d\mu_x + \int_X f^-(x) d\mu_x$$

is finite. Hence  $|f|$  is integrable.

If  $|f(x)|$  is integrable, using a similar argument  $\int_X f(x) d\mu_x$  is finite so that  $f(x)$  is integrable.  $\blacksquare$

**Remark 2.4** The Lebesgue integral is an important concept for abstract measure spaces  $(X, \mathcal{A}, \mu)$ . In calculus the definition of the Riemann integral is related to

the algebraic properties of the real line. For example, if we define  $\mu$  as the Borel measure (see Example 2.2), then,

$$\int_{[a,b]} f(x) d\mu_x = \int_a^b f(x) dx$$

where the left-hand side is the Lebesgue integral and the right-hand side is the Riemann integral. This equality is true for any bounded real-valued Riemann integrable function  $f$  defined on a closed interval  $[a, b]$ , that is if  $f$  is Riemann integrable then  $f$  is Lebesgue integrable (Cohn 1980 Theorem 2.5.1).■

The Lebesgue integral has some important properties that we will often use. They are as follows:

(L1) If  $f, g : X \rightarrow R$  are measurable,  $g$  is integrable, and  $|f(x)| \leq g(x)$ , then  $f$  is integrable and

$$\left| \int_X f(x) d\mu_x \right| \leq \int_X g(x) d\mu_x.$$

(L2)

$$\int_X |f(x)| d\mu_x = 0 \text{ if and only if } f(x) = 0 \text{ a.e.}$$

(L3) If  $f_1, f_2 : X \rightarrow R$  are integrable functions, then for  $\lambda_1, \lambda_2 \in R$  the linear combination  $\lambda_1 f_1 + \lambda_2 f_2$  is integrable and

$$\begin{aligned} & \int_X [\lambda_1 f_1(x) + \lambda_2 f_2(x)] d\mu_x \\ &= \lambda_1 \int_X f_1(x) d\mu_x + \lambda_2 \int_X f_2(x) d\mu_x. \end{aligned}$$

(L4) Let  $f, g : X \rightarrow R$  be measurable functions and  $f_n : X \rightarrow R$  be measurable functions such that  $|f_n(x)| \leq g(x)$  and  $f_n(x)$  converges to  $f(x)$  almost everywhere. If  $g$  is integrable, then  $f$  and  $f_n$  are also integrable and

$$\lim_{n \rightarrow \infty} \int_X f_n(x) d\mu_x = \int_X f(x) d\mu_x.$$

(L5) Let  $f : X \rightarrow R$  be an integrable function and the sets  $A_i \in \mathcal{A}, i = 1, 2, \dots$ , be disjoint. If  $A = \bigcup_i A_i$ , then

$$\sum_i \int_{A_i} f(x) d\mu_x = \int_A f(x) d\mu_x.$$

**Remark 2.5** The properties described in (L4) are often referred to as the **Lebesgue dominated convergence theorem** ( $|f_n(x)| \leq g(x)$ ). It is called the **Lebesgue monotone convergence theorem** if  $|f_n(x)| \leq g(x)$  with an integrable  $g$  is replaced by  $0 \leq f_1(x) \leq f_2(x) \dots$ . ■

**Remark 2.6** Another observation is that from our construction of the Lebesgue integral, is that for every integrable function  $f$  there is a sequence of simple functions

$$f_n(x) = \sum_i \lambda_{i,n} I_{A_{i,n}}(x)$$

such that

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ a.e. and } |f_n(x)| \leq |f(x)|.$$

Thus by the Lebesgue dominated convergence theorem (L4), we have

$$\lim_{n \rightarrow \infty} \int_X f_n(x) d\mu_x = \int_X f(x) d\mu_x. \blacksquare$$



We note that the double subscript  $i, n$  refers to the double sequence where  $i, n \in \mathbb{N}$ , the set of natural numbers. The subscript  $n$  represents the summation of the sequence of simple functions  $f_n$ , and the subscript  $i$  represents the summation of each simple function in its defined representation.

We use this observation to simplify some of the proofs in our analysis. We do this by showing that if a relation holds for characteristic functions,  $I_{A_i}$  ( $A_i$  disjoint in  $\mathcal{A}$ ), then it must hold for linear simple functions  $f = \sum_i a_i A_i$ . Therefore we,

- (1) need to only verify some formula for simple functions, and then,
- (2) pass to the limit.

**Remark 2.7** From the properties and the definition of the Lebesgue integral it follows that if  $f : X \rightarrow \mathbb{R}$  is a nonnegative integrable function, then  $\mu_f(A)$ , defined by

$$\mu_f(A) = \int_A f(x) d\mu_x,$$

is a nonnegative, additive and finite measure. Further, from property (L2) if  $\mu(A) = 0$ , then

$$\mu_f(A) = \int_A I_A(x) f(x) d\mu_x = 0$$

since  $I_A(x)f(x) = 0$  a.e. Thus  $\mu_f(A)$  satisfies the properties of a measure and  $\mu_f(A) = 0$  whenever  $\mu(A) = 0$ . This observation that every integrable nonnegative function defines a finite measure can be reversed by the following theorem, which is of fundamental importance for the development of the theory of the Frobenius-Perron operator. ■

**Theorem 2.1** [Radon-Nikodym theorem] Let  $(X, \mathcal{A}, \mu)$  be a measure space where  $\mu$  is a finite measure. Let  $\nu$  be a second finite measure with the property that  $\nu(A) = 0$  for all  $A \in \mathcal{A}$  such that  $\mu(A) = 0$ . Then there exists a nonnegative integrable function  $f : X \rightarrow \mathbb{R}$  such that

$$\nu(A) = \int_A f(x) d\mu_x \text{ for all } A \in \mathcal{A}.$$

The function  $f$  is in fact unique as can be seen from the following proposition.

**Proposition 2.1** If  $f_1$  and  $f_2$  are integrable functions such that

$$\int_A f_1(x) d\mu_x = \int_A f_2(x) d\mu_x \text{ for } A \in \mathcal{A}$$

then  $f_1 = f_2$  a.e.

Next we introduce the concept of an  $L^p$  space.

**Definition 2.17** Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $p$  be a real number,  $1 \leq p < \infty$ . The family of all real-valued measurable functions  $f : X \rightarrow \mathbb{R}$  satisfying

$$\int_X |f(x)|^p d\mu_x < \infty \tag{2.1}$$

is called the  $L^p(X, \mathcal{A}, \mu)$  space.

Note that if  $p = 1$  then the  $L^1$  space consists of all integrable functions.

The integral appearing in Equation(2.1) is important for an element  $f \in L^p$ . Thus it is assigned the special notation

$$\|f\|_{L^p} = \left[ \int_X |f(x)|^p d\mu_x \right]^{\frac{1}{p}} \quad (2.2)$$

and is called the  **$L^p$  norm** of  $f$ .

**Remark 2.8** When property L2 of the Lebesgue integral is applied to  $|f|^p$ , it follows that the condition  $\|f\|_{L^p} = 0$  is equivalent to  $f(x) = 0$  a.e. Or, more precisely,  $\|f\|_{L^p} = 0$  if and only if  $f$  is a zero element in  $L^p$  (which is an element represented by all functions equal to zero almost everywhere).■

Two important properties of the norm are

(N1) The norm is homogeneous, that is

$$\|\alpha f\|_{L^p} = |\alpha| \cdot \|f\|_{L^p} \text{ for } f \in L^p, \alpha \in \mathbb{R}.$$

(N2) The norm satisfies the triangle inequality, that is

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p} \text{ for } f, g \in L^p.$$

If we think of  $f, g$  and  $f + g$  as vectors, we can consider a triangle with sides  $f, g$  and  $f + g$ . Then property (N2) simply means that the length of the side  $(f + g)$  is not longer than the sum of the lengths of the other two sides. Furthermore, from (N1) it follows that for every  $f \in L^1$  and real  $\alpha$ , the product  $\alpha f$  belongs to  $L^p$ , and from (N2) for every  $f, g \in L^1$  the sum  $f + g$  also belongs to  $L^p$ . Together properties (N1) and (N2) imply that  $L^p$  is a vector space.

Since the value of  $\|f\|_{L^p}$  is interpreted as the length of  $f$ , we say that

$$\|f - g\|_{L^p} = \left[ \int_X |f(x) - g(x)|^p d\mu_x \right]^{\frac{1}{p}}$$

is the  $L^p$  **distance** between  $f$  and  $g$ .

However the product  $fg$  of two functions  $f, g \in L^p$  is not necessarily in  $L^p$ , for example,  $f(x) = x^{-\frac{1}{2}}$  is integrable on  $[0, 1]$  but  $[f(x)]^2 = x^{-1}$  is not.

This leads us to introduce the space adjoint to  $L^p$ .

**Remark 2.9** Let  $(X, \mathcal{A}, \mu)$  be a measure space. The **space adjoint** to  $L^p(X, \mathcal{A}, \mu)$  is  $L^{p'}(X, \mathcal{A}, \mu)$  where

$$\frac{1}{p} + \frac{1}{p'} = 1. \blacksquare$$

**Remark 2.10** If  $p = 1$ , then the adjoint space consists of all measurable bounded almost everywhere functions, and is denoted by  $L^\infty$ .  $\blacksquare$

For  $f \in L^1, g \in L^\infty$ , we take the  $L^\infty$  norm of  $g$  to be the smallest constant  $c$  such that

$$|g(x)| \leq c$$

for almost all  $x \in X$ . This constant is denoted by  $\text{ess sup } |g(x)|$ , and is called the **essential supremum** of  $g$ .

From the Cauchy-Hölder inequality (Cohn 1980, Proposition 3.3.2) if  $f \in L^p$  and  $g \in L^{p'}$  then  $fg \in L^1$  i.e  $fg$  is integrable and satisfies

$$| \langle f, g \rangle | \leq \|f\|_{L^p} \cdot \|g\|_{L^{p'}}$$

where

$$\langle f, g \rangle = \int_X f(x)g(x)d\mu_x$$

is called the **duality pairing** of two functions.

**Remark 2.11** As we shall usually work in  $L^1$ , we will not indicate the space in which the norm is taken unless it is not  $L^1$ . Thus we will write  $\|f\|$  instead of  $\|f\|_{L^1}$ . Further we observe from property (L3) that in  $L^1$  the norm has the property that the triangle inequality is sometimes an equality.■

The following corollary provides a simplification of the Radon-Nikodym theorem in the space  $L^1$ .

**Corollary 2.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $\nu$  a finite measure on  $\mathcal{A}$  such that  $\nu(A) = 0$  whenever  $\mu(A) = 0$ , then there exists a unique element  $f \in L^1$  such that*

$$\nu(A) = \int_A f(x)d\mu_x \text{ for } A \in \mathcal{A}.$$

**Definition 2.18** *The **Cartesian product** of two arbitrary sets  $A_1$  and  $A_2$  is the set of all pairs  $(x_1, x_2)$  such that  $x_1 \in A_1$  and  $x_2 \in A_2$ . It is written as*

$$A_1 \times A_2 = \{(x_1, x_2) : x_1 \in A_1, x_2 \in A_2\}$$

*The **Cartesian product** of the sets  $A_1, \dots, A_d$  is the set of all sequences  $(x_1, \dots, x_d)$  such that  $x_i \in A_i, i = 1, \dots, d$ , or*

$$A_1 \times \dots \times A_d = \{(x_1, \dots, x_d) : x_i \in A_i \text{ for } i = 1, \dots, d\}$$

An important consequence following the concept of the Cartesian product is that if a structure is defined on each of the factors  $A_i$ , for example, a measure, then it is possible to extend that property to the Cartesian product. Thus, given  $d$  measure spaces  $(X_i, \mathcal{A}_i, \mu_i), i = 1, \dots, d$ , we define

$$X = X_1 \times \dots \times X_d, \quad (2.3)$$

$\mathcal{A}$  to be the smallest  $\sigma$ - algebra of subsets of  $X$  containing all sets of the form

$$A_1 \times \dots \times A_d \text{ with } A_i \in \mathcal{A}_i, i = 1, \dots, d, \quad (2.4)$$

and

$$\mu(A_1 \times \dots \times A_d) = \mu_1(A_1) \dots \mu_d(A_d) \text{ for } A_i \in \mathcal{A}_i. \quad (2.5)$$

However, these sets by themselves they do not define a measure space  $(X, \mathcal{A}, \mu)$ . There is no problem with either  $X$  or  $\mathcal{A}$ , but  $\mu$  is defined only on special sets, namely  $A = A_1 \times \dots \times A_d$ , that do not form a  $\sigma$ -algebra. The following theorem shows that  $\mu$ , defined by Equation (2.5) can be extended to the entire  $\sigma$  - algebra  $\mathcal{A}$ .

**Theorem 2.2** *Let  $(X_i, \mathcal{A}_i, \mu_i), i = 1, \dots, d$  be measure spaces defined by Equation (2.3), (2.4), and (2.5) respectively, then there exists a unique extension of  $\mu$  to a measure defined on  $\mathcal{A}$ .*

The measure space  $(X, \mathcal{A}, \mu)$  in Theorem 2.2 is called the **product of measure spaces**  $(X_1, \mathcal{A}_1, \mu_1), \dots, (X_d, \mathcal{A}_d, \mu_d)$ , or the **product space**. The measure  $\mu$  is called the **product measure**.

From Equation (2.5) it follows that

$$\mu(X_1 \times \dots \times X_d) = \mu(X_1) \cdots \mu(X_d).$$

Thus, if all measure spaces  $(X_i, \mathcal{A}_i, \mu_i)$  are finite or probabilistic, then  $(X, \mathcal{A}, \mu)$  will also be finite or probabilistic. The next theorem shows that integrals on the product of measure spaces are related to integrals on the individual factors.

**Theorem 2.3** [Fubini's theorem] *Let  $(X, \mathcal{A}, \mu)$  be the product space formed by  $(X_1, \mathcal{A}_1, \mu_1)$  and  $(X_2, \mathcal{A}_2, \mu_2)$ , and let a  $\mu$ -integrable function  $f : X \rightarrow R$  be given. Then, for almost every  $x_1$ , the function  $f(x_1, x_2)$  is  $\mu_2$  integrable with respect to  $x_2$ . Furthermore the function*

$$\int_{\check{X}_2} f(x_1, x_2) d\mu_{x_2}$$

*of the variable  $x_1$  is  $\mu_1$  integrable and*

$$\int_{\check{X}_1} \left\{ \int_{\check{X}_2} f(x_1, x_2) d\mu_{x_2} \right\} d\mu_{x_1} = \int_X f(x_1, x_2) d\mu_{(x_1, x_2)}. \quad (2.6)$$

*(We note that  $d\mu_{x_1}$  and  $d\mu_{x_2}$  can also be written as  $\mu(dx_1)$  and  $\mu(dx_2)$  respectively.)*

*If  $(X, \mathcal{A}, \mu)$  is the product of the measure spaces  $(X_i, \mathcal{A}_i, d\mu_i), i = 1, \dots, d$ , and  $f : X \rightarrow R$  is  $\mu$  integrable, then*

$$\begin{aligned} & \int \cdots \int_X f(x_1, \dots, x_d) d\mu_{(x_1 \dots x_d)} \\ &= \int_{\check{X}_1} \left\{ \cdots \int_{\check{X}_{d-1}} \left[ \int_{\check{X}_d} f(x_1, \dots, x_d) d\mu_{x_d} \right] d\mu_{x_{d-1}} \cdots \right\} d\mu_{x_1}. \end{aligned}$$

**Remark 2.12** As we noted in Example 2.2, the 'natural' Borel measure on the real line  $R$  is defined on the smallest  $\sigma$ -algebra  $\mathcal{B}$  that contains all intervals. For every interval  $[a, b]$  this measure satisfies  $\mu([a, b]) = b - a$ . Having the structure  $(R, \mathcal{B}, \mu)$  we define by Theorem 2.2 the product space  $(R^d, \mathcal{B}^d, \mu^d)$ , where

$$R^d = R \times \cdots \times R, (d \text{ times})$$

$\mathcal{B}^d$  is the smallest  $\sigma$ -algebra containing all sets of the form

$$A_1 \times \cdots \times A_d \text{ with } A_i \in \mathcal{B},$$

and

$$\mu^d(A_1 \times \cdots \times A_d) = \mu(A_1) \cdots \mu(A_d).$$

The measure  $\mu^d$  is again called the Borel measure. It can be shown that  $\mathcal{B}^d$  may be alternatively defined as either the smallest  $\sigma$ -algebra containing all the rectangles

$$[a_1, b_1] \times \cdots \times [a_d, b_d],$$

or as the smallest  $\sigma$ -algebra containing all the open subsets of  $R^d$ . The former is a consequence of Proposition 1.1.4 (Cohn 1980), whilst the latter follows from the fact that if we generate a Borel  $\sigma$ -algebra from open intervals, then every open interval since it is the union of sequences of the form  $[a, b)$  is a Borel set. Furthermore, every interval  $[a, b)$  is the intersection of a sequence of open intervals. It therefore follows that



$$\mu^d([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d),$$

which is the formula for the volume of an  $n$ -dimensional box.

The same construction may be repeated by starting, not from the whole real line  $R$ , but from the unit interval  $[0, 1]$  or from any other finite interval. Thus, from Theorem 2.2, we will obtain the Borel measure on the unit square  $[0, 1] \times [0, 1]$  or on the  $d$ -dimensional cube

$$[0, 1]^d = [0, 1] \times \cdots \times [0, 1] \quad (d \text{ times}).$$

In all cases we will omit the superscript  $d$  on  $\mathcal{B}^d$  and  $\mu^d$  and write  $(R^d, \mathcal{B}, \mu)$  instead of  $(R^d, \mathcal{B}^d, \mu^d)$ . Furthermore, in all cases when the space is  $R$ ,  $R^d$ , or any subset of these ( $[0, 1]$ ,  $[0, 1]^d$ , etc.) and the measure space and  $\sigma$ -algebra are not specified, we will assume the measure space is taken with the Borel  $\sigma$ -algebra and Borel measure. Finally, all the integrals on  $R$  or  $R^d$  taken with respect to the Borel measure will be written with  $dx$  instead of  $d\mu_x$ . ■

In order to define a monotonic measure we first define **partial order**:

a **partial order** on a set  $X$  is a relation  $\leq$  that is reflexive ( $x \leq x$  holds for each  $x \in X$ ), antisymmetric (if  $x \leq y$  and  $y \leq x$ , then  $x = y$ ), and transitive (if  $x \leq y$  and  $y \leq z$ , then  $x \leq z$ ). If  $\leq$  is a partial order on a set  $X$ , then  $x < y$  means that  $x$  and  $y$  satisfy  $x \leq y$  but are not equal.

Hence from the additivity property of a measure (Definition 2.2c) it follows that every measure is **monotonic**, that is, if  $A$  and  $B$  are measurable sets and  $A \subset B$  then  $\mu(A) \leq \mu(B)$ .

This is so since,

the sets  $A$  and  $B - A$  are disjoint and satisfy  $B = A \cup (B - A)$ , thus the additivity of  $\mu$  implies that

$$\mu(B) = \mu(A) + \mu(B - A).$$

Since  $\mu(B - A) \geq 0$  this implies that  $\mu(A) \leq \mu(B)$ .

## 2.3 Convergence of Sequences of Functions

We now define, using the notions of norms and duality pairings, three different types of convergence for a sequence of functions in  $L^p$  spaces.

**Definition 2.19** *A sequence of functions  $\{f_n\}, f_n \in L^p, 1 \leq p < \infty$ , is called weakly Cesàro convergent to  $f \in L^p$  if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \langle f_k, g \rangle = \langle f, g \rangle \text{ for all } g \in L^{p'}.$$

**Definition 2.20** *A sequence of functions  $\{f_n\}, f_n \in L^p, 1 \leq p < \infty$ , is called weakly convergent to  $f \in L^p$  if*

$$\lim_{n \rightarrow \infty} \langle f_n, g \rangle = \langle f, g \rangle \text{ for all } g \in L^{p'}.$$

**Definition 2.21** *A sequence of functions  $\{f_n\}, f_n \in L^p, 1 \leq p \leq \infty$ , is called strongly convergent to  $f \in L^p$  if*

$$\lim_{n \rightarrow \infty} \|f_n - f\|_{L^p} = 0.$$

We observe that strong convergence implies weak convergence. In fact from the Cauchy-Hölder inequality, we have

$$| \langle f_n - f, g \rangle | \leq \|f_n - f\|_{L^p} \cdot \|g\|_{L^{p'}},$$

and thus since  $\|f_n - f\|_{L^p} \rightarrow 0$  as  $n \rightarrow \infty$  so must  $\langle f_n - f, g \rangle$  which implies the weak convergence of  $\{f_n\}$  to  $f \in L^p$ . However, to prove the condition for weak convergence, it is sufficient to check weak convergence for a restricted class of functions. We start with the following definitions.

**Definition 2.22** *A set  $D_0 \subset D$  is called **dense** in  $D$  if, for every  $h \in D$  and  $\epsilon > 0$ , there is a  $g \in D_0$  such that  $\|h - g\| < \epsilon$ .*

**Definition 2.23** *A subspace  $K \subset L^p$  is called **linearly dense** if for each  $f \in L^p$  and  $\epsilon > 0$  there are  $g_1, \dots, g_n \in K$  and constants  $\lambda_1, \dots, \lambda_n$  such that*

$$\|f - g\|_{L^p} \leq \epsilon,$$

where

$$g = \sum_{i=1}^n \lambda_i g_i.$$

With the use of linearly dense sets, it is possible to simplify the proof of weak convergence. If the sequence  $\{f_n\}$  is bounded in norm, that is,  $\|f_n\|_{L^p} \leq c < \infty$ , and if  $K$  is linearly dense in  $L^{p'}$ , then it is sufficient to check weak convergence in Definition 2.20 for any  $g \in K$ . (Kreyszig 1978 Uniform Boundedness theorem)

It is also possible to compare convergence in different  $L^p$  spaces by using the following proposition:

**Proposition 2.2** *If  $(X, \mathcal{A}, \mu)$  is a finite measure space and  $1 \leq p_1 < p_2 \leq \infty$ , then*

$$\|f\|_{L^{p_1}} \leq c \|f\|_{L^{p_2}} \text{ for every } f \in L^{p_2}$$

*where  $c$  depends on  $\mu(X)$ . Thus every element of  $L^{p_2}$  belongs to  $L^{p_1}$ , and strong convergence in  $L^{p_2}$  implies strong convergence in  $L^{p_1}$ .*

The **Cauchy condition for convergence** states that if  $\{f_n\}$  is strongly convergent in  $L^p$  to  $f$ , then

$$\lim_{n \rightarrow \infty} \|f_{n+k} - f_n\|_{L^p} = 0 \text{ uniformly for all } k \geq 0. \quad (2.7)$$

It can be proved that all  $L^p$  spaces ( $1 \leq p \leq \infty$ ) have the property that the above condition (2.7) is also sufficient for convergence. This is stated more precisely in the following theorem.

**Theorem 2.4** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and let  $\{f_n\}, f_n \in L^p(X, \mathcal{A}, \mu)$  be a sequence such that Equation (2.7) holds. Then there exists an element  $f \in L^p(X, \mathcal{A}, \mu)$  such that  $\{f_n\}$  converges strongly to  $f$ , that is Definition 2.21 is satisfied.*

The fact that Theorem 2.4 holds for  $L^p$  spaces is referred to by saying that  $L^p$  spaces are **complete**.

Finally to close this section we state the following theorem which will be used in this analysis.

**Theorem 2.5** [*Hahn-Banach theorem*] *Let  $X$  be a normed linear space, let  $Z$  be a linear subspace of  $X$ , and  $f_0$  be a continuous linear functional on  $Z$ . Then there is*

a continuous linear functional  $f$  on  $X$  such that  $\|f\| = \|f_0\|$  and such that  $f_0$  is the restriction of  $f$  to  $Z$ .

## 2.4 Probability Theory

In this section we review some material from probability theory that is necessary to understand the Poisson processes that are discussed in Section 8.2.

The key notion of probability theory is that of a *probability space*  $(\Omega, \mathcal{F}, \text{prob})$ , where  $\Omega$  is a nonempty set called the space of all *elementary events*,  $\mathcal{F}$  is a  $\sigma$ -algebra of all subsets of  $\Omega$ , which are called *events*, and "prob" is a normalized measure on  $\mathcal{F}$ . The equation

$$\text{prob}(A) = p, A \in \mathcal{F}$$

means that the probability of event  $A$  is  $p$ . Since prob is a measure it follows that

$$\text{prob}\left(\bigcup_i A_i\right) = \sum_i \text{prob}(A_i),$$

where the  $A_i \in \mathcal{F}$  are mutually disjoint, that is,  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ .

**Definition 2.24** A sequence of events  $A_1, A_2, \dots$  (finite or not) are called **independent** if, for any increasing sequence of integers  $k_1 < k_2 < \dots < k_n$ ,

$$\text{prob}(A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_n}) = \text{prob}(A_{k_1}) \cdots \text{prob}(A_{k_n}).$$

This simply means that the probability of all the events  $A_{k_i}$  occurring is the product of the probabilities that each will occur separately.

**Definition 2.25** A random variable  $\xi$  is a measurable transformation from  $\Omega$  into  $R$ . More precisely,  $\xi : \Omega \rightarrow R$  is a random variable if, for any Borel set  $B \subset R$ ,

$$\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{F}.$$

This set is usually written as  $\{\xi \in B\}$ . Thus, for any Borel set  $B \subset R$ ,  $\text{prob}\{\xi \in B\}$  is well defined.

A function  $f \in D(R)$  is called a **density** of the random variable  $\xi$  if

$$\text{prob}\{\xi \in B\} = \int_B f(x)dx$$

for any Borel set  $B \subset R$ .

Let  $\xi_1, \xi_2, \dots$  be a sequence of random variables. We say the  $\xi_i$  are **independent** if, for any sequence of Borel sets  $B_1, B_2, \dots$  the events

$$\{\xi_1 \in B_1\}, \{\xi_2 \in B_2\}, \dots$$

are independent. Thus a finite sequence of independent random variables satisfies

$$\text{prob}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \text{prob}\{\xi_1 \in B_1\} \dots \text{prob}\{\xi_n \in B_n\},$$

and the probability that all events  $\{\xi_i \in B_i\}$  will occur is given by the product of the probabilities that each will occur separately.

Using the above definitions we now define a stochastic process.

**Definition 2.26** A stochastic process  $\{\xi_t\}$  is a family of random variables that depends on a parameter  $t$ , which we usually call time.

If  $t$  assumes only integer values,  $t = 1, 2, \dots$ , then the stochastic process reduces to a sequence  $\{\xi_n\}$  of random variables called a **discrete time stochastic process**. However, if  $t \in R$  or  $R^+$  then the stochastic process is called a **continuous time stochastic process**.

A stochastic process  $\{\xi_t\}$  is a function of two variables, namely, time  $t$  and event  $\omega$ . If the time is fixed, then  $\xi_t$  is simply a random variable. However, if  $\xi$  is fixed, then the mapping  $t \rightarrow \xi_t(\omega)$  is called the **sample path** of the stochastic process.

In the following definition we describe two important properties that stochastic processes may have.

**Definition 2.27** *A continuous time process  $\{\xi_t\}_{t \geq 0}$  has **independent increments** if, for any sequence of times  $t_0 < t_1 < \dots < t_n$ , the random variables*

$$\xi_{t_1} - \xi_{t_0}, \xi_{t_2} - \xi_{t_1}, \dots, \xi_{t_n} - \xi_{t_{n-1}}$$

*are independent. Further, if for any  $t_1$  and  $t_2$  and Borel set  $B \subset R$ ,*

$$\text{prob}\{\xi_{t_2+t'} - \xi_{t_1+t'} \in B\}$$

*does not depend on  $t'$ , then the continuous time stochastic process  $\{\xi_t\}$  has **stationary independent increments**.*

We are now able to define counting process which will enable us to define and discuss a Poisson process.

**Definition 2.28** *A stochastic process  $\{\xi_t\}$  is called a **counting process** if its sample paths are non-decreasing functions of time, with integer values.*

Counting processes will be denoted by  $\{N_t\}_{t \geq 0}$ .



## 3 The Relationship between Densities and Dynamical Systems

### 3.1 Introduction

Our main objective is to study the types of irregular behaviour displayed by any dynamical system as it evolves in time. We examine the conditions governing such behaviours and the changes, if any, that are observed. However, often for a large system containing many elements it is not practical to account for the evolution of each element within our system.

When the quantity in the evolution that we analyse has a probabilistic interpretation, for example the probability of finding a particle in a certain state at a certain time, we can adopt a probabilistic approach by defining the evolution of suitable averages called (probability) densities. In what follows we shall discuss the notion of the evolution of densities under the operation of deterministic transformations in a dynamical system.

## 3.2 Defining a Dynamical System

When we are talking about a system we mean a variable or a collection of variables describing the state of the system. The same real (physical, biological etc.) system may be represented in many ways.

If we are interested in, say, the evolution of the average temperature of a body, then at each time the state of the system will be described by a single variable and thus the system will be one-dimensional. If we want to know the temperature at each point of the body, then for each time the state of the system is described by a function of three spatial variables, and, in such a case the system is infinite dimensional since the set, say, of all continuous functions is not a finite dimensional space. We describe the system by a variable taken from some set which we call the **state space**. In other words, the state space is that set consisting of all the distinct possible states of a system.

A **dynamical system** is one whose states  $x \in X$  ( $X$  is the state space) change with some parameter  $t$  (time). Two main types of dynamical systems occur in our study: if the time variable  $t \in Z^+$  (or  $Z$ ), then we refer to the system as a **discrete dynamical system**, whereas if  $t \in R^+$  (or  $R$ ) we refer to the system as a **continuous dynamical system**.

Discrete dynamical systems can be represented as the iteration of a function

$$x_{t+1} = f(x_t), \quad t \in Z \text{ or } t \in N.$$

When  $t$  is continuous, the dynamics are usually described by a differential equation

$$\frac{dx}{dt} = \dot{x} = A(x), \quad t \in R \text{ or } R^+.$$

The functions  $f$  and  $A$ , respectively, describe mechanisms forcing the evolution of the system.

Theories for discrete and continuous dynamical systems are to some extent parallel.

### 3.3 Defining a Density

As an example let us consider the system described by the transformation  $S : X \rightarrow X$  defined by the quadratic map

$$S(x) = 4x(1 - x)$$

Then,  $S$  maps the closed unit interval  $[0, 1]$  onto itself so that  $[0, 1]$  is the state space of the system.

We next pick an initial point  $x^0 \in [0, 1]$  so that successive states of the system at times  $1, 2, \dots$  are given by the trajectory

$$x^0, S(x^0), S^2(x^0) = S(S(x^0)), \dots$$

The following observations have been made of the typical trajectory corresponding to a given initial state: (Lasota and Mackey 1985 and Smith 1998).

- (1) it is visibly chaotic or erratic for almost all  $x^0$ .
- (2) it shows sensitive dependence on initial conditions.

To overcome these problems, we shall, instead of examining the trajectories of single

points, define suitable averages of the states of the system at each given time in the evolution of the system. We refer to this representation of the averages of the states as **densities**.

