UNIVERSITY OF CALIFORNIA

Santa Barbara

Ergodic Theory and the Control of Mixing in Fluid Flows

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mechanical and Environmental Engineering

by

Domenico D'Alessandro

Committee in Charge:
Professor Mohammed Dahleh, co-Chair
Professor Igor Mezić, co-Chair
Professor Bassam Bamieh
Professor Laura Giarre'
Professor Petar V. Kokotović
Professor Anna Stefanopoulou

June 1999

The dissertation of
Domenico D'Alessandro is approved:

May 1999

Acknowledgements

It is a pleasure to acknowledge the help and support in this study of my co-advisor Prof. Mohammed Dahleh. I am very thankful to him for having given me the opportunity of pursuing graduate studies at UCSB and for his advice and guidance during the whole period of study. My co-advisor Prof. Igor Mezić introduced me to the field of Ergodic Theory that was for me a fascinating discovery.

A number of other people have contributed in various ways to this work and to my graduate studies. Prof. P. V. Kokotović taught me two very interesting courses in Nonlinear Control. His knowledge and enthusiasm are an example for every researcher. I also wish to thank Prof. C. Akemann; his course in Measure Theory was largely applied here. Prof. B. Bamieh and Prof. A. Stefanopoulou also gave me advice in many occasions. A special mention goes to Prof. Marie Dahleh, whose course in Fluid Dynamics was very important for this work. She also made my experience as a teaching assistant a useful and an enjoyable one.

This work would have been much harder without the help and support of my friends and lab-mates Jayati Ghosh, Mariateresa Napoli and Vasu Salapaka. They shared with me many important moments through this endeavour and very often listened to me with patience and understanding. I would like to acknowledge, in particular, several useful technical discussions with Vasu. Sami Ashhab and Murti Salapaka, at the beginning of my period of study, provided, as more experienced students, important suggestions and advice. Also, I would like to thank Adriano Batista for his help in the courses in Physics. A special mention should go to Paolo Mercorelli who shared with me many hard working days during the Summer of 1997 and to Gregory Hagen for helping me in the editing of the figures of this thesis. I also would like to mention David Betz, Aaron Brown, Mihailo Jovanović, Niklas Karlsson, Sascha Klein, Kerry Sears and Marion Volpert that contributed, in different ways, to create a friendly atmosphere at work. My thanks go to all of these people along

with a wish for the best in their future endeavours.

Abstract

Ergodic Theory and the Control of Mixing in Fluid Flows

by

Domenico D'Alessandro

The mixing of fluids is one of the most widespread processes in nature that underly many natural phenomena, and, at the same time, plays a major role in many industrial applications. Although the mixing processes are very different in different situations, important factors and mechanisms for mixing can indeed be identified and controlled in a systematic fashion. The development of a unified theory for the quantification and manipulation of mixing parameters is the aim of the emerging theory of *Control of Mixing*.

In this thesis, a theory of mixing control is described where the model of the mixing process is based on concepts of Ergodic Theory. A criterion is proposed for the quantification of mixing using ergodic theoretic Kolmogorov-Sinai *entropy*. In the process, a set of standard ergodic theory tools, originally developed for dynamical systems described by a single transformation, are extended to *sequences* of transformations.

A motion planning problem will be dealt with, in detail, for the so-called *prototypical flow*. This flow is important because it describes, in a simplified way, the basic phenomena involved in mechanical mixing, namely *stretching* and *folding*. The problem consists of choosing a sequence of velocity profiles which achieves the best mixing. A closed form solution is derived in terms of maximization of entropy. The entropy criterion is then tested on other mixers of physical interest, via numerical experiments. In particular the *Partitioned Pipe Mixer* will be dealt with, in detail, in different configurations.

Contents

Chapter 1

Introduction

In many technological processes a fundamental stage involves the mixing of fluids. For example, mixing between two or more fluids in a fuel-air mixture in combustion engines, enhances the efficiency of the combustion process. Mixing between two chemicals enhances the speed of a chemical reaction. Active control can be introduced in these processes in order to force the fluid to follow prescribed trajectories so that mixing is fast and uniform. Overall, the design of optimal mixing protocols is a problem of both fundamental and practical importance.

Until recently, mixing has been treated on a case by case basis, where the specifications for the steering of fluid flows were usually given by heuristics. On the surface this may seem quite reasonable since mixing processes in nature are very different from each other. As a simple example, consider the difference between an explosion and the mixing of food. The first process is very fast, it is affected by chemical reactions, and it is inherently difficult to control, whereas the latter process is quite slow, chemical reactions play almost no role, and the motion of the fluid can be affected to give good mixing in a reasonable time. Despite these differences in various mixing situations, important general factors and mechanisms can be identified and we can abstract from some other physical processes in order to construct a mathematical theory for mixing.

The aim of this thesis is to introduce concepts and tools from *Ergodic Theory* in the analysis and control of mixing fluids. The model that we adopt is a *measure theoretic model* in which the whole space where the mixing process occurs is modeled as a *measure space*, the region occupied by the fluid as a *measurable set* and the transformation acting on the fluid as a *measure preserving automorphism.* In this context, quantitative measures from ergodic theory are applied to the specification of control laws in fluid mixing.

Compared to other approaches, ergodic theory has the advantage of being 'quantitative' in the sense that it is possible to associate to a particular mixing protocol a quantity which characterizes its capability of mixing well. The quantity that will be proposed in this work is the ergodic theoretic Kolmogorov-Sinai *entropy.* It is also possible to use the ergodic theoretic concept of *mixing* to define in mathematically precise terms when a given protocol mixes well.

The connection between the mathematical machinery of ergodic theory and fluid dynamics will be done in Chapter 4 of the thesis. Until that point the treatment will be mostly mathematical. In particular, Chapter 2 will introduce the main mathematical background from measure theory and dynamical systems theory needed for the following chapters.

Chapter 3 will be a concise introduction to Ergodic Theory. Some new mathematical results will be presented here, regarding sequences of transformations on a measure space. Standard ergodic theory is developed for a *single* measure preserving transformation on a measure space. In the treatment developed here we consider the more general setting of *sequences* of measure preserving transformations. Classical results such as the *Pointwise Ergodic Theorem*, the *Mean Ergodic Theorem* and the *Recurrence Theorem* will be extended to the setting of sequences under suitable assumptions. Other new results concern the definition of entropy for sequences. In this case, theorems are proven that allow us to reduce the computation of entropy to the standard case and therefore to use standard results in the literature. The extension to sequences is physically motivated by the fact that time invariant incompressible flows are unable to undergo mixing. It is therefore necessary to consider time varying flows and in mathematical terms, a time varying set-up. This point is discussed in detail in Chapter 4.

In Chapter 5 we take up a motion planning problem for a *prototypical flow* which models the basic *stretching* and *folding* mechanisms involved in mechanical mixing. The flow is constrained to move on a two-dimensional torus according to a sequence of purely horizontal and vertical velocity profiles. The considered problem consists of choosing the latter velocity profiles in order to achieve good mixing. The concepts and tools of ergodic theory, developed in the previous chapters, are used here to formalize the problem in mathematical terms. In particular, the problem will be posed in terms of maximization of entropy. We will see that this optimization problem admits a nontrivial closed form solution which describes the best mixing sequence according to the established criteria.

Other mixers are dealt with in Chapter 6. For these mixers, a numerical analysis is necessary to verify the mixing capability of a given protocol. The analysis in this chapter is meant to test, on mixers of physical interest, the postulate that higher entropy corresponds to better mixing. In particular we consider in detail the

Partitioned Pipe Mixer in different configurations and numerically show that higher entropy always corresponds to better mixing.

This thesis is a unique work in a literature where mixing has been dealt with in a 'qualitative' way or on a case by case scenario. In an important paper [1], Aref addressed the topic of mixing from the point of view of dynamical systems theory. The book by Ottino [30] is a first attempt in introducing dynamical systems concepts in the study of mixing, in a systematic way. There is some important more recent work dealing with the control of mixing. In [24], the use of aperiodic sequences of transformation was proposed in order to enhance mixing. In [22, 23] the authors searched for conditions for the nonexistence of low period elliptic fixed points of associated maps. These points are in general responsible for the formation of large islands and consequently poor mixing. In [13], the authors advocated the use of aperiodic mixing protocols that destroy the symmetries in the flow.