Again, let us consider the transformation  $S : [0, 1] \rightarrow [0, 1]$  and pick a large number  $N$  of initial states

$$x_1^0, x_2^0, \dots, x_N^0.$$

To each of these states we apply the map  $S$  to obtain  $N$  new states denoted by

$$x_1^1 = S(x_1^0), x_2^1 = S(x_2^0), \dots, x_N^1 = S(x_N^0).$$

To define the densities of the states  $x_i^0$  and  $x_i^1, i = 1, 2, \dots, N$  we use the **characteristic function** (see Definition 2.11). We say that a function  $f_0(x)$  is the **density function** for the initial states  $x_1^0; \dots; x_N^0$  if for every (not too small) interval  $\Delta_0 \subset [0; 1]$  we have

$$\int_{\Delta_0} f_0(u) du \simeq \frac{1}{N} \sum_{j=1}^N I_{\Delta_0}(x_j^0) \quad (3.1)$$

Similarly the density function  $f_1(x)$  for the states  $x_1^1, \dots, x_N^1$  satisfies for  $\Delta \subset [0, 1]$

$$\int_{\Delta} f_1(u) du \simeq \frac{1}{N} \sum_{j=1}^N I_{\Delta}(x_j^1) \quad (3.2)$$

Next we try to establish a relationship between  $f_0$  and  $f_1$ . This relationship will enable us to devise the means to obtain successive densities representing the subsequent states of our system. We use the notion of the **counterimage** (see Definition 2.7) of an interval  $\Delta \subset [0, 1]$  under the operation of the map  $S$ .

We note that for  $\Delta \subset [0, 1]$

$$x_j^1 \in \Delta \iff x_j^0 \in S^{-1}(\Delta),$$

and consequently that

$$I_\Delta(S(x)) = I_{S^{-1}(\Delta)}(x). \quad (3.3)$$

Hence from (3.3) we may rewrite (3.2) as

$$\int_{\Delta} f_1(u) du \simeq \frac{1}{N} \sum_{j=1}^N I_{S^{-1}(\Delta)}(x_j^0) \quad (3.4)$$

Since  $\Delta_0$  and  $\Delta$  have been arbitrary up to this point we may pick  $\Delta_0 = S^{-1}(\Delta)$ .

From this the right hand side of (3.1) and (3.4) are equal and therefore

$$\int_{\Delta} f_1(u) du = \int_{S^{-1}(\Delta)} f_0(u) du \quad (3.5)$$

Thus (3.5) is the relationship that we sought between  $f_0$  and  $f_1$  and it tells us how a density of initial states  $f_0$  will be transformed by a given map  $S$  into a new density  $f_1$ .

Now, if  $\Delta$  is an interval, say,  $\Delta = [a, x]$ , it is possible to obtain an explicit representation for  $f_1$ . In this case (3.5) becomes

$$\int_a^x f_1(u) du = \int_{S^{-1}([a, x])} f(u) du$$

and differentiating w.r.t  $x$  we obtain

$$f_1(x) = \frac{d}{dx} \int_{S^{-1}([a,x])} f(u) du \quad (3.6)$$

Since  $f_0$  is arbitrary and  $f_1$  depends on  $f_0$  we may rewrite (3.6) as

$$Pf(x) = \frac{d}{dx} \int_{S^{-1}([a,x])} f(u) du \quad (3.7)$$

Thus, (3.7) explicitly defines the so called Frobenius-Perron operator  $P : L^1 \rightarrow L^1$  (which we shall discuss in detail in the next Chapter) corresponding to a transformation  $S$  and it tells us how  $S$  transforms a given density  $f$  into a new density  $Pf$ . Successive densities are then obtained by iterates of  $P$ , namely  $P(Pf) = P^2f, P(P^2f) = P^3f, \dots$

## 4 Markov Operators

In this chapter we motivate our use of the framework of Markov operators by highlighting its main characteristics. Subsequently we introduce two other operators which play an important role in our study: the Frobenius-Perron operator and its adjoint, the Koopman operator.

### 4.1 Markov Operators

We begin with a brief discussion of Markov chains.

Consider a random movement of a mouse through a maze of rooms by assuming that the mouse changes rooms at times  $n = 1, 2, \dots$ . Let  $X_n$  = the number of the room occupied by the mouse at time  $n$ . Then the probability that the mouse will be in room  $X_n$  at time  $n$  depends on his location at time  $n - 1$  and not on his earlier times.

Hence if we speak of an element in a dynamical system as being in state  $i$  at time  $n - 1$  we mean that  $X_{n-1} = i$ , but, since its movement between states is random we have no means of predicting its state at time  $n$ . From this we understand that the value of the  $X_{n-1}$  depends on the state which is subject to chance. Thus the value

of  $X_{n-1}$  varies in some 'random' manner.

We can therefore predict the probability of an element being in a certain state at a certain time, but this is conditional to the elements' location at time  $n - 1$ . This conditional probability  $\mathcal{P}$  is given by

$$\mathcal{P}[X_n = j | X_{n-1} = i]$$

Thus if the state of an element before the time  $n - 1$  does not alter this conditional probability, we then say that the dynamical system satisfies the Markov property which is defined as follows:

**Definition 4.1** *A stochastic process  $\{X_k\}, k = 1, 2, \dots$  with state space  $S = \{1, 2, \dots\}$  is said to satisfy the **Markov property** if for every  $n$  and all states  $i_1, i_2, \dots, i_n$  we have that*

$$\mathcal{P}[X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1] = \mathcal{P}[X_n = i_n | X_{n-1} = i_{n-1}]$$

In other words the impact of the past on the future evolution of the system is concentrated in the state present at the last moment at which the system was observed. This can also be expressed as saying that the system is "memoryless". Once in a certain state at a given time, the way in which the system reached that state does not affect its' future evolution. This does not however mean a complete absence of memory since the Markov system retains its most recent recollection of the past. This simply means that the present value of the state will depend only on its value in its immediate past. And in so saying, is, one step away from complete independence.

In the analysis of a typical dynamical system, we showed (see Chapter 3) that it was



necessary to represent the states at a given time as a (probability) density. There is therefore a need to introduce a suitable operator that would serve to generate an evolution of the density of states within the Markov process. This operator has to ensure that the two important properties:

- (1) each state is defined by a nonnegative function and
- (2) the overall quantity of the elements in the system remains preserved

These are the main characteristics of the Markov operator which is defined as follows:

**Definition 4.2** *Let  $(X, A, \mu)$  be a measure space. A **Markov operator** is a linear operator  $P : L^1 \rightarrow L^1$  such that*

1.  $Pf \geq 0$  for  $f \geq 0, f \in L^1$ ; and
2.  $\|Pf\| = \|f\|$  for  $f \geq 0, f \in L^1$

Hence the Markov operator is nonnegative and preserves the total number of elements in our system.

Markov operators have many properties. We illustrate the first with the following:

We say that a Markov operator is **monotonic** if for  $f, g \in L^1$

$$Pf(x) \geq Pg(x) \text{ whenever } f(x) \geq g(x).$$

Further properties of the Markov operator are stated in the following proposition.

**Proposition 4.1** *Let  $(X, A, \mu)$  be a measure space and  $P$  a Markov operator, then, for every  $f \in L^1$  we have*

$$(M1) \quad (Pf(x))^+ \leq Pf^+(x)$$

$$(M2) \ (Pf(x))^- \leq Pf^-(x)$$

$$(M3) \ |Pf(x)| \leq P|f(x)|$$

$$(M4) \ ||Pf|| \leq ||f||$$

**Proof:** (M1) From the linearity of the Markov operator and the definitions of  $f^+$  and  $f^-$  (see Section 2.2) it follows that

$$(Pf)^+ = (Pf^+ - Pf^-)^+ = \max(0, Pf^+ - Pf^-) \leq \max(0, Pf^+) = Pf^+$$

(M2) Using a similar argument we obtain property (M2).

(M3) From (M1) and (M2) we have

$$\begin{aligned} |Pf| &= (Pf)^+ + (Pf)^- \leq Pf^+ + Pf^- \\ &= P(f^+ + f^-) = P|f(x)| \end{aligned}$$

(M4) To obtain (M4) using Definition 2.17 we integrate (M3) over  $X$  to get

$$\begin{aligned} ||Pf|| &= \int_X |Pf(x)| d\mu_x \leq \int_X P|f(x)| d\mu_x \\ &= \int_X |f(x)| d\mu_x = ||f|| \blacksquare \end{aligned}$$

We next state a definition of a density in the space  $L^1$ .

**Definition 4.3** *Let  $(X, A, \mu)$  be a measure space and the set  $D(X, A, \mu)$  be defined by*

*$D(X, A, \mu) = \{f \in L^1(X, A, \mu) | f \geq 0 \text{ and } ||f|| = 1\}$ . Any function  $f \in D(X, A, \mu)$  is called a **density**.*

**Definition 4.4** The normalized measure  $\mu_f$  for  $f \in D(X, A, \mu)$  is said to be **absolutely continuous** with respect to  $\mu$  if

$$\mu_f(A) = \int_A f(x) d\mu_x \text{ for } A \in \mathcal{A} \text{ and } f \text{ is called the density of } \mu_f(A).$$

**Definition 4.5** The **support** of a function  $g$  is the set of all  $x$  such that  $g(x) \neq 0$ , that is,  $\text{supp } g = \{x | g(x) \neq 0\}$

NOTE This definition is slightly different from the one used usually which uses topological notions.

The following example illustrates a relationship between the Markov operator and the support of a function. For this example we need to note the following definition.

A **stochastic / transition matrix** is an  $n \times n$  matrix whose entries are all nonnegative and whose columns add up to 1. Another result is that for the set  $X = \{1, \dots, N\}$  with the counting measure, then, for any Markov operator  $P : L^1(X) \rightarrow L^1(X)$  we have that

$$(Pf)_i = \sum_{j=1}^N p_{ij} f_j, \text{ where } i = 1, \dots, N$$

and where  $(p_{ij})$  is the stochastic matrix, that is,  $p_{ij} \geq 0$ ,  $\sum_{i=1}^N p_{ij} = 1$ .

**Example 4.1** Let  $P : L^1(X) \rightarrow L^1(X)$  be a Markov operator. We show that for every nonnegative  $f, g \in L^1$  the condition  $\text{supp } f \subset \text{supp } g$  implies  $\text{supp } Pf \subset \text{supp } Pg$ .

**Proof:** Let  $P : L^1(X) \rightarrow L^1(X)$  be a Markov operator defined by the stochastic matrix as stated above. Then

$$(Pf)_i = \sum_{j=1}^N p_{ij} f_j \text{ and } (Pg)_i = \sum_{j=1}^N p_{ij} g_j \quad (4.1)$$

where

$$p_{ij} \geq 0, \sum_{i=1}^N p_{ij} = 1, f(j) = f_j \text{ and } g(j) = g_j. \quad (4.2)$$

Let  $f, g \in L^1(X)$  such that  $f \geq 0$  and  $g \geq 0$ . Then by definition of the Markov operator  $Pf \geq 0$  and  $Pg \geq 0$ .

Suppose  $\text{supp } f \subset \text{supp } g$ . Then for any  $k \in X = \{1, \dots, N\}$  if  $k \in \text{supp } f$ , then  $k \in \text{supp } g$ . That is, if  $f_k \neq 0$ , then  $g_k \neq 0$ .

We show that if  $(Pf)_k \in \text{supp } Pf$  then  $(Pg)_k \in \text{supp } Pg$ , that is for any  $k \in 1, \dots, n$  if  $(Pf)_k > 0$  then  $(Pg)_k > 0$ .

Suppose that this is not true. Then there exists  $k \in \{1, \dots, N\}$  such that  $(Pf)_k \neq 0$  but  $(Pg)_k = 0$ . Hence for some  $k$  it follows that

$$p_{k1}f_1 + \dots + p_{kN}f_N \neq 0 \quad (4.3)$$

but,

$$p_{k1}g_1 + \dots + p_{kN}g_N = 0 \quad (4.4)$$

Since each term of (4.4) represents an entry in the stochastic matrix, it can be equal to zero or greater than zero. This means that in order to obtain a summation to zero every term of (4.4) must be equal to zero.

This therefore implies that for all  $i \in \{1, \dots, N\}$  we have that  $p_{ki} = 0$  or  $g_i = 0$ . Then, for any given  $i$ , either  $p_{ki} = 0$ , or  $g_i = 0$ . In the second case, from the assumption on supports,  $f_i = 0$ . Thus, each entry in the sum in (4.3) is equal to zero, which is a contradiction. Hence  $\text{supp} Pf \subset \text{supp} Pg$ .

Property (M4) of the Markov operator is called a contraction (Lasota and Mackey (1985)).

From this we observe that for any  $f \in L^1$  we have

$$\|P^n f\| = \|P(P^{n-1} f)\| \leq \|P^{n-1} f\|$$

and, thus, for any two  $f_1, f_2 \in L^1, f_1 \neq f_2$

$$\begin{aligned} \|P^n f_1 - P^n f_2\| &= \|P^n(f_1 - f_2)\| \\ &\leq \|P^{n-1}(f_1 - f_2)\| \\ &= \|P^{n-1} f_1 - P^{n-1} f_2\| \end{aligned} \tag{4.5}$$

Equation (4.5) demonstrates that during the process of iteration of two individual functions the distance between them decreases. This can be seen as some indication that our system is becoming more stable. This is known as the **stability property** of iterates of the Markov operator.

The contractive property (M4) can be a strict inequality as illustrated by the following proposition.

**Proposition 4.2**  $\|Pf\| = \|f\|$  if and only if  $Pf^+$  and  $Pf^-$  have disjoint supports.

**Proof:**  $\Rightarrow$ : Using the inequality  $|Pf^+(x) - Pf^-(x)| \leq |Pf^+(x)| + |Pf^-(x)|$

we observe that this inequality will clearly be strong if both  $Pf^+(x) > 0$  and

$Pf^-(x) > 0$ , and the equality will hold if  $Pf^+(x) = 0$  or  $Pf^-(x) = 0$ .

Thus by integrating over the space  $X$  we obtain

$$\int_X |Pf^+(x) - Pf^-(x)| d\mu_x = \int_X |Pf^+(x)| d\mu_x + \int_X |Pf^-(x)| d\mu_x \quad (4.6)$$

if and only if there does not exist a set  $A \in \mathcal{A}$ ,  $\mu(A) > 0$  such that  $Pf^+(x) > 0$  and  $Pf^-(x) > 0$  for  $x \in A$ , that is,  $Pf^+(x)$  and  $Pf^-(x)$  have disjoint supports.

$\Leftarrow$ : Suppose  $Pf^+$  and  $Pf^-$  have disjoint supports. Then since  $f = f^+ - f^-$ , the left hand integral in (4.6) is simply  $\|Pf\|$ .

Further, from Definition 4.2 the right hand integral is

$$\begin{aligned} \|Pf^+\| + \|Pf^-\| &= \|f^+\| + \|f^-\| \\ &= \|f\| \end{aligned}$$

and so we have  $\|Pf\| = \|f\|$ . ■

Having developed some of the more important and elementary properties of the Markov operator, we now introduce the concept of a fixed point.

**Definition 4.6** *If  $P$  is a Markov operator and for some  $f \in L^1$ ,  $Pf = f$ , then  $f$  is called a fixed point of  $P$ .*

Using this definition and properties (M1) and (M2) of Proposition 4.1 we have the following result.

**Proposition 4.3** *If  $Pf = f$  then  $Pf^+ = f^+$  and  $Pf^- = f^-$ .*

**Proof:** Suppose  $Pf = f$ . Therefore from Proposition 4.1 we have

$f^+ = (Pf)^+ \leq Pf^+$  and  $f^- = (Pf)^- \leq Pf^-$ , hence

$$\begin{aligned}
& \int_X [Pf^+(x) - f^+(x)]d\mu_x + \int_X [Pf^-(x) - f^-(x)]d\mu_x \\
&= \int_X [Pf^+(x) + Pf^-(x)]d\mu_x - \int_X [f^+(x) + f^-(x)]d\mu_x \\
&= \int_X P|f(x)|d\mu_x - \int_X |f(x)|d\mu_x \\
&= ||Pf|| - ||f||
\end{aligned}$$

Hence from the contractive property of  $P$  it follows that

$||Pf|| \leq ||f||$ ; which implies that  $||Pf|| - ||f|| \leq 0$ .

Since both the integrands  $(Pf^+ - f^+)$  and  $(Pf^- - f^-)$  are non-negative then  $||Pf|| - ||f||$  can only be equal to zero; so that  $Pf^+ = f^+$  and  $Pf^- = f^-$ . ■

We now define a **stationary density**.

**Definition 4.7** Let  $(X, A, \mu)$  be a measure space and  $P$  a Markov operator. Any  $f \in D$  that satisfies  $Pf = f$  is called a **stationary density** of  $P$ .

## 4.2 The Frobenius-Perron Operator

In Chapter 3 we briefly introduced the Frobenius-Perron operator. This operator is a special types of Markov operator. We will now examine its utility in our analysis

to describe the types of chaotic behaviour that we observe.

Suppose  $S : X \rightarrow X$  is a non-singular transformation on a measure space  $(X, \mathcal{A}, \mu)$ .

We now construct the definition for an operator  $P : L^1 \rightarrow L^1$ .

Let  $f \geq 0$  and  $f \in L^1$ . Consider the following

$$\int_{S^{-1}(A)} f(x) d\mu_x \quad (4.7)$$

Since  $S^{-1}(\cup_i A_i) = \cup_i S^{-1}(A_i)$ , it follows from the property L5 of the Lebesgue integral that the integral in (4.7) defines a finite measure.

We denote this measure by  $\nu(A)$  so that,  $\nu(A) = \int_{S^{-1}(A)} f(x) d\mu_x$ .

Therefore by Corollary 2.1 there exists a unique element in  $L^1$ , which we denote by  $Pf$ , such that for  $A \in \mathcal{A}$

$$\begin{aligned} \nu(A) &= \int_A Pf(x) d\mu_x \\ &= \int_{S^{-1}(A)} f(x) d\mu_x \end{aligned} \quad (4.8)$$

Next we choose  $f \in L^1$  to be arbitrary, that is, not necessarily nonnegative. Then  $f = f^+ - f^-$  so that,

$$Pf = Pf^+ - Pf^-$$

.

From this and (4.8) we obtain for  $A \in \mathcal{A}$

$$\int_A Pf(x) d\mu_x = \int_A Pf^+ d\mu_x - \int_A Pf^- d\mu_x$$



$$\begin{aligned}
\int_A Pf(x)d\mu_x &= \int_{S^{-1}(A)} f^+ d\mu_x - \int_{S^{-1}(A)} f^- d\mu_x \\
\int_A Pf(x)d\mu_x &= \int_{S^{-1}(A)} f(x)d\mu_x
\end{aligned} \tag{4.9}$$

It follows from Proposition 2.1 and the non-singularity of  $S$  that, (4.9) uniquely defines an operator  $P$ .

We now define the Frobenius-Perron operator associated with the transformation  $S$ .

**Definition 4.8** *Let  $(X, \mathcal{A}, \mu)$  be a measure space. If  $S : X \rightarrow X$  is a non-singular transformation, then the unique operator  $P : L^1 \rightarrow L^1$  defined by (4.9) is called the **Frobenius-Perron operator** corresponding to  $S$ .*

From (4.9)  $P$  has the following properties:

(FP1)  $P(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 P f_1 + \lambda_2 P f_2$ ; for all  $f_1, f_2 \in L^1, \lambda_1, \lambda_2 \in R$ , so  $P$  is a linear operator.

(FP2)  $Pf \geq 0$  if  $f \geq 0$ ; and

(FP3)  $\int_X Pf(x)d\mu_x = \int_X f(x)d\mu_x$

(FP4) If  $S_n = S \cdots S$  (n times) and  $P_n$  is the operator corresponding to  $S_n$ , then  $P_n = P^n$ , where  $P$  is the Frobenius-Perron operator corresponding to  $S$ .

We observe that properties (FP1) and (FP2) follow from Definition 4.8. Properties (FP3) and (FP4) follow from (FP2) and (2.2), and (4.9) respectively.

In special cases when the transformation  $S$  is differentiable and invertible, we can obtain an explicit form for  $Pf$ . It follows that if  $X = [a, b]$  is an interval on the real line  $R$  and  $A = [a, x]$  then (4.9) becomes

$$\int_a^x Pf(s)ds = \int_{S^{-1}([a,x])} f(s)ds$$

and by differentiating

$$Pf(x) = \frac{d}{dx} \int_{S^{-1}([a,x])} f(s)ds \quad (4.10)$$

If  $S$  is differentiable and invertible, then  $S$  must be monotone. Suppose  $S$  is increasing and  $S^{-1}$  has a continuous derivative. Then

$$S^{-1}([a, x]) = [S^{-1}(a), S^{-1}(x)],$$

and applying the chain rule to the right hand side of (4.10)

$$\begin{aligned} Pf(x) &= \frac{d}{dx} \int_{S^{-1}(a)}^{S^{-1}(x)} f(s)ds \\ &= f(S^{-1}(x)) \frac{d}{dx} [S^{-1}(x)]. \end{aligned}$$

If  $S$  is decreasing, then the sign of the right hand side is reversed. Thus, in the general one-dimensional case, for  $S$  differentiable and invertible with continuous derivative we have,

$$Pf(x) = f(S^{-1}(x)) \left| \frac{d}{dx} [S^{-1}(x)] \right| \quad (4.11)$$

As an application of (4.11) consider the following example.

**Example 4.2** Let  $S(x) = \exp(x)$ , then  $S^{-1}(x) = \ln x$  and

$\frac{d}{dx} S^{-1}(x) = \frac{1}{x}$ . Hence from (4.11) where  $P$  is the Frobenius-Perron operator associ-

ated with  $S$ , we have

$$Pf(x) = \frac{1}{x}f(\ln x)$$

Next we consider an initial  $f$  given by

$$f(x) = \frac{1}{2}I_{[-1,1]}(x)$$

so that,

$$Pf(x) = \frac{1}{2x}I_{[e^{-1},e]}(x)$$

We observe two important points from Example 4.2:

(1) For an initial  $f$  supported on a set  $[a, b]$ ,  $Pf$  will be supported on  $[S(a), S(b)]$ , and,

(2)  $Pf$  is small where  $\frac{dS}{dx}$  is large and vice versa.

We generalize the first observation with the following proposition.

**Proposition 4.4** *Let  $S : X \rightarrow X$  be a non-singular transformation and  $P$  the associated Frobenius-Perron operator. Suppose  $f \geq 0, f \in L^1$ . Then*

$$\text{supp } f \subset S^{-1}(\text{supp } Pf)$$

and, more generally, for every set  $A \in \mathcal{A}$ ,  $Pf(x) = 0$  for  $x \in A \Leftrightarrow f(x) = 0$  for  $x \in S^{-1}(A)$ .

**Proof:**  $\Rightarrow$ : From the definition of the Frobenius-Perron operator we have

$$\int_A Pf(x)d\mu_x = \int_{S^{-1}(A)} f(x)d\mu_x$$

or

$$\int_X I_A(x) Pf(x) d\mu_x = \int_X I_{S^{-1}(A)} f(x) d\mu_x$$

Hence, if  $Pf(x) = 0$  on  $A$ , then, by property (L2) of the Lebesgue integral,  $f(x) = 0$  for  $x \in S^{-1}(A)$  and vice versa.

Now set  $A = X \setminus \text{supp}(Pf)$ , then we have for  $Pf(x) = 0$  for  $x \in A$ ,  $f(x) = 0$  for  $x \in S^{-1}(A)$ .

This implies that  $\text{supp } f \subset X \setminus S^{-1}(A)$ . But  $S^{-1}(A) = X \setminus \text{supp } S^{-1}(Pf)$ .

Hence  $\text{supp } f \subset S^{-1}(\text{supp}(Pf))$ . ■

We can derive a useful generalization of (4.11) for the general case  $X = \mathbb{R}^d$  and where  $S : X \rightarrow X$  is invertible. We first state and prove a change of variables theorem based on the Radon-Nikodym theorem.

**Theorem 4.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $f \in L^1 \cap L^\infty$  (that is a bounded integrable function), and  $S : X \rightarrow X$  a non-singular transformation. Then for every  $A \in \mathcal{A}$ ,*

$$\begin{aligned} \int_{S^{-1}(A)} f(S(x)) d\mu_x &= \int_A f(x) \mu S^{-1}(dx) \\ &= \int_A f(x) J^{-1}(x) d\mu_x \end{aligned} \quad (4.12)$$

where  $\mu S^{-1}$  denotes the measure  $\mu S^{-1}(B) = \mu(S^{-1}(B))$ , for  $B \in \mathcal{A}$  and  $J^{-1}$  is the density of  $\mu S^{-1}$  with respect to  $\mu$ , that is,

$$\mu(S^{-1}(B)) = \int_B J^{-1}(x) d\mu_x \text{ for } B \in \mathcal{A}$$

**Proof:** From the Lebesgue dominated convergence theorem and since  $\mu_f(A) =$

$\int_A f(x) d\mu_x$  where  $f$  is a nonnegative measurable function, we first let  $f(x) = I_B(x)$  so that  $f(S(x)) = I_B(S(x)) = I_{S^{-1}(B)}(x)$  and, hence,

$$\begin{aligned}
 \int_{S^{-1}(A)} f(S(x)) d\mu_x &= \int_X I_{S^{-1}(A)}(x) f(S(x)) d\mu_x \\
 &= \int_X I_{S^{-1}(A)}(x) I_{S^{-1}(B)}(x) d\mu_x \\
 &= \mu(S^{-1}(A) \cap S^{-1}(B)) \\
 &= \mu(S^{-1}(A \cap B)).
 \end{aligned}$$

For the second integral of (4.12) it follows that

$$\begin{aligned}
 \int_A f(x) \mu S^{-1} dx &= \int_X I_A(x) I_B(x) \mu S^{-1} dx \\
 &= \mu(S^{-1}(A \cap B))
 \end{aligned}$$

For the third integral of (4.12) we have

$$\begin{aligned}
 \int_A f(x) J^{-1}(x) d\mu_x &= \int_A I_B(x) J^{-1}(x) d\mu_x \\
 &= \int_{A \cap B} J^{-1}(x) d\mu_x \\
 &= \mu(S^{-1}(A \cap B))
 \end{aligned}$$

For functions other than  $f(x) = I_B(x)$  we repeat the above for simple functions  $f(x)$  and apply the linearity of property (L3) of the Lebesgue integral. ■

Using this we now prove an extension of (4.11) from this change of variables theorem.

**Corollary 4.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  an invertible non-singular transformation, and  $P$  the associated Frobenius-Perron operator. Then for every  $f \in L^1 \cap L^\infty$*

$$Pf(x) = f(S^{-1}(x))J^{-1}(x)$$

**Proof:** From the definition of  $P$ , we have for  $A \in \mathcal{A}$

$$\int_A Pf(x)d\mu_x = \int_{S^{-1}(A)} f(x)d\mu_x$$

Next we change the variables in the right-hand integral with  $y = S(x)$  and from Theorem 4.1, so that we obtain

$$\int_{S^{-1}(A)} f(x)d\mu_x = \int_A f(S^{-1}(y))J^{-1}(y)d\mu_y.$$

Thus we have

$$\int_A Pf(x)d\mu_x = \int_A f(S^{-1}(x))J^{-1}(x)d\mu_x$$

and, hence, from Proposition 2.1,

$$Pf(x) = f(S^{-1}(x))J^{-1}(x) \blacksquare$$

### 4.3 The Koopman Operator

Finally we introduce a third operator which is closely related to the Frobenius-Perron operator. We define some of its properties and demonstrate an inter-relationship between these two operators with an example.

**Definition 4.9** Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  a non-singular transformation, and  $f \in L^\infty$ . The operator  $U : L^\infty \rightarrow L^\infty$  defined by

$$Uf(x) = f(S(x))$$

is called the **Koopman operator** with respect to  $S$ .

This operator was first introduced by Koopman (1931).  $S$  is non-singular, hence  $U$  is well defined since  $f_1(x) = f_2(x)$  a.e implies  $f_1(S(x)) = f_2(S(x))$  a.e. The operator  $U$  has some important properties:

$$(K1) \quad U(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 Uf_1 + \lambda_2 Uf_2,$$

for all  $f_1, f_2 \in L^\infty, \lambda_1, \lambda_2 \in \mathbb{R}$  (that is  $U$  is a linear operator);

$$(K2) \quad \text{For every } f \in L^\infty,$$

$$\|Uf\|_{L^\infty} \leq \|f\|_{L^\infty}$$

that is,  $U$  is a contraction on  $L^\infty$ ;

$$(K3) \quad \text{For every } f \in L^1, g \in L^\infty,$$

$$\langle Pf, g \rangle = \langle f, Ug \rangle$$

so that  $U$  is adjoint to the Frobenius-Perron operator  $P$ .

**Proof:** (K1) This follows by the linearity of  $U$ .

(K2) From the definition of the norm since

$$|f(x)| \leq \|f\|_{L^\infty}$$

a.e., this implies that

$$|f(S(x))| \leq \|f\|_{L^\infty}$$

a.e. Thus since  $Uf(x) = f(S(x))$  it follows from the latter inequality that

$$\|Uf\|_{L^\infty} \leq \|f\|_{L^\infty}.$$

(K3) We first prove property (K3) with  $g = I_A$ . Then,

$$\langle Pf, g \rangle = \int_X Pf(x)I_A(x)d\mu_x = \int_A Pf(x)d\mu_x,$$

while from the definition of the Koopman operator

$$\begin{aligned} \langle f, Ug \rangle &= \int_X f(x)UI_A(x)d\mu_x \\ &= \int_X f(x)I_A(S(x))d\mu_x \\ &= \int_{S^{-1}(A)} f(x)d\mu_x. \end{aligned}$$

From the equation defining the Frobenius-Perron operator  $P$ :

$$\int_A Pf(x)d\mu_x = \int_{S^{-1}(A)} f(x)d\mu_x;$$

it follows that  $\langle Pf, g \rangle = \langle f, Ug \rangle$ . Since (K3) is true for  $g(x) = I_A$  it is true for any simple function  $g(x)$ . Thus by Remark (2.6) property (K3) must be true for all  $g \in L^\infty$ . ■

We conclude this section with the following result.

**Definition 4.10** *Let  $(X, \mathcal{A}, \mu)$  be a measure space. A Markov operator  $P : L^1 \rightarrow L^1$  is called **deterministic** if its adjoint  $U = P^*$  has the following property:*



For every  $A \in \mathcal{A}$  the function  $UI_A$  is a characteristic function, i.e.,  $UI_A = I_B$  for some  $B \in \mathcal{A}$ .

**Example 4.3** We show that the Frobenius-Perron operator is a deterministic operator.

**Proof:** Let  $P$  be any Frobenius-Perron operator with  $U = P^*$  the adjoint operator of  $P$ . By definition of the adjoint operator,

$$\begin{aligned} \langle Pf, g \rangle &= \langle f, P^*g \rangle \\ \Rightarrow \int_X Pf(x)g(x)d\mu_x &= \int_X f(x)P^*g(x)d\mu_x \\ \Rightarrow \int_X Pf(x)g(x)d\mu_x &= \int_X f(x)Ug(x)d\mu_x. \end{aligned}$$

Let  $g = I_A \in L^\infty$ , where  $A \in \mathcal{A}$ . Then

$$\begin{aligned} \int_X Pf(I_A(x))d\mu_x &= \int_X f(x)UI_A(x)d\mu_x \\ \int_A Pf(x)d\mu_x &= \int_X f(x)UI_A(x)d\mu_x \\ \int_{S^{-1}(A)} f(x)d\mu_X &= \int_X f(x)UI_A(x)d\mu_x, \end{aligned} \tag{4.13}$$

We now show that:

$$UI_A(x) = \begin{cases} 1 & \text{if } x \in S^{-1}(A), \\ 0 & \text{if } x \notin S^{-1}(A) \end{cases}$$

Let  $S^{-1}(A) = B \in \mathcal{A}$  and  $UI_A = h(x)$ .

The proof then follows in two parts.

PART 1

To show:  $h(x) = 0$  a.e in  $X - B \Leftrightarrow$  for all  $\tilde{B} \subset X - B$  such that

$$\tilde{B} = \{x : h(x) \neq 0\} \Rightarrow \mu(\tilde{B}) = 0.$$

On the contrary, suppose that there exists  $\tilde{B} \subset X - B$  such that  $\tilde{B} = \{x : h(x) \neq 0\}$  and  $\mu(\tilde{B}) > 0$ . (We do not consider  $\mu(\tilde{B}) < 0$  since in this analysis we define a measure as in Definition 2.2.)

We then choose  $f(x) = \begin{cases} \operatorname{sgn} h(x) & \text{if } x \in \tilde{B}, \\ 0 & \text{if } x \notin \tilde{B}. \end{cases}$

Then

$$\int_{\tilde{B}} |h(x)| d\mu_x = \int_{\tilde{B}} f(x) d\mu_x \quad (4.14)$$

Since  $\tilde{B} \subset X - B$  we have that  $X - \tilde{B}$  contains  $B$ , if  $x \in B$  then  $x \in X - \tilde{B}$  i.e.  $x \notin \tilde{B}$  and it then follows that  $\int_{\tilde{B}} f(x) d\mu_x = 0$ .

Hence we obtain in (4.14) that  $\int_{\tilde{B}} |h(x)| d\mu_x = 0$ .

But since the integral of a positive function over a set of positive (non-zero) measure is positive we have that  $\int_{\tilde{B}} |h(x)| d\mu_x > 0$ .

Hence (4.14) is contradictory since the left hand side is greater than zero and the right hand side is equal to zero; thereby contradicting our assumption that  $\mu(\tilde{B}) > 0$ .

It therefore follows that for any  $\tilde{B} \subset X - B$  such that  $\tilde{B} = \{x : h(x) \neq 0\}$  then  $\mu(\tilde{B}) = 0$ . This implies that  $h(x) = 0$  a.e. in  $X - B$ , that is  $h(x) = 0$  a.e if  $x \notin B$ .

## PART 2

To show:  $h(x) = 1$  a.e in  $B \Leftrightarrow$  for all  $B' \subset B$  such that  $B' = \{x : h(x) \neq 1\}$  then  $\mu(B') = 0$ .

Again, by contradiction suppose that this is not true. That is suppose there exists  $B' \subset B$  such that  $B' = \{h(x) \neq 1\}$  and  $\mu(B') > 0$ . (Again as stated above we do not consider the case  $\mu(B') < 0$ .)

For  $x \in B$  we have

$$\begin{aligned} \int_B f(x)h(x)d\mu_x &= \int_B f(x)d\mu_x \\ \int_B f(x)h(x)d\mu_x - \int_B f(x)d\mu_x &= 0 \\ \int_B (f(x)h(x) - f(x))d\mu_x &= 0 \\ \int_B f(x)(h(x) - 1)d\mu_x &= 0. \end{aligned}$$

Let  $h(x) - 1 = H(x)$  to obtain  $\int_B f(x)H(x)d\mu_x = 0$ .

$$\text{Choose } f(x) = \begin{cases} \text{sgn}H(x) & \text{if } x \in B', \\ 0 & \text{if } x \notin B' \end{cases}$$

We obtain a contradiction since the integral of a positive function over a set of positive (non-zero) measure is positive. Thus  $\forall B' \subset B$  such that  $B' = \{x : h(x) \neq 1\} \Rightarrow \mu(B') = 0$ . This implies that  $h(x) = 1$  a.e. for  $x \in B$ .

From part 1 and part 2 and (4.13) we have that

$$UI_A = I_{S^{-1}(A)} = I_B = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases}$$

Hence the Frobenius-Perron operator is a deterministic operator.

## 4.4 Frobenius-Perron and Koopman operators in the Space $L^p$

In Chapter 5 we shall study the irregular behaviour that the transformation  $S$  associated with the Frobenius-Perron and Koopman operators may display. These results are stated for  $L^1$  and  $L^\infty$  spaces. The same results can also be proven using adjoint spaces  $L^p$  and  $L^{p'}$  instead of  $L^1$  and  $L^\infty$ , respectively. We note that all these results have been proven in a measure space  $(X, \mathcal{A}, \mu)$  where  $S$  is a measure-preserving transformation.

We now aim to associate to  $S$  a morphism of  $L^p$  spaces. To do this we show that the Frobenius-Perron and Koopman operators are isometric in  $L^p$ .

Let  $L^0(X, \mathcal{A}, \mu)$  = the space of all measurable functions of  $(X, \mathcal{A}, \mu)$ . We show that  $UL^p(X_2, \mathcal{A}_2, \mu_2) \subset L^p(X_1, \mathcal{A}_1, \mu_1)$ .

**Lemma 4.1** *Let  $(X_1, \mathcal{A}_1, \mu_1)$  and  $(X_2, \mathcal{A}_2, \mu_2)$  be normalized measure spaces and let  $S : X_1 \rightarrow X_2$  be a measure-preserving transformation. If  $F \in L^0(X_2, \mathcal{A}_2, \mu_2)$  then  $\int UF d\mu_1 = \int F d\mu_2$ .*

**Proof:** Let  $F \in L^0(X_2, \mathcal{A}_2, d\mu_2)$  and  $F \geq 0$ .

If  $F$  is a simple function, then, since simple functions are measurable and  $S$  is measure-preserving, the result then follows.

Choose a sequence of simple functions  $F_n \uparrow F$ . Then, by the definition of  $U$ ,  $UF_n$  are also simple functions increasing to  $UF$ , so,

$$\int_{X_2} F d\mu_2 = \sum_{i=1}^n a_i \mu_2(A_i)$$

$$\begin{aligned}
&= \sum_{i=1}^n a_i \mu_1(S^{-1}(A_i)) \\
&= \int_{X_1} \sum_{i=1}^n a_i I_{S^{-1}(A_i)} d\mu_1 \\
&= \int_{X_1} U \left( \sum_{i=1}^n a_i I_{A_i} \right) d\mu_1 \\
&= \int_{X_1} U F d\mu_1
\end{aligned}$$

Hence  $UL^0(X_2, \mathcal{A}_2, d\mu_2) \subset L^0(X_1, \mathcal{A}_1, \mu_1)$ . ■

Using the above lemma we now show:

$UL^p(X_2, \mathcal{A}_2, \mu_2) \subset L^p(X_1, \mathcal{A}_1, \mu_1)$  and

$\|Uf\|_{L^p} = \|f\|_{L^p}$  for all  $f \in L^p(X_2, \mathcal{A}_2, \mu_2)$ .

**Theorem 4.2** *Let  $p \geq 1$ . Then  $UL^p(X_2, \mathcal{A}_2, \mu_2) \subset L^p(X_1, \mathcal{A}_1, \mu_1)$  and  $\|Uf\|_{L^p} = \|f\|_{L^p}$  for all  $f \in L^p(X_2, \mathcal{A}_2, \mu_2)$ .*

**Proof.** Let  $f \in L^p(X_2, \mathcal{A}_2, \mu_2)$ .

Let  $F(x) = |f(x)|^p$  as in Lemma 4.1. Then, from Definition 4.9

$$\begin{aligned}
UF &= U(|f|^p) \\
&= |f|^p(S(x)) \\
&= |f(S(x))|^p \\
&= |Uf|^p,
\end{aligned}$$

so that

$$\begin{aligned}
\int_{X_1} UF d\mu_1 &= \int_{X_1} U|f|^p d\mu_1 \\
&= \int_{X_1} |Uf|^p d\mu_1
\end{aligned}$$

$$= ||Uf||_{L^p}^p$$

and

$$\begin{aligned} \int_{X_2} F d\mu_2 &= \int_{X_2} |f|^p d\mu_2 \\ &= ||f||_{L^p}^p \end{aligned}$$

Using the above arguments and Lemma 4.1 this implies that  $||Uf||_{L^p} = ||f||_{L^p}$  ■

For  $p = 1$ ,  $||Uf||_{L^1} = ||f||_{L^1}$ . Since the Frobenius-Perron operator is adjoint to the Koopman operator, the Frobenius-Perron operator is also isometric in  $L^p$  (Walters 1982). We therefore have shown that the Frobenius-Perron and Koopman operator associated with  $S$  present a morphism of  $L^p$  spaces.

## 5 Various Types of Chaotic Behaviour

There are many possible definitions of chaos, ranging from measure-theoretic notions of randomness in ergodic theory to the topological approach. Our main objective in this chapter is to understand what it means for a dynamical system to be chaotic from a measure-theoretic point of view. We study three types of transformations that exhibit gradually stronger chaotic properties: ergodicity, mixing and exactness.

### 5.1 Measure-Preserving Transformations

In this section we introduce and discuss measure-preserving transformations and we shall use their asymptotic properties to describe the three levels of chaotic behaviour.

**Definition 5.1** *Let  $(X, \mathcal{A}, \mu)$  be measure space. A measurable transformation  $S : X \rightarrow X$  is called **measure-preserving** if*

$$\mu(S^{-1}(A)) = \mu(A) \quad \text{for all } A \in \mathcal{A}. \quad (5.1)$$

If  $S : X \rightarrow X$  is a measure-preserving transformation on  $(X, \mathcal{A}, \mu)$  we say that  $S$  preserves  $\mu$  or that  $\mu$  is **invariant** under  $S$ .

Further, from equation (5.1) we note that every measure-preserving transformation is non-singular.

The following theorem on invariant measures provides some equivalent characteristics of this concept.

**Theorem 5.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  a nonsingular transformation, and  $P$  the Frobenius-Perron operator corresponding to  $S$ . Consider a nonnegative  $f \in L^1$ . Then a measure  $\mu_f$  given by*

$$\mu_f(A) = \int_A f(x) d\mu_x$$

*is invariant if and only if  $f$  is a fixed point of  $P$ , that is  $Pf = f$ .*

**Proof:** "  $\Rightarrow$  " : Suppose  $\mu_f$  is an invariant measure. Then by definition

$$\mu_f(A) = \mu_f(S^{-1}(A)), \quad \forall A \in \mathcal{A};$$

so that

$$\int_A f(x) d\mu_x = \int_{S^{-1}(A)} f(x) d\mu_x, \quad \forall A \in \mathcal{A} \quad (5.2)$$

From the definition of the Frobenius-Perron operator, we have that

$$\int_A Pf(x) d\mu_x = \int_{S^{-1}(A)} f(x) d\mu_x \text{ for } A \in \mathcal{A} \quad (5.3)$$

Hence from equations (5.2) and (5.3) we obtain

$$\int_A Pf(x) d\mu_x = \int_{S^{-1}(A)} f(x) d\mu_x$$



$$= \int_A f(x) d\mu_x;$$

which from Proposition 2.1 implies that  $Pf = f$ .

"  $\Rightarrow$ :" Suppose that  $f$  is a fixed point of  $P$ , that is,  $Pf = f$  for some  $f \in L^1, f \geq 0$ ; and let a measure  $\mu_f$  be defined by  $\mu_f = \int_A f(x) d\mu_x$ . Hence from the definition of the Frobenius-Perron operator we obtain

$$\begin{aligned} \mu_f(S^{-1}(A)) &= \int_{S^{-1}(A)} f(x) d\mu_x = \int_A Pf(x) d\mu_x \\ &= \int_A f(x) d\mu_x \\ &= \mu_f(A) \end{aligned} \quad ;$$

thereby proving that  $\mu_f$  is an invariant measure.■

**Remark 5.1** We note the initial measure  $\mu$  is invariant if and only if  $P1 = 1$ .■

Next we discuss two examples of invariant measures. (See Remark 2.12.)

**Example 5.1** Consider the  $r$ -adic transformation  $S[0, 1] \rightarrow [0, 1]$  given by

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

where  $r > 1$  is an integer on the measure space  $([0, 1], \mathcal{B}, \mu)$  where  $\mathcal{B}$  is the Borel  $\sigma$ -algebra and  $\mu$  is the Borel measure (Refer to Example 2.2). Then for any  $[0, x] \subset [0, 1]$  and any integer  $r > 1$  the counterimage of  $[0, x]$  under  $S$  is given by

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

and from (4.11) the Frobenius-Perron operator corresponding to  $S$  is given by

$$\begin{aligned}
Pf(x) &= \frac{d}{dx} \sum_{i=0}^{r-1} \int_{\frac{i}{r}}^{\frac{i}{r} + \frac{x}{r}} f(u) du = \frac{1}{r} \sum_{i=0}^{r-1} f\left(\frac{i}{r} + \frac{x}{r}\right) \\
&= \frac{1}{r} \left[ f\left(\frac{x}{r}\right) + f\left(\frac{1}{r} + \frac{x}{r}\right) + \dots + f\left(\frac{r-1}{r} + \frac{x}{r}\right) \right]
\end{aligned}$$

Thus for  $f(x) \equiv 1$  we have

$$\begin{aligned}
P1 &= \frac{1}{r} [1 + \dots + 1] = \frac{r \times 1}{r} \\
&= 1
\end{aligned}$$

and hence by Remark 5.1 the Borel measure is invariant under the  $r$ -adic transformation.

**Example 5.2** [The baker transformation] This transformation is known as the baker transformation since it mimics some aspects of kneading dough. Let  $(X, \mathcal{B}, \mu)$  be a measure space where  $X = [0, 1] \times [0, 1]$  (refer to Definition 2.18) is a unit square in a plane,  $\mathcal{B}$  is the Borel  $\sigma$ -algebra  $\mathcal{B}$  generated by all possible rectangles of the form  $[0, a] \times [0, b]$  and  $\mu$  the Borel measure is the unique measure on  $\mathcal{B}$  such that

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

See Remark 2.12. (We therefore observe that the Borel measure is a generalization of the concept of area.) Next we define a transformation  $S : X \rightarrow X$  by

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

The first application of  $S$  involves a compression of  $X$  in the  $y$  direction by  $\frac{1}{2}$  and a stretching of  $X$  in the  $x$  direction by the factor 2. The transformation is then completed by vertically dividing the compressed and stretched rectangle into two equal

parts and then placing the right-hand part on top of the left-hand part. The baker transformation is measurable since the counterimage of any rectangle is again a rectangle or a pair of rectangles with the same total area.

To prove that the Borel measure under the baker transformation is invariant we calculate the Frobenius-Perron operator for it. For this we distinguish two cases: we consider two intervals  $0 \leq y < \frac{1}{2}$  and  $\frac{1}{2} \leq y \leq 1$ .

CASE (1): For  $0 \leq x < 1$  and  $0 \leq y < \frac{1}{2}$  we obtain

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

so that from (4.11) in the plane  $R^2$  we obtain

$$\int_0^{\frac{x}{2}} ds \int_0^{2y} Pf(s, t) dt = \int \int_{S^{-1}([0, \frac{x}{2}] \times [0, 2y])} f(s, t) ds dt.$$

Differentiating first respect to  $x$  and then with respect to  $y$  we obtain

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

CASE (2): For  $\frac{1}{2} \leq y \leq 1$ , we find that since

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

we obtain that

$$\begin{aligned}
Pf(x, y) &= \frac{\partial^2}{\partial x \partial y} \left\{ \int_0^{\frac{x}{2}} ds \int_0^1 f(s, t) dt + \int_{\frac{1}{2}}^{\frac{1}{2} + \frac{x}{2}} ds \int_0^{2y-1} f(s, t) dt \right\} \\
&= f\left(\frac{1}{2} + \frac{1}{2}x, 2y - 1\right), \quad \frac{1}{2} \leq y \leq 1.
\end{aligned}$$

$$\text{Thus, finally, } Pf(x, y) = \begin{cases} f(\frac{1}{2}x, 2y), & 0 \leq y < \frac{1}{2} \\ f(\frac{1}{2} + \frac{1}{2}x, 2y - 1), & \frac{1}{2} \leq y \leq 1 \end{cases}$$

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

## 5.2 Ergodicity

The first type of irregular behaviour that we consider is ergodicity. We shall define this concept and discuss important results that may be used in proving the existence of this behaviour.

Let  $(X, \mathcal{A}, \mu)$  be a probability space and  $S : X \rightarrow X$  a measure preserving transformation. (Hence  $S$  is a non-singular transformation with invariant measure.) Then a transformation  $S$  is called **decomposable** if it can be decomposed into two disjoint invariant sets of positive measure, that is,  $X = A \cup B$ , where  $B = X \setminus A$ ,  $\mu(A) > 0, \mu(B) > 0$  and  $A$  and  $B$  are invariant under  $S$  which means that points of  $A$  never enter into  $B$  and vice versa.

If  $S^{-1}(A) = A$  for  $A \in \mathcal{A}$ , then also  $S^{-1}(X \setminus A) = X \setminus A$ , and we could therefore study  $S$  by studying the two simpler transformations  $S|_A$  and  $S|_{X \setminus A}$ . If  $0 < \mu(A) < 1$  this has simplified the study of  $S$ . If  $\mu(A) = 0$  or  $\mu(X \setminus A) = 0$  we can ignore  $A$  (or  $X \setminus A$ )

and we have not significantly simplified  $S$  since neglecting a set of measure zero is allowed in measure theory. However this is not a very interesting transformation. This raises the idea of studying those transformations that cannot be decomposed as above and of trying to express every measure-preserving transformation in terms of the **indecomposable** transformations. The indecomposable transformations are called ergodic.

A simple example to demonstrate the above discussion is as follows:

**Example 5.3** *Let*

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

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

We now have the following definition

**Definition 5.2** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and let a non-singular transformation  $S : X \rightarrow X$  be given. Then  $S$  is called **ergodic** if every invariant set  $A \in \mathcal{A}$  is such that either  $\mu(A) = 0$  or  $\mu(X \setminus A) = 0$ ; that is,  $S$  is ergodic if all the invariant sets are trivial subsets of  $X$ .*

From this definition it follows that any ergodic transformation  $S$  must be studied on the entire space  $X$ . Determining ergodicity on the basis of Definition 5.2 is, in general, difficult except for simple examples on finite spaces. Thus, we observe the

transformation in Example 5.3 is not ergodic on the space  $X$  of integers, but is ergodic on the sets of even and odd integers.

The next example, is however, not as simple. We briefly introduce it and will prove it is an ergodic transformation in detail in Section 5.4.

**Example 5.4** *Consider a circle of radius 1, and let  $S$  be a rotation through an angle  $\phi$ . This transformation is equivalent to the map  $S : [0, 2\pi) \rightarrow [0, 2\pi)$  defined by*

$$S(x) = x + \phi \pmod{2\pi}.$$

*This transformation is not ergodic for  $\frac{\phi}{2\pi}$  rational but is ergodic for  $\frac{\phi}{2\pi}$  irrational.*

This example will be proven later in this chapter when we have more techniques at our disposal.

There are several ways of stating the ergodicity condition as we do in the next theorem. (Walters 1982) (Note: For  $A, B \subset X$ ,  $A \Delta B$  denotes the **symmetric difference**  $A \setminus B \cup B \setminus A$ .)

**Theorem 5.2** *Let  $S : X \rightarrow X$  be a measure-preserving transformation in the probability space  $(X, \mathcal{A}, \mu)$ . Then the following statements are equivalent:*

- (1)  *$S$  is ergodic.*
- (2) *The only  $A \in \mathcal{A}$  such that  $\mu((S^{-1}(A)) \Delta A) = 0$  are those with  $\mu(A) = 0$  or  $\mu(A) = 1$ .*
- (3) *For every  $A \in \mathcal{A}$  with  $\mu(A) > 0$  we have  $\mu(\bigcup_{n=1}^{\infty} S^{-n}(A)) = 1$ .*
- (4) *For every  $A, B \in \mathcal{A}$  such that  $\mu(A) > 0, \mu(B) > 0$  there exists  $n > 0$  with  $\mu(S^{-n}(A) \cap B) > 0$*

**Proof:** "(1)  $\Rightarrow$  (2)" Let  $A \in \mathcal{A}$  and  $\mu((S^{-1}(A)) \Delta A) = 0$ . We shall construct a set  $A_{\infty}$  with  $S^{-1}A_{\infty} = A_{\infty}$  and  $\mu(A \Delta A_{\infty}) = 0$ . For each  $n \geq 0$  we have  $\mu((S^{-n}A) \Delta A) =$

0 since

$$\begin{aligned}(S^{-n}A) \Delta A &\subset \bigcup_{i=0}^{n-1} (S^{-(i+1)}A) \Delta (S^{-i}A) \\ &= \bigcup_{i=0}^{n-1} S^{-i}((S^{-1}A) \Delta A),\end{aligned}$$

and hence by property (c) of Definition 2.2  $\mu((S^{-n}A) \Delta A) \leq n\mu(S^{-1}A \Delta A)$ .

Let  $A_\infty = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} S^{-i}(A)$ . From the above it follows that

$$\mu(A \Delta \bigcap_{i=n}^{\infty} S^{-i}A) \leq \sum_{i=n}^{\infty} \mu(A \Delta S^{-i}A) = 0$$

for each  $n \geq 0$ .

Since the sets  $\bigcup_{i=n}^{\infty} S^{-i}A$  decrease with  $n$  and each has measure equal to  $A$  we have  $\mu(A_\infty \Delta A) = 0$  and hence  $\mu(A_\infty) = \mu(A)$ .

Therefore we have obtained a set  $A_\infty$  with  $S^{-1}A_\infty = A_\infty$  and  $\mu(A_\infty \Delta A) = 0$ . By the ergodicity of  $S$  we must have  $\mu(A_\infty) = 0$  or  $1$  and hence  $\mu(A) = 0$  or  $1$ .

"(2)  $\Rightarrow$  (3)" : Let  $A \in \mathcal{A}$  and  $\mu(A) > 0$ . Let  $A_1 = \bigcup_{n=1}^{\infty} S^{-n}A$ . We have  $S^{-1}A_1 \subset A_1$  and since  $\mu(S^{-1}A_1) = \mu(A_1)$  we have  $\mu(S^{-1}A_1 \Delta A_1) = 0$ .

From (2) we get  $\mu(A_1) = 0$  or  $1$ . But we cannot have  $\mu(A_1) = 0$  since  $S^{-1}A \subset A_1$  and  $\mu(S^{-1}A) = \mu(A) > 0$ . Therefore  $\mu(A_1) = 1$ .

"(3)  $\Rightarrow$  (4)" : Let  $\mu(A) > 0$  and  $\mu(B) > 0$ . From (3) we have  $\mu(\bigcup_{n=1}^{\infty} S^{-n}A) = 1$ , so that

$$0 < \mu(B) = \mu(B \cap \bigcup_{n=1}^{\infty} S^{-n}A) = \mu(\bigcup_{n=1}^{\infty} B \cap S^{-n}A).$$

Therefore  $\mu(B \cap S^{-n}A) > 0$  for some  $n \geq 1$ .

"(4)  $\Rightarrow$  (1)" : Suppose  $A \in \mathcal{A}$  and  $S^{-1}A = A$ . If  $0 < \mu(A) < 1$ , then

$$0 = \mu(A \cap (X \setminus A)) = \mu(S^{-n}A \cap (X \setminus A))$$

for all  $n \geq 1$ , which contradicts (4).■

In order to study more examples we have the following theorem:

**Theorem 5.3** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $S : X \rightarrow X$  a nonsingular transformation. Then  $S$  is ergodic if and only if for every measurable function  $f : X \rightarrow \mathbb{R}$ ,*

$$f(S(x)) = f(x) \text{ for almost all } x \in X \tag{5.4}$$

*implies that  $f$  is constant almost everywhere.*

**Proof:** " $\Rightarrow$ " : Suppose that  $S$  is an ergodic nonsingular transformation on a measure space  $(X, \mathcal{A}, \mu)$ . Suppose that we have a function  $f$  satisfying (5.4), but which is not constant a.e. Then there is some  $r$  such that the sets

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

have positive measure. These sets are also invariant since

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

and

$$S^{-1}(B) = \{x | S(x) \in B\} = \{x | f(S(x)) > r\} = \{x | f(x) > r\}$$



$$= B.$$

Since  $A$  and  $B$  are invariant sets of positive measure, this implies that  $S$  is not ergodic, which is a contradiction. Thus every function satisfying (5.4) must be constant.

" $\Leftarrow$ " : Suppose that  $f$  is a measurable function satisfying (5.4) which is constant a.e. but that  $S$  is not ergodic. Then, from Definition 5.2 there exists a nontrivial set  $A \in \mathcal{A}$  which is invariant. Let  $f = I_A$ , and, since  $A$  is nontrivial,  $f$  cannot be a constant function. Furthermore, since  $A = S^{-1}(A)$ , we have

$$\begin{aligned} f(S(x)) &= I_A(S(x)) = I_{S^{-1}(A)}(x) = I_A(x) \\ &= f(x) \text{ a.e} \end{aligned}$$

showing that (5.4) is satisfied by a nonconstant function, which is a contradiction. Hence  $S$  is ergodic. ■

An immediate consequence of Theorem 5.3 in combination with the definition of the Koopman operator is the following corollary.

**Corollary 5.1** *Let  $(X, \mathcal{A}, \mu)$  be measure space,  $S : X \rightarrow X$  a nonsingular transformation, and  $U$  the Koopman operator with respect to  $S$ . Then  $S$  is ergodic if and only if all the fixed points of  $U$  are constant functions.*

In addition to Theorem 5.3 and Corollary 5.1, we have another result which is of use in proving the ergodicity of  $S$  in which we use the Frobenius-Perron operator:

**Theorem 5.4** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  a nonsingular transformation, and  $P$  the Frobenius-Perron operator corresponding to  $S$ . Then*

(1) *If  $S$  is ergodic, then there is at most one stationary density  $f_*$  of  $P$ .*

(2) Further, if there is a unique stationary density  $f_*$  of  $P$  and  $f_*(x) > 0$  a.e, then  $S$  is ergodic.

**Proof:** (1) Suppose that  $S$  is ergodic but that we have two different stationary densities  $f_1$  and  $f_2$  of  $P$ . Set  $g = f_1 - f_2$  so that since  $P$  is linear (from (FP1)) we have that

$$\begin{aligned} Pg &= P(f_1 - f_2) = Pf_1 - Pf_2 = f_1 - f_2 \\ &= g \end{aligned}$$

Then from Proposition 4.3 it follows that

$$Pg^+ = g^+ \text{ and } Pg^- = g^- \quad (5.5)$$

Further, since  $f_1$  and  $f_2$  are not only different but are also densities, we have

$$g^+ \neq 0 \text{ and } g^- \neq 0 \quad (5.6)$$

Indeed since if we suppose to the contrary that  $g^+ = 0$  or  $g^- = 0$ , then if  $g^+ = 0$  by the definition of  $g$  and  $g^+$  we must have that  $f_1 \leq f_2$ . By our assumption  $f_1 \neq f_2$ . Hence  $f_1 < f_2$ . This implies that there exists a set  $A$  such that  $A = \{f_1, f_2 : f_1 < f_2\}$  and  $\mu(A) > 0$ . Since  $f_1, f_2$  are densities,  $f_1, f_2 \geq 0$  and  $\|f_1\| = \|f_2\| = 1$ , which implies that  $\mu(A) = \|f_1\| - \|f_2\| = 0$ , contradicting  $\mu(A) > 0$ . Hence  $g^+ \neq 0$ . By using a similar argument we can show that  $g^- \neq 0$ .

Thus, for any function  $g$ ,  $\text{supp } g = \{x | g(x) \neq 0\}$  we set

$$A = \text{supp } g^+ = \{x | g^+(x) > 0\}$$

and

$$B = \text{supp } g^- = \{x | g^-(x) > 0\}$$

We observe that  $A$  and  $B$  are clearly disjoint sets and from (5.6) that they both have positive (non-zero) measure. From (5.5) and Proposition 4.4.

$$\text{supp } g^+ = A \subset S^{-1}(\text{supp } Pg^+) = S^{-1}(\text{supp } g^+) = S^{-1}(A)$$

$$\text{supp } g^- = B \subset S^{-1}(\text{supp } Pg^-) = S^{-1}(\text{supp } g^-) = S^{-1}(B);$$

hence  $A \subset S^{-1}(A)$  and  $B \subset S^{-1}(B)$ . Since  $A$  and  $B$  are disjoint sets, then  $S^{-1}(A)$  and  $S^{-1}(B)$  are also disjoint. By induction it follows that

$$A \subset S^{-1}(A) \subset S^{-2}(A) \subset \dots \subset S^{-n}(A) \text{ and } B \subset S^{-1}(B) \subset S^{-2}(B) \subset \dots \subset S^{-n}(B) \quad (5.7)$$

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

$$\bar{A} = \bigcup_{n=0}^{\infty} S^{-n}(A) \text{ and } \bar{B} = \bigcup_{n=0}^{\infty} S^{-n}(B) \quad (5.8)$$

Consequently, from the above argument, we have that  $\bar{A}$  and  $\bar{B}$  are also disjoint, and, furthermore they are invariant since

$$S^{-1}(\bar{A}) = \bigcup_{n=1}^{\infty} S^{-n}(A) = \bigcup_{n=0}^{\infty} S^{-n}(A) = \bar{A} \text{ and } S^{-1}(\bar{B}) = \bigcup_{n=1}^{\infty} S^{-n}(B) = \bigcup_{n=0}^{\infty} S^{-n}(B) = \bar{B} \quad (5.9)$$

Also, since  $A$  and  $B$  both have positive (non-zero) measure, from (5.7), (5.8) and (5.9) we obtain that  $\bar{A}$  and  $\bar{B}$  are also of positive (non-zero) measure. This means that we have two nontrivial invariant sets, namely  $\bar{A}$  and  $\bar{B}$ , which by Definition 5.2, contradicts the ergodicity of  $S$ . Therefore if  $S$  is ergodic, there is at most one

stationary density  $f_*$  of  $P$ .