Chapter 2

Mathematical Background

2.1 Measure theory and abstract integration

2.1.1 Structures of sets

Definition 2.1.1. *Given a set M, a σ-algebra (or σ–field) on M is a collection $\mathcal{S}$ of subsets of M, with the following properties:*

a)

$$M \in \mathcal{S};$$

b)

$$A \in \mathcal{S} \to A^c \in \mathcal{S};$$

c) If A_n is a countable collection of sets in $\mathcal{S}$, then

$$\cup_{n=1}^{\infty} A_n \in \mathcal{S}.$$

The sets in $\mathcal{S}$ are called *measurable sets*, a pair $(M, \mathcal{S})$ is called a *measurable space.* If 'countable' is replaced by 'finite' in c), we have an *algebra* (or *field*). It follows from a) and b) that the empty set $\emptyset$ is also in $\mathcal{S}$, and from b) and c) that a countable intersection of sets in $\mathcal{S}$ is still in $\mathcal{S}$ since $\cap_n^{\infty} A_n = (\cup_n^{\infty} A_n^c)^c$. Also $A - B = A \cap B^c$ is in $\mathcal{S}$, if A and B are in $\mathcal{S}$.

Definition 2.1.2. *Given a set M a topology on M is a collection $\mathcal{T}$ of subsets of M, with the following properties*

a) $\emptyset$ and M are in $\mathcal{T}$;

b) If $A_1, ..., A_n$ is a finite collection of sets in $\mathcal{T}$, then

$$\cap_{j=1}^{n} A_j \in \mathcal{T};$$

c) If A_i is a collection of sets in $\mathcal{T}$, with the index i varying in a (finite, countable or uncountable) set I, then

$$\cup_{i \in I} A_i \in \mathcal{T}.$$

The sets in $\mathcal{T}$ are called *open sets*, the complement of an open set is said *closed*, a pair $(M, \mathcal{T})$ is called a *topological space*.

Given any collection $\mathcal{F}$ of subsets of a set M, there always exists a $\sigma-$algebra $\mathcal{S}(\mathcal{F})$ which is the *smallest* $\sigma-$algebra containing $\mathcal{F}$. This $\sigma-$algebra contains $\emptyset$, M, the sets in $\mathcal{F}$, their complements and all the countable unions of these sets. In these cases, we say that $\mathcal{F}$ *generates* (or *is a generator of*) $\mathcal{S}(\mathcal{F})$. In particular, if $\mathcal{F}$ is a topology, the associated $\sigma-$algebra $\mathcal{S}(\mathcal{F})$ is called *Borel $\sigma-$algebra* and the sets in it *Borel sets*. A generator is called a *π-generator* if it is closed under finite intersection. In particular, an algebra is a π-generator.

2.1.2 Measures

Definition 2.1.3. *A positive measure is a function μ which assigns to every element A in a $\sigma-$algebra $\mathcal{S}$ a value in $[0, \infty]$, with the following additional properties*

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

b)

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n),$$

for every countable collection of disjoint measurable sets A_n in $\mathcal{S}$.

A measurable space $(M, \mathcal{S})$, endowed with a measure μ is called a *measure space*. If $\mu(M) = 1$, the space is said to be a *probability space* and the measure a *probability measure*. Every measure space $(M, \mathcal{S}, \mu)$ with finite $\mu(M) < \infty$ can be turned into a probability space just by redefining the measure μ by $\mu' := \frac{\mu}{\mu(M)}$.

As an example of a measure space, consider M a finite set of elements $\{1, 2, ..., n\}$. We can define a $\sigma-$algebra $\mathcal{S}$ as the collection of all the subsets of M, namely $\mathcal{S} := 2^M$. Notice that this is in fact an algebra. A probability measure can be defined by assigning a weight $p_i > 0$ to every element $i \in M$, in such a way that $\sum_{i=1}^{n} p_i = 1$. In a general measure space an element x such that $\mu(x) > 0$ is called an *atom*. Another measure that will be important in the sequel is the *Lebesgue measure* defined in the Euclidean space $\mathbf{R}^n$ (or on one submanifold M of $\mathbf{R}^n$). Its definition involves some technical work but we can think of it simply as the volume in $\mathbf{R}^n$, where the $\sigma-$algebra of measurable sets is the one generated by *rectangles* of the form $R := \Pi_{i=1}^{n}(a_i, b_i)$. The Lebesgue measure in an $n-$dimensional Euclidean space will be sometime denoted by $dx_1 dx_2 ... dx_n$. The definition of Lebesgue measure can be extended to arbitrary submanifolds of $\mathbf{R}^n$ of dimension m. The open sets generating the $\sigma-$ algebra are the inverse images of open sets of $\mathbf{R}^m$ under the parametrization of the manifold and the measure is given by $dx_1 \wedge \cdots \wedge dx_m$.

Given a measurable space $(M, \mathcal{S})$ and two measures defined on it ν and μ, μ is said to be *absolutely continuous* with respect to ν, if, for each $A \in \mathcal{S}$, such that $\nu(A) = 0$, we have $\mu(A) = 0$.

2.1.3 Mappings

A mapping Φ of a topological space $(M, \mathcal{T})$ into a topological space $(N, \mathcal{R})$ is called *continuous* if $\Phi^{-1}(E)$ is an open set in $\mathcal{T}$ whenever E is open in $\mathcal{R}$. It is called an *open* mapping if $\Phi(E)$ is an open set in $\mathcal{R}$ whenever E is open in $\mathcal{T}$.

A mapping Φ of a measurable space $(M, \mathcal{S})$ into a measurable space $(N, \mathcal{A})$ is called *measurable* if $\Phi^{-1}(E)$ is a measurable set in $\mathcal{S}$, whenever E is measurable in $\mathcal{A}$. It is called *bimeasurable* if also $\Phi(E)$ is a measurable set in $\mathcal{A}$ when E is measurable in $\mathcal{S}$.

A mapping Φ of a measurable space $(M, \mathcal{S})$ into a topological space $(N, \mathcal{T})$ is called *measurable* if $\Phi^{-1}(E)$ is a measurable set in $\mathcal{S}$ whenever E is open in $\mathcal{T}$. If, in particular, $(M, \mathcal{S})$ is a Borel measurable space and $\Phi : (M, \mathcal{S}) \to (N, \mathcal{T})$ is a continuous mapping then Φ is also measurable, since all of the open sets in the corresponding topology are in $\mathcal{S}$ by definition, and Φ is said to be *Borel measurable*.

A one to one and onto mapping between two topological spaces which is continuous and open is said to be an *homeomorphism* and the two spaces are said to be *homeomorphic*.

A one to one and onto, bimeasurable mapping between two measurable spaces is said to be an *isomorphism* of measurable spaces and the two spaces are said to be *isomorphic*.

A measurable mapping Φ of a measure space $(M, \mathcal{S}, \mu)$ into a measure space $(N, \mathcal{A}, \nu)$ is said to be *measure preserving* if $\mu(\Phi^{-1}(E)) = \nu(E)$, for each measurable set E in $\mathcal{A}$. If $\mathcal{F}$ is a $\pi-$generator of $\mathcal{S}$ and $\Phi : M \to M$ is a measurable transformation, and $\mu(\Phi^{-1}(A)) = \mu(A)$ for each $A \in \mathcal{F}$, then $\mu(\Phi^{-1}A) = \mu(A)$, for each $A \in \mathcal{S}$.

An isomorphism of a measure space onto itself which is also measure preserving is said to be an *automorphism*. In the context of measure spaces, sets of measure zero do not matter therefore we shall refer to as an automorphism also in the case of mappings that satisfy all the requirements of the definition except on a set of measure zero.

Special type of mappings are *functions* f which map topological or measurable spaces into the set of real $\mathbf{R}$ (or complex $\mathbb{C}$) numbers endowed with the *usual topology*, namely the topology whose open sets are arbitrary unions of open intervals in $\mathbf{R}$ (or in $\mathbb{C}$). For functions the same terminology (continuous, open, measurable, etc) is adopted.

A sequence of measurable functions $f_n : M \to \mathbf{R}$ ($\mathbb{C}$) is said to converge almost everywhere to a measurable function $f : M \to \mathbf{R}$ ($\mathbb{C}$), if for every $x \in M$, except perhaps on a subset of M of measure 0, we have $\lim_{n\to\infty} f_n(x) = f(x)$. f_n is said to converge to f in measure if, for every $\epsilon > 0$, we have $\lim_{n\to\infty} \mu(\{x \in M | |f_n(x) - f(x)| \geq \epsilon\}) = 0$. Convergence in measure implies convergence almost everywhere for a subsequence of $\{f_n\}$. Moreover, if $\mu(M) < \infty$, then convergence almost everywhere implies convergence in measure.

2.1.4 Lebesgue integral

A particularly important case of a function defined on a measure space $(M, \mathcal{S}, \mu)$ is the characteristic function χ_E of a measurable set E, defined by $\chi_E(x) = 1$, if $x \in E$ and $\chi_E(x) = 0$, if $x \notin E$. This function is clearly measurable since the inverse image of every open set is either E, E^c, $\emptyset$ or M, that are all measurable sets. The next simplest example of a measurable function is a function s with values in $[0, \infty)$ which only assumes a finite number n of values $\{\alpha_1, .., \alpha_n\}$. If the sets $A_i := \{x \in M | s(x) = \alpha_i\}$ are measurable, s can be written as a linear combination of characteristic functions χ_{A_i}

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}, \tag{2.1}$$

and therefore it is also measurable. A function s so defined is called a *simple function.* Given a measurable set E in M and a simple function s as in (2.1), the integral of s on E is defined as

$$\int_E s d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E). \tag{2.2}$$

Given a measurable function f, defined on a measure space $(M, \mathcal{S}, \mu)$, with values in $[0, \infty)$, and a measurable set E in M, the integral of f on E is defined as

$$\int_E f d\mu := \sup_s \int_E s d\mu, \tag{2.3}$$

where the *sup* is taken over all of the simple functions s, with $0 \leq s \leq f$. This definition can be generalized to a measurable function f with values in $\mathbf{R}$ by writing f as the sum of its positive and negative part $f = f^+ + f^-$ and defining

$$\int_E f d\mu := \int_E f^+ d\mu - \int_E f^- d\mu. \tag{2.4}$$

Analogously one can extend the definition to complex function by considering real and imaginary parts. The integral of a measurable function as defined in (2.3) is called *Lebesgue integral* and generalizes the standard Riemann integral of elementary calculus with which it shares many important properties such as *linearity* and *monotonicity*. The following are the basic results concerning convergence and Lebesgue integration.

From now on a property will be said to be valid almost everywhere (a.e) if it is valid everywhere except on a set of measure zero.

Theorem 2.1.4. (Lebesgue Monotone Convergence Theorem) *Consider a sequence of measurable functions* $\{f_n\}$ *with*

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots < \infty$$

for a.e. $x \in M$ *and assume that for a.e.* $x \in M$ $\lim_{n\to\infty} f_n(x) = f(x)$. *Then,* f *is measurable and we also have*

$$\lim_{n\to\infty} \int_M f_n d\mu = \int_M f d\mu. \tag{2.5}$$

From now on we will drop the adjective 'measurable' when dealing with functions.

Lemma 2.1.5. (Fatou's Lemma) *Let* $\{f_n\}$ *be a sequence of functions with values in* $[0, \infty)$ *and assume*

$$\liminf_{n\to\infty} \int_M f_n d\mu < \infty. \tag{2.6}$$

Then,

$$\int_M \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_M f_n d\mu. \tag{2.7}$$

In the following we shall call *integrable* a function f such that $|\int_M f d\mu| < \infty$.

Theorem 2.1.6. (Lebesgue Dominated Convergence Theorem) *Let* f_n, $n = 1, 2, \dots$ *a sequence of complex valued functions with*

$$\lim_{n\to\infty} f_n(x) = f(x), \quad a.e. \tag{2.8}$$

Assume there exists a nonnegative valued integrable function $g(x)$ such that

$$|f_n(x)| \leq g(x), \quad n = 1, 2, ..., \tag{2.9}$$

for almost every $x \in M$. Then,

$$\lim_{n\to\infty} \int_M |f_n - f| d\mu = 0. \tag{2.10}$$

From (2.10) and the property of the integral

$$\int_M |f| d\mu \geq |\int_M f d\mu|, \tag{2.11}$$

one also gets (2.5), under the assumptions of Theorem 2.1.6.

If $(M, \mathcal{S}, \mu)$ is a measure space and $\Phi : M \to M$ a one to one and onto measurable map, then Φ naturally carries the measure μ onto a measure $\nu = \mu(\Phi^{-1})$ on $(M, \mathcal{S})$. Then Φ is measure preserving if and only if, for every measurable function f on $(M, \mathcal{S})$,

$$\int_M f d\nu = \int_M f \circ \Phi d\mu. \tag{2.12}$$

In particular, if $M \subseteq \mathbf{R}^n$, μ is the Lebesgue measure $\mu := dx_1 dx_2 \cdots dx_n$ and Φ is differentiable, then we have

$$\int_M f dx_1 dx_2 \cdots dx_n = \int_M f \circ \Phi |det D_x \Phi| dx_1 dx_2 \cdots dx_n, \tag{2.13}$$

where $D_x\Phi$ is the Jacobian of Φ. Φ is measure preserving if and only if $|det D_x \Phi| = 1$.

2.1.5 L^p Spaces

Given a measure space $(M, \mathcal{S}, \mu)$, the space of the complex valued functions defined on M such that

$$\int_M |f|^p d\mu < \infty,$$

for some integer p, with $1 \leq p < \infty$, is denoted by $L^p(M)$. The quantity

$$||f||_p := (\int_M |f|^p d\mu)^{\frac{1}{p}}$$

is called the *p-norm* of f. If there exists a finite B such that

$$|f(x)| \leq B, \quad a.e. \quad x \in M, \tag{2.14}$$

then we say that f is in $L^\infty(M)$ and we denote by $L^\infty(M)$ the space of all the functions with this property. The smallest B such that (2.14) holds is called the ∞-*norm* of f and it is denoted by $||f||_\infty$. If $f, g \in L^p(M)$, with $1 \le p \le \infty$, then one has the following *Minkowski's inequality*

$$||f+g||_p \le ||f||_p + ||g||_p. \tag{2.15}$$

Moreover, if p and q are *conjugate exponents*, namely $\frac{1}{p} + \frac{1}{q} = 1$ (where $\frac{1}{\infty} = 0$) then one has the following *Holder's inequality*

$$||fg||_1 \le ||f||_p ||g||_q. \tag{2.16}$$

Under the assumption that $\mu(M) = 1$ we also have that if $0 < r < s \le \infty$

$$||f||_r \le ||f||_s. \tag{2.17}$$

For $1 \le p \le \infty$ the p–norm is in fact a norm on the elements of $L^p(M)$ if one identifies functions that are different only on sets of measure zero. If f is in $L^p(M)$, and α is a real (or complex) number, αf is also in $L^p(M)$, and if g is another function in $L^p(M)$, inequality (2.15) shows that $f+g$ is also in $L^p(M)$ so that $L^p(M)$ is a *vector space* endowed with a *norm*. Every vector space V with a norm $||\cdot||$ is also a *metric space*, in that one can define a *distance function* (or *metric*) $d(f,g)$ between any two elements f and g as $d(f,g) := ||f-g||$. An *open ball* with center f and radius $r > 0$ is the set of all the g such that $d(f,g) < r$, and a topology on V can be defined as the collection of subsets of V which are arbitrary unions of open balls. This topology is called the topology *induced* by the metric. In general a metric space does not need be a vector space if we can define a distance function on it. A sequence f_n in a metric space is said to be *Cauchy* if, for every $\epsilon > 0$, there exists a $N > 0$ such that $d(f_n, f_m) < \epsilon$, whenever $n, m > N$. A metric space is said to be *complete* if every Cauchy sequence converges to an element of the space. Each one of the space L^p spaces above introduced is complete. Moreover $L^\infty(M)$ is dense in $L^1(M)$, meaning that for each $f \in L^1(M)$, and each $\epsilon > 0$ there exists a $g \in L^\infty(M)$ such that $||f-g||_1 < \epsilon$.

The fact of being a normed vector space which is also complete qualifies L^p as a *Banach space*. A special case happens when $p = 2$. In this case one can define an *inner product* $< \cdot, \cdot >$ between any two elements of $L^2(M)$ as

$$< f, g > := \int_M f\bar{g}d\mu. \tag{2.18}$$

Using (2.16) and the fact that f and g are both in $L^2(M)$, it is easily seen that the value in (2.18) is always finite so that the above definition is well posed. A Banach space with an inner product is an *Hilbert space.*

2.2 Dynamical Systems

We introduce now our main object of study.

Definition 2.2.1. *A (sequential) abstract dynamical system is a quadruple* $(M, \mathcal{S}, \mu, \{\Phi_t\})$*, where* $(M, \mathcal{S}, \mu)$ *is a probability space and* $\{\Phi_t\}$ *is a sequence of automorphisms of* M*, parametrized by* $t \in \mathbb{G}$ *,* $\mathbb{G}$ *being a group.*

The parameter t, in the above definition, is usually interpreted as time, and therefore it varies in $\mathbb{R}$ (*continuous dynamical systems*) or $\mathbb{Z}$ (*discrete dynamical systems*). In the first case, it is usually assumed that Φ_t is the *flow* associated to a differential equation on the compact manifold M

$$\dot{x} = f(x,t) \qquad x(0) = x_0. \tag{2.19}$$

In this case one has for Φ_t the *group property* namely

$$\Phi_{t+s} = \Phi_t \circ \Phi_s \quad \forall s, t \in \mathbb{R}\,. \tag{2.20}$$

In the discrete time case, one usually assumes that Φ_t is the same Φ for each $t \in \mathbb{Z}$ and considers the system to be given by $(M, \mathcal{S}, \mu, \Phi)$. In this work, we will provide some new results for discrete time systems whose evolution law is allowed to change at each time step. We will sometimes call this case *sequential* to emphasize the generalization with respect to the standard case. We will see, through examples, how the structure of the time-independent case is lost in the general case and we will extend some classical results under suitable assumptions.