(2) Suppose that  $Pf_* = f_*$  where  $f_* > 0$  is the unique stationary density but suppose also, by contradiction, that  $S$  is not ergodic. Therefore there exists a nontrivial invariant set  $A$  such that

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

and for  $B = X \setminus A$

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

For these two sets  $A$  and  $B$  we may write  $f_* = I_A f_* + I_B f_*$ , so that

$$I_A f_* + I_B f_* = P(I_A f_*) + P(I_B f_*) \quad (5.10)$$

But  $I_B f_* = 0$  on the set  $X \setminus B = A = S^{-1}(A)$ . Thus by Proposition 4.4

$P(I_B f_*)$  is equal to zero in the set  $A = X \setminus A$ , and, similarly  $P(I_A f_*) = 0$  in  $B = X \setminus A$ . Thus, from (5.10) we have

$$I_A f_* = P(I_A f_*) \quad \text{and} \quad I_B f_* = P(I_B f_*)$$

Since  $f_* > 0$  on  $A$  and  $B$ , we may replace  $I_A f_*$  with  $f_A = \frac{I_A f_*}{\|I_A f_*\|}$ , and  $I_B f_*$  with  $f_B = \frac{I_B f_*}{\|I_B f_*\|}$  to obtain

$$f_A = P f_A \quad \text{and} \quad f_B = P f_B.$$

Thus we observe that in assuming that  $S$  is not ergodic, we have constructed two stationary densities of  $P$ , contradicting the existence of a unique stationary density.

Thus, if there is a unique positive stationary density  $f_*$  of  $P$ , then  $S$  is ergodic. ■

The first major result in ergodic theory was proven in (Birkhoff 1931). We state without proof the Birkhoff Individual Ergodic Theorem.

**Theorem 5.5** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  a measurable transformation, and  $f : X \rightarrow \mathbb{R}$  an integrable function. If the measure  $\mu$  is invariant, then there exists an integrable function  $f^*$  such that*

$$f^*(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) \text{ for almost all } x \in X \blacksquare \quad (5.11)$$

Although the limit  $f^*(x)$  is generally difficult to determine, it can be shown that  $f^*(x)$  satisfies

$$f^*(x) = f^*(S(x)) \text{ for almost all } x \in X, \quad (5.12)$$

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

$$\int_X f^*(x) d\mu_x = \int_X f(x) d\mu_x \quad (5.13)$$

Theorem 5.5 is used to give information concerning the asymptotic behaviour of trajectories. However, since our emphasis is on densities and not on individual trajectories, we will seldom use this theorem. With the notion of ergodicity we now derive an important and often quoted extension of the Birkhoff individual ergodic theorem.

**Theorem 5.6** *Let  $(X, \mathcal{A}, \mu)$  be a finite measure space and  $S : X \rightarrow X$  be measure preserving and ergodic. Then, for any integrable  $f$ , the average of  $f$  along the trajectory of  $S$  is equal almost everywhere to the average of  $f$  over the space  $X$ , that*

is,

$$f^*(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) = \frac{1}{\mu(X)} \int_X f(x) d\mu_x \text{ a.e.} \quad (5.14)$$

**Proof:** From (5.12) and Theorem 5.3 it follows that  $f^*$  is constant almost everywhere. Hence, from (5.13), we have

$$\begin{aligned} \int_X f^*(x) d\mu_x &= f^* \int_X d\mu_x = f^* \mu(X) \\ &= \int_X f(x) d\mu_x, \end{aligned}$$

so that

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

Thus from (5.11) and the preceding formula we obtain (5.14). ■

One of the most quoted consequences of this theorem is the following corollary. It says that every set of non-zero measure is visited infinitely often by the iterates of almost every  $x \in X$ . This result is a special case of the Poincaré recurrence theorem.

**Theorem 5.7** *Let  $(X, \mathcal{A}, \mu)$  be a finite measure space and  $S : X \rightarrow X$  be measure preserving and ergodic. Then for any set  $A \in \mathcal{A}, \mu(A) > 0$ , and almost all  $x \in X$ , the fraction of the points  $\{S^k(x)\}$  in  $A$  as  $k \rightarrow \infty$  is given by  $\frac{\mu(A)}{\mu(X)}$ .*

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(S^k(x))$$

However from (5.14) this is simply  $\frac{\mu(A)}{\mu(X)}$ . ■

### 5.3 Mixing and Exactness

We now introduce two other types of behaviour which display a higher degree of irregularity, namely, mixing and exactness.

**Definition 5.3** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space, and  $S : X \rightarrow X$  a measure preserving transformation. Then  $S$  is called **mixing** if*

$$\lim_{n \rightarrow \infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B) \text{ for all } A, B \in \mathcal{A} \quad (5.15)$$

This type of behaviour maybe interpreted as follows: consider points  $x$  belonging to the set  $A \cap S^{-n}(B)$ . These are the points such that  $x \in A$  and  $S^n(x) \in B$ . Thus, from (5.15), as  $n \rightarrow \infty$  the measure of the set of such points is  $\mu(A)\mu(B)$ . This means that the fraction of the points starting in  $A$  that ended up in  $B$  after  $n$  iterations ( $n$  must be a large number) is given by the product of the measures of  $A$  and  $B$  and is independent of the position of  $A$  and  $B$  in  $X$ . Intuitively we may describe mixing as meaning that for any set  $B$  the sequence of sets  $S^{-n}(B)$  becomes, asymptotically, independent of any other set  $A$ .

**Remark 5.2** One can observe that if a transformation  $S$  is mixing, then, it is ergodic: suppose that  $B \in \mathcal{A}$  is an invariant set. Then  $B = S^{-1}(B)$  and, even further,

$B = S^{-n}(B)$  by induction. Let  $A = X \setminus B$  so that  $\mu(A \cap B) = \mu(A \cap S^{-n}(B)) = 0$ . However, from (5.15), we must have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(A \cap S^{-n}(B)) &= \mu(A)\mu(B) \\ &= (1 - \mu(B))\mu(B), \end{aligned}$$

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

There are many examples of mixing transformations, namely, the baker, Anosov, and  $r$ -adic. However proving that a given transformation is mixing by using Definition 5.3 is difficult. In the next section we introduce easier and more powerful tools for this purpose.

The next type of irregular behaviour that we may observe demonstrates that if in a normalized space we start with a set  $A$  of initial conditions of nonzero measure, then after a large number of iterations of the transformation  $S$  the points will spread and completely fill the space  $X$ . Such a behaviour maybe defined as follows:

**Definition 5.4** *Let  $(X, \mathcal{A}, \mu)$  be normalized measure space and  $S : X \rightarrow X$  be a measure-preserving transformation such that  $S(A) \in \mathcal{A}$  for each  $A \in \mathcal{A}$ . If*

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

*then  $S$  is called **exact**.*

**Remark 5.3** (a) It can be proven that exactness of  $S$  implies that  $S$  is mixing.

(b) Another important fact is that invertible transformations cannot be exact. This



is so since for any invertible measure-preserving transformation  $S$ , we have

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

and by induction  $\mu(S^{-n}(A)) = \mu(A)$ , which then does not satisfy Definition 5.4. ■

## 5.4 Characterizing Chaotic Behaviour

In this section we reformulate the concepts of ergodicity, mixing, and exactness in terms of the behaviour of sequences of iterates of Frobenius-Perron and Koopman operators and show how they can be used to determine whether a given transformation is ergodic, mixing, or exact. The techniques used in classifying these behaviours use the notions of weak Cesáro, weak and strong convergences, which we had defined in Section 2.3.

We will first state and prove the main theorem of this section and then show its utility in determining the existence of these irregular behaviours.

**Theorem 5.8** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space,  $S : X \rightarrow X$  a measure-preserving transformation, and  $P$  the Frobenius-Perron operator corresponding to  $S$ . Then*

- (1)  *$S$  is ergodic if and only if the sequence  $\{P^n f\}$  is weakly Cesáro convergent to 1 for all  $f \in D$ .*
- (2)  *$S$  is mixing if and only if  $\{P^n f\}$  is weakly convergent to 1 for all  $f \in D$ .*
- (3)  *$S$  is exact if only if  $\{P^n f\}$  is strongly convergent to 1 for all  $f \in D$ .*

Before we prove this theorem we note the following:

**Remark 5.4** Since  $P$  is linear, the convergence of  $\{P^n f\}$  to 1 for  $f \in D$  is equivalent to the convergence of  $\{P^n f\}$  to  $\langle f, 1 \rangle$  for every  $f \in L^1$ . This observation is valid for weak, strong, and weak Cesáro convergence. ■

We now restate Theorem 5.8 in the equivalent form.

**Corollary 5.2** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space,  $S : X \rightarrow X$  a measure-preserving transformation, and  $P$  the Frobenius-Perron operator corresponding to  $S$ . Then*

(1)  *$S$  is ergodic if and only if*

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

(2)  *$S$  is mixing if and only if*

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

(3)  *$S$  is exact if and only if*

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

**Proof of Theorem 5.8:**

(1) "  $\Rightarrow$  " : Since  $S$  is measure-preserving, we have that  $P1 = 1$  (see Remark 5.1). Suppose that  $S$  is ergodic, then by Theorem 5.4  $f_*(x) = 1$  is the unique stationary density of  $P$  and, from Theorem 5.4, 5.5, and 5.6 (Birkhoff individual theorem) and the definition of the Koopman operator we obtain that for almost all  $x \in X$

$$1 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} U^k f(x)$$

and that for  $f \in L^1, g \in L^\infty$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle f, U^k g \rangle &= \langle f, 1 \rangle \langle 1, g \rangle \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle P^k f, g \rangle, \end{aligned}$$

thereby proving that  $\{P^k f\}$  is Cesaro convergent to 1 for all  $f \in D$ .

"  $\Leftarrow$  " : Again, since  $S$  is measure-preserving,  $P1 = 1$ . Then, by applying  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n-1} P^k f = 1$  to a stationary density  $f$  gives  $f = 1$ . Thus  $f_*(x) = 1$  is the unique stationary density of  $P$  and by Theorem 5.4,  $S$  is ergodic.

(2)"  $\Rightarrow$  " : Suppose that  $S$  is mixing transformation. Hence by definition this implies that

$$\lim_{n \rightarrow \infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B) \quad \forall A, B \in \mathcal{A}$$

The mixing condition can be written in the integral form as follows:

$$\lim_{n \rightarrow \infty} \int_X I_A(x) I_B(S^n(x)) d\mu_x = \int_X I_A(x) d\mu_x \int_X I_B(x) d\mu_x.$$

Subsequently from the definition of the Koopman operator and duality pairing we obtain

$$\lim_{n \rightarrow \infty} \langle I_A, U^n I_B \rangle = \langle I_A, 1 \rangle \langle 1, I_B \rangle \quad (5.16)$$

Since the Koopman operator is adjoint to the Frobenius-Perron operator, we can rewrite (5.16) as

$$\lim_{n \rightarrow \infty} \langle P^n I_A, I_B \rangle = \langle I_A, 1 \rangle \langle 1, I_B \rangle$$

or

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

for  $f = I_A$  and  $g = I_B$ . Therefore, since this relation holds for characteristic functions it must hold for simple functions that are linear in  $f$  and  $g$

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

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

$$\begin{aligned} & | \langle P^n f, g \rangle | - \langle f, 1 \rangle \langle 1, g \rangle \\ & \leq | \langle P^n f, g \rangle - \langle P^n f_k, g_k \rangle | \\ & + | \langle P^n f_k, g_k \rangle - \langle f_k, 1 \rangle \langle 1, g_k \rangle | \\ & + | \langle f_k, 1 \rangle \langle 1, g_k \rangle - \langle f, 1 \rangle \langle 1, g \rangle | \end{aligned} \tag{5.17}$$

If  $\|f_k - f\| \leq \epsilon$  and  $\|g_k - g\|_{L^\infty} \leq \epsilon$ , then the first and last terms on the right-hand side of (5.17) satisfy

(a)

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

$$\begin{aligned}
&\leq \epsilon \|g\|_{L^\infty} + \epsilon \|f_k\| \\
&\leq \epsilon (\|g\|_{L^\infty} + \|f\| + \epsilon)
\end{aligned}$$

and similarly

(b)

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

Finally, for fixed  $k$  the middle term of (5.17) ,

(c)

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

converges to zero as  $n \rightarrow \infty$ . Thus the terms in (a) and (b) are arbitrarily small for small  $\epsilon$  and the term in (c) converges to zero as  $n \rightarrow \infty$ , which shows that the right-hand side of (5.17) can be as small as we wish for large  $n$ . This proves that mixing implies the convergence of  $\langle P^n f, g \rangle$  to  $\langle f, 1 \rangle \langle 1, g \rangle$  for all  $f \in L^1$  and  $g \in L^\infty$ .

"  $\Leftarrow$  " : Suppose that  $\lim_{n \rightarrow \infty} \langle P^n f, g \rangle = \langle f, 1 \rangle \langle 1, g \rangle$  for  $f \in L^1, g \in L^\infty$

Set  $f = I_A$  and  $g = I_B$  and rewrite the above convergence in integral form to obtain

$$\begin{aligned}
\int_X I_A(x) d\mu_x \int_X I_B(x) d\mu_x &= \lim_{n \rightarrow \infty} \int_X P^n I_A(x) I_B(x) d\mu_x \\
&= \lim_{n \rightarrow \infty} \int_X I_A(x) U^n I_B(x) d\mu_x \\
&= \lim_{n \rightarrow \infty} \int_X I_A(x) I_B(S^n(x)) d\mu_x
\end{aligned}$$

and hence

$$\lim_{n \rightarrow \infty} \mu(A \cap S^{-n}(B)) = \mu(A)\mu(B);$$

showing that  $S$  is mixing.

"  $\Rightarrow$  " : Suppose that  $\{P^n f\}$  is strongly convergent to  $\langle f, 1 \rangle$  for all  $f \in L^1$ . We shall show this implies that  $S$  is exact. Assume that  $\mu(A) > 0$  and define

$$f_A(x) = \frac{1}{\mu(A)} I_A(x)$$

Then, by our definition  $f_A(x) \geq 0$  and  $0 < f_A(x) < 1$  and therefore is a density. If the sequence  $\{r_n\}$  is defined by

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

then it follows that the sequence is convergent to zero. Furthermore, from the definition of  $r_n$ , we have

$$\begin{aligned} \mu(S^n(A)) &= \int_{S^n(A)} d\mu_x \\ &= \int_{S^n(A)} P^n f_A(x) d\mu_x - \int_{S^n(A)} (P^n f_A(x) - 1) d\mu_x \\ &\geq \int_{S^n(A)} P^n f_A(x) d\mu_x - r_n \end{aligned} \tag{5.18}$$

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

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

and, since  $S^{-n}(S^n(A))$  contains  $A$ , the last integral above is equal to 1. Thus inequality (5.18) gives

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

but  $r_n \rightarrow 0$  as  $n \rightarrow \infty$  and  $\mu$  is a normalized measure which implies that

$$\lim_{n \rightarrow \infty} \mu(S^n(A)) = 1 \quad \blacksquare$$

For the proof of the converse see (Lin 1971). We will not prove it since we not use this fact in our analysis and the proof uses different techniques.

Next we present a reformulation of Corollary 5.2 in terms of the Koopman operator with the following proposition. We use the fact that the Koopman and Frobenius-Perron operators are adjoint. However this reformulation cannot be extended to condition (c) of Corollary 5.2 since it is not expressed in terms of a duality pairing.

**Proposition 5.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $S : X \rightarrow X$  a measure-preserving transformation, and  $U$  the Koopman operator corresponding to  $S$ . Then*

*(a)  $S$  is ergodic if and only if*

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

*(b)  $S$  is mixing if and only if*

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

**Proof:** The proof of this proposition is similar to that for Theorem 5.8 except that we use property (K3)

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

so that conditions (1) and (2) of Corollary 5.2 and (a) and (b) of Proposition 5.1 are identical. ■

We shall discuss examples of ergodic, exact and mixing transformations.

### Discussion of examples

Theorem 5.8 and Corollary 5.2 were stated in terms of  $L^1$  and  $L^\infty$  spaces in which the Frobenius-Perron operator acts. We can prove the same results using adjoint spaces  $L^p$  and  $L^{p'}$  also. (Refer to Section 4.4 on convergence in  $L^p$  spaces). Our discussion in Section 4.4 shows that in applying the conditions of Theorem 5.8, Corollary 5.2, and Proposition 5.1 it is not necessary to check for the validity for all  $f \in L^p$  and  $L^{p'}$ . We simply check these conditions for  $f$  and  $g$  in linearly dense subsets of  $L^p$  and  $L^{p'}$  respectively. By using the concept of linearly dense sets, we can simplify the proof of weak convergence. It is sufficient to check for weak convergence from its definition for any  $g \in K \subset L^{p'}$  by

- (1) Showing that the sequence  $\{f_n\}$  is bounded in norm, that is,  $\|f_n\|_{L^p} \leq c < \infty$ , and;
- (2)  $K$  is linearly dense in  $L^{p'}$ .

Using the above discussion we now discuss the following examples of transformations that display ergodicity, exactness and mixing respectively.

**Example 5.5** *We show that the rotational transformation*

$$S(x) = x + \phi \pmod{2\pi}$$



is ergodic when  $\frac{\phi}{2\pi}$  is irrational.

Let  $(X, \mathcal{B}, \mu)$  be measure space and  $S : X \rightarrow X$  a measure-preserving transformation that preserves the Borel measure  $\mu$  and the normalized measure  $\frac{\mu}{2\pi}$ . Let the set consisting of the functions  $K = \{\sin kx, \cos lx | k, l = 0, 1, \dots\}$  be our linearly dense set in  $L^2([0, 2\pi])$ .

From Proposition 5.1(a) we therefore need to show that for each  $g \in K$ , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} U^k g(x) = \langle 1, g \rangle \quad (5.19)$$

uniformly for all  $x$  thereby proving that this condition is satisfied for all  $g$ . Firstly, we write  $\sin kx$  and  $\cos kx$  in the following forms

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

where  $i = \sqrt{-1}$ . Consequently, it is sufficient to verify (5.19) only for

$$\begin{aligned} g(x) &= \cos kx + i \sin kx \\ &= e^{ikx} \end{aligned}$$

where  $k$  is an arbitrary (not necessarily positive) integer. Hence we consider two cases:  $k \neq 0$  and  $k = 0$ .

Case (1) For  $k \neq 0$ , we have, from the definition of the Koopman operator

$$U^l g(x) = g(S^l(x)) = e^{ik(x+l\phi)} \quad (5.20)$$

Using this we define the sequence  $u_n(x)$  as follows

$$u_n(x) = \frac{1}{n} \sum_{l=0}^{n-1} U^l g(x) \quad (5.21)$$

and determine if: (1)  $u_n(x)$  is bounded in norm, that is, if  $\|u_n(x)\|_{L^2}$  is bounded, and;

(2)  $u_n(x)$  satisfies (5.19)

Substituting (5.20) into (5.21) we obtain

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

and

$$\begin{aligned} \|u_n(x)\|_{L^2} &= \left[ \int_0^{2\pi} |u_n(x)|^2 d\mu_x \right]^{\frac{1}{2}} \\ &= \left[ \int_0^{2\pi} \frac{1}{n^2} e^{i2kx} \frac{|e^{ink\phi} - 1|^2}{|e^{ik\phi} - 1|^2} \frac{dx}{2\pi} \right]^{\frac{1}{2}} \\ &= \frac{|e^{ink\phi} - 1|}{n|e^{ik\phi} - 1|} \left( \int_0^{2\pi} e^{i2kx} \frac{dx}{2\pi} \right)^{\frac{1}{2}} \\ &= \frac{2}{n|e^{ik\phi} - 1|}. \end{aligned}$$

Thus  $\{u_n(x)\}$  is bounded in  $L^2$  norm thereby showing that our proof is valid in  $L^2$  and hence that  $u_n(x)$  converges in  $L^2$  to zero as  $n \rightarrow \infty$ . Furthermore,

$$\langle 1, g \rangle = \int_0^{2\pi} g(x) d\mu_x = \int_0^{2\pi} e^{ikx} \frac{dx}{2\pi}$$

$$\begin{aligned}
&= \frac{1}{ik} [e^{ikx}]_0^{2\pi} = \frac{1}{ik} [e^{2\pi ik} - 1] \\
&= 0
\end{aligned}$$

and condition (a) of Proposition 5.1 is satisfied when  $k \neq 0$ .

Case (2) For  $k = 0$  we have since  $g(x) \equiv 1$  is a fixed point. Hence,

$$\begin{aligned}
g(x) &= e^{ikx} = 1; \\
u_n(x) &= \frac{1}{n} \sum_{l=0}^{n-1} U^l g(x) = \frac{1}{n} \sum_{l=0}^{n-1} 1 = 1
\end{aligned}$$

and,

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

Thus from (1) and (2) we observe that (5.19) is satisfied proving that  $S$  is an ergodic transformation.

**Example 5.6** In this example we demonstrate that the  $r$ -adic transformation

$$S(x) = rx \pmod{1}$$

is exact.

By using Corollary 5.2 we shall show that  $\{P^n f\}$  converges strongly to  $\langle f, 1 \rangle$  for a linearly dense set in  $L^p([0, 1])$ . Let that linearly dense set  $K$  = the set of all continuous functions on  $[0, 1]$ . (Lasota and Mackey 1985) Then from Example 5.1

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

and thus by induction

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

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

$$\lim_{n \rightarrow \infty} P^n f(x) = \int_0^1 f(s) ds,$$

uniformly in  $x$ , which, by definition, is just  $\langle f, 1 \rangle$ . Thus  $S$  is an exact transformation.

**Example 5.7** Here we show that the Anosov diffeomorphism

$$S(x, y) = (x + y, x + 2y) \pmod{1}$$

is a mixing transformation. Using Proposition 5.1(b) we prove that  $U^n g(x, y) = g(S^n(x, y))$  converges weakly to  $\langle 1, g \rangle$  for  $g$  in a linearly dense set in  $L^{p'}([0, 1] \times [0, 1])$ .

Suppose that  $g(x, y)$  is periodic in  $x$  and  $y$  with period 1. Hence

$$\begin{aligned} Ug(x, y) &= g(S(x, y)) = g(x + y, x + 2y) \\ U^2 g(x, y) &= g(S^2(x, y)) = g(2x + 3y, 3x + 5y) \\ U^3 g(x, y) &= g(S^3(x, y)) = g(5x + 8y, 8x + 13y), \end{aligned}$$

and so on. By induction we can therefore conclude that

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

where the  $a_n$  are the Fibonacci numbers given by  $a_0 = a_1 = 1, a_{n+1} = a_n + a_{n-1}$ .

Thus if we take

$$g(x, y) = \exp [2\pi i(kx + ly)],$$

and

$$f(x, y) = \exp [-2\pi i(px + qy)],$$

where the  $k, l, p, q$  are integers, we obtain that

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

and by integrating the exponential function we have

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

We now determine the conditions under which  $ka_{2n-2} + la_{2n-1} - p$  and  $ka_{2n-1} + la_{2n} - q$  are zero. This means that we need to show that either  $ka_{2n-2} + la_{2n-1} - p$  or  $ka_{2n-1} + la_{2n} - q$  is not equal to zero if at least one of  $k, l, p$ , or  $q$  is not zero:

(1) If  $k = l = 0$  but  $p \neq 0$  or  $q \neq 0$  then clearly both of these two expressions will be equal to zero.

(2) Suppose  $k \neq 0$  or  $l \neq 0$ , then if  $k \neq 0$  and  $ka_{2n-2} + la_{2n-1} - p = 0$  for infinitely many  $n$ . Then we have

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

.

From a proven result in (Hardy and Wright 1959) and (Frleigh and Beauregard 1990) namely

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

hence

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

Since  $k$  and  $l$  are integers this limit cannot be zero. Thus  $ka_{2n-2} + la_{2n-1} - p \neq 0$

for large  $n$ . Similarly  $ka_{2n-1} + la_{2n} - q \neq 0$  for large  $n$ , so that for

$ka_{2n-2} + la_{2n-1} - p = ka_{2n-1} + la_{2n} - q = 0$  we must have that  $k = l = p = q = 0$

Therefore, for large  $n$ ,

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

But

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

so that

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

Thus

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

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

Finally, we note that since the Frobenius-Perron is a special type of Markov operator we can extend the notions of ergodicity, mixing, and exactness for transformations to Markov operators as follows:

**Definition 5.5** Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and  $P : L^1(X, \mathcal{A}, \mu) \rightarrow L^1(X, \mathcal{A}, \mu)$  a Markov operator with stationary density 1, that is,  $P1 = 1$ . Then we say:

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

## 5.5 Other Types of Chaotic Behaviour

The three types of chaotic behaviors that we have discussed are not the only the only types. However they are probably the most important, but it is also possible to find some intermediate types, which we will not discuss in detail in this analysis.

For example, between ergodicity and mixing there is a class of weakly mixing transformations defined as follows:

**Definition 5.6** Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and  $S : X \rightarrow X$  a measure-preserving transformation. Then  $S$  is **weakly mixing** if

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

Another type of transformation that may display chaotic behaviour is the class of Kolmogorov automorphisms, abbreviated K-automorphisms. These are invertible transformations and therefore cannot be exact but are stronger than mixing. They are defined as follows:

**Definition 5.7** Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and let  $S : X \rightarrow X$  be an invertible transformation such that  $S$  and  $S^{-1}$  are measurable and measure-preserving. The transformation  $S$  is called a **K-automorphism** if there exists a  $\sigma$ -algebra  $A_0 \subset \mathcal{A}$  such that the following three conditions are satisfied:

(a)  $S^{-1}(A_0) \subset A_0$  ;

(b) the  $\sigma$ -algebra

$$\bigcap_{n=0}^{\infty} S^{-n}(A_0)$$

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

(c) the smallest  $\sigma$ -algebra containing

$$\bigcup_{n=0}^{\infty} S^n(A_0)$$

is identical to  $\mathcal{A}$ .



More information concerning  $K$ -automorphisms can be found in the books by (Walters 1982) and by (Parry 1981).

## 6 Asymptotic Stability of Markov Operators

### 6.1 Introduction

An important notion in the study of dynamical systems is the stability or persistence of the system under small changes or perturbations. This is the concept of stability which is very important in applications of the theory of dynamical systems, since if the dynamical system in question is not stable, then small errors or approximations made in the model have a chance of dramatically changing the real solution to the system. That is, our solution could be radically wrong or unstable. If, on the other hand, the dynamical system in question is stable, then the small errors introduced by approximations and experimental errors may not matter at all: the solution to the model system maybe equivalent to the actual solution.

We shall observe that in the midst of this random behaviour within a system, we nevertheless may have an element of regularity in the form of asymptotic stability. Our focal point in this chapter is, in particular, the asymptotic stability of the Markov operator. We introduce this concept, examine ways and techniques that can be used to demonstrate this property and then discuss related examples.

## 6.2 Asymptotic Properties of the Averages $\{A_n f\}$

In this section we examine the convergence of the sequences  $\{A_n f\}$  of averages and show how this may be used to determine the existence of a stationary density of  $P$ .

We begin with two definitions which establish sufficient conditions for the existence of convergent subsequences. Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $\mathcal{F}$  a set of functions in  $L^p$ .

**Definition 6.1** *The set  $\mathcal{F}$  is called **strongly precompact** if every sequence of functions  $\{f_n\}, f_n \in \mathcal{F}$ , contains a subsequence  $\{f_{\alpha_n}\}$  that converges strongly to an  $\bar{f} \in L^p$ .*

**Definition 6.2** *The set  $\mathcal{F}$  is called **weakly precompact** if every sequence of functions  $\{f_n\}, f_n \in \mathcal{F}$ , contains a subsequence  $\{f_{\alpha_n}\}$  that converges weakly to an  $\bar{f} \in L^p$ .*

**Remark 6.1** The prefix "pre-" is used since we take  $\bar{f} \in L^p$  rather than  $\bar{f} \in \mathcal{F}$ . The two definitions are often applied to sets consisting of sequences of functions. Hence the precompactness of  $\mathcal{F} = \{f_n\}$  simply means that every subsequence  $\{f_{\alpha_n}\}$  contains a convergent subsequence. ■

There are several simple and general criteria useful for demonstrating the weak precompactness of sets in  $L^p$  (see Dunford and Schwartz 1957). The two criteria that we will use are as follows:

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

$$|f(x)| \leq g(x) \text{ for } x \in X \text{ a.e.} \tag{6.1}$$

is weakly precompact in  $L^1$ .

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

$$\|f\|_{L^p} \leq M \tag{6.2}$$

is weakly precompact in  $L^1$ .

We note that in *Criterion 2* the set of all functions  $f \in L^1$  is weakly precompact in  $L^1$  if it is bounded in  $L^p$  norm.

Next we define sequence of averages.

**Definition 6.3** Let  $(X, \mathcal{A}, \mu)$  be a measure space and a Markov operator  $P : L^1 \rightarrow L^1$ . The sequence  $\{A_n f\}$  of averages is defined by

$$A_n f = \frac{1}{n} \sum_{k=0}^{n-1} P^k f, \text{ for } f \in L^1 \tag{6.3}$$

The two following propositions are useful tools in proving our next theorem.

**Proposition 6.1** For all  $f \in L^1$ ,

$$\lim_{n \rightarrow \infty} \|A_n f - A_n P f\| = 0.$$

**Proof:** From Definition 6.3 we have

$$\begin{aligned}
A_n f - A_n P f &= \frac{1}{n} \sum_{k=0}^{n-1} P^k (f - P f) \\
&= \frac{1}{n} (f - P^n f)
\end{aligned}$$

and thus

$$\|A_n f - A_n P f\| \leq \frac{1}{n} (\|f\| + \|P^n f\|).$$

Using the contractive property of the Markov operator that  $\|P^n f\| \leq \|f\|$ , we have

$$\|A_n f - A_n P f\| \leq \frac{2}{n} \|f\| \rightarrow 0$$

as  $n \rightarrow \infty$ . ■

**Proposition 6.2** *If, for  $f \in L^1$ , there is a subsequence  $\{A_{\alpha_n}\}$  of the sequence  $\{A_n f\}$  that converges weakly to  $f_* \in L^1$ , then  $P f_* = f_*$ .*

**Proof:** We first apply  $P$  to the sequence  $\{A_{\alpha_n} f\}$  so that, since  $P A_{\alpha_n} f = A_{\alpha_n} P f$ , then  $\{A_{\alpha_n} P f\}$  converges weakly to  $P f_*$ . From Proposition 6.1, since  $\{A_{\alpha_n} P f\}$  has the same limit as  $\{A_{\alpha_n} f\}$ , we have  $P f_* = f_*$ . ■

We now state and prove a special case of the Kakutani-Yosida abstract ergodic theorem (see Dunford and Schwartz 1957). This theorem is useful in defining a condition for the existence of a stationary density for a given Markov operator  $P$

by simply demonstrating the weak convergence of the sequence  $\{A_n f\}$ . As we shall observe in Section 6.4 the existence of a stationary density serves as a basis in defining the concept of asymptotic stability. If  $P$  is a Frobenius-Perron operator, then, in this case it is sufficient to prove the existence of an invariant measure.

**Theorem 6.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a Markov operator. If for a given  $f \in L^1$  the sequence  $\{A_n f\}$  is weakly precompact, then it converges strongly to some  $f_* \in L^1$  that is a fixed point of  $P$ , namely,  $Pf_* = f_*$ . Furthermore, if  $f \in D$ , then  $f_* \in D$ , so that  $f_*$  is a stationary density.*

**Proof:** Suppose that for  $f \in L^1$  the sequence  $\{A_n f\}$  is weakly precompact, then from Definition 6.1 there exists a subsequence  $\{A_{\alpha_n} f\}$  that converges weakly to some  $f_* \in L^1$ . Further, from Proposition 6.2 we know  $Pf_* = f_*$ .

Next we show that  $\{A_n f\}$  converges strongly to  $f_* \in L^1$ . We write  $f \in L^1$  in the form

$$f = (f - f_*) + f_* \tag{6.4}$$

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

$$f - f_* = Pg - g + r, \tag{6.5}$$

where  $g \in L^1$  and  $\|r\| < \epsilon$ . Hence from (6.4) and (6.5), we have

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

Since  $Pf_* = f_*$ ,  $A_n f_* = f_*$ , we obtain

$$\|A_n f - f_*\| = \|A_n(f - f_*)\|$$

$$\leq \|A_n(Pg - g)\| + \|A_nr\|.$$

From Proposition 6.1, we know that  $\|A_n(Pg - g)\|$  is strongly convergent to zero as  $n \rightarrow \infty$  and by our assumptions  $\|A_nr\| \leq \|r\| < \epsilon$ . Thus, for sufficiently large  $n$ , we must have

$$\|A_nf - f_*\| \leq \epsilon.$$

Since  $\epsilon$  is arbitrary, this therefore implies that  $\{A_nf\}$  is strongly to  $f_*$ .

Finally, to show that if  $f \in D$ , then  $f_* \in D$ , we recall that from Definition 4.3 that  $f \in D$  means that  $f \geq 0$  and  $\|f\| = 1$ .

Therefore since  $P$  is a Markov operator from Definition 4.3  $Pf \geq 0$  and  $\|Pf\| = 1$ . From Proposition 6.1 we consequently have that  $A_nf \geq 0$  and  $\|A_nf\| = 1$  and, since  $\{A_nf\}$  is strongly convergent to  $f_*$ , it follows that  $f_* \in D$ . This completes the proof that under the supposition that (6.5) is valid for every  $\epsilon$ .

In proving this assumption, we use a simplified version of the Hahn-Banach theorem. Suppose that for some  $\epsilon$  there does not exist an  $r$  such that (6.5) is true. Then, in this case we have  $f - f_* \notin \overline{(P - I)L^1(X)}$  and, thus, by the Hahn-Banach theorem, there must exist a  $g_0 \in L^\infty$  such that

$$\langle f - f_*, g_0 \rangle \neq 0 \tag{6.6}$$

and

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

In particular,

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

Thus

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

and by induction we must, therefore, have

$$\langle P^j f, g_0 \rangle = \langle f, g_0 \rangle. \quad (6.7)$$

As a result

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

or

$$\langle A_n f, g_0 \rangle = \langle f, g_0 \rangle. \quad (6.8)$$

Since we assumed  $\{A_{\alpha_n} f, g_0\}$  to converge weakly to  $f_*$ , we have

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

and, from (6.8),

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

which gives

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

This clearly contradicts (6.6), and therefore we conclude that the (6.5) is, indeed, always possible. ■



We now state two simple and useful corollaries to Theorem 6.1.

**Corollary 6.1** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a Markov operator. If, for some  $f \in D$  there is a  $g \in L^1$  such that*

$$P^n f \leq g$$

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

**Proof:** Since,  $P^n f \leq g$  we have that

$$\begin{aligned} 0 &\leq A_n f = \frac{1}{n} \sum_{k=0}^{n-1} P^k f \\ &\leq g \end{aligned}$$

and thus,  $|A_n f| \leq g$ . By applying our first criterion for weak precompactness (6.1), we know that  $\{A_n f\}$  is weakly precompact. Then the result that  $f_*$  is a stationary density follows from Theorem 6.1. ■

**Corollary 6.2** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a Markov operator. If for some  $f \in D$  there exists  $M > 0$  and  $p > 1$  such that*

$$\|P^n f\|_{L^p} \leq M$$

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

**Proof:** We have

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

Hence, by our second criterion for weak precompactness (6.2),  $\{A_n f\}$  is weakly precompact, and again from Theorem 6.1 we obtain that  $f_*$  is a stationary density. ■

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

**Theorem 6.2** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a Markov operator with a unique stationary density  $f_*$ . If  $f_*(x) > 0$  for all  $x \in X$ , then*

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

**Proof:** We firstly assume that  $\frac{f}{f_*}$  is bounded, and, then set  $c = \sup(\frac{f}{f_*})$ , to obtain that

$$P^n f \leq P^n c f_* = c P^n f_* = c f_* \text{ and } A_n f \leq c A_n f_* = c f_*.$$

Thus the sequence  $\{A_n f\}$  is weakly precompact and from Theorem 6.1, is convergent

to a stationary density. Since  $f_*$  is the unique stationary density,  $\{A_n f\}$  must converge to  $f_*$ . Hence the theorem is proved when  $\frac{f}{f_*}$  is bounded.

In the second case when  $\frac{f}{f_*}$  is not bounded, write  $f_c = \min(f, cf_*)$ . We then have

$$f = \frac{1}{\|f_c\|} f_c + r_c, \quad (6.9)$$

where

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

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

$$\lim_{c \rightarrow \infty} f_c(x) = f(x) \quad \text{for all } x,$$

which follows from the fact that  $cf_* \rightarrow \infty$  as  $c \rightarrow \infty$ . Then, applying the definition of  $f_c$ , we observe that,  $f_c(x) \leq f(x)$ . Thus, from the Lebesgue dominated convergence theorem,  $\|f_c - f\| \rightarrow 0$  and  $\|f_c\| \rightarrow \|f\| = 1$  as  $c \rightarrow \infty$ . Hence the remainder  $r_c$  is strongly convergent to zero as  $c \rightarrow \infty$ . By choosing  $\epsilon > 0$  we can find a value  $c$  such that  $\|r_c\| < \frac{\epsilon}{2}$ . Then

$$\|A_n r_c\| \leq \|r_c\| < \frac{\epsilon}{2}. \quad (6.10)$$

However, since  $\frac{f_c}{\|f_c\|}$  is a density bounded by  $c\|f_c\|^{-1}f_*$ , from the first part of our proof we have,

$$\|A_n(\frac{1}{\|f_c\|} f_c) - f_*\| \leq \frac{\epsilon}{2} \quad (6.11)$$

for sufficiently large  $n$ . Combining inequalities (6.10) and (6.11) with the decomposition (6.9), we obtain

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

for sufficiently large  $n$  thereby proving the convergence of  $\{A_n f\}$  to  $f_*$ . ■

In the case that  $P$  is the Frobenius-Perron operator corresponding to a non-singular transformation  $S$ , Theorem 6.2 offers a convenient criterion for ergodicity. From Theorem 5.4, the ergodicity of  $S$  is equivalent to the uniqueness of the solution to  $Pf = f$ . Using this relationship, we can now prove the following corollary which implies that if  $S$  is ergodic then the sequence of averages  $\{A_n f\}$  converges to 1.

**Corollary 6.3** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space,  $S : X \rightarrow X$  a measure-preserving transformation, and  $P$  the corresponding Frobenius-Perron operator. Then  $S$  is ergodic if and only if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k f = 1 \text{ for every } f \in D. \quad (6.12)$$

**Proof:** "  $\Rightarrow$  " : Since  $S$  is measure-preserving, we have from Theorem 5.1 that  $P1 = 1$ . If  $S$  is ergodic, then from Theorem 5.4,  $f_*(x) \equiv 1$  is the unique stationary density of  $P$  and, from Theorem 6.2 and Definition 6.3 of the sequence of averages  $\{A_n f\}$ , the convergence of (6.12) follows.

"  $\Leftarrow$  " : Suppose that the convergence (6.12) holds. Then by applying (6.12) to a stationary density  $f$  gives  $f = 1$ . Thus  $f_*(x) = 1$  is the unique stationary density of  $P$  and again from Theorem 5.4, the transformation  $S$  is ergodic. ■

### 6.3 The Existence of a Stationary Density

In this section we show that almost any kind of upper bound on the iterates  $P^n f$  of a Markov operator  $P$  is sufficient to establish that  $\{P^n f\}$  will also have very regular (asymptotically periodic) behaviour. Subsequently we examine ways determining the ergodicity, mixing, and exactness for the Markov operator  $P$  which displays asymptotically periodic behaviour in a normalized measure space  $(X, \mathcal{A}, \mu)$ .

We begin by developing the concept of an upper bound by introducing the definitions of the convergence of a sequence of functions to a set.

**Definition 6.4** *Let  $\mathcal{F}$  be nonempty set in  $L^1$  and let  $g \in L^1$ . Then the **distance**  $d$  between  $g$  and  $\mathcal{F}$  is defined as the "shortest of the distances" between  $g$  and the elements  $f \in \mathcal{F}$ , or, more precisely,*

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

**Definition 6.5** *A sequence of functions  $\{f_n\}, f_n \in L^1$ , is **convergent to a set**  $\mathcal{F} \subset L^1$  if*

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

**Definition 6.6** *A Markov operator  $P$  will be called **strongly (weakly) constrictive** if there exists a strongly (weakly) precompact set  $\mathcal{F}$  such that*

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

A property that holds for strongly (weakly) constrictive Markov operators is stated in the following theorems:

**Theorem 6.3** *Let  $(X, \mathcal{A}, \mu)$  be a measure space, and  $P : L^1 \rightarrow L^1$  a Markov operator. If  $P$  is weakly constrictive, then it is strongly constrictive. ■*

We shall henceforth refer to a strongly (weakly) Markov operator as simply just a constrictive operator.

Another property which is useful in proving ergodicity, mixing, and exactness is proven in (Lasota, Li, and Yorke 1984), is as follows:

**Theorem 6.4** *[Spectral decomposition theorem] Let  $P$  be constrictive Markov operator. Then there is an integer  $r$ , two sequences of nonnegative functions  $g_i \in D$  and  $k_i \in L^\infty, i = 1, \dots, r$ , and an operator  $Q : L^1 \rightarrow L^1$  such that for all  $f \in L^1, Pf$  may be written in the form*

$$Pf(x) = \sum_{i=1}^r \lambda_i(f) g_i(x) + Qf(x), \quad (6.13)$$

where

$$\lambda_i(f) = \int_X f(x) k_i(x) d\mu_x. \quad (6.14)$$

The functions  $g_i$  and operator  $Q$  have the following properties:

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

**Remark 6.2** We observe from (6.13) that the operator  $Q$  is automatically determined if we know the functions  $g_i$  and  $k_i$ , that is,

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

From (6.13) of Theorem 6.4 for  $Pf$ , it immediately follows that the structure of  $P^n f$  is given by

$$P^n f(x) = \sum_{i=1}^r \lambda_i(f)g_{\alpha^n(i)}(x) + Q_n f(x), \quad (6.15)$$

where  $Q_n = P^{n-1}Q$ , and  $\alpha^n(i) = \alpha(\alpha^{n-1}(i)) = \dots$ , and  $\|Q_n f\| \rightarrow 0$  as  $n \rightarrow \infty$ . The terms under the summation in (6.15) are just permuted with each application of  $P$ , and since  $r$  is finite, the sequence

$$\sum_{i=1}^r \lambda_i(f)g_{\alpha^n(i)}(x) \quad (6.16)$$

must be periodic with a period  $\tau \leq r!$ . Since  $\{\alpha^n(1), \dots, \alpha^n(r)\}$  is a permutation of  $\{1, \dots, r\}$ , there must exist a unique  $i$  corresponding to each  $\alpha^n(i)$ . Thus it is clear that summation (6.16) may be written as

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

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

In the next two propositions we demonstrate that every constrictive Markov operator has a stationary density and then give an explicit representation for  $P^n f$  when this stationary density is constant.

**Proposition 6.3** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  be a constrictive Markov operator. Then  $P$  has a stationary density.*

**Proof:** Let a density  $f$  be defined by

$$f(x) = \frac{1}{r} \sum_{i=1}^r g_i(x), \quad (6.17)$$

where  $r$  and  $g_i$  are defined as in Theorem 6.4. From property (2) of Theorem 6.4 and (6.17) we have

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

and thus  $Pf = f$ , hence proving that  $P$  has a stationary density. ■

Let us now assume that the measure  $\mu$  is normalized [ $\mu(X) = 1$ ] and examine the consequences for the representation of  $P^n f$  when we have a constant stationary density  $f = I_X$ . Note that if  $P$  is a Frobenius-Perron operator, then this is equivalent to  $\mu$  being invariant.

**Proposition 6.4** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a constrictive Markov operator. If  $P$  has a constant stationary density, then the representation for  $P^n f$  in Theorem 6.4 takes the form*

$$P^n f(x) = \sum_{i=1}^r \lambda_{\alpha^{-n}(i)}(f) \bar{I}_{A_i}(x) + Q_n f(x) \text{ for all } f \in L^1, \quad (6.18)$$

where

$$\bar{I}_{A_i}(x) = \frac{1}{\mu(A_i)} I_{A_i}(x).$$

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

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



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

**Proof:** Suppose that  $P$  has constant stationary density  $f = I_X$ . This then implies that  $PI_X = I_X$  so that  $P^n I_X = I_X$ . However, if  $P$  is constrictive, then, from Theorem 6.4,

$$P^n I_X(x) = \sum_{i=1}^r \lambda_{\alpha^{-n}(i)}(I_X) g_i(x) + Q_n I_X(x). \quad (6.19)$$

From our discussions following Theorem 6.4, we know that the summation in (6.19) is periodic. Let  $\tau$  be the period of the summation portion of  $P^n$  (remember that  $\tau \leq r!$ ) so that  $\alpha^{-n\tau}(i) = i$  and

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

Passing to the limit as  $n \rightarrow \infty$  and using the fact that  $I_X$  is a stationary density, we have

$$I_X(x) = \sum_{i=1}^r \lambda_i(I_X) g_i(x). \quad (6.20)$$

However, since the functions  $g_i$  are supported on disjoint sets, therefore, from (6.20), we must have each  $g_i$  constant or, more specifically,

$$g_i(x) \equiv \frac{1}{\lambda_i(I_X)} I_{A_i}(x),$$

where  $A_i \subset X$  denotes the support of  $g_i$ , that is, the set of all  $x$  such that  $g_i(x) \neq 0$ . From (6.20) it also follows that  $\bigcup_i A_i = X$ .

Next we apply the operator  $P^n$  to (6.20) to obtain

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

and, similarly as above, we have

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

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

$$\int_{A_i} g_i(x) d\mu_x = 1 = \frac{\mu(A_i)}{\lambda_i(I_X)}.$$

Thus,  $\mu(A_i) = \lambda_i(I_X)$  and

$$g_i(x) = \frac{1}{\mu(A_i)} I_{A_i}(x).$$

We also have that,  $\mu(A_{\alpha^n(i)}) = \mu(A_i)$  for all  $n$ . ■

We shall now show that if the explicit representation in (6.13) for  $Pf$  for a given Markov operator  $P$  is known, then it is easy to check for the existence of invariant measures and hence to determine ergodicity, mixing, or exactness. However, we seldom have an explicit representation for a given Markov operator  $P$ , but in the remainder of this section we show that the mere existence of the representation in (6.13) allows us to deduce some interesting properties.

We shall assume throughout that  $\mu(X) = 1$  and that  $PI_X = I_X$ . Further, we note that a permutation  $\{\alpha(1), \dots, \alpha(r)\}$  of the set  $\{1, \dots, r\}$  (Refer to Theorem 6.4) for which there is no invariant subset is called a **cycle** or **cyclical permutation**.

**Theorem 6.5** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and  $P : L^1 \rightarrow L^1$  a constrictive Markov operator. Then  $P$  is ergodic if and only if the permutation  $\{\alpha(1), \dots, \alpha(r)\}$  of the sequence  $\{1, \dots, r\}$  is cyclical.*

**Proof:** " $\Leftarrow$ " : Firstly we recall from (6.3) that the average  $A_n f$  is defined by

$$A_n f(x) = \frac{1}{n} \sum_{j=0}^{n-1} P^j f(x).$$

Hence, from (6.18),  $A_n f$  can be written as

$$A_n f(x) = \sum_{i=1}^r \left[ \frac{1}{n} \sum_{j=0}^{n-1} \lambda_{\alpha^{-j}(i)}(f) \right] \bar{I}_{A_i}(x) + \bar{Q}_n f(x),$$

where the remainder is given by

$$\bar{Q}_n f = \frac{1}{n} \sum_{j=0}^{n-1} Q_j f \quad Q_0 f = - \sum_{i=1}^r \lambda_i(f) \bar{I}_{A_i} + f.$$

We now consider the coefficients

$$\frac{1}{n} \sum_{j=0}^{n-1} \lambda_{\alpha^{-j}(i)}(f). \tag{6.21}$$

Since, as we showed in the discussion following Theorem 6.4, the sequence  $\{\lambda_{\alpha^{-j}(i)}\}$  is periodic in  $j$ , the summation in (6.21) will always have a limit as  $n \rightarrow \infty$ . Let this limit be  $\bar{\lambda}_i(f)$ . Suppose that there are no invariant subsets of  $\{1, \dots, r\}$  under the permutation  $\alpha$ . Then the limits  $\bar{\lambda}_i(f)$  must be independent of  $i$  since every piece of the summation in (6.21) of length  $r$  for different  $i$  consists of the same numbers

but in a different order. Thus,

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

Further, since  $\alpha$  is cyclical, Proposition 6.4 implies that  $\mu(A_i) = \mu(A_j) = \frac{1}{r}$  for all  $i, j$  and  $\bar{I}_{A_i} = r I_{A_i}$ , so that

$$\lim_{n \rightarrow \infty} A_n f = r \bar{\lambda}(f).$$

Hence, for  $f \in D$ ,  $\bar{\lambda}(f) = \frac{1}{r}$ . Thus, we have proved that if the permutation  $\{\alpha(1), \dots, \alpha(r)\}$  of  $\{1, \dots, r\}$  is cyclical, then  $\{P^n f\}$  is Cesàro convergent to 1, so that from Definition 5.5 is ergodic.

"  $\Rightarrow$  " : Suppose that  $P$  is ergodic and that  $\{\alpha(i)\}$  is not a cyclical permutation. Then  $\{\alpha(i)\}$  has an invariant subset  $I$ . As an initial  $f$  take

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

where

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

Then

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

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

**Theorem 6.6** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and  $P : L^1 \rightarrow L^1$  a constrictive Markov operator. If  $r = 1$  in the representation (6.13) for  $P$ , then  $P$  is exact.*

**Proof:** Suppose  $r = 1$ , so that from (6.18) we have

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

and, then,

$$\lim_{n \rightarrow \infty} P^n f = \lambda(f)I_X.$$

In particular, when  $f \in D$   $\lambda(f) \equiv 1$  since  $P$  preserves the norm. Hence, for all  $f \in D$ ,  $\{P^n f\}$  converges strongly to 1, and from Definition 5.5  $P$  is therefore exact (and, hence, also mixing).■

The converse to this theorem is interesting, since we prove that  $P$  mixing implies that  $r = 1$ .

**Theorem 6.7** *Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space and  $P : L^1 \rightarrow L^1$  a constrictive Markov operator. If  $P$  is mixing, then in the representation in (6.13)  $r = 1$ .*

**Proof:** Suppose  $P$  is mixing but  $r > 1$  and take an initial  $f \in D$  given by

$$f(x) = c_1 I_{A_1}(x), \quad \text{where } c_1 = \frac{1}{\mu(A_1)}.$$

Therefore

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

where  $A(n) = A_{\alpha^n(1)}$ . Since  $P$  was assumed to be mixing, again from Definition 5.5  $\{P^n f\}$  converges weakly to 1. However, we note that

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

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

## 6.4 Asymptotic Stability of $\{P^n\}$

We now establish the conditions under which we obtain asymptotic stability of the Markov operator and then show how the lower bound function technique maybe used to demonstrate asymptotic stability within our system. Here we assume the general case in which  $(X, \mathcal{A}, \mu)$  is taken to be an arbitrary measure space.

**Definition 6.7** *Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  a Markov operator. Then  $\{P^n\}$  is said to be **asymptotically stable** if there exists a unique  $f_* \in D$  such that  $Pf_* = f_*$  and*

$$\lim_{n \rightarrow \infty} \|P^n f - f_*\| = 0 \quad \text{for every } f \in D.$$

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

**Definition 6.8** Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $P : L^1 \rightarrow L^1$  the Frobenius-Perron operator corresponding to a nonsingular transformation  $S : X \rightarrow X$ . If  $\{P^n\}$  is asymptotically stable, then the transformation  $S$  is called **statistically stable**.

The next theorem is a direct consequence of the spectral decomposition theorem. It stipulates one of the prerequisites for the asymptotic stability of  $\{P^n\}$  where  $P$  is the Markov operator.

**Theorem 6.8** Let  $P$  be a constrictive Markov operator. Suppose there is a set  $A \subset X$ ,  $\mu(A) > 0$ , with the property that for every  $f \in D$  there is an integer  $n_0(f)$  such that

$$P^n f(x) > 0 \tag{6.22}$$

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

**Proof:** Since  $P$  is constrictive we can apply (6.13) of Theorem 6.4 and write  $Pf$  in the form

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

We will first show that  $r = 1$ .

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

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

Hence, clearly  $P^{nr}$  is not positive on the set  $A$  since  $A$  is not contained in the support of  $g_{i_0}$ . But, this result contradicts (6.22) of the theorem and, thus, we must have  $r = 1$ .

Since  $r = 1$ , (6.15) reduces to

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

so that

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

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

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

Thus  $\lim_{n \rightarrow \infty} P^n f = g$  for all  $f \in D$ , that is,  $\lim_{n \rightarrow \infty} \|P^n f - f_*\| = 0$  for all  $f \in D$ , thereby proving that  $\{P^n f\}$  is asymptotically stable. ■

However, the disadvantage of this theorem is that it requires checking two different criteria that:

- (1)  $P$  is constrictive, and
- (2) there exists a set  $A \subset X$ ,  $\mu(A) > 0$ .

We can, by a slight modification of the assumption that  $P^n f$  is positive on a set  $A$ , completely eliminate the necessity of assuming  $P$  to be constrictive. To do this, we first introduce the notion of a lower-bound function.



**Definition 6.9** A function  $h \in L^1$  is called a **lower-bound function** for a Markov operator  $P : L^1 \rightarrow L^1$  if

$$\lim_{n \rightarrow \infty} \|(P^n f - h)^-\| = 0 \text{ for every } f \in D. \quad (6.23)$$

Condition (6.23) may be rewritten as

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

where  $\|\epsilon_n\| \rightarrow 0$  as  $n \rightarrow \infty$  or, even more explicitly, as

$$P^n f \geq h - \epsilon_n.$$

**Remark 6.3** From the above we observe that  $\|h\| \leq 1 + \|\epsilon_n\|$  so, in particular  $\|h\| \leq 1$ . ■

**Remark 6.4** If  $\lim_{n \rightarrow \infty} \|P^n f - h\| \rightarrow 0$ , then  $\|(P^n f - h)^+ - (P^n f - h)^-\| \rightarrow 0$  so that  $\lim_{n \rightarrow \infty} \|(P^n f - h)^-\| = 0$ . ■

Thus, a lower-bound function  $h$  is one such that, for every density  $f$ , successive iterates of that density by  $P$  are eventually almost above  $h$ . Hence, any nonpositive function is a lower-bound function, but, since  $f \in D$  and thus  $P^n f \in D$  and all densities are positive, a negative lower-bound function is of no interest. Hence we have a second definition.

**Definition 6.10** A lower-bound function  $h$  is called **nontrivial** if  $h \geq 0$  and  $\|h\| > 0$ .

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

**Theorem 6.9** *Let  $P : L^1 \rightarrow L^1$  be a Markov operator.  $\{P^n\}$  is asymptotically stable if and only if there is a nontrivial lower-bound function for  $P$ .*

**Proof:** "  $\Rightarrow$  " : Suppose  $P : L^1 \rightarrow L^1$  is a Markov operator and that  $\{P^n\}$  is asymptotically stable. Then from Definition 6.7 this means that there exists a unique  $f_* \in D$  such that  $Pf_* = f_*$  and that

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

Next we observe that  $\|P^n f - f_*\| = \|(P^n f - f_*)^+\| + \|(P^n f - f_*)^-\| \geq \|(P^n f - f_*)^-\|$  so  $h = f_*$  can be taken as lower-bound function. It is also a nontrivial lower-bound as  $\|f_*\| = 1 > 0$  and  $f_* \geq 0$ . We hence obtain that

$$\lim_{n \rightarrow \infty} \|P^n f - h\| = 0 \text{ for every } f \in D$$

which from Remark 6.4 implies Condition (6.23) so that  $h$  is a nontrivial lower-bound function for  $P$ .

"  $\Leftarrow$  " : The proof of the "if" part will be done in two steps.

*STEP 1.* Let  $f_1, f_2 \in D$ . We first show that

$$\lim_{n \rightarrow \infty} \|P^n(f_1 - f_2)\| = 0 \tag{6.24}$$

and then proceed to construct the function  $f$ . Since every Markov operator is contractive (see (4.5)):  $\|Pf\| \leq \|f\|$ , so that

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

Hence for every pair of densities  $f_1, f_2 \in D$ , the  $\|P^n(f_1 - f_2)\|$  is a decreasing function of  $n$ .

If  $f_1 = f_2$ , then (6.24) is trivial. Assume then that  $g = f_1 - f_2$  and note that, since  $f_1, f_2 \in D$ ,  $g > 0$ , and from the definition of  $g, g^+$  and  $g^-$ ,

$$c = \|g^+\| = \|g^-\| = \frac{1}{2}\|g\| > 0.$$

(Note  $c > 0$  since  $g^+ \neq 0$  and  $g^- \neq 0$  by using a similar argument to that in the proof Theorem 5.4.) Since  $c > 0$ , we have from  $g = g^+ - g^-$  that

$$\|P^n g\| = c \|(P^n \left(\frac{g^+}{c}\right) - h) - (P^n \left(\frac{g^-}{c}\right) - h)\|, \quad (6.25)$$

where  $h$  is the nontrivial lower bound for  $P$ . Since  $\frac{g^+}{c}$  and  $\frac{g^-}{c}$  belong to  $D$ , from (6.23), there must exist an integer  $n_1$  such that for all  $n \geq n_1$

$$\|(P^n \left(\frac{g^+}{c}\right) - h)^-\| \leq \frac{1}{4}\|h\|$$

and

$$\|(P^n \left(\frac{g^-}{c}\right) - h)^-\| \leq \frac{1}{4}\|h\|.$$

We now try to establish upper bounds for  $\|P^n(\frac{g^+}{c}) - h\|$  and  $\|P^n(\frac{g^-}{c}) - h\|$ .

To do this, we first note that, for any pair of nonnegative real numbers  $a$  and  $b$ ,

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

Hence, by applying this we obtain

$$\begin{aligned} \|P^n \left( \frac{g^+}{c} \right) - h\| &= \int_X |P^n \left( \frac{g^+}{c} \right) (x) - h(x)| d\mu_x \\ &= \int_X P^n \left( \frac{g^+}{c} \right) (x) d\mu_x - \int_X h(x) d\mu_x + 2 \int_X (P^n \left( \frac{g^+}{c} \right) (x) - h(x))^- d\mu_x \\ &= \|P^n \left( \frac{g^+}{c} \right)\| - \|h\| + 2\|(P^n \left( \frac{g^+}{c} \right) - h)^-\| \\ &\leq 1 - \|h\| + 2 \cdot \frac{1}{4} \|h\| \\ &= 1 - \frac{1}{2} \|h\| \text{ for } n \geq n_1 \end{aligned}$$

Similarly we obtain,

$$\|P^n \left( \frac{g^-}{c} \right) - h\| \leq 1 - \frac{1}{2} \|h\| \text{ for } n \geq n_1$$

Note from Remark 6.3 that  $1 - \frac{1}{2} \|h\| > 0$ . Thus (6.25) gives

$$\begin{aligned} \|P^n g\| &\leq c\|P^n \left( \frac{g^+}{c} \right) - h\| + c\|P^n \left( \frac{g^-}{c} \right) - h\| \\ &\leq c(2 - \|h\|) \\ &= \|g\|(1 - \frac{1}{2} \|h\|) \text{ for } n \geq n_1. \end{aligned}$$

Hence from this equation, for any  $f_1, f_2 \in D$ , we can find an integer  $n_1$  such that

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

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

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

After  $k$  repetition of this procedure, we have

$$||P^{n_1+\dots+n_k}(f_1 - f_2)|| \leq ||f_1 - f_2||\left(1 - \frac{1}{2}||h||\right)^k,$$

and we observe that since  $0 < 1 - \frac{1}{2}||h|| < 1$ ,  $(1 - \frac{1}{2}||h||)^k \rightarrow 0$  as  $k \rightarrow \infty$ . Therefore by the "Sandwich theorem"  $\lim_{n \rightarrow \infty} ||P^{n_1+\dots+n_k}(f_1 - f_2)|| = 0$ , and since  $||P^n(f_1 - f_2)||$  is a decreasing and bounded sequence of  $n$  this then implies (6.24).

*STEP 2* Next we construct a maximal lower-bound function for  $P$ . Thus, let

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

Since by assumption, from the first part of this proof, there is a nontrivial  $h$ , we must have  $0 < \rho \leq 1$ . We then observe that for any two lower-bound functions

$h_1, h_2$ , the function  $h = \max (h_1, h_2)$  is also a lower-bound function. To see this, note that

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

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

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

is also a lower-bound function since

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

and by the Lebesgue dominated convergence theorem,

$$||h_j - h_*|| = \int_X h(x) d\mu_x - \int_X h_j(x) d\mu_x \rightarrow 0 \text{ as } j \rightarrow \infty.$$

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

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

which implies that  $h \leq h_*$ .

Observe that, from Property M2, since  $(Pf)^- \leq Pf^-$ , for every  $m$  and  $n(n > m)$ ,

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

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

Finally, from (6.23), we have

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

which implies from Definition 6.7 that  $\{P^n\}$  is asymptotically stable. ■

In checking the conditions of Theorem 6.9, it is necessary to show that (6.23) is satisfied for all  $f \in D$  and, this is difficult to do. In fact it is sufficient to check that this is true only for an arbitrary class of functions  $f \in D_0 \subset D$ , where the set  $D_0$  is dense in  $D$ . (See Definition 2.23)

## 6.5 Asymptotic Stability of the Markov Operator defined by a Stochastic Kernel

In this section we show that Theorem 6.9 can be applied to operators  $P$  defined by stochastic kernels and, in fact, gives a simple sufficient condition for the asymptotic stability of  $\{P^n\}$ .

**Definition 6.11** Let  $(X, \mathcal{A}, \mu)$  be a measure space and  $K : X \times X \rightarrow \mathbb{R}$  be measurable function that satisfies

$$0 \leq K(x, y) \quad (6.26)$$

and

$$\int_X K(x, y) dx = 1 \quad [dx = d\mu_x] \quad (6.27)$$

Any function  $K$  satisfying (6.26) and (6.27) is called a **stochastic kernel**.