When dealing with discrete time systems, we will use the following notation: ${}^t\Phi := \Phi_t \circ \cdots \circ \Phi_1$, ${}^0\Phi := Id$. When Φ_t is the same Φ, for every t, then ${}^t\Phi = \Phi^t$. ${}^t\Phi_k := \Phi_t \circ \cdots \Phi_k$, for $t \geq k$.

A discrete time *periodic dynamical systems* is a system of the form $(M, \mathcal{S}, \mu, \{\Phi_t\})$, such that there exists an integer T, called the *period*, with $\Phi_{t+T} = \Phi_t$, $(mod \mu)$, for each $t \geq 1$. This is the simplest case of sequential discrete-time dynamical system and we will see that the basic results of the standard case with single transformation can be extended directly to this case. We will also see in the next chapter that the ergodic properties of periodic sequences can be studied in terms of the composite transformation that characterizes it namely, if T is the period, $\bar{\Phi} := \Phi_T \circ \cdots \Phi_1$.

2.2.1 Variational equations and Liouville's Theorem

The Jacobian of the flow $D_x\Phi_t$ of a a system of n differential equations such as (2.19) satisfies a matrix linear time varying differential equation, the so called *variational equation*:

$$\frac{d}{dt}D_x\Phi_t = D_xf(x(t),t)D_x\Phi_t, \qquad D_x\Phi_0 = I_{n\times n}, \tag{2.21}$$

where the matrix of coefficients on the right hand side is the Jacobian of the vector field f in (2.19), with $x(t)$ varying along the trajectory solution of (2.19), namely $x(t) = \Phi_t(x_0)$. In order to show this, recall that by the definition of flow, for each $x \in M$, we have

$$\frac{d}{dt}\Phi_t(x) = f(\Phi_t(x),t). \tag{2.22}$$

Differentiating this with respect to x, we obtain

$$D_x\frac{d}{dt}\Phi_t(x) = D_x(f(\Phi_t(x),t)); \tag{2.23}$$

Interchanging the order of differentiation in the left hand side and using the chain rule in the right hand side of (2.23), we obtain the differential equation in (2.21).

Concerning, matrix linear time varying differential equation of the form

$$\dot{X}(t) = A(t)X(t), \qquad X(0) = X_0, \tag{2.24}$$

the following result, referred to as *Liouville's Lemma*, is often useful.

Lemma 2.2.2. (Liouville's Lemma) *Consider a linear time varying differential equation (2.24), with $A(t)$ continuous, then we have*

$$detX(t) = detX(t_0)e^{\int_0^t Tr(A(s))ds}, \tag{2.25}$$

where $Tr(A(s))$ denotes the trace of the matrix $A(s)$.

Proof. If $X(t)$ is not a fundamental solution of (2.24), then $det(X(t)) = 0$, for each $t \geq 0$ and (2.25) is verified. If it is a fundamental solution, then $det(X(t)) \neq 0$, for each $t \geq 0$, and we can define the matrix $B(t) := X^{-1}(t)A(t)X(t)$, with $A(t)X(t) = X(t)B(t)$ and $Tr(B(t)) = Tr(A(t))$. Define $z(t) := detX(t)$. We have

$$\dot{z} = \sum_{j=1}^{n} det[x_1(t)|\cdots|x_{j-1}(t)|\dot{x}_j(t)|x_{j+1}(t)|\cdots|x_n(t)]$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} det[x_1(t)|\cdots|x_{j-1}(t)|A(t)x_j(t)|x_{j+1}(t)|\cdots|x_n(t)] \\
&= \sum_{j=1}^{n} det[x_1(t)|\cdots|x_{j-1}(t)|X(t)b_j(t)|x_{j+1}(t)|\cdots|x_n(t)] \\
&= \sum_{j=1}^{n} det[x_1(t)|\cdots|x_{j-1}(t)|b_{jj}(t)x_j|x_{j+1}(t)|\cdots|x_n(t)] \\
&= (\sum_{j=1}^{n} b_{jj}(t))det(X(t)) \\
&= (Tr(A(t)))z(t).
\end{aligned} \tag{2.26}
$$

From (2.26), (2.25) follows immediately. □

Liouville's Lemma along with the variational equation (2.21) can be used to determine the property of a flow associated to a differential equation (2.19) to preserve the Lebesgue measure. To this aim, recall that, given a (time varying) vector field $f(x,t)$, its divergence is defined as the scalar quantity

$$
div(f(x,t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x,t), \tag{2.27}
$$

and notice that this quantity, computed along the solution of (2.19), is actually the trace of the coefficient matrix in the variational equation (2.21). Applying Liouville's Lemma to the variational equation, one concludes that, in the case $div(f(x,t)) = 0$, the determinant of the Jacobian of the flow is constant and equal to one for each $t \geq 0$, and therefore, the flow, in this case, preserves the Lebesgue measure. In conclusion, we have the following Liouville's theorem.

Theorem 2.2.3. *The flow associated to a divergence free vector field, preserves the Lebesgue measure in* $\mathbf{R}^n$.

The importance of Liouville's Theorem in dynamical systems theory, is usually associated to Hamiltonian dynamics, where the evolution of the system is described by equations of the type

$$
\begin{aligned}
\dot{q} &= \frac{\partial H(p,q,t)}{\partial p}, \\
\dot{p} &= -\frac{\partial H(p,q,t)}{\partial q},
\end{aligned} \tag{2.28}
$$

for $p \in \mathbf{R}^m$, $q \in \mathbf{R}^m$, and a suitable *Hamiltonian function* $H(p,q,t)$. In this case the vector field is divergence free and Liouville's Theorem applies.

2.2.2 Examples of dynamical systems

Example 2.2.4 (Examples in Probability)

Given a probability space $(M, \mathcal{S}, \mu)$, a *random variable* is a measurable function ω from M to the set of real numbers $\mathbf{R}$ endowed with the Borel σ−algebra $\mathcal{B}$ generated by the usual topology. The space $(M, \mathcal{S}, \mu)$ is also called the *space of events* and ω represents a *coding* of the events in $\mathcal{S}$ into the elements of $\mathbf{R}$. For example, in the experiment of tossing a coin, we can assign a 0 to the event of having a 'Head' and a 1 to a 'Tail'. A random variable ω also induces a probability measure ν on $(\mathbf{R}, \mathcal{B})$ in the natural way: $\nu(E) = \mu(\omega^{-1}(E))$, where E is a measurable set of $\mathcal{B}$.

In order to introduce the concept of *stochastic process*, we have to proceed to the construction of a measurable space as the product of a countable number of measurable spaces. Recall, we denote by $\mathcal{B}$ the Borel σ−algebra associated with the usual topology of $\mathbf{R}$.

Consider $\mathbf{R}^{\mathbf{Z}}$ which is the bi-infinite product of copies of $\mathbf{R}$ or, equivalently, the set of all the bilateral sequences $\Omega := \{..\omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, ...\}$ with real values. Consider the subsets of $\mathbf{R}^{\mathbf{Z}}$ of the form

$$C := \{\Omega \in \mathbf{R}^{\mathbf{Z}} \, | \omega_j \in A\}, \tag{2.29}$$

for some $j \in \mathbf{Z}$ and some $A \in \mathcal{B}$. We call these sets *elementary cylinders*. Elementary cylinders generate a σ−algebra on $\mathbf{R}^{\mathbf{Z}}$ that we denote by $\mathcal{R}^{\mathbf{Z}}$. More complicated cylinders belong to this σ−algebra, which have the form

$$C := \{\Omega \in \mathbf{R}^{\mathbf{Z}} \, | \omega_{j_1} \in A_1, \omega_{j_2} \in A_2, ..., \omega_{j_k} \in A_k\}. \tag{2.30}$$

The set C in (2.30) is the intersection of k elementary cylinders. The cylinders (2.30) generate $\mathcal{R}^{\mathbf{Z}}$ and actually they are a π−generator since the intersection of a finite number of them is still a cylinder of the form (2.30).

Given a probability space of events $(M, \mathcal{S}, \mu)$, a stochastic process is defined by a measurable mapping $T : (M, \mathcal{S}) \to (\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}})$. An event $x \in M$ is mapped in the bi-infinite sequence $T(x) := (..., \omega_{-2}(x), \omega_{-1}(x), \omega_0(x), \omega_1(x), \omega_2(x), ...)$. The functions ω_j, $j = 0, \pm 1, \pm 2, \pm 3, ..$ are such that, for any measurable set $A_j \in \mathcal{B}$,

$$\omega_j^{-1}(A_j) = T^{-1}\{\Omega \in \mathbf{R}^{\mathbf{Z}} \, | \omega_j \in A_j\}, \tag{2.31}$$

is also measurable, since it is the inverse image under T of an elementary cylinder, and T is measurable by definition. The functions ω_j are therefore random variables.

The measure μ in $(M, \mathcal{S}, \mu)$ naturally induces a measure ν, on $(\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}})$ given by $\nu(E) = \mu(T^{-1}(E))$, for each measurable set E in $\mathcal{R}^{\mathbf{Z}}$. ν is also called the *probability distribution* of $(\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}})$.

A *stochastic process* $(\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}}, \nu)$ is said to be *stationary* if the measure of every cylinder is invariant under shift. More specifically, define the (left) shift transformation Sh, on the sequence Ω of $\mathbf{R}^{\mathbf{Z}}$, by

$$(Sh(\Omega))_n = \omega_{n+1}, \quad \forall n \in \mathbf{Z}. \tag{2.32}$$

The action of the shift on a cylinder (2.30) is described by

$$\begin{aligned} Sh(C) : &= Sh(\{\Omega \in \mathrm{R}^{\mathbf{Z}} | \omega_{j_1+1} \in A_1, \omega_{j_2+1} \in A_2, ..., \omega_{j_k+1} \in A_k\}) \\ &= \{\Omega \in \mathbf{R}^{\mathbf{Z}} | \omega_{j_1} \in A_1, \omega_{j_2} \in A_2, ..., \omega_{j_k} \in A_k\}. \end{aligned} \tag{2.33}$$

A stochastic process is stationary if $\nu(Sh(C)) = \nu(C)$ and by induction $\nu(Sh^l(C)) = \nu(C)$, for each $l \geq 0$. In other terms, the joint probability distribution ν of a $k-$uple of random variables only depends on the reciprocal distance of the random variables in the sequence and not on their position in the sequence itself. Recalling that the class of cylinders is a $\pi-$generator of the $\sigma-$algebra $\mathcal{R}^{\mathbf{Z}}$ we conclude that Sh is a measure preserving transformation for $(\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}}, \nu)$ (see Subsection 2.1.1).

In conclusion, a stationary stochastic process is defined as a dynamical system characterized by a space of random sequences, the $\sigma-$ algebra generated by cylinders and a shift-invariant probability measure, otherwise arbitrary. The measure preserving transformation is given by the shift so that the dynamical system is written as $(\mathbf{R}^{\mathbf{Z}}, \mathcal{R}^{\mathbf{Z}}, \nu, Sh)$. This is one of the main objects of study in Probability Theory.

Example 2.2.5 (Examples in Statistical Mechanics)

The most important system of interest in statistical mechanics consists of N (very large) noninteracting particles in a volume V. The state of the system is specified by $3N$ *position* coordinates $q := q_1, q_2, ..., q_{3N}$ and $3N$ *momentum* coordinates $p := p_1, p_2, ..., p_{3N}$. The coordinates (q, p) specify a point in the *phase space* which is the set M on which the system evolves. The evolution of the system is described by

$$\dot{q}_i = \frac{\partial H(q,p)}{\partial p_i} \quad i = 1, ..., 3N \tag{2.34}$$

$$\dot{p}_i = -\frac{\partial H(q,p)}{\partial q_i} \quad i = 1, ..., 3N \tag{2.35}$$

where $H := H(p, q)$ is the Hamiltonian function of the motion which gives the total energy in terms of the coordinates p and q. It is easily verified, from equations (2.34), (2.35), that the system is divergence free so that, from Liouville's Theorem 2.2.3, we have that the flow Φ_t preserves the Lebesgue measure $\lambda := dq_1 dq_2 \cdots dq_{3N} dp_1 dp_2 \cdots dp_{3N}$. We denote the $\sigma-$ algebra underlying the Lebesgue measure λ by $\mathcal{S}$ and the dynamical system under consideration is given by $\{M, \mathcal{S}, \lambda, \{\Phi_t\}\}$.

Example 2.2.6 (Examples in Information Theory)

The output of a transmission channel can be modeled by a bi-infinite sequence in $\mathring{A}^{\mathbb{Z}}$, where $\mathring{A} := \{1, ..., n\}$ is the set of n possible symbols called the *alphabet.* Every symbol has an a priori strictly positive probability of being transmitted p_i, with $\sum_{i=1}^{n} p_i = 1$. From these probabilities one can construct a measure ν on the $\sigma-$algebra $\mathcal{S}$ generated by cylinders of the form

$$C = \{\omega \in \mathring{A}^{\mathbb{Z}} \, | \omega_{i_1} = i_1, ..., \omega_{i_k} = i_k\}. \tag{2.36}$$

If the measure ν on cylinders is invariant under the left shift Sh, then the quadruple $\{\mathring{A}^{\mathbb{Z}}, \mathcal{S}, \nu, Sh\}$ define a dynamical system in the sense of the Definition 2.2.1 (compare to the discussion for Example 2.2.4).

2.2.3 Koopman operators in Hilbert spaces

Every measure preserving transformation Φ, on a measure space $(M, \mathcal{S}, \mu)$, determines a linear operator $U : L^2(M) \to L^2(M)$. In particular, if $f \in L^2(M)$, we have $Uf := f \circ \Phi$. Such an operator is said *Koopman operator.* A Koopman operator is an *isometry* in $L^2(M)$, since, for every $f \in L^2(M)$, we have

$$||Uf||_2^2 := \int_M |Uf|^2 d\mu := \int_M |f \circ \Phi|^2 d\mu = \int_M |f|^2 d\mu, \tag{2.37}$$

where we have used (2.12), with $\nu := \mu(\Phi^{-1}) = \mu$, since Φ is measure preserving. A special case is when the measure preserving transformation Φ_t comes from a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$. This simple identification makes it possible to study the properties of a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, in terms of the associated operators U_t. In this case, being Φ_t an automorphism, U_t is also invertible and we have $U_t^{-1} f = f \circ \Phi_t^{-1}$. Recall that a Hilbert space is, by definition, endowed with an inner product and the adjoint U^* of an operator U on it, is such that $< f, Ug >=< U^* f, g >$ for each f and g. In the case of invertible Koopman operators the adjoint is equal to the inverse.

2.2.4 Distance between dynamical systems

Consider two discrete time (or continuous time) dynamical systems acting on the same space M, with the same $\sigma-$algebra $\mathcal{S}$, and the same measure μ, namely

$(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ and $(M, \mathcal{S}, \mu, \{\Phi_t^2\})$. If for each $t \in \mathbf{Z}$ $(t \in \mathbf{R})$ Φ_t^1 and Φ_t^2 are the same except on a set of measure μ zero, the two systems from a measure theoretic point of view are the same and we can say that their *distance* is zero. In order to define a distance between dynamical systems, it is convenient to identify in classes of equivalence automorphisms that differ only on sets of measure zero and to deal only with classes of equivalence.

The group $\mathcal{F}$ of equivalence classes of invertible measure preserving transformations on M which differ on a set of measure zero can be turned into a metric space by the introduction of the metric

$$\rho_1(\Phi_1, \Phi_2) = \mu\{x \in M | \Phi_1 x \neq \Phi_2 x\}. \tag{2.38}$$

The metric (2.38) is equivalent to

$$\rho_2(\Phi_1, \Phi_2) = \sup\{\mu(\Phi_1 A \Delta \Phi_2 A) | A \in \mathcal{S}\}, \tag{2.39}$$

and to

$$\begin{aligned}\rho_3(\Phi_1, \Phi_2) = \; & sup\{\frac{1}{||f||_2}||f \circ \Phi^1 - f \circ \Phi^2||_2 | f \in L^2(M)\} \qquad (2.40)\\ = \; & \sup\{\frac{1}{||f||_2}||U_1 f - U_2 f||_2 | f \in L^2(M)\},\end{aligned}$$

where U_1 and U_2 are the Koopman operators associated to Φ_1 and Φ_2. The equivalence has to be intended as that convergence in the topologies induced by ρ_1, ρ_2 and ρ_3 are equivalent properties. In the following, we shall sometimes consider pairs of sequences of transformations $\{\Phi_t^1\}$, $\{\Phi_t^2\}$ and denote by

$$W_t := \{x \in M | \Phi_t^1 x \neq \Phi_t^2 x\}. \tag{2.41}$$

According to the above defined metric ρ_1 (ρ_2, ρ_3), we will say that Φ_t^1 tends to Φ_t^2, if $\lim_{t\to\infty} \mu(W_t) = 0$. This topology on $\mathcal{F}$ is called *strong topology*, and the metric ρ_1 is called *Rohlin metric.*

The *weak topology* on $\mathcal{F}$ is defined in the following way: $\{\Phi_t^2\}$ tends to $\{\Phi_t^1\}$ for $t \to \infty$ if for each $A \in \mathcal{S}$,

$$\lim_{t\to\infty} \mu(\Phi_t^2 A \Delta \Phi_t^1 A) = 0, \tag{2.42}$$

or, equivalently, for each $f \in L^2(M)$

$$\lim_{t\to\infty} \frac{1}{||f||_2}||f \circ \Phi_t^1 - f \circ \Phi_t^2||_2 = \lim_{t\to\infty} \frac{1}{||f||_2}||U_1 f - U_2 f||_2 = 0. \tag{2.43}$$

Clearly, convergence in the strong topology implies convergence in the weak topology.