We now define an integral operator  $P$  by

$$Pf(x) = \int_X K(x, y)f(y)dy \quad \text{for } f \in L^1. \quad (6.28)$$

We observe the following properties of the operator  $P$ :

(1)  $P$  is linear and nonnegative. Since we have that

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

therefore  $P$  is a Markov operator. Particularly if  $X$  is a finite space and  $\mu$  is a counting measure, then  $P$  is a stochastic matrix (refer to Example 6.1.)

(2)  $P_a P_b$  is also an integral operator.

Consider two Markov operators  $P_a$  and  $P_b$  and their corresponding stochastic ker-



nels,  $K_a$  and  $K_b$ . Thus,

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

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

$$K(x, y) = \int_X K_a(x, z)K_b(z, y)dz. \quad (6.29)$$

We denote this composed kernel  $K$  by

$$K = K_a * K_b. \quad (6.30)$$

and note that the composition has the properties:

- (a)  $K_a * (K_b * K_c) = (K_a * K_b) * K_c$  (associative law) and
- (b) Any kernel formed by the composition of stochastic kernels is stochastic.

We are able to state the following corollary.

**Corollary 6.4** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $K : X \times X \rightarrow R$  a stochastic kernel, i.e.  $K$  satisfies (6.26) and (6.27), and  $P$  the corresponding Markov operator defined by (6.28). Denote by  $K_n$  the kernel corresponding to  $P^n$ . If, for some  $m$ ,*

$$\int_X \inf_y K_m(x, y)dx > 0, \quad (6.31)$$

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

**Proof:** Let us fix  $m$  for which (6.31) holds. From the definition of  $K_m$ , for every  $f \in D(X)$  we have that

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

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

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

so that

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

If we set

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

then, since  $K_n$  is a stochastic kernel we have

$$P^{n+m} f(x) \geq h(x) \int_X \left\{ \int_X K_n(z, y) dz \right\} f(y) dy$$

$$= h(x) \int_X f(y) dy$$

Futhermore, since  $f \in D(X)$ ,

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

and therefore,

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

Thus

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

which implies that condition (6.23) holds, thereby proving that  $\{P^n\}$  is asymptotically stable.■

Although condition (6.31) on the kernel is quite simple, it is seldom satisfied when  $K(x, y)$  is defined on an unbounded space. In such a case, where condition (6.31) is not satisfied, an alternative approach offers a way to examine the asymptotic properties of iterates of densities by Markov operators.

**Definition 6.12** *Let  $G$  be an unbounded measurable subset of a  $d$ -dimensional Euclidean space  $R^d$ ,  $G \subset R^d$ . We call any continuous nonnegative function  $V : G \rightarrow R$  satisfying for  $x \in G$ ,*

$$\lim_{|x| \rightarrow \infty} V(x) = \infty \tag{6.32}$$

**a Liapunov function.**

With this definition we now introduce the **Chebyshev inequality** through the following proposition.

**Proposition 6.5** *Let  $(X, \mathcal{A}, \mu)$  be a measure space,  $V : X \rightarrow R$  an arbitrary non-negative measurable function, and for all  $f \in D$  set*

$$E(V|f) = \int_X V(x)f(x)d\mu_x.$$

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

$$\int_{G_a} f(x)d\mu_x \geq 1 - \frac{E(V|f)}{a} \quad (6.33)$$

*(the Chebyshev inequality).*

**Proof:** Since  $V$  and  $f$  are nonnegative it follows that,

$$\begin{aligned} E(V|f) &\geq \int_{X \setminus G_a} V(x)f(x)d\mu_x \\ &\geq a \int_{X \setminus G_a} f(x)d\mu_x \\ &\geq a \left\{ 1 - \int_{G_a} f(x)d\mu_x \right\} \end{aligned}$$

thereby proving the Chebyshev inequality.■

Now with the lower-bound Theorem 6.9 and the Chebyshev inequality, it is possible to prove the following theorem.

**Theorem 6.10** *If the kernel  $K(x, y)$  satisfies*

$$\int_G \inf_{|y| \leq r} K(x, y) dx > 0 \text{ for every } r \geq 0, \quad (6.34)$$

*and has a Liapunov function  $V : G \rightarrow R$  such that*

$$\int_G K(x, y) V(x) dx \leq \alpha V(y) + \beta, \quad 0 \leq \alpha < 1, \beta \geq 0. \quad (6.35)$$

*then for the Markov operator  $P : L^1(G) \rightarrow L^1(G)$ , defined by (6.28)  $\{P^n\}$  is asymptotically stable.*

**Proof** We first define the function

$$E_n(V|f) = \int_G V(x) P^n f(x) dx \quad (6.36)$$

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

$$E_n(V|f) = \int_G V(x) dx \int_G K(x, y) P^{n-1} f(y) dy = \int_G P^{n-1} f(y) dy \int_G K(x, y) V(x) dx. \quad (6.37)$$

Substituting (6.35) into (6.37) yields

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

Hence by an induction argument, we observe that from this equation we obtain that

$$E_n(V|f) \leq \frac{\beta}{(1-\alpha)} + \alpha^n E_0(V|f).$$

Even though  $E_0(V|f)$  is dependent on our initial choice of  $f$ , we observe that, for every  $f$  such that

$$E_0(V|f) < \infty, \tag{6.38}$$

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

$$E_n(V|f) \leq \frac{\beta}{(1-\alpha)} + 1 \text{ for all } n \geq n_0. \tag{6.39}$$

Now let

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

so that from the Chebyshev inequality we have

$$\int_{G_a} P^n f(x) dx \geq 1 - \frac{E_n(V|f)}{a}. \tag{6.40}$$

Further, set

$$a > 1 + \frac{\beta}{(1-\alpha)}.$$

From (6.39) we have

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

and thus (6.40) becomes

$$\int_{G_a} P^n f(x) dx \geq 1 - \frac{1}{a} \left( 1 + \frac{\beta}{1-\alpha} \right) = \epsilon > 0 \text{ for } n \geq n_0.$$

Since  $V(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ , there is an  $r > 0$  such that  $V(x) > a$  for  $|x| > r$ .

Thus the set  $G_a$  is entirely contained in the ball  $|x| \leq r$ , and we may write

$$\begin{aligned} P^{n+1}f(x) &= \int_G K(x,y) P^n f(y) dy \\ &\geq \int_{G_a} K(x,y) P^n f(y) dy \\ &\geq \inf_{y \in G_a} K(x,y) \int_{G_a} P^n f(y) dy \\ &\geq \inf_{|y| \leq r} K(x,y) \int_{G_a} P^n f(y) dy \\ &\geq \epsilon \inf_{|y| \leq r} K(x,y) \end{aligned} \tag{6.41}$$

for all  $n \geq n_0$ . By setting

$$h(x) = \epsilon \inf_{|y| \leq r} K(x,y)$$

in inequality (6.41) we therefore have from assumption (6.34), that

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

Finally, as a result of the continuity of  $V$ , the set  $D_0 \subset D$  of all  $f$  such that (6.38) is satisfied is dense in  $D$ . Thus all the conditions of Theorem 6.9 are satisfied, proving that  $\{P^n f\}$  is asymptotically stable. ■

## 6.6 Examples of Asymptotic Stability

Here we shall present two examples in which the Markov operator displays asymptotic stability. In the first example we discuss an alternative concept of asymptotic stability namely the existence of absorbing states.

**Example 6.1** *Let  $X = \{1, \dots, n\}$  be a finite set,  $\mu$  a counting measure and  $P$  a stochastic matrix. We now prove that in an absorbing chain with a single absorbing state, for any density  $f$ ,  $P^n f$  approaches the density containing 1 in the component that corresponds to the absorbing state and zeros elsewhere.*

For the purpose of this example the events are described by the components of the density  $f = (f_1, \dots, f_n)$ ,  $f_i \geq 0$ ,  $\sum_i^n f_i = 1$ . The components of the vector are called states. (Fraleigh and Beauregard 1990)

**Definition 6.13** *A state in a Markov chain is called **absorbing** if it is impossible to leave that state over the next time period.*



**Definition 6.14** A Markov chain is called **absorbing** if:

- (1) it contains at least one absorbing state.
- (2) it is possible to get from any state to an absorbing state in some number of time periods.

Let

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{k1} & \cdots & p_{kn} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

be any stochastic matrix, and

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_k \\ \vdots \\ f_n \end{pmatrix}$$

be any initial density. Let  $k$  be an absorbing state. Then it follows that

$$P = \begin{pmatrix} p_{11} & \cdots & 0 & \cdots & p_{1n} \\ \vdots & & & & \vdots \\ p_{k1} & \cdots & 1 & \cdots & p_{kn} \\ \vdots & & & & \vdots \\ p_{n1} & \cdots & 0 & \cdots & p_{nn} \end{pmatrix},$$

where the  $k$ -th column contains 1 in the  $k$ -th row and 0's everywhere else.

From Definition 6.14(2) we know that there exists  $m > 0$  such that it is possible to reach the absorbing state from any state in  $m$  time periods. Raising  $P$  to the power  $m$ , we get

$$P^m f = \begin{pmatrix} p'_{11} & \cdots & 0 & \cdots & p'_{1n} \\ \vdots & & & & \vdots \\ p'_{k1} & \cdots & 1 & \cdots & p'_{kn} \\ \vdots & & & & \vdots \\ p'_{n1} & \cdots & 0 & \cdots & p'_{nn} \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_k \\ \vdots \\ f_n \end{pmatrix}.$$

From the property of  $m$  let  $q = p'_{ki}$  be the smallest entry in the row of  $P^m$  corresponding to the absorbing state. We prove that  $q \neq 0$ .

Firstly, we note that the absorbing state is given by

$$p'_{k1}f_1 + \cdots + p'_{k-1}f_{k-1} + 1 \cdot f_k + p'_{k+1}f_{k+1} + \cdots + p'_{kn}f_n.$$

Suppose that  $q = 0$ . This then implies that  $p'_{ki} = 0$ , and that there is no representative from the state  $i$  in the absorbing state  $k$  after  $m$  time intervals, which is contrary to the definition of  $m$ . Hence  $q \neq 0$ .

We now prove that as  $n \rightarrow \infty$ ,

$$(P^m)^n f \rightarrow \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

Define, for any  $r \geq 1$

$$(P^m)^r f = \begin{pmatrix} f_1^{(r)} \\ \vdots \\ f_n^{(r)} \end{pmatrix},$$

then, using the definitions of  $q$  and the stochastic matrix we have the estimate

$$\begin{aligned}
f_k^{(1)} &= p'_{k1}f_1 + \dots + 1 \cdot f_k + \dots + p'_{kn}f_n \\
&\geq f_k + q(f_1 + \dots + f_{k-1} + f_{k+1} + \dots + f_n) \\
&= f_k + q(1 - f_k) \\
&= f_k + q - qf_k \\
&= f_k(1 - q) + q \\
&= f_k p + q
\end{aligned} \tag{6.42}$$

Similarly

$$(P^m)^2 f = \begin{pmatrix} f_1^{(2)} \\ \vdots \\ f_k^{(2)} \\ \vdots \\ f_n^{(2)} \end{pmatrix}$$

so that from (6.42) we obtain

$$\begin{aligned}
f_k^{(2)} &= p'_{k1}f_1^{(1)} + \dots + 1 \cdot f_k^{(1)} + \dots + p'_{kn}f_n^{(1)} \\
&\geq f_k^{(1)} + q(1 - f_k^{(1)}) \\
&= f_k^{(1)} p + q \\
&\geq f_k p^2 + qp + q
\end{aligned} \tag{6.43}$$

Thus we make the induction conjecture:

$$f_k^{(r)} \geq f_k p^r + q(p^{r-1} + p^{r-2} + \dots + 1) \quad (6.44)$$

Suppose that this true for some  $r > 1$ .

We therefore prove that :

$$f_k^{(r+1)} \geq f_k p^{r+1} + q(p^r + p^{r-1} + \dots + 1).$$

Hence, from Equation (6.44) we have

$$\begin{aligned} f_k^{(r+1)} &= p'_{k1} f_1^{(r)} + \dots + p'_{k,k-1} f_{k-1}^{(r)} + 1 \cdot f_k^{(r)} + \dots + p'_{kn} f_n^{(r)} \\ &\geq f_k^{(r)} + q(1 - f_k^{(r)}) \\ &= f_k^{(r)} p + q \\ &\geq (f_k p^r + q(p^{r-1} + p^{r-2} + \dots + 1))p + q \\ &= f_k p^{r+1} + q(p^r + p^{r-1} + \dots + 1) \end{aligned} \quad (6.45)$$

Since  $p^{(r)} \rightarrow 0$  as  $r \rightarrow \infty$  we obtain from (6.44) that

$$\begin{aligned} \lim_{r \rightarrow \infty} p_k^r &\geq 0 + q \frac{1}{1-p} = q \frac{1}{q} \\ &= 1 \end{aligned}$$

In addition to this from the definition of the stochastic matrix we therefore obtain that

$$1 \geq p_k^{(r)} \geq 1,$$

and from the "sandwich theorem" we can conclude that  $p_k^{(r)} = 1$  as  $r \rightarrow \infty$ . Our final result is that as  $r \rightarrow \infty$

$$(P^m)^r f \rightarrow \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}. \quad (6.46)$$

It can be shown that the convergence (6.46) holds not only for  $\{P^{mr}f\}_{r \in N}$  but for the sequence  $\{P^n f\}_{n \in N}$ . Thus in the notation of this chapter, we see that if  $P$  is a Markov operator generating an absorbing Markov chain, then  $\{P^n\}$  is asymptotically stable by Definition 6.7.

**Example 6.2** *An example of the applicability of results in this chapter, is a simple model for the cell cycle (Lasota and Mackey 1984).*

**Proof** Firstly, it is assumed that there exists an intracellular substance (mitogen), necessary for mitosis and that the rate of change of mitogen is governed by

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

with solution  $m(r, t)$ . The rate  $g$  is a  $C^1$  function on  $[0, \infty)$  and  $g(x) > 0$  for  $x > 0$ .

Secondly, we assume that the probability of mitosis in the interval  $[t, t + \Delta t]$  is given by  $\phi(m(t))\Delta t + o(\Delta t)$ , where  $\phi$  is a  $C^1$  function on  $[0, \infty)$  such that

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

where  $q(x) = \frac{\phi(x)}{g(x)}$ .

Finally, we assume that at mitosis each daughter cell receives exactly one-half of the mitogen present in the mother cell.

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

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

where

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

Then, clearly  $K(x, r)$  satisfies conditions (6.26) and (6.27) and is, thus, a stochastic kernel. Hence the operator  $P : L^1 \rightarrow L^1$  defined by

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

is a Markov operator.

We use Theorem 6.10 to show that there is a unique stationary density  $f_* \in D$  to which  $\{P^n f\}$  converges strongly. First we examine the integral

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

Integrating by parts, we have

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

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

$$q(x) \geq \epsilon \text{ for } x \geq d,$$

and, as a consequence,

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

Futhermore,

$$\begin{aligned} \int_{\frac{r}{2}}^{\infty} \exp \left[ -\int_r^{2x} q(y)dy \right] dx &\leq \int_{\frac{r}{2}}^{\infty} \exp \{ -\epsilon[2x - \max(r, d)] \} dx \\ &= \frac{1}{2\epsilon} \exp \{ -\epsilon[r - \max(r, d)] \} \\ &\leq \frac{1}{2\epsilon} \exp(\epsilon d). \end{aligned}$$

Consequently we obtain

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

so that the kernel satisfies inequality (6.35) of Theorem 6.10. Hence it only remains to show that  $K$  satisfies (6.34).

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

$$\begin{aligned} K(x, r) &= 2q(2x) \exp \left[ - \int_r^{2x} q(y) dy \right] \\ &\geq 2q(2x) \exp \left[ - \int_0^{2x} q(y) dy \right] \quad \text{for } 0 \leq r \leq r_0, x \geq \frac{1}{2}r \end{aligned}$$

and, as a consequence,

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

Further,

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

and, hence  $K(x, r)$  satisfies condition (6.34) of Theorem 6.10. Thus, in the simple model of cell division, we know that there is a globally asymptotically stable distribution of mitogen.



## **7 An Introduction to Continuous Time Systems**

### **7.1 Introduction**

We have, thus far, introduced some of the basic techniques in the study of irregular behaviour from the theory of dynamical systems in a setting that was simple. This setting was that of an iterative procedure in a discrete time dynamical system. Our goal in this chapter is now to introduce the concept of continuous time systems since these systems are much easier to understand once the basic behaviour of discrete time systems has been established.

We will also present an extension of many results previously found for discrete time systems, as well as the development of tools and techniques specifically designed for studying continuous time systems.

### **7.2 Continuous Time Dynamical Systems**

In this section we define a continuous time dynamical system. We also establish the concepts of invariance, ergodicity, mixing, and exactness for continuous dynamical systems. In what follows we will be concerned with continuous dynamical systems.

We now state a more precise definition of a continuous dynamical system in a measure space that is also equipped with a topology. Let  $X$  be a topological Hausdorff space and  $\mathcal{A}$  the  $\sigma$ -algebra of Borel sets. (Note that a topological space is called a **Hausdorff space** if every two distinct points have nonintersecting neighbourhoods.)

**Definition 7.1** *A dynamical system  $\{S_t\}_{t \in \mathbb{R}}$  on  $X$  is a family of transformations  $S_t : X \rightarrow X, t \in \mathbb{R}$ , 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  $t, t' \in \mathbb{R}$ ;
- (c) *the mapping  $(t, x) \rightarrow S_t(x)$  from  $X \times \mathbb{R}$  into  $X$  is continuous.*

**Definition 7.2** *A semidynamical system  $\{S_t\}_{t \geq 0}$  on  $X$  is a family of transformations  $S_t : X \rightarrow X, t \in \mathbb{R}^+$ , 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  $t, t' \in \mathbb{R}^+$ ;
- (c) *The mapping  $(t, x) \rightarrow S_t(x)$  from  $X \times \mathbb{R}^+$  into  $X$  is continuous.*

**Remark 7.1** The only difference between dynamical and semidynamical lies in property (b). We shall discuss only systems for  $t \geq 0$ , that is, semidynamical systems. Thus by a continuous dynamical system we will understand a family of functions  $\{S_t\}_{t \geq 0}$  that satisfies properties (a) to (c) of Definition 7.2. ■

**Remark 7.2** Property (b) simply means that the dynamics governing the evolution of the system are the same on the interval  $[0, t']$  and  $[t, t + t']$ . From the algebraic point of view it means that the (semi-)dynamical system has a semigroup structure and therefore in the linear case, it is called a semigroup of operators. ■

Next we define an invariant measure and an invariant set for a continuous dynamical system. Firstly we note from the continuity property (c) of Definition 7.2 that all

our transformations  $\{S_t\}$  are measurable, that is, for all  $A \in \mathcal{A}, t \geq 0$ ,

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

We now state the following definition.

**Definition 7.3** *A measure  $\mu$  is called **invariant** under a family  $\{S_t\}$  of measurable transformations  $S_t : X \rightarrow X$  if*

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

For a given finite invariant measure  $\mu$ , we can now formulate a continuous time analogue of the Birkhoff individual ergodic theorem.

**Theorem 7.1** *Let  $\mu$  be a finite invariant measure with respect to the continuous dynamical system  $\{S_t\}_{t \geq 0}$ , and let  $f : X \rightarrow R$  be an arbitrary integrable function. Then the limit*

$$f^*(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(S_t(x)) dt \tag{7.1}$$

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

**Proof:** We prove this in a similar way to the corresponding discrete time result in Theorem 5.5. We shall not however provide a proof for an arbitrary integrable function, but shall assume that for almost all  $x \in X$  the integrand  $f(S_t(x))$  is a

bounded measurable function of  $t$  and shall set

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

and assume at first that  $T$  is an integer,  $T = n$ . From property (b) of Definition 7.2 we obtain

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

Then the integral on the right-hand side of (7.1) 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,  $S_k = S_1 \circ S_{k-1} = S_1 \circ \dots \circ S_1 = S_1^k$ , so that

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

and the right-hand side exists from Theorem 5.5. We 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  $T \rightarrow \infty$ , then from Theorem 5.6 the first term on the right-hand side converges to  $f^*(x)$ , as we have previously shown, whereas the second term converges to zero since  $f(S_t(x))$  is bounded. This is so since  $n < T$  therefore  $T - 1 < n < T$  and  $n \rightarrow \infty$  as  $T \rightarrow \infty$ . ■

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

- (1)  $f^*(S_t(x)) = f^*(x)$ , a.e in  $x$  for every  $t \geq 0$  and
- (2)  $\int_X f^*(x) dx = \int_X f(x) dx$ .

We now define ergodicity, mixing and exactness for continuous dynamical systems. To begin we note that for a continuous dynamical system  $\{S_t\}_{t \geq 0}$  a set  $A \in \mathcal{A}$  is called *invariant* if

$$S_t^{-1}(A) = A \text{ for } t \geq 0 \tag{7.2}$$

By using this notion of invariant sets, we can now define ergodicity for continuous dynamical systems. ■

**Definition 7.4** A continuous dynamical system  $\{S_t\}_{t \geq 0}$  consisting of nonsingular transformations  $S_t : X \rightarrow X$  is called **ergodic** if every invariant set  $A \in \mathcal{A}$  is such that either  $\mu(A) = 0$  or  $\mu(X \setminus A) = 0$ , that is, every invariant set  $A$  is trivial.

**Example 7.1** As an example of ergodicity we consider rotations on the unit circle, that we introduced in Example 5.5. Let  $X = [0, 2\pi)$  and

$$S_t(x) = x + \omega t \pmod{2\pi}. \quad (7.3)$$

$S_t$  is measure preserving with respect to the Borel measure on the circle and, for  $\omega \neq 0$ , it is also ergodic.

To prove this we first pick  $t = t_0$  such that  $\frac{\omega t_0}{2\pi}$  is irrational. Then the transformation  $S_{t_0} : X \rightarrow X$  is ergodic, as was shown in Example 5.5. Therefore 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 from Definition 5.2. Thus any set  $A$  that satisfies (7.2) must likewise be trivial, and the semidynamical system  $\{S_t\}_{t \geq 0}$  with  $S_t$  given by (7.8) is ergodic.

**Definition 7.5** A continuous dynamical system  $\{S_t\}_{t \geq 0}$  on a measure space  $(X, \mathcal{A}, \mu)$  with a normalized invariant measure  $\mu$  is called **mixing** if

$$\lim_{t \rightarrow \infty} \mu(A \cap S_t^{-1}(B)) = \mu(A)\mu(B) \text{ for all } A, B \in \mathcal{A} \quad (7.4)$$

The example of the continuous dynamical system  $\{S_t\}_{t \geq 0}$  consisting of rotation on the unit circle as defined in (7.3) is not mixing. This result can be proven as follows:

Consider any two nontrivial disjoint sets  $A, B \in \mathcal{A}$ .

Then,  $A \cap S_t^{-1}(B)$  is always zero for  $\omega t = 2\pi n$  ( $n$  an integer), so that the left-hand side of (7.4) is always zero. On the other hand the right-hand side  $\mu(A)\mu(B) \neq 0$ , thereby showing that (7.4) does not hold in this case.

Since continuous dynamical systems are defined for  $t \geq 0$ , they cannot be invertible and may therefore display exactness.

**Definition 7.6** Let  $(X, \mathcal{A}, \mu)$  be a normalized measure space. A measure-preserving continuous dynamical system  $\{S_t\}_{t \geq 0}$  such that  $S_t(A) \in \mathcal{A}$  for  $A \in \mathcal{A}$  is called **exact**

if

$$\lim_{t \rightarrow \infty} \mu(S_t(A)) = 1 \text{ for all } A \in \mathcal{A}, \mu(A) > 0. \quad (7.5)$$

**Remark 7.3** As in discrete time systems, exactness of  $\{S_t\}_{t \geq 0}$  implies that  $\{S_t\}_{t \geq 0}$  is mixing. ■

### 7.3 Semigroups of the Frobenius-Perron and Koopman Operators

As we have observed in the discrete time case many properties of dynamical systems containing a large number of elements are more easily studied using suitable averages e.g. (probability) densities. In the discrete time case our system is generated by the iterates of a linear operator. In the continuous time case this approach leads to a semigroup of linear operators which thereby enables us to apply the techniques of linear functional analysis in our study.

We now introduce the concept of the semigroup of Frobenius-Perron operators.

Suppose that a measure  $\mu$  on  $X$  is given and that all transformations  $S_t$  of a continuous dynamical system  $\{S_t\}_{t \geq 0}$  are nonsingular, that is,

$$\mu(A) = 0 \Rightarrow \mu(S_t^{-1}(A)) = 0$$

for each  $A \in \mathcal{A}$ .

Then, as in (4.9), the condition

$$\int_A P_t f(x) d\mu_x = \int_{S_t^{-1}(A)} f(x) d\mu_x \text{ for } A \in \mathcal{A} \quad (7.6)$$

for each fixed  $t \geq 0$  uniquely defines the Frobenius-Perron operator  $P_t : L^1(X) \rightarrow L^1(X)$ , corresponding to the transformation  $S_t$ . Then, from (7.6)  $P_t$  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  $f_1, f_2 \in L^1, \lambda_1, \lambda_2 \in R$ , that is,  $P_t$  is a linear operator;

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

(FP3)  $\int_X P_t f(x) d\mu_x = \int_X f(x) d\mu_x$  for all  $f \in L^1$ .

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

It is also observed that the entire family of Frobenius-Perron operators  $P_t : L^1(X) \rightarrow L^1(X)$  satisfies properties (a) and (b) of Definition 7.2. This is so since:

(a) From  $S_0(x) = x$ , we have  $S_0^{-1}(A) = A$ , it follows that for any  $A$

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

which implies that

$$P_0 f = f \text{ for all } f \in L^1(X). \quad (7.7)$$

(b) To show property (b) we first note that since  $S_{t+t'} = S_t(S_{t'})$ , then  $S_{t+t'}^{-1} = S_{t'}^{-1}(S_t^{-1})$

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



$$\begin{aligned}
&= \int_{S_t^{-1}(A)} P_{t'} f(x) d\mu_x \\
&= \int_A P_t(P_{t'} f(x)) d\mu_x
\end{aligned}$$

This implies that

$$P_{t+t'} f = P_t(P_{t'} f) \text{ for all } f \in L^1(X), t, t' \geq 0 \quad (7.8)$$

and thus  $P_t$  satisfies the semigroup property of Definition 7.2.

Hence, it can be called a semigroup of Frobenius-Perron operators.

**Definition 7.7** *Let  $(X, \mathcal{A}, \mu)$  be a measure space. A family of operators satisfying properties FP1, FP2, FP3, (7.7) and (7.8) is called a **stochastic semigroup**. Further, if, for every  $f \in L^1$  and  $t_0 \geq 0$ ,*

$$\lim_{t \rightarrow t_0} \|P_t f - P_{t_0} f\| = 0,$$

*then we say that this semigroup is continuous.*

A very useful and important property of stochastic semigroups arising as a result of the contractive property of the Markov operator is that

$$\|P_t f_1 - P_t f_2\| \leq \|f_1 - f_2\| \text{ for } f_1, f_2 \in L^1, \quad (7.9)$$

and thus from the semigroup property in (7.8), the function  $t \rightarrow \|P_t f_1 - P_t f_2\|$  is a nonincreasing function of  $t$ , since from (7.9) we observe that

$$\|P_{t+t'} f_1 - P_{t+t'} f_2\| = \|P_{t'}(P_t f_1 - P_t f_2)\| \leq \|P_t f_1 - P_t f_2\|.$$

We now use this property to prove a continuous time analogue of Theorem 6.9 demonstrating the asymptotic stability of  $\{P_t\}_{t \geq 0}$  for large times.

**Theorem 7.2** *Let  $\{P_t\}_{t \geq 0}$  be a semigroup of Markov operators, not necessarily continuous. Suppose that there exists an  $h \in L^1, h(x) \geq 0, \|h\| > 0$  such that*

$$\lim_{t \rightarrow \infty} \|(P_t f - h)^-\| = 0 \text{ for every } f \in D. \quad (7.10)$$

*Then there exists a unique density  $f_*$  such that  $P_t f_* = f_*$  for all  $t \geq 0$ . Furthermore,*

$$\lim_{t \rightarrow \infty} P_t f = f_* \text{ for every } f \in D \quad (7.11)$$

**Proof:** Take any  $t_0 > 0$  and define  $P = P_{t_0}$  so that  $P_{nt_0} = P^n$ . Then, from (7.10)

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

Hence, from Theorem 6.9, there is a unique  $f_* \in D$  such that  $P f_* = f_*$  and

$$\lim_{n \rightarrow \infty} P^n f = f_* \text{ for every } f \in D$$

Thus we have shown that  $P_t f_* = f_*$  for the set  $\{t_0, 2t_0, \dots\}$ . We shall prove that  $P_t f_* = f_*$  for all  $t$ . For this we choose a particular time  $t'$ , set  $f_1 = P_{t'} f_*$ , and note that  $f_* = P^n f_* = P_{nt_0} f_*$ . Therefore,

$$\begin{aligned} \|P_{t'} f_* - f_*\| &= \|P_{t'}(P_{nt_0} f_*) - f_*\| \\ &= \|P_{nt_0}(P_{t'} f_*) - f_*\| \end{aligned}$$

$$\begin{aligned}
&= ||P^n(P_{t'}f_*) - f_*|| \\
&= ||P^n f_1 - f_*||.
\end{aligned} \tag{7.12}$$

Thus, since,

$$\lim_{n \rightarrow \infty} ||P^n f_1 - f_*|| = 0$$

and the left-hand side of (7.12) is independent of  $n$ , we must have  $||P_{t'}f_* - f_*|| = 0$ , so,  $P_{t'}f_* = f_*$ . Since  $t'$  is arbitrary, we must have  $P_t f_* = f_*$  for all  $t \geq 0$ .

Finally, to show (7.11), we choose a function  $f \in D$  so that  $||P_t f - f_*|| = ||P_t f - P_t f_*||$  is a non-increasing function of  $t$ . Then choose a subsequence  $t_n = nt_0$ . Since  $\lim_{n \rightarrow \infty} ||P_{t_n} f - f_*|| = 0$  it follows that we have a non-increasing function that converges to zero on a subsequence, therefore

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

Next, we define the semigroup of Koopman operators.

Let  $f \in L^\infty$ . Then, from Definition 4.9 of the Koopman operator the function  $U_t f$ , defined by

$$U_t f(x) = f(S_t(x)), \tag{7.13}$$

is again a function in  $L^\infty$ . Equation (7.13) 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 (7.13), satisfies all the properties of the discrete time Koopman operator in Definition 4.9.

We now prove that  $\{U_t\}_{t \geq 0}$  is a semigroup.

(a) We observe that,  $U_0 f(x) = f(S_0(x))$ , or

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

(b) Furthermore, from (7.13) we have that

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

which implies that

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

Hence from (a) and (b) it follows that  $\{U_t\}_{t \geq 0}$  is a semigroup.