2.3 Notes and References

There are several excellent introductory books to measure theory and abstract integration such as [38] and [6]. A text in dynamical systems theory with a more advanced and abstract point of view is [17]. The abstract definition of dynamical systems presented here follows [3]. The book [2] contains a derivation of Liouville's theorem using symplectic geometry and integration on manifolds that would complement the one given in this chapter. Standard texts in probability theory and stochastic processes are [19] and [31] and a good text in statistical mechanics at the first year graduate level is [35]. The book [5] is a classic text to look at information theory from a measure theoretic point of view such as the one presented here with particular attention to ergodic theory which will be introduced in next chapter. The strong and weak metric is discussed in detail in [37] and some further properties are proven. These metrics were also discussed in [14] [36] in conjunction with the problem of establishing 'genericity' for some ergodic properties of dynamical systems.

Chapter 3

Ergodic Theory

This chapter is an introduction to ergodic theory. We will introduce the basic concepts that will be used in the following chapters to formulate and solve control problems involving mixing of fluids.

Ergodic theory is the study of long term average behaviour of dynamical systems. In many problems of practical interest, we are not interested in the specific trajectories described by a system under the action of a transformation but only a global description of the trajectories has to be given. An example is the study of mixing, as treated in this work, where the basic question is whether the trajectories of the system are such that points initially belonging to prescribed sets get homogeneously distributed. In this type of study a *global* or *average* point of view is required.

Basic topics in ergodic theory include questions of *existence of averages* computed along trajectories dealt with in the Ergodic Theorems (Section 1) and the study of recurrence properties such as *ergodicity* and *mixing* dealt with in Section 2. The study of *entropy* a quantity which measures the randomness introduced by a transformation is introduced to Section 3.

Compared to other introductions to ergodic theory we will take here a more general *time varying* point of view. In fact, we will introduce new results concerning the nonstandard case where the transformation acting on the measure space, in the definition of the dynamical system, possibly changes at each time step.

3.1 Ergodic Theorems

The ergodic theorems deal with questions of convergence of averages of functions computed along trajectories of dynamical systems. More specifically, consider a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ and a function f defined on M. The *average of f along n steps of the trajectory starting at x* is defined as

$$A_n(x) = \frac{1}{n}\sum_{k=0}^{n} f({}^k\Phi x), \tag{3.1}$$

in the discrete time case, and as

$$A_T(x) = \frac{1}{T}\int_0^T f(\Phi_t x), \tag{3.2}$$

in the continuous time case. A_n and A_T are also called the *Ergodic Averages.* The question in the ergodic theorems consists of checking if, for a fixed $x \in M$, the limit

$$\bar{f}(x) := \lim_{n\to\infty} A_n(x), \tag{3.3}$$

in the discrete time case, or

$$\bar{f}(x) := \lim_{T\to\infty} A_T(x), \tag{3.4}$$

in the continuous time case, exist. It is in fact not always true that the above limit exists, for a given x, and it may exist only in the mean, namely it may exist a function $\bar{f}$ such that $\lim_{n/T\to\infty}\int_M |A_{n/T} - \bar{f}|d\mu = 0$. The Pointwise Ergodic Theorem and the Mean Ergodic Theorem answer these questions in the continuous time case and in the discrete time case with measure preserving transformation, for every $t \in \mathbb{Z}$. Extensions can be given to the general case and will be discussed below.

We will discuss and present proof for the discrete time case in the following. The standard results, namely the ones concerning a constant automorphism Φ have a natural counterpart in the continuous time case when $\{\Phi_t\}$ is the flow associated to a differential equation.

Physically the question of the ergodic theorems arises when one tries to measure macroscopic quantities, for example temperature, of a system in evolution. Consider, as an illustration, Example 2.2.5. Every point in the phase space represents a state of the system characterized by specific values for position and momentum variables. If one needs to measure the temperature, one can average the temperature along a very long period of time, following the evolution of the system. Is this a plausible measurement procedure? In other terms, do the average values converge to a given value? Would the measurement be independent of the initial state of the system? The answer to the latter question is provided by the concept of Ergodicity, discussed in the next section. The answer to the first one by the ergodic theorems.

3.1.1 The Mean Ergodic Theorem

Theorem 3.1.1. (von Neumann Mean Ergodic Theorem) *Consider a discrete time dynamical system of the form* $(M, \mathcal{S}, \mu, \Phi)$, *a function* f *in* $L^2(M)$ *and the associated ergodic averages given by*

$$A_n(x) := \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x). \tag{3.5}$$

There exists a function $\bar{f} \in L^2(M)$, *such that*

$$\lim_{n\to\infty}\int_M |A_n - \bar{f}|^2 d\mu := \lim_{n\to\infty} ||A_n - \bar{f}||_2^2 = 0. \tag{3.6}$$

Moreover $\bar{f}$ *is invariant under* Φ, *namely* $\bar{f} = \bar{f} \circ \Phi$.

Proof. We denote by $\mathcal{N}$ the closure of the linear span of $\{g - g\circ\Phi | g \in L^2(M)\}$ and by $\mathcal{M}$ the set of all functions in $L^2(M)$ invariant under Φ. Recall, from elementary Hilbert spaces theory, that given a closed subspace $\mathcal{N}$, the Hilbert space $\mathcal{H}$ can be written as the direct sum of $\mathcal{N}$ and $\mathcal{N}^\perp$, the orthogonal complement of $\mathcal{N}$, namely the vector subspace $\{g \in \mathcal{H} | < g, h >= 0, \forall h \in \mathcal{N}\}$. The first step of the proof consists of showing that with the above definitions $\mathcal{M} = \mathcal{N}^\perp$.

We first show that $\mathcal{N}^\perp \subseteq M$. To this aim, for each $h \in \mathcal{N}^\perp$, we need to show that $h \circ \Phi = h$. For each $g \in L^2(M)$, we have

$$0 =< h, g - g\circ\Phi >=< h, g > - < h, g\circ\Phi >=< h, g > - < h\circ\Phi^{-1}, g >=< h - h\circ\Phi^{-1}, g >, \tag{3.7}$$

and since this has to be true for each $g \in L^2(M)$, we have $h = h \circ \Phi^{-1}$. Applying Φ on both sides of this equation, we get $h \in \mathcal{M}$. Similarly, to show that $\mathcal{M} \subseteq \mathcal{N}^\perp$, we have that if $h \in \mathcal{M}$, $h \circ \Phi^{-1} = h$, and, for each $g \in L^2(M)$

$$< h, g - g \circ \Phi >=< h, g > - < h, g \circ \Phi >=< h - h \circ \Phi^{-1}, g >= 0. \tag{3.8}$$

The next step of the proof is to show that, if $f \in \mathcal{N}$, then $\frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k)$ converges to zero in $L^2(M)$. If $f = g - g \circ \Phi$, we have

$$||\frac{1}{n}\sum_{k=0}^{n-1}(g - g\circ\Phi)(\Phi^k)||_2 = ||\frac{1}{n}(g - g\circ\Phi^n)||_2 \le \frac{2}{n}||g||_2, \tag{3.9}$$

where we used Minkowski's inequality (2.15) and the fact that Φ is measure preserving. Letting n go to infinity in (3.9), we obtain the claim, in this case. If $f \in \mathcal{N}$ is the limit of a sequence of functions $f_i = g_i - g_i \circ \Phi$, then we have

$$||\frac{1}{n}\sum_{k=0}^{n-1} f \circ \Phi^k||_2 \le ||\frac{1}{n}\sum_{k=0}^{n-1}(f - f_i)\circ\Phi^k||_2 + ||\frac{1}{n}\sum_{k=0}^{n-1} f_i \circ \Phi^k||_2, \tag{3.10}$$

and we can choose i and n to make the first and the second term, respectively, arbitrarily small.

In conclusion, we can write every $f \in L^2(M)$ as $f = f_0 + \bar{f}$, with $f_0 \in \mathcal{N}$ and $\bar{f} \in \mathcal{M}$, namely $\bar{f}$ invariant under Φ. We have

$$||\frac{1}{n}(\sum_{k=0}^{n-1}(f \circ \Phi^k) - \bar{f})|| = ||\frac{1}{n}\sum_{k=0}^{n-1}(f - \bar{f}) \circ \Phi^k|| = ||\frac{1}{n}\sum_{k=0}^{n-1} f_0 \circ^k \Phi||, \tag{3.11}$$

which tends to zero as n tends to infinity since $f_0 \in \mathcal{N}$, for what proved before. □

Remark 3.1.2. The above theorem can be generalized in several directions. In particular the theorem is a special case of a more general result about contraction operators in Hilbert spaces. Recall that a contraction operator U, on a Hilbert space $\mathcal{H}$, is an operator such that $||Uf|| \leq ||f||$ for each $f \in \mathcal{H}$. In particular, we have that, in this case, the average $\frac{1}{n}\sum_{k=0}^{n-1} U^k f$ converges, in the norm of the space, when n goes to infinity, to an element $\bar{f}$ of $\mathcal{M} := \{f \in \mathcal{H} : Uf = f\}$. The above result is recovered replacing the operator U by the Koopman operator induced by the dynamical system under consideration. The proof is essentially the same, with a slight modification of the first step, due to the fact that a contraction operator U need not be invertible.

In the following theorem, we will provide a generalization of the Mean Ergodic Theorem to sequences of measure preserving transformations, namely to a time varying discrete time system of the form $(M, \mathcal{S}, \mu, \{\Phi_t\})$, under the following additional assumptions:

$H1$: *The sequence of automorphisms Φ_t converges weakly to a single transformation* Φ (see definitions in Section 2.2.4)

$H2$ *The limit automorphism Φ defined in $H1$ is ergodic in the sense that each function f on M, invariant with respect to Φ, namely $f \circ \Phi = f$, is constant almost everywhere.*

The convergence of the ergodic averages for discrete time, time varying dynamical systems, under convergence assumptions, will be also investigated in the next subsection, when we will deal with pointwise convergence in the Pointwise Ergodic Theorem. In that case, we will require *strong convergence* of a system to another to conclude convergence of the ergodic averages. The concept of ergodicity will be dealt with in detail in the next section.

Theorem 3.1.3. (Mean Ergodic Theorem for converging sequences of transformations) *Consider a sequence of measure preserving transformations converging to a* *single ergodic transformation Φ in the weak topology defined in (2.42). Then, for every function $f \in L^2(M)$, the ergodic average*

$$\frac{1}{n}\sum_{k=0}^{n} f({}^k\Phi x) \tag{3.12}$$

converges in $L^2(M)$ to a function $\bar{f}$ constant almost everywhere.

Proof. The proof follows the same steps as the proof of the Mean Ergodic Theorem 3.1.1. In particular, one first proves, in exactly the same way, that the closure of $\mathcal{N} := \{g - g \circ \Phi | g \in L^2(M)\}$ and $\mathcal{M} := \{f \in L^2(M) | f = f \circ \Phi\}$, that in this case corresponds to set of functions which are constant almost everywhere, are orthogonal complementary subspaces in $L^2(M)$.

Then one proves that if $f \in \mathcal{N}$, then $\frac{1}{n}\sum_{k=0}^{n-1} f({}^k\Phi)$ converges to zero in $L^2(M)$. If $f = g - g \circ \Phi$, we have

$$\begin{aligned}
&||\tfrac{1}{n}\textstyle\sum_{k=0}^{n-1}(g - g\circ\Phi)({}^k\Phi)||_2 \\
&= \tfrac{1}{n}||g - g(\Phi \circ^{n-1} \Phi) + \textstyle\sum_{k=1}^{n-1} g \circ^k \Phi - \sum_{k=0}^{n-2} g \circ \Phi \circ^k \Phi||_2 \\
&\leq \tfrac{1}{n}||g - g \circ \Phi \circ^{n-1} \Phi||_2 + \tfrac{1}{n}||\textstyle\sum_{k=1}^{n-1} g \circ \Phi_k \circ^{k-1} \Phi - g \circ \Phi \circ^{k-1} \Phi||_2 \\
&\leq \tfrac{1}{n}||g - g \circ \Phi \circ^{n-1} \Phi||_2 + \tfrac{1}{n}\textstyle\sum_{k=1}^{n-1} ||g \circ \Phi_k - g \circ \Phi||_2.
\end{aligned} \tag{3.13}$$

In the right hand side of the above expression, both the first and the second term go to zero when n goes to infinity. In particular, for the second term, we used the convergence assumption. If $f \in \mathcal{N}$, and f is the limit of a sequence of functions $f_i = g_i - g_i \circ \Phi$, then we have

$$||\frac{1}{n}\sum_{k=0}^{n-1} f \circ^k \Phi||_2 \leq ||\frac{1}{n}\sum_{k=0}^{n-1}(f - f_i) \circ^k \Phi||_2 + ||\frac{1}{n}\sum_{k=0}^{n-1} f_i \circ^k \Phi||_2, \tag{3.14}$$

and we can choose i and n to make the first and the second term, respectively, arbitrarily small.

In conclusion, we can write every $f \in L^2(M)$ as $f = f_0 + \bar{f}$, with $f_0 \in \mathcal{N}$ and $\bar{f} \in \mathcal{M}$. Notice that $\bar{f}$ is constant a.e., therefore $\bar{f} \circ^k \Phi = \bar{f}$ a.e., for each $k \geq 0$, and we have

$$||\frac{1}{n}\sum_{k=0}^{n-1}(f - \bar{f}) \circ^k \Phi||_2 = ||\frac{1}{n}\sum_{k=0}^{n-1} f_0 \circ^k \Phi||_2, \tag{3.15}$$

which goes to zero as n goes to infinity. □

Notice that applying (2.17) to $A_n - \bar{f}$ in Theorems 3.1.1 and 3.1.3, we have that

$$||A_n - \bar{f}||_2 \geq ||A_n - \bar{f}||_1, \tag{3.16}$$

and letting n go to infinity, we have that convergence in $L^2(M)$ implies convergence in $L^1(M)$. Using (2.11) we obtain

$$\begin{aligned}\int_M \left|\frac{1}{n}\sum_{k=0}^{n-1} f^k\Phi - \bar{f}\right| d\mu &\geq \left|\int_M \frac{1}{n}\sum_{k=0}^{n-1} f^k\Phi - \bar{f} d\mu\right| = \\ \left|\int_M \frac{1}{n}\sum_{k=0}^{n-1} f^k\Phi d\mu - \int_M \bar{f} d\mu\right| &= \left|\int_M f d\mu - \int_M \bar{f} d\mu\right|,\end{aligned} \tag{3.17}$$

and letting n go to infinity in the above expression, and using the convergence in $L^1(M)$, we obtain

$$\int_M f d\mu = \int_M \bar{f} d\mu. \tag{3.18}$$

In particular, under the assumptions of Theorem 3.1.3, the constant $\bar{f}$ is equal to the space average of the function f.

3.1.2 The Pointwise Ergodic Theorem

In this subsection, we will deal with pointwise convergence of the ergodic averages (3.1). In other terms, fixed x in the state space M, we will investigate the convergence of the sequence (3.1). It is one of the main results in ergodic theory that this limit actually exists for every $f \in L^1(M)$, for almost every $x \in M$, and $f \in L^1(M)$, in the case of systems described by a single measure preserving transformation, namely $(M, \mathcal{S}, \mu, \Phi)$. This is the content of the celebrated Birkhoff Pointwise Ergodic Theorem. In physical terms, this theorem gives a justification to the procedure of measuring a quantity in a system in evolution by averaging the measures over a very long interval of time. The ergodic theorem does not hold, in general, if a single transformation is replaced by a sequence of transformations, not even if these transformations converge in the strong sense of (2.38).

This subsection is devoted to a proof of the Pointwise Ergodic Theorem as well as to a discussion of the theorem for sequences, with examples and sufficient conditions for the convergence of the ergodic averages in terms of distance between dynamical systems. Among the many proofs of the Pointwise Ergodic theorem that have been proposed so far, we present the one based on the Maximal Ergodic Theorem, since the latter result is of interest by itself.