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

$$\langle P_t f, g \rangle = \langle f, U_t g \rangle \text{ for all } f \in L^1, g \in L^\infty \text{ and } t \geq 0 \quad (7.14)$$

However, the family of Koopman operators is, in general, not a stochastic semigroup since  $U_t$  does not map  $L^1$  into itself unless  $S_t$  is measure-preserving (though it does map  $L^\infty$  into itself). In general the Koopman operator preserves neither the  $L^1$  nor  $L^\infty$  norm but satisfies the inequality

$$\|U_t\|_{L^\infty} = \text{ess sup } |U_t f| \leq \text{ess sup } |f| = \|f\|_{L^\infty}.$$

The next definition provides a common notion for families of operators such as  $\{P_t\}$  and  $\{U_t\}$ . Let  $L$  be a space  $L^p$ . Let  $L$  be a space  $L^p$  or in general an arbitrary Banach space.

**Definition 7.8** *A family  $\{T_t\}_{t \geq 0}$  of operators  $T_t : L \rightarrow L$ , defined for  $t \geq 0$ , is called a **semigroup of contracting linear operators** (or a **semigroup of contractions**) if for each  $t, t' \geq 0$ ,  $T_t$  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  $f_1, f_2 \in L, \lambda_1, \lambda_2 \in R$ , that is,  $T_t$  is a linear operator;
- (b)  $\|T_t f\|_L \leq \|f\|_L$  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

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

then this semigroup is called **continuous**.

One may observe that the study of continuous time systems is more difficult than in discrete time systems. This is due to the continuity of processes over time. Corresponding 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, we are able to define an important tool in the study of continuous time problems. It is called the infinitesimal generator and is defined as follows:

**Definition 7.9** Let  $\{T_t\}_{t \geq 0}$  be a semigroup of contractions. We define  $\mathcal{D}(A)$  to be the set of all  $f \in L$  such that the limit

$$Af = \lim_{t \rightarrow 0} \frac{T_t f - f}{t} \quad (7.15)$$

exists, where the limit is considered in the sense of strong convergence. (Refer to Definition 2.21). Thus (7.15) is equivalent to

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

The operator  $A : \mathcal{D}(A) \rightarrow L$  is called the **infinitesimal generator**.

From (7.15) we observe that the subspace  $\mathcal{D}(A)$  is a linear subspace and the operator  $A : \mathcal{D}(A) \rightarrow L$  is linear.

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

We now state a theorem that presents a relation between semigroups of contractions and infinitesimal generators. We first define the strong derivative of a function.

**Definition 7.10** Let  $u : \Delta \rightarrow L$  be a function, where  $\Delta \subset \mathbb{R}$ , and  $t_0 \in \Delta$ . We define the **strong derivative**  $u'(t_0)$  by

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

where the limit is considered in the sense of strong convergence.

Hence we see that the value of the infinitesimal generator for  $f \in \mathcal{D}(A)$ ,  $Af$ , is simply the derivative of the function  $u(t) = T_t f$  at  $t = 0$ . The following theorem

provides a more explicit relation between the strong derivative and the infinitesimal generator.

**Theorem 7.3** *Let  $\{T_t\}_{t \geq 0}$  be continuous semigroup of contractions acting on  $L$ , and  $A : \mathcal{D}(A) \rightarrow L$  the corresponding infinitesimal generator. Further, let  $u(t) = T_t f$  for fixed  $f \in \mathcal{D}(A)$ . Then  $u(t)$  has the following properties:*

- (1)  $u(t) \in \mathcal{D}(A)$  for  $t \geq 0$ ;
- (2)  $u'(t)$  exists for  $t \geq 0$ ; and
- (3)  $u(t)$  satisfies the differential equation

$$u'(t) = Au(t) \text{ for } t \geq 0 \quad (7.16)$$

and the initial condition

$$u(0) = f. \quad (7.17)$$

**Proof:** From our assumption the properties (1) - (3) are satisfied for  $t = 0$ .

We then consider  $t > 0$ . Let  $t_0 > 0$  be fixed. From the definition of  $u(t)$ , we have

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

Since  $T_t = T_{t-t_0} T_{t_0}$  for  $t > t_0$ , this differential equation may be written as

$$\frac{u(t) - u(t_0)}{t - t_0} = T_{t_0} \left( \frac{T_{t-t_0} f - f}{t - t_0} \right) \text{ for } t > t_0. \quad (7.18)$$

Since  $f \in \mathcal{D}(A)$ , the limit of

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

exists as  $t \rightarrow t_0$  and gives  $Af$ . Thus the limit of (7.18) also exists and is equal to  $T_{t_0}Af$ . Similarly, if  $t < t_0$ , we have  $T_{t_0} = T_t T_{t_0-t}$  and, as a consequence,

$$\frac{u(t) - u(t_0)}{t - t_0} = T_t \left( \frac{T_{t_0-t}f - f}{t_0 - t} \right) \text{ for } t < t_0 \quad (7.19)$$

and

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

Again, since  $T_tAf$  converges to  $T_{t_0}Af$  as  $t \rightarrow t_0$ , the limit of (7.19) exists as  $t \rightarrow 0$  and is equal to  $T_{t_0}Af$ . Thus the existence of the derivative  $u'(t_0)$  is proved. We can therefore rewrite the (7.18) in the form

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

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

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

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

**Remark 7.4** The main property of the set  $\mathcal{D}(A)$  that follows directly from this



theorem is that, for  $f \in \mathcal{D}(A)$ , the function  $u(t) = T_t f$  is a solution of (7.16) and (7.17).■

Since we have now developed the concept of the semigroup of the Frobenius-Perron operators and the notion of infinitesimal generators we can now examine applications of these semigroups in determining the invariance of a measure and the ergodicity of a transformation.

**Theorem 7.4** *Let  $(X, \mathcal{A}, \mu)$  be a measure space, and  $S_t : X \rightarrow X$  a family of non-singular transformations. Also let  $P_t : L^1 \rightarrow L^1$  be the Frobenius-Perron operator corresponding to  $\{S_t\}_{t \geq 0}$ . Then the measure*

$$\mu_f(A) = \int_A f(x) d\mu_x$$

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

**Proof:** The proof is similar to that in Theorem 5.1. If  $\mu_f$  is invariant, then

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

which, from the definition of  $P_t$ , implies that  $P_t f = f$ .

To prove the converse we apply the equality  $P_t f = f$  to  $\mu_f(A) = \int_A f(x) d\mu_x$  to obtain that  $\mu_f$  is invariant.■

**Remark 7.5** Now we examine the inter-dependence between  $A_{FP}$ , the infinitesimal generator for the semigroup of Frobenius-Perron operators, and the existence of an invariant measure.

Suppose that  $\mu_f$  is invariant. Then, from Theorem 7.4 we know that  $P_t f = f$ , and since

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

then  $A_{FP}f = 0$ . Thus, if  $\mu_f$  is invariant then  $A_{FP}f = 0$ . However, to show that if  $A_{FP}f = 0$ , then  $\mu_f$  is invariant is not so easy since the proof requires that we pass from the infinitesimal generator to the semigroup. The solution to this problem lies in establishing the way in which semigroups are constructed from infinitesimal generators. We shall describe this in detail in the next section. ■

Next we investigate the way in which the semigroup of the Koopman operator is used to study the ergodicity of a continuous dynamical system  $\{S_t\}_{t \geq 0}$ . We state this in the following theorem.

**Theorem 7.5** *A continuous dynamical system  $\{S_t\}_{t \geq 0}$  with nonsingular transformations  $S_t : X \rightarrow X$  is ergodic if and only if the fixed points of  $\{U_t\}_{t \geq 0}$  are constant functions.*

**Proof:** The proof is similar to that of Theorem 5.3.

"  $\Rightarrow$  " : Suppose that  $\{S_t\}_{t \geq 0}$  is not ergodic. Then, there is an invariant nontrivial subset  $C \subset X$ , that is,

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

By setting  $f = I_C$ , we have

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

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

"  $\Leftarrow$  " : Suppose that 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. 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.

## 7.4 The Hille-Yosida Theorem

We now state and discuss the applications of an important result that forms part of our analysis: the Hille-Yosida theorem. This result serves primarily to demonstrate the existence of a semigroup corresponding to a given linear operator.

**Theorem 7.6** [Hille-Yosida] *Let  $A : \mathcal{D}(A) \rightarrow L$  be a linear operator, where  $\mathcal{D}(A) \subset L$  is a linear subspace of  $L$ . In order for  $A$  to be an infinitesimal generator for a continuous semigroup of contractions, it is necessary and sufficient that the following three conditions are satisfied:*

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

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

$$\lambda g - Ag = f; \quad (7.20)$$

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

$$\|\lambda g - Ag\|_L \geq \lambda \|g\|_L. \quad (7.21)$$

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

$$T_t f = \lim_{\lambda \rightarrow \infty} e^{tA_\lambda} f, f \in L, \quad (7.22)$$

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

(The meaning of  $e^{tA_\lambda}$  is explained in (7.31))

The proof of this theorem can be found in (Dynkin 1965 or Dunford and Schwartz 1957).

We now derive an expression which represents  $e^{tA_\lambda}$  by writing the operator  $A_\lambda = \lambda A R_\lambda$  in different forms, each of which is useful in different applications.

Firstly, if we substitute  $g = R_\lambda f$  into (7.20) we have

$$\lambda R_\lambda f - A R_\lambda f = f \text{ for } f \in L. \quad (7.23)$$

Next we apply the operator  $R_\lambda$  to both sides of (7.20) and by also using  $g = R_\lambda f$  we obtain

$$\lambda R_\lambda g - R_\lambda A g = g \text{ for } g \in \mathcal{D}(A). \quad (7.24)$$

Then from (7.23) and (7.24) we have

$$R_\lambda A f = A R_\lambda f \text{ for } f \in \mathcal{D}(A). \quad (7.25)$$

Re-arranging (7.23) we also obtain

$$A R_\lambda f = (\lambda R_\lambda - I) f \text{ for } f \in L. \quad (7.26)$$

Where  $I$  is the identity operator ( $I f \equiv f$  for all  $f$ ). Thus we have three possible representations for  $A_\lambda$ :

(1) the original definition

$$A_\lambda = \lambda A R_\lambda; \quad (7.27)$$

(2) from (7.26)

$$A_\lambda = \lambda(\lambda R_\lambda - I); \quad (7.28)$$

(3) from (7.25)

$$A_\lambda = \lambda R_\lambda A. \quad (7.29)$$

The representations in (7.27) and (7.28) hold in the entire space  $L$ , whereas (7.29) holds in  $\mathcal{D}(A)$ .

Now, from conditions (b) and (c) of the Hille-Yosida theorem, and also by using  $g = R_\lambda f$ , it follows that

$$\|f\|_L \geq \lambda \|R_\lambda f\|_L. \quad (7.30)$$

Hence from (7.28) and (7.30),

$$\begin{aligned} \|A_\lambda f\|_L &= \|\lambda^2 R_\lambda f - \lambda f\|_L \leq \|\lambda^2 R_\lambda f\|_L + \|\lambda f\|_L \\ &\leq 2\lambda \|f\|_L, \end{aligned}$$

so that the operator  $e^{tA_\lambda}$  is defined as the series

$$e^{tA_\lambda} f = \sum_{n=0}^{\infty} \frac{t^n}{n!} A_\lambda^n f, \quad (7.31)$$

which is strongly convergent for all  $t$ . (Strong convergence of the above series follows since  $|tA_\lambda| < \infty$ .)

The Hille-Yosida has many applications.

Suppose we have an operator  $A : \mathcal{D}(A) \rightarrow L$  that satisfies conditions (a) to (c) of the Hille-Yosida theorem, and such that the solution  $g = R_\lambda f \geq 0$  for  $f \geq 0$ .

(a) Then, as we shall show next,  $T_t f \geq 0$  for every  $f \geq 0$ .

By substituting from (7.28) we have

$$e^{tA_\lambda} f = e^{\lambda(\lambda R_\lambda - I)t} f = e^{-t\lambda} (e^{t\lambda^2 R_\lambda} f), \quad (7.32)$$

where from (7.31)

$$e^{t\lambda^2 R_\lambda} f = \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} (\lambda R_\lambda)^n f. \quad (7.33)$$

Further, for any  $f \geq 0$ ,  $R_\lambda f \geq 0$ . Therefore, by induction,  $R_\lambda^n f \geq 0$ .

Thus, from (7.33), since  $\lambda > 0$  and  $t \geq 0$ ,  $\exp(t\lambda^2 R_\lambda) f \geq 0$  and so, from (7.32),  $e^{tA_\lambda} f \geq 0$ . From (7.22), we finally have  $T_t f \geq 0$  since it is the limit of nonnegative functions.

(b) Now suppose that the operator  $\lambda R_\lambda$  preserves the integral, that is,

$$\lambda \int_X R_\lambda f(x) d\mu_x = \int_X f(x) d\mu_x \text{ for all } f \in L^1, \lambda > 0. \quad (7.34)$$

We now show that these properties imply that

$$\int_X T_t f(x) d\mu_x = \int_X f(x) d\mu_x, f \geq 0, t \geq 0.$$

Since the series in (7.33) is strongly convergent, and by using (7.34), we obtain

$$\begin{aligned} \int_X e^{t\lambda^2 R_\lambda} f(x) d\mu_x &= \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} \int_X (\lambda R_\lambda)^n f(x) d\mu_x \\ &= \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} \int_X f(x) d\mu_x \\ &= e^{t\lambda} \int_X f(x) d\mu_x. \end{aligned} \quad (7.35)$$

Then, from (7.32) and (7.35),

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

thereby showing that  $\|T_t f\| = \|f\|$ .

Hence, from the above results in (a) and (b) we have the following corollary.

**Corollary 7.1** *Let  $A : \mathcal{D}(A) \rightarrow L^1$  be an operator satisfying conditions (a) to (c) of the Hille-Yosida theorem. If the solutions  $g = R_\lambda f$  of (7.20) is such that  $\lambda R_\lambda$  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 to be checked, as condition (c) is automatically satisfied for any Markov operator. This is so since if we set  $f = \lambda g - Ag$  and write inequality (7.21) in the form

$$\|f\| \geq \|\lambda R_\lambda f\|,$$

then, from the contractive property of the Markov operator the above inequality is always satisfied if  $\lambda R_\lambda$  is a Markov operator.

Another important application of the Hille-Yosida theorem is that it provides an immediate and simple way to demonstrate that  $A_{FP} = 0$  is a sufficient condition that  $\mu_f$  is an invariant measure as discussed in Remark 7.5.

Thus,  $Af = 0$  implies, from (7.29), that  $A_\lambda f = 0$  and from (7.31)

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

This, combined with (7.22) and Theorem 7.4 gives

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

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



Finally, we investigate one more application of the Hille-Yosida theorem. Thus far, one of the main uses of this theorem was to demonstrate the existence of a semigroup corresponding to an infinitesimal generator  $A$ , and to study the properties of this semigroup and the resolvent equation (7.20). Our next objective is to determine the semigroup corresponding to an infinitesimal generator  $A$ .

Let  $X = R$  and  $L = L^1(R)$ , and for  $f \in L^1$ , consider the infinitesimal generator

$$Af = \frac{d^2 f}{dx^2} \quad (7.36)$$

Let  $\mathcal{D}(A)$  be the set of all  $f \in L^1$  such that  $f''(x)$  exists almost everywhere, is integrable on  $R$  and is such that

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

Hence  $\mathcal{D}(A)$  is the set of all  $f$  such that  $f'$  is absolutely continuous and  $f''$  is integrable on  $R$ . This then implies that  $f'$  can always be determined since  $f'$  is differential and integrable.

We will show that there exists a unique semigroup corresponding to the infinitesimal generator  $A$ . To do so we must show that the three conditions of the Hille-Yosida theorem are satisfied.

(1)  $\mathcal{D}(A)$  is dense in  $L^1$ .

(2) From (7.21) and (7.36) the resolvent equation has the form

$$\lambda g - \frac{d^2 g}{dx^2} = f \quad (7.37)$$

which is a second order differential equation in the unknown function  $g$ . Using standard arguments, the general solution of (7.37) may be written as

$$g(x) = C_1 e^{-\alpha x} + C_2 e^{\alpha x} + \frac{1}{2\alpha} \int_{x_0}^x e^{-\alpha(x-y)} f(y) dy - \frac{1}{2\alpha} \int_{x_1}^x e^{\alpha(x-y)} f(y) dy$$

where  $\alpha = \sqrt{\lambda}$ , and  $C_1, C_2, x_0$  and  $x_1$  are arbitrary constants. Next we observe that since the functions  $e^{-\alpha(x-y)}$  and  $e^{\alpha(x-y)}$  are continuous functions on the intervals  $[x_0, x]$  and  $[x, x_1]$  respectively, they are therefore integrable over these intervals.

Furthermore, for  $f(y) \in L^1(-\infty, +\infty)$

(1) if  $y \leq x$  and  $\alpha > 0$ , then  $-\alpha(x-y) \leq 0$ , so that

$$|e^{-\alpha(x-y)}| \leq 1 \Rightarrow |e^{-\alpha(x-y)} f(y)| \leq |f(y)|. \text{ Hence,}$$

$$\lim_{x_0 \rightarrow -\infty} \int_{x_0}^x e^{-\alpha(x-y)} f(y) dy = \int_{-\infty}^x e^{-\alpha(x-y)} f(y) dy.$$

Similarly,

(2) if  $y > x$ , and  $\alpha < 0$ , then  $\alpha(x-y) > 0$  so that

$$|e^{\alpha(x-y)}| > 1 \Rightarrow |e^{\alpha(x-y)} f(y)| > |f(y)|, \text{ and therefore}$$

$$\lim_{x_1 \rightarrow +\infty} \int_x^{x_1} e^{\alpha(x-y)} f(y) dy = \int_x^{+\infty} e^{\alpha(x-y)} f(y) dy.$$

Then, by the use of improper integrals we are able to assign  $x_0 = -\infty$  and  $x_1 = +\infty$ .

Now set

$$K(x-y) = \left(\frac{1}{2\alpha}\right) e^{-\alpha|x-y|}. \quad (7.38)$$

so that the solution to (7.37) can be written in the more compact form

$$g(x) = C_1 e^{-\alpha x} + C_2 e^{\alpha x} + \int_{-\infty}^{\infty} K(x-y) f(y) dy. \quad (7.39)$$

It follows that  $K(x-y)f(y)$  is integrable over  $R \times R$ , so by (7.38), Fubini's Theorem 2.3 and the equality  $\alpha = \sqrt{\lambda}$

$$\int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} K(x-y) f(y) dy \right) dx = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} K(x-y) f(y) dy$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} K(x-y)dx \int_{-\infty}^{\infty} f(y)dy \\
&= \frac{1}{\lambda} \int_{-\infty}^{\infty} f(y)dy.
\end{aligned} \tag{7.40}$$

Thus, since neither  $e^{-\alpha x}$  nor  $e^{\alpha x}$  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 (7.36) has a unique solution  $g \in L^1$  given by

$$g(x) = R_{\lambda}f(x) = \int_{-\infty}^{\infty} K(x-y)f(y)dy, \tag{7.41}$$

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

Combining (7.40) and (7.41) it follows that the operator  $\lambda R_{\lambda}$  preserves the integral. Moreover,  $\lambda R_{\lambda}f \geq 0$  if  $f \geq 0$ , so that  $\lambda R_{\lambda}$  is a Markov operator. Thus condition (c) of the Hille-Yosida theorem is satisfied, and we have shown that the operator  $\frac{d^2}{dx^2}$  is an infinitesimal generator of a continuous semigroup  $\{T_t\}_{t \geq 0}$  of Markov operators, where

$$T_t f = \lim_{\lambda \rightarrow \infty} e^{-t\lambda} \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} (\lambda R_{\lambda})^n f \tag{7.42}$$

and  $R_{\lambda}$  is defined by (7.38) and (7.41).

## 8 An Application

### 8.1 Introduction

Our main objective in this chapter is to demonstrate the use of the theory that we have developed in our study thus far. We discuss a system that is governed by a Poisson process and show that the equation governing its dynamics can be alternatively derived using methods of linear kinetic theory. We call this equation the abstract linear Boltzmann equation of kinetic theory. Finally we examine some properties of the linear Boltzmann equation which enable us to describe asymptotic behaviour of the system.

### 8.2 The Abstract linear Boltzmann Equation

In this section we provide two approaches to derive the linear Boltzmann equation. One is based on the analysis of a deterministic discrete time process coupled with a Poisson process and the other by using a hypothetical system of  $N$  particles.

Refer to Section 2.4 so as to recall key concepts namely stationary independent increments, random variables, stochastic process and counting process.

**Definition 8.1** *A Poisson process is a counting process  $\{N_t\}_{t \geq 0}$  with stationary independent increments satisfying:*

(a)  $N_0 = 0$ ;

(b)  $\lim_{t \rightarrow 0} (\frac{1}{t}) \text{prob} \{N_t \geq 2\} = 0$ ;

(c) *The limit*

$$\lambda = \lim_{t \rightarrow 0} \frac{1}{t} \text{prob} \{N_t = 1\}$$

*exists and is positive; and*

(d)  $\text{prob} \{N_t = k\}$  *as functions of  $t$  are continuous.*

As an example of a Poisson process we shall consider a decay of a radioactive substance placed in a chamber equipped with a device for detecting and counting the total number of atomic disintegrations  $N_t$  that have occurred up to a time  $t$ . It is also assumed that the amount of substance must be sufficiently large such that during the time of observation there is no significant decrease in mass. This process has stationary independent increments, that is, the number of disintegrations that occur during any given interval of time is independent of the number occurring during any other interval. It therefore satisfies the Markovian property.

In addition we observe that for this example we may explain conditions (a) to (c) of Definition 8.1 as follows:

(a) simply means that we have zero (no) disintegrations when we start to count at  $t = 0$ .

(b) means that we cannot have two or more disintegrations in a very short space of time.

(c) means that we will most definitely have at least one disintegration in a short space of time.

We now derive an explicit formula for

$$p_k(t) = \text{prob} \{N_t = k\} \quad (8.1)$$

We do this in two steps:

STEP(1) We derive an ordinary differential equation for  $p_k(t)$ .

Firstly we rewrite conditions (a) to (c) of Definition 8.1 using the notation of (8.1):

$$p_0(0) = 1, \quad (8.2)$$

$$\lim_{t \rightarrow 0} \frac{1}{t} \sum_{i=2}^{\infty} p_i(t) = 0, \quad (8.3)$$

and

$$\lambda = \lim_{t \rightarrow 0} \frac{1}{t} p_1(t). \quad (8.4)$$

To obtain the differential equation for  $p_k(t)$ , we start with  $p_0(t)$  and note that

$$\begin{aligned} p_0(t+h) &= \text{prob} \{N_{t+h} = 0\} \\ &= \text{prob} \{N_{t+h} - N_t + N_t - N_0 = 0\} \end{aligned}$$

Then, since  $N_t$  is not decreasing in time, it follows that  $(N_{t+h} - N_t) + (N_t - N_0) = 0$  if and only if  $(N_{t+h} - N_t = 0)$  and  $(N_t - N_0) = 0$ . Thus, using the property of stationary independent increments we have

$$\begin{aligned} p_0(t+h) &= \text{prob} \{(N_{t+h} - N_t) = 0 \text{ and } (N_t - N_0) = 0\} \\ &= \text{prob} \{N_{t+h} - N_t = 0\} \text{prob} \{N_t - N_0 = 0\} \end{aligned}$$

$$\begin{aligned}
&= \text{prob} \{N_h - N_0 = 0\} \text{prob} \{N_t - N_0 = 0\} \\
&= p_0(h)p_0(t)
\end{aligned} \tag{8.5}$$

From (8.5) above we obtain

$$\frac{p_0(t+h) - p_0(t)}{h} = \frac{p_0(h) - 1}{h} p_0(t) \tag{8.6}$$

Since  $\sum_{i=0}^{\infty} p_i(t) = 1$ , we have

$$\begin{aligned}
\frac{p_0(h) - 1}{h} &= \frac{p_0(h) - \sum_{i=0}^{\infty} p_i(t)}{h} \\
&= -\frac{p_1(h)}{h} - \frac{1}{h} \sum_{i=2}^{\infty} p_i(h) \\
&= -\frac{1}{h} \sum_{i=1}^{\infty} p_i(h),
\end{aligned}$$

and thus by taking the limit as  $h \rightarrow 0$  of both sides of (8.6) we obtain

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) \tag{8.7}$$

where from (8.3) and (8.4)  $\lambda = \lim_{t \rightarrow 0} \frac{1}{t} \sum_{i=1}^{\infty} p_i(t)$ .

We can now derive the differential equation for  $p_k(t)$  in a similar manner. Thus

$$\begin{aligned}
p_k(t+h) &= \text{prob}\{N_{t+h} = k\} \\
&= \text{prob}\{N_{t+h} - N_t + N_t - N_0 = k\} \\
&= \text{prob}\{N_t - N_0 = k \text{ and } N_{t+h} - N_t = 0\} \\
&+ \text{prob}\{N_t - N_0 = k-1 \text{ and } N_{t+h} - N_t = 1\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=2}^k \text{prob}\{N_t - N_0 = k - i \text{ and } N_{t+h} - N_t = i\} \\
& = p_k(t)p_0(h) + p_{k-1}(t)p_1(h) + \sum_{i=2}^k p_{k-i}(t)p_i(h).
\end{aligned}$$

Again as above in (8.6) we have

$$\frac{p_k(t+h) - p_k(t)}{h} = \frac{p_0(h) - 1}{h} p_k(t) + \frac{p_1(h)}{h} p_{k-1}(t) + \frac{1}{h} \sum_{i=2}^k p_{k-i}(t) p_i(h),$$

and taking the limit as  $h \rightarrow 0$  we obtain

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t). \quad (8.8)$$

STEP(2) Our next step is to solve the differential equations (8.7) and (8.8).

The initial conditions for  $p_0(t)$  and  $p_k(t)$ ,  $k \geq 1$ , are just  $p_0(0) = 1$  (by definition), and this immediately gives  $p_k(0) = 0$  for all  $k \geq 1$ . Thus, from (8.7) we have

$$p_0(t) = e^{-\lambda t}. \quad (8.9)$$

Substituting this into (8.8) when  $k = 1$  gives

$$\frac{dp_1(t)}{dt} = -\lambda p_1(t) + \lambda e^{-\lambda t},$$

whose solution is

$$p_1(t) = \lambda t e^{-\lambda t}.$$

Repeating this procedure for  $k = 2, 3, \dots$  we find, by induction, that



$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This therefore shows  $p_k(t)$  as a function of  $t$  and hence the dependence of the process on time.

Subsequently, we now consider the following problem:

given an initial distribution of points  $x \in X$ , with density  $f$ , how does this distribution evolve in time? We denote this time-dependent density by  $u(t, x)$  and set  $u(0, x) = f(x)$ .

Let  $S : X \rightarrow X$  be a non-singular transformation on a measure space  $(X, A, \mu)$  coupled with a Poisson process  $\{N_t\}_{t \geq 0}$  such that the times at which the transformation  $S$  operate are dependent on the Poisson process, that is, each point  $x \in X$  is transformed into  $S^{N_t}(x)$ .

Since the dynamics of our system is stochastic, this again allows us to build our derivations based on calculating the probability of an arbitrary outcome. We begin by calculating the probability that  $S^{N_t}(x) \in A$  for a given set  $A \subset X$  and time  $t > 0$ . This probability depends on two factors: the initial density  $f$  and the counting process  $\{N_t\}$ .

This means that we need to calculate the measure of the set

$$\{(\omega, x) : S^{N_t(\omega)}(x) \in A\} \tag{8.10}$$

To do this we define the **product space** (see Definition 2.18)  $\Omega \times X$  given by

$$\Omega \times X = \{(\omega, x) : \omega \in \Omega, x \in X\}$$

that contains all sets of the form (8.10). In the space  $\Omega \times X$  we define a **product measure** (see Theorem 2.2) that, for the sets  $C \times A, C \in \mathcal{F}, A \in \mathcal{A}$ , is given by  $\text{Prob}(C \times A)$  or

$$\text{Prob}(C \times A) = \text{prob}(C)\mu_f(A), \quad (8.11)$$

where, as usual,

$$\mu_f(A) = \int_A f(x) d\mu_x.$$

This measure is denoted by "Prob" since it is a probability measure. We now calculate the measure of the set in (8.10). This set maybe written as the union of disjoint sets as follows:

$$\begin{aligned} \{(\omega, x) : S^{N_t(\omega)}(x) \in A\} &= \bigcup_{k=0}^{\infty} \{N_t(\omega) = k, S^k(x) \in A\} \\ &= \bigcup_{k=0}^{\infty} \{N_t(\omega) = k\} \times \{S^k(x) \in A\}. \end{aligned}$$

Thus using (8.1) and the definition of the Frobenius-Perron operator, the measure of this set is

$$\begin{aligned} \text{Prob}\{S^{N_t} \in A\} &= \sum_{k=0}^{\infty} \text{Prob}\{N_t(\omega) = k, S^k(x) \in A\} \\ &= \sum_{k=0}^{\infty} \text{prob}\{N_t(\omega) = k\} \mu_f(x \in S^{-k}(A)) \\ &= \sum_{k=0}^{\infty} p_k(t) \int_{S^{-k}(A)} f(x) d\mu_x \end{aligned}$$

$$= \sum_{k=0}^{\infty} p_k(t) \int_A P^k f(x) d\mu_x \quad (8.12)$$

so that

$$\text{Prob}\{S^{N_t} \in A\} = \int_A \sum_{k=0}^{\infty} p_k(t) P^k f(x) d\mu_x \quad (8.13)$$

for  $A \in \mathcal{A}$ , where as before,  $P$  denotes the Frobenius-Perron operator associated with the non-singular transformation  $S : X \rightarrow X$ . (Note that we can change the order of integration and summation in (8.13) since  $\|P^k f\| = 1$  and  $\sum_{k=0}^{\infty} p_k(t) = 1$ .) Then, the integrand on the right-hand side of (8.13) is the density  $u(t, x)$ :

$$u(t, x) = \sum_{k=0}^{\infty} p_k(t) P^k f(x). \quad (8.14)$$

The sequence on the right-hand side of (8.14) is strongly convergent in  $L^1$ . Differentiating (8.14) with respect to  $t$  and using (8.8), we have

$$\begin{aligned} \frac{\partial u(t, x)}{\partial t} &= \sum_{k=0}^{\infty} \frac{dp_k(t)}{dt} P^k f(x) \\ &= -\lambda \sum_{k=0}^{\infty} p_k(t) P^k f(x) + \lambda \sum_{k=1}^{\infty} p_{k-1}(t) P^k f(x). \end{aligned}$$

Since the last two series are strongly convergent in  $L^1$ , the initial differentiation was correct. Thus by substituting (8.8) and (8.14) we have

$$\begin{aligned} \frac{\partial u(t, x)}{\partial t} &= -\lambda u(t, x) + \lambda \sum_{k=0}^{\infty} p_k(t) P^{k+1} f(x) \\ &= -\lambda u(t, x) + \lambda P \sum_{k=0}^{\infty} p_k(t) P^k f(x) \\ &= -\lambda u(t, x) + \lambda P u(t, x). \end{aligned}$$

Therefore  $u(t, x)$  satisfies the differential equation

$$\frac{\partial u(t, x)}{\partial t} = -\lambda u(t, x) + \lambda P u(t, x) \quad (8.15)$$

with, from (8.14) and (8.2), the initial condition

$$\begin{aligned} u(0, x) &= p_0(0)P^0 f(x) + \sum_{k=1}^{\infty} p_k(0)P^k f(x) \\ &= f(x) + 0 \\ &= f(x). \end{aligned}$$

Next we aim to obtain a more "intuitive" derivation of the equation for  $u(t, x)$  based on the kinetic theory of interacting particles.

Suppose that we have a system consisting of  $N$  particles enclosed in a container, where  $N$  is a large number. Each particle may change its velocity  $x = (v_1, v_2, v_3)$  from  $x$  to  $S(x)$  only by colliding with the walls of the container. (We are considering the linear case.)

We want to determine how the velocity distribution of particles evolves with time. Thus we must determine the function  $u(t, x)$  such that

$$N \int_A u(t, x) dx$$

is the number of particles having, at time  $t$ , velocities in the set  $A$ .

The change in the number of particles, whose velocity is in  $A$ , between  $t$  and  $t + \Delta t$  is given by

$$N \int_A u(t + \Delta t, x) dx - N \int_A u(t, x) dx. \quad (8.16)$$

From our assumption, such a change can only take place through collisions with the walls of the container. Choose  $\Delta t$  to be sufficiently small so that a negligible number of particles make two or more collisions with a wall during  $\Delta t$ . Thus, the number of particles striking the wall during a time  $\Delta t$  with a velocity in  $A$  before the collision and therefore having velocities in  $S(A)$  after the collision is

$$N \lambda \Delta t \int_A u(t, x) dx, \quad (8.17)$$

where  $\lambda N$  is the number of particles striking the walls of the container per unit time. In this idealized, abstract example we neglect the important physical fact that the faster particles are striking the walls of the container more frequently than the slower particles.

Conversely, to find the number of particles whose velocity is in  $A$  after the collision, we must calculate the number having velocities in the set  $S^{-1}(A)$  before the collision. (That is, the state (in this case velocity) of the particle at  $t = n - 1$ .) Again, we assume  $\Delta t$  to be sufficiently small to make the number of double collisions by single particles negligible, we have

$$N \lambda \Delta t \int_{S^{-1}(A)} u(t, x) dx. \quad (8.18)$$

Hence the total number of particles with velocity in the set  $A$  over a short time  $\Delta t$  is given by the difference between (8.17) and (8.18)

$$N\lambda\Delta t \int_{S^{-1}(A)} u(t, x)dx - N\lambda\Delta t \int_A u(t, x)dx. \quad (8.19)$$

By combining (8.16) with (8.19), we have

$$N \int_A [u(t + \Delta t, x) - u(t, x)] dx = \lambda N \Delta t \left\{ \int_{S^{-1}(A)} u(t, x)dx - \int_A u(t, x)dx \right\},$$

and, since

$$\int_{S^{-1}(A)} u(t, x)dx = \int_A Pu(t, x)dx,$$

where  $P$  is the Frobenius-Perron operator associated with  $S$ , we have

$$\begin{aligned} & N \int_A [u(t + \Delta t, x) - u(t, x)]dx \\ &= \lambda N \Delta t \int_A [-u(t, x) + Pu(t, x)]dx. \end{aligned} \quad (8.20)$$

Equation (8.20) above is exact to within an error that is small compared to  $\Delta t$ . By dividing (8.20) through by  $\Delta t$  and passing to the limit  $\Delta t \rightarrow 0$ , we obtain

$$\int_A \frac{\partial u(t, x)}{\partial t} dx = \lambda \int_A [-u(t, x) + Pu(t, x)] dx,$$

which gives

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

Thus we again arrived at Equation (8.15).

In these two approaches we see how uncertainty (and therefore probability) appears in the study of deterministic systems.

In this example we adopted a simplifying assumption that the particle upon striking the wall, changed its velocity from  $x$  to  $S(x)$ , where  $S : X \rightarrow X$  is a point-to-point transformation. A more realistic point of view would be to assume that the change in velocity is not uniquely determined but another probabilistic event. Thus, if before the collision the particles have a velocity distribution with density  $g$ , then after the collision they have a distribution with density  $Pg$ , where  $P : L^1(X) \rightarrow L^1(X)$  is a Markov operator.

As before, we assume that  $u(t, x)$  is the density of the distribution of particles having velocity  $x$  at time  $t$ , so

$$N \int_A u(t, x) dx$$

is the number of particles with velocities in  $A$ . Once again,

$$\lambda N \Delta t \int_A u(t, x) dx$$

is the number of particles with velocity in  $A$  that will collide with the walls in a time  $\Delta t$ , whereas

$$\lambda N \Delta t \int_A Pu(t, x) dx$$

is the number of particles whose velocities go into  $A$  because of collisions over a time  $\Delta t$ . Thus,

$$-\lambda N \Delta t \int_A u(t, x) dx + \lambda N \Delta t \int_A Pu(t, x) dx$$

is the net change, due to collisions over a time  $\Delta t$  in the number of particles whose velocities are in  $A$ .

Combining this result with (8.16), we immediately obtain the balance equation (8.20), which once again leads to (8.15). The only difference is that  $P$  is no longer a Frobenius-Perron operator corresponding to a one-to-one deterministic transformation  $S$ , but is an arbitrary Markov operator.

**Remark 8.1** Since in the above derivations we have used arguments that are used to derive the Boltzmann equation, we call (8.15) a **linear abstract Boltzmann equation** corresponding to a collision (Markov) operator  $P$ . Note that  $x$  corresponds to the particle velocity and not to position. Furthermore, the equation is called linear since we assume that our only source of change for particle velocity is collisions with the wall, that drift and external forces are not considered. ■

### 8.3 Solutions of the Linear Boltzmann Equation

In order to solve the linear Boltzmann equation we let the solution be the function  $u(t, x)$ .



Then, by writing (8.15) in the form

$$\frac{du}{dt} = (P - I)u, \quad (8.21)$$

where  $P$  is a given Markov operator and  $I$  is the identity operator, we can apply the Hille-Yosida Theorem 7.6 to the study of (8.15).

We show that conditions (a) to (c) of the Hille-Yosida theorem are satisfied by the linear Boltzmann equation using the operator  $A = (P - I)$ .

(a) Firstly, since  $A = P - I$  is defined on the whole space  $L^1$ ,  $\mathcal{D}(A) = L^1$  and property (a) satisfied.

(b) To show property (b) we rewrite the resolvent equation  $\lambda f - Af = g$  using  $A = P - I$  to obtain

$$\begin{aligned} g &= \lambda f - (P - I)f \\ &= \lambda f - Pf + f \\ &= (\lambda + 1)f - Pf \end{aligned} \quad (8.22)$$

(8.22) may then be solved by using the method of successive approximations. Starting from an arbitrary  $f_0$  we define  $f_n$  by

$$(\lambda + 1)f_n - Pf_{n-1} = g,$$

where  $f_n \rightarrow f$  and  $f_{n-1} \rightarrow f$  so that,

$$f_n = \frac{1}{(\lambda + 1)^n} P^n f_0 + \sum_{k=1}^n \frac{1}{(\lambda + 1)^k} P^{k-1} g. \quad (8.23)$$

in which  $(\frac{1}{\lambda+1}) \rightarrow 0$  and  $\|P^n f_0\| \leq \|f\|$ .

From the contractive property of the Markov operator  $\|P^k g\| \leq \|g\|$  and since  $|\frac{1}{\lambda+1}| < 1$ , the series in (8.23) is therefore convergent, and the unique solution  $f$  of the resolvent equation (8.22) above is

$$f \equiv R_\lambda g = \lim_{n \rightarrow \infty} f_n = \sum_{k=1}^{\infty} \frac{1}{(\lambda+1)^k} P^{k-1} g. \quad (8.24)$$

We note that the method of successive approximations applied to an equation such as (8.22) will always result in a solution (8.23) that converges to a unique limit, as  $n \rightarrow \infty$ , when  $\|P\| \leq \lambda + 1$ .

(c) Finally, to show that property of the Hille-Yosida theorem is satisfied, we integrate (8.24) over the entire space  $X$ . It is to be noted that  $\sum_{k=0}^{\infty} \frac{1}{(1+\lambda)^k} = \frac{1}{\lambda} + 1$ .

$$\begin{aligned} \int_X R_\lambda g(x) d\mu_x &= \sum_{k=1}^{\infty} \frac{1}{(\lambda+1)^k} \int_X P^{k-1} g(x) d\mu_x \\ &= \sum_{k=1}^{\infty} \frac{1}{(\lambda+1)^k} \int_X g(x) d\mu_x \\ &= \frac{1}{\lambda} \int_X g(x) d\mu_x \\ &= \frac{1}{\lambda}, \end{aligned}$$

where we used the integral preserving property of the Markov operator. Thus,

$$\int_X \lambda R_\lambda g(x) d\mu_x = 1,$$

and, since  $\lambda R_\lambda$  is linear, nonnegative, and also preserves the integral, it is a Markov operator. Thus, from Corollary 7.1 condition (c) is satisfied.

Therefore, by the Hille-Yosida theorem, the linear Boltzmann equation (8.15) generates a continuous semigroup of Markov operators,  $\{\hat{P}_t\}_{t \geq 0}$ .

To determine an explicit formula for  $\hat{P}_t$ , we first write

$$\begin{aligned} A_\lambda f &= \lambda A R_\lambda f = \lambda(P - I) R_\lambda f \\ &= \lambda(P - I) \sum_{k=1}^{\infty} \frac{1}{(\lambda + 1)^k} P^{k-1} f \\ &= \lambda \sum_{k=1}^{\infty} \frac{1}{(\lambda + 1)^k} P^k f - \lambda \sum_{k=1}^{\infty} \frac{1}{(\lambda + 1)^k} P^{k-1} f, \end{aligned}$$

so that

$$\lim_{\lambda \rightarrow \infty} A_\lambda f = Pf - f.$$

Thus, by the Hille-Yosida theorem and (7.21), the unique semigroup corresponding to  $A = P - I$  is given by

$$\hat{P}_t f = e^{t(P-I)} f, \tag{8.25}$$

and the unique solution to (8.15) with initial condition  $u(0, x) = f(x)$  is

$$u(t, x) = e^{t(P-I)} f(x). \tag{8.26}$$

**Remark 8.2** Although we have determined the solution of (8.15) using the Hille-Yosida theorem, it is possible to obtain the same result by applying the method of successive approximations to (8.21). However our derivation illustrates the techniques involved in using the Hille-Yosida theorem and also establishes that (8.15) generates a continuous semigroup of Markov operators.■

In addition to the existence and uniqueness of the solution of (8.15) other properties of  $\hat{P}_t$  may be demonstrated.

*Property 1.* From inequality (7.9) it follows that, given  $f_1, f_2 \in L^1$ , the norm

$$||\hat{P}_t f_1 - \hat{P}_t f_2||$$

is an nonincreasing function of time  $t$ .

*Property 2.* If for some  $f \in L^1$  the limit

$$f_* = \lim_{t \rightarrow \infty} \hat{P}_t f \quad (8.27)$$

exists, then for the same  $f_*$

$$\lim_{t \rightarrow \infty} \hat{P}_t(Pf) = f_* \quad (8.28)$$

*Property 3.* The operators  $P$  and  $\hat{P}_t$  commute, that is,  $P\hat{P}_t f = \hat{P}_t P f$  for all  $f \in L^1$ .

*Property 4.* If for some  $f \in L^1$  the limit (8.27),  $f_* = \lim_{t \rightarrow \infty} \hat{P}_t f$ , exists, then  $f_*$  is a fixed point of the Markov operator  $P$ , that is,  $Pf_* = f_*$ .

*Property 5.* If  $Pf_* = f_*$  for some  $f_* \in L^1$ , then also  $\hat{P}_t f_* = f_*$ .

Our objective in determining behaviour to our evolving system is to be able to identify, using the results that we have established thus far, patterns of behaviour: be they regular or irregular. We shall now show that  $\hat{P}_t$  always converges to a limit.

Recall our definitions of precompactness Definition 6.1 and Definition 6.2.

**Theorem 8.1** *If the trajectory  $\{\hat{P}_t f\}$  is weakly precompact, then there exists a fixed point of  $P$ .*

**Proof:** Suppose that  $\{\hat{P}_t f\}$  is weakly precompact, then there exists a sequence  $\{t_n\}$  such that

$$\lim_{n \rightarrow \infty} \hat{P}_{t_n} f = f_* \text{ weakly} \quad (8.29)$$

exists. From Property 2 this implies the weak convergence of

$$\lim_{n \rightarrow \infty} \hat{P}_{t_n}(Pf) = \lim_{n \rightarrow \infty} P(\hat{P}_{t_n} f) = Pf_*. \quad (8.30)$$

However, from (8.28) we have

$$\lim_{n \rightarrow \infty} \hat{P}_{t_n}(f - Pf) = 0,$$

and, thus, from (8.29) and (8.30), we have

$$Pf_* = f_*,$$

proving that there exists a fixed point of  $P$ . We also observe from Property 5 of  $\hat{P}_t$  that this implies that  $\hat{P}_t f_* = f_*$ . ■

**Theorem 8.2** *For a given  $f \in L^1$ , if the trajectory  $\{\hat{P}_t f\}$  is weakly precompact, then  $\hat{P}_t f$  strongly converges to a limit.*

**Proof:** The proof is similar to that of Theorem 6.1 in the discrete time case.

Let  $f \in L^1$ . Suppose that the trajectory  $\{\hat{P}_t f\}$  is weakly precompact. Then from Theorem 8.1  $\hat{P}_t f$  converges weakly to an  $f_*$  that is a fixed point of  $P$  and  $\hat{P}_t$ . Write  $f$  in the form

$$f = f - f_* + f_*.$$

Suppose that for every  $\epsilon > 0$  the function  $f - f_*$  may be written in the form

$$f - f_* = Pg - g + r, \quad (8.31)$$

where  $g \in L^1$  and  $\|r\| \leq \epsilon$ .

We prove that this representation is possible. From (8.31) we have

$$\begin{aligned} \hat{P}_t f &= \hat{P}_t(f - f_* + f_*) \\ &= \hat{P}_t(Pg - g) + \hat{P}_t f_* + \hat{P}_t r. \end{aligned}$$

But  $\hat{P}_t f_* = f_*$  and, thus,

$$\|\hat{P}_t f - f_*\| \leq \|\hat{P}_t(Pg - g)\| + \|\hat{P}_t r\|.$$

From (8.28) it follows that  $\lim_{t \rightarrow \infty} \hat{P}_t(f - Pf) = 0$ , so that, the first term on the right-hand side approaches zero as  $t \rightarrow \infty$ , whereas the second term is not greater than  $\epsilon$ . Thus,

$$\|\hat{P}_t f - f_*\| \leq 2\epsilon$$

for  $t$  sufficiently large, and, since  $\epsilon$  is arbitrary,

$$\lim_{t \rightarrow \infty} \|\hat{P}_t f - f_*\| = 0,$$

thereby proving that  $\hat{P}_t f$  converges strongly to  $f_*$  when (8.31) is true.

Suppose that (8.31) is not true. Then this implies that

$$f - f_* \notin \overline{(P - I)L^1(X)}.$$

This, in turn, implies that from the Hahn-Banach theorem (see Proposition 6.2) that there is a  $g_0 \in L^\infty$  such that

$$\langle f - f_*, g_0 \rangle \neq 0 \quad (8.32)$$

and

$$\langle h, g_0 \rangle \neq 0$$

for all  $h \in \overline{(P - I)L^1(X)}$ . In particular

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

since  $(P - I)P^n f \in (P - I)L^1(X)$ , so

$$\langle P^{n+1} f, g_0 \rangle = \langle P^n f, g_0 \rangle$$

for  $n = 0, 1, \dots$ . Thus, by induction, we have

$$\langle P^n f, g_0 \rangle = \langle f, g_0 \rangle. \quad (8.33)$$

Furthermore, since by definition  $\hat{P}_t f = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P^n f$  where  $e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1$ , we may multiply both sides of (8.33) by  $e^{-t} \frac{t^n}{n!}$  and sum over  $n$  to obtain

$$\left\langle e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P^n f, g_0 \right\rangle = \langle f, g_0 \rangle,$$

or

$$\langle \hat{P}_t f, g_0 \rangle = \langle f, g_0 \rangle. \quad (8.34)$$

Substituting  $t = T_n$  and taking the limit as  $t \rightarrow \infty$  in (8.34) gives

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

and, thus,

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

contradicting (8.32). Thus (8.31) is true. ■

## 8.4 The Effect of the Properties of the Markov Operator on Solutions of the Linear Boltzmann Equation

To conclude this chapter we state and demonstrate some important results regarding the effect of the properties of the Markov operator  $P$  on the convergence of solutions  $\hat{P}_t$ .

We begin with the following corollary.



**Corollary 8.1** *If for  $f \in L^1$  there exists a  $g \in L^1$  such that*

$$|\hat{P}_t f| \leq g, t \geq 0 \quad (8.35)$$

*then the (strong) limit*

$$\lim_{t \rightarrow \infty} \hat{P}_t f \quad (8.36)$$

*exists. That is, either  $\hat{P}_t f$  is not bounded by any integrable function or  $\hat{P}_t$  is strongly convergent.*

**Proof:** From the first criterion for precompactness (see Section 6.2) it follows that  $\{\hat{P}_t f\}$  is weakly precompact. Thus, from Theorem 8.1 the limit (8.36) exists. ■

The next corollary is a direct consequence from a property of the Markov operator.

**Corollary 8.2** *If the Markov operator has a positive fixed point  $f_*$ ,  $f_*(x) > 0$  a.e., then the strong limit,  $\lim_{t \rightarrow \infty} \hat{P}_t f$ , exists for all  $f \in L^1$ .*

**Proof:** The proof follows in two steps.

Firstly we note that when the initial function  $f$  satisfies

$$|f| \leq c f_* \quad (8.37)$$

for some sufficiently large constant  $c > 0$ , and since  $f_*$  is a fixed point of  $P$  we have

$$\begin{aligned} |P^n f| &\leq P^n(c f_*) = c P^n f_* \\ &= c f_* \end{aligned}$$

Multiply both sides by  $e^{-t} \frac{t^n}{n!}$  and sum the result over  $n$  to get

$$\begin{aligned} \left| e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} P^n f \right| &\leq c e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} f_* \\ &= c f_* \end{aligned}$$

The left-hand side of this inequality is just  $|\hat{P}_t f|$ , so that

$$|\hat{P}_t f| \leq c f_*,$$

and, since  $\hat{P}_t f$  is bounded, Corollary 8.1 implies that the strong limit  $\lim_{t \rightarrow \infty} \hat{P}_t f$  exists.

Secondly, we consider the more general case when the initial function  $f$  does not satisfy (8.37). Define a new function by

$$f_c(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq c f_*(x); \\ 0 & \text{if } |f(x)| > c f_*(x). \end{cases}$$

It follows from the Lebesgue dominated convergence theorem that

$$\lim_{c \rightarrow \infty} \|f_c - f\| = 0.$$

Thus, by writing  $f = f_c + f - f_c$ , we have

$$\hat{P}_t f = \hat{P}_t f_c + \hat{P}_t (f - f_c).$$

Since  $f_c$  satisfies

$$|f_c| \leq cf_*,$$

from (8.37) we know that in this case  $\{\hat{P}_t f_c\}$  converges strongly. Now take  $\epsilon > 0$ . Since  $\{\hat{P}_t f_c\}$  is strongly convergent, there is a  $t_0 > 0$ , which in general depends on  $c$ , such that

$$\|\hat{P}_{t+t'} f_c - \hat{P}_t f_c\| \leq \epsilon \text{ for } t \geq t_0, t' \geq 0. \quad (8.38)$$

Further,

$$\|\hat{P}_t f - \hat{P}_t f_c\| \leq \|f - f_c\| \leq \epsilon \text{ for } t \geq 0 \quad (8.39)$$

for a fixed but sufficiently large  $c$ . From (8.38) and (8.39) it follows that

$$\|\hat{P}_{t+t'} f - \hat{P}_t f\| \leq 3\epsilon \text{ for } t \geq t_0, t' \geq 0,$$

which is the Cauchy condition for  $\{\hat{P}_t f\}$ . Thus  $\{\hat{P}_t f\}$  also converges strongly. ■

**Remark 8.3** Although the existence of a limit is interesting, from the point of view of applications we would like to know what the limit is and whether it is unique. In the next corollary we give a sufficient condition for this. ■

**Corollary 8.3** *Assume that in the set of all densities  $f \in D$  the equation  $Pf = f$  has a unique solution  $f_*$  and  $f_*(x) > 0$  a.e. Then, for an initial density,  $f \in D$*

$$\lim_{t \rightarrow \infty} \hat{P}_t f = f_*, \quad (8.40)$$

*and the convergence is strong.*

**Proof:** From Corollary 8.2 the  $\lim_{t \rightarrow \infty} \hat{P}_t f$  exists and is also a nonnegative normalized function. However, from Property 4 of  $\hat{P}_t$  (see Section 8.3), we know that this limit is a fixed point of the Markov operator  $P$ . Since, by our assumption, the fixed point is unique it must be  $f_*$ . ■

We now compare and relate these results to those which identify and define the irregular behaviour of dynamical systems as discussed in Chapter 5.

In the special case that  $P$  is a Frobenius-Perron operator for a non-singular transformation  $S : X \rightarrow X$ , the condition  $Pf_* = f_*$  (from Theorem 5.1) is equivalent to the fact that the measure

$$\mu_{f_*}(A) = \int_A f_*(x) d\mu_x$$

is invariant with respect to  $S$ . Thus, in this case, from Corollary 8.2 the existence of an invariant measure  $\mu_{f_*}$  with a density  $f_*(x) > 0$  is sufficient for the existence of the strong limit (8.36) of the solution of the linear Boltzmann equation (8.15). Recall from Theorem 5.5 that for ergodic transformations  $f_*$  is unique, we can thus summarize these results in the following corollary.

**Corollary 8.4** *Suppose  $S : X \rightarrow X$  is a nonsingular transformation and  $P$  is the corresponding Frobenius-Perron operator. Then with respect to the trajectories  $\{\hat{P}_t f\}$  that generate the solutions of the linear Boltzmann equation (8.15) we have that:*

(1) *If there exists an absolutely continuous invariant measure  $\mu_{f_*}$  with a positive density  $f_*(x) > 0$  a.e., then for every  $f \in L^1$  the strong limit,  $\lim_{t \rightarrow \infty} \hat{P}_t f$  exists, and*

(2) *If, in addition, the transformation  $S$  is ergodic, then*

$$\lim_{t \rightarrow \infty} \hat{P}_t f = f_* \quad (8.41)$$

for all  $f \in D$ .

Now consider the special case where  $(X, \mathcal{A}, \mu)$  is a finite measure space and  $S : X \rightarrow X$  is a measure preserving transformation. Since  $S$  is measure preserving,  $f_*$  exists and is given by

$$f_*(x) = \frac{1}{\mu(X)} \text{ for } x \in X. \quad (8.42)$$

Thus  $\lim_{t \rightarrow \infty} \hat{P}_t f$  always exists. Furthermore, this limit is unique, that is,

$$\lim_{t \rightarrow \infty} \hat{P}_t f = f_* = \frac{1}{\mu(X)}$$

if and only if  $S$  is ergodic. (see Theorem 5.5)

Finally, in closing this chapter we now compare this conclusion with that of exactness. We recall that from Definition 5.5, a Markov operator  $P : L^1 \rightarrow L^1$  is exact if and only if the sequence  $\{P^n f\}$  has a strong limit that is a constant for every  $f \in L^1$ . Although we have never talked about exactness when describing the behaviour of stochastic semigroups, for the situation when (8.42) holds, then, the behaviour of the trajectory  $\{\hat{P}_t f\}$  is similar to our original definition of exactness. Figuratively speaking, in such a case we can say that  $S$  is ergodic if and only if  $\{\hat{P}_t\}_{t \geq 0}$  is exact.

Note also that mixing implies ergodicity, exactness implies mixing and hence ergodicity. Now we have also, informally, deduced that ergodicity implies that exactness, thereby making these three types of irregular behaviour equivalent to each other.

## 9 Conclusion

We shall now briefly summarize the results discussed in this dissertation.

Our main objective was to present an analysis of some types of irregular behaviour that may be displayed by dynamical systems. To this end we have, by using measure-theoretic notions, described these behaviours and the conditions under which they exist in terms of the properties as displayed by non-singular transformations and the asymptotic properties of measure-preserving transformations.

In the first part we focused on discrete dynamical systems generated by pointwise deterministic transformations and introduced the concepts of ergodicity, mixing and exactness to describe their long-time properties. The central idea behind the analysis of these notions was to replace the pointwise description of the evolution by the evolution of the corresponding density function and thus to embed the process in the space of integrable functions and to replace the direct analysis of the ergodicity, mixing and exactness of the original transformation by the investigation of some analytic properties of the associated isometries called Frobenius-Perron and Koopman operators. The main advantage of this method is that the analysis of a possibly nonlinear transformation was reduced to the analysis of associated linear, though in an infinite-dimensional space.

The Frobenius-Perron and Koopman operators are particular examples of Markov

operators and therefore the above results allowed to define the notions of ergodicity, mixing and exactness for all Markov operators. Following this we considered two types of asymptotic behaviour of a general Markov operator. The first was that of asymptotic periodicity that turned out to be closely related to the notion of a constrictive Markov operator, that is, the Markov operator with a precompact attracting set. The other type was that of asymptotic stability where we required that this attracting set reduced to a point. These two concepts were subsequently related to the ergodicity, mixing and exactness of the Markov operator in question.

Next we proceeded from discrete to continuous dynamical systems and described the semigroup theory which is used to solve abstract Cauchy problem. We also discussed continuous analogues of the notions introduced and analysed earlier for discrete systems.

As the final example we discussed the abstract linear Boltzmann equation that can be used to model some dynamical system involving an element of randomness. We were able to draw a number of conclusions regarding the asymptotic behaviour of solutions to this equation.

However, although perceive the discussed behaviours as being "chaotic", we have not conclusively stipulated its relation to that of the topological definition of a chaotic dynamical system as stated by e.g. Devaney. An interesting subject for future research would be to compare the measure-theoretic and topological definitions of "chaos" so as to form a link between these two approaches by definitely deducing whether or not they are equivalent.

## 10 References

1. Anderson W J 1991 *Continuous-Time Markov Chains* (New York: Springer-Verlag)
2. Banasiak J 2000 *An introduction to chaotic behaviour of linear dynamical systems with applications* (School of Mathematical and Statistical Sciences, University of Natal, Durban)
3. Bellini-Morante A 1994 *A Concise Guide to Semigroups and Evolution Equations* (Singapore: World Scientific Publishing)
4. Birkhoff G D 1931 *Proof of the Ergodic theorem* (USA: Proc. Natl. Acad. Sci.) **17** 656-60
5. Brown J R 1976 *Ergodic Theory and Topological Dynamics* (New York: Academic Press)
6. Cohn D L 1980 *Measure Theory* (Boston: Stuttgart)
7. Coleman R 1974 *Stochastic Processes* (Norwich: George Allen and Unwin Ltd.)
8. de Barra G 1981 *Measure Theory and Integration* (New York: John Wiley and Sons)
9. Devaney R L 1989 *An Introduction to Chaotic Dynamical Systems* 2nd edn. (New York: Addison-Wesley)



10. Dunford N and Schwartz J T 1957 *Linear Operators. Part1. General Theory* (New York: Wiley)
11. Dynkin E G 1965 *Markov Processes* (New York: Springer-Verlag)
12. Fraleigh S H and Beauregard R A 1990 *Linear Algebra* 2nd edn. (New York: Addison-Wesley)
13. Friedberg S H, Insel A J and Spence L E 1989 *Linear Algebra* 2nd edn. (Engelwood, New Jersey: Prentice Hall)
14. Friedman N A 1970 *Introduction to Ergodic Theory* (New York: Van Nostrand-Reinhold Company)
15. Gray R M 1988 *Probability, Random Processes and Ergodic Properties* (New York: Springer-Verlag)
16. Hall N ed. 1992 *The new Scientist Guide to Chaos* (England: Penguin Books Ltd.)
17. Halmos P R 1950 *Measure Theory* (New York: Van Nostrand-Reinhold)
18. Hardy G H and Wright EM 1959 *An Introduction to the Theory of Numbers* 4th edn. (London: Oxford University Press)
19. Iosifescu M 1980 *Finite Markov Processes and Their Applications* (New York: John Wiley and Sons Ltd.)
20. Isaacson D L and Madsen R W 1976 *Markov Chains-Theory and Applications* (New York: John Wiley and Sons)
21. Kennedy J G 1957 *Introduction to Finite Mathematics* (Engelwood, New Jersey: Prentice-Hall)

22. Kingman J F C and Taylor S J 1966 *Introduction to Measure and Probability* (Cambridge: Cambridge University Press)
23. Kirchgraber U and Stoffer D 1989 *On the definition of Chaos* (Z. Angew. Math. Mech.) **69** 175-185
24. Koopman B O 1931 *Hamiltonian systems and transformations in Hilbert space* (USA: Proc. Natl. Acad. Sci.) **17** 315-318
25. Kreyszig E 1978 *Introductory Functional Analysis with Applications* (New York: John Wiley and Sons Ltd.)
26. Lasota A, Li T Y and Yorke J A 1984 *Asymptotic periodicity of the iterates of Markov operators* (Trans. Amer. Math. Soc.) **286** 751-764
27. Lasota A and Mackey M C 1984 *Globally asymptotic properties of proliferating cell populations* (J. Math. Biology) **19** 43-62
28. Lasota A and Mackey M C 1985 *Probabilistic properties of Deterministic Systems* (Cambridge: Cambridge University Press)
29. Lasota A and Mackey M C 1994 *Chaos, Fractals and Noise* 2nd edn. (New York: Springer-Verlag)
30. Lasota A and Yorke J A 1982 *Exact dynamical systems and the Frobenius-Perron operator* (Trans. Am. Math. Soc.) **273** 375-384
31. Li T Y and Yorke J A 1975 *Period three implies chaos* (Amer. Math. Monthly) **82** 985-992
32. Lin M 1971 *Mixing for Markov operators* (Z. Wahrscheinlichkeitstheorie Verw. Gebiete) **19** 231-242

- 33. Mika J R and Banasiak J 1995 *Singularly Perturbed Evolution Equations with Applications to Kinetic Theory* (Singapore: World Scientific Publishing Co. Pte. Ltd.)
- 34. Parry W 1981 *Topics in Ergodic Theory* (Cambridge: Cambridge University Press)
- 35. Parthasarathy K 1977 *Introduction to Probability and Measure* (Delhi: Mc Millan)
- 36. Roberts F S 1976 *Discrete Mathematical Models* (Engelwood, New Jersey: Prentice-Hall)
- 37. Smith P 1998 *Explaining Chaos* (Cambridge: Cambridge University Press)
- 38. Walters P 1982 *An Introduction to Ergodic Theory* (New York: Springer-Verlag)
- 39. Wiggins S 1990 *An Introduction to Applied Non-linear Dynamical Systems and Chaos* (New York: Springer-Verlag)