Theorem 3.1.4. (Maximal Ergodic Theorem) *Consider a dynamical system* $(M, \mathcal{S}, \mu, \Phi)$*, and assume* $f \in L^1(M)$*. Consider the associated ergodic averages* $A_n(x) := \frac{1}{n}\sum_{k=0}^{n-1} f\,(\Phi^k x)$*, the function*

$$A^*(x) = \sup_{n\geq 1} A_n(x), \tag{3.19}$$

and the set $\{A^* > 0\} := \{x \in M | A^*(x) > 0\}$ *Then*

$$\int_{\{A^*>0\}} f d\mu \geq 0. \tag{3.20}$$

The function A^*defined in (3.19) is sometimes called *Maximal Ergodic function.*

Proof. First define the functions

$$S_n := \sup_{1 \leq j \leq n} \sum_{k=0}^{j-1} f \circ \Phi^k, \tag{3.21}$$

and notice that

$$f = S_1 \leq S_2 \leq \cdots, \tag{3.22}$$

and that

$$\{A^* > 0\} = \cup_{n=1}^{\infty} (\{x \in M | S_n(x) > 0\} := \{S_n > 0\}). \tag{3.23}$$

In the first step of the proof, we prove the following inequality:

$$S_n \leq f + S_n^+ \circ \Phi, \tag{3.24}$$

$\forall n \geq 1$, where $S_n^+ := max\{0, S_n\}$. In order to do this, it is sufficient from (3.22) to prove that $S_{n+1} \leq f + S_n^+ \circ \Phi$. We have, for $1 < j \leq n+1$,

$$\sum_{k=0}^{j-1} f\Phi^k = f + (\sum_{k=0}^{j-2} f \circ \Phi^k) \circ \Phi \leq f + S_{j-1} \circ \Phi \leq f + S_n \circ \Phi \leq f + S_n^+ \circ \Phi, \tag{3.25}$$

and for $j = 1$

$$S_1 = f \leq f + S_n^+ \circ \Phi; \tag{3.26}$$

Using (3.26) and (3.25) in the definition (3.21) we get (3.24).

In the second step of the proof, we use inequality (3.24) to prove (3.20). We have,

$$\int_{\{S_n>0\}} f d\mu \geq \int_{\{S_n>0\}} S_n d\mu - \int_{\{S_n>0\}} S_n^+ \circ \Phi d\mu =$$

$$\int S_n^+ d\mu - \int_{\{S_n>0\}} S_n^+ \circ \Phi d\mu \geq \int_M S_n^+ d\mu - \int_M S_n^+ \circ \Phi d\mu = 0. \tag{3.27}$$

The claim follows observing that

$$\int_{A^*>0} f d\mu = \int_M (\lim_{n \to \infty} f \chi_{S_n}) d\mu, \tag{3.28}$$

where χ_{S_n} is the characteristic function of the set S_n and applying the Lebesgue Monotone Convergence Theorem 2.1.4, to the sequence of functions $f\chi_{S_n}$, along with (3.20). □

Theorem 3.1.5. (Birkhoff Pointwise Ergodic Theorem) *Consider a dynamical system of the form $(M, \mathcal{S}, \mu, \Phi)$, and a function $f \in L^1(M)$. Then, for almost every $x \in M$, the following limit exists*

$$\lim_{n\to\infty} A_n(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x) \tag{3.29}$$

Proof. The statemnet of the theorem is equivalent to the fact that the set

$$E = \{x \in M | \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x) < \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x)\} \tag{3.30}$$

has measure zero. Notice that the set E is the countable union over rationals α and β of sets of the form

$$E_{\alpha\beta} = \{x \in M | \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x) < \alpha < \beta < \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k x)\}, \tag{3.31}$$

and the claim of the theorem is proven if we prove that $E_{\alpha,\beta}$ has measure zero for each α and β. It is easily proven that

$$\liminf_{n\to\infty} A_n(x) = \liminf_{n\to\infty} A_n(\Phi x), \tag{3.32}$$

and

$$\limsup_{n\to\infty} A_n(x) = \limsup_{n\to\infty} A_n(\Phi x), \tag{3.33}$$

so that the set $E_{\alpha,\beta}$ is invariant under Φ. Moreover, if $x \in E_{\alpha,\beta}$, then x is such that $A^*(x) \geq \beta$, namely $E_{\alpha,\beta} \subseteq \{A^* > \beta\}$. Applying the Maximal Ergodic Theorem 3.1.4 to the dynamical system defined by the restriction of Φ to $E_{\alpha,\beta}$ and with the normalized measure $\frac{\mu}{\mu(E_{\alpha,\beta})}$ and considering the function $f - \beta$, we obtain

$$\int_{E_{\alpha,\beta}} f d\mu \geq \beta\mu(E_{\alpha,\beta}). \tag{3.34}$$

Applying the same argument to the function $-f$ one gets

$$\int_{E_{\alpha,\beta}} -f d\mu \geq -\alpha\mu(E_{\alpha,\beta}) \tag{3.35}$$

or, equivalently,

$$\int_{E_{\alpha,\beta}} f d\mu \leq \alpha\mu(E_{\alpha,\beta}). \tag{3.36}$$

From (3.34) and (3.36), one gets

$$\beta\mu(E_{\alpha,\beta}) \leq \alpha\mu(E_{\alpha,\beta}), \tag{3.37}$$

which is possible, since $\alpha < \beta$, only if $\mu(E_{\alpha,\beta}) = 0$. □

The rest of this subsection will be devoted to the investigation of the convergence of the ergodic averages for a general, possibly time varying discrete time dynamical system, namely a dynamical system of the form $(M, \mathcal{S}, \mu, \{\Phi_t\})$. We start this discussion with a simple example.

Example 3.1.6. Assume that M consists of two atoms A and B, with the natural $\sigma-$ algebra $\mathcal{S} := \{A, B, \emptyset, M\}$, μ is the measure $\mu(A) = \mu(B) = \frac{1}{2}$ and consider Φ the transformation that switches A with B. Consider the sequence of transformations $\{\Phi_t\}$ with $\Phi_t = \Phi$ when $t = \sum_{j=0}^{k} 2^j$, $k = 0, 1, ...$ and $\Phi_t = Id$, where Id denotes the identity, otherwise. If f is the function with $f(A) = 1$, $f(B) = 0$, we have that the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f({}^k\Phi x)$, for $x = A$, is equal to the limit when n goes to infinity of $d_n(A) := \frac{s_n(A)}{n}$ where $s_n(A)$ is the number of times the trajectory passes through the point A after $n-1$ iterations. The function d_n is sometimes referred to as the $n-$ *density* of the system and its upper (lower) limit, when n goes to infinity, as the *upper (lower) density*. In the above case, for $n = \sum_{j=0}^{k} 2^j$, $k = 0, 1, ...$

$$d_n(A) = \frac{1}{2}\frac{\sum_{j=0}^{k}(2^j + (-2)^j)}{\sum_{j=0}^{k} 2^j} = \frac{1}{2}(1 + \frac{1}{3}\frac{1-(-2)^{k+1}}{2^{k+1}-1}); \tag{3.38}$$

the limit of the above expression when k goes to infinity is equal to $\frac{1}{3}$ for odd k and $\frac{2}{3}$ for even k, showing that the n-density does not admit a limit when n goes to infinity.

The following result shows that, if we can prove pointwise convergence for the averages (3.1), then also convergence in $L^1(M)$ follows. Notice that we have already proved L^1-convergence as a consequence of the Mean Ergodic Theorems 3.1.1 and 3.1.3.

Theorem 3.1.7. *Consider a sequential dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, and a function $f \in L^1(M)$. If the ergodic average $A_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} f({}^k\Phi x)$ converges a.e. to a function $\bar{f}(x)$, then it converges in $L^1(M)$. As a consequence,*

$$\int_M \bar{f}(x) d\mu = \int_M f(x) d\mu. \tag{3.39}$$

Proof. First, notice that applying Fatou's Lemma 2.1.5 to the sequences of functions A_n, we have

$$||\bar{f}||_1 \le \int_M \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |f({}^k\Phi x)|d\mu \le \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int_M |f({}^k\Phi x)|d\mu = ||f||_1 \qquad (3.40)$$

Therefore $\bar{f}$ is also in $L_1(M)$ and $||\bar{f}||_1 \le ||f||_1$. That A_n converges to $\bar{f}$ in $L^1(M)$ follows immediately from the Lebesgue Dominated Convergence Theorem 2.1.6 if f is in $L_\infty(M)$. If this is not the case, for any $g \in L^\infty(M)$, we have

$$||\frac{1}{n}\sum_{k=0}^{n-1} f^k\Phi - \bar{f}||_1 \le ||\frac{1}{n}\sum_{k=0}^{n-1}(f^k\Phi - g^k\Phi)||_1 + ||\frac{1}{n}\sum_{k=0}^{n-1} g^k\Phi - \bar{g}||_1 + ||\bar{g} - \bar{f}||_1. \quad (3.41)$$

The first term is less or equal than $||g - f||_1$, and so is $||\bar{g} - \bar{f}||_1$ from what proved before. The second term converges to zero by the above Lebesgue Dominated Convergence Theorem argument since $g \in L^\infty$. Therefore, the right hand side of the above expression can be made arbitrarily small by a suitable choice of $g \in L^\infty(M)$ and for n sufficiently large. The equality (3.39) follows easily from L^1 convergence (see (3.17), (3.18)). □

In the following, we will sometimes call *B-regular* (from Birkhoff) a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, or the associated sequence of measure preserving transformations $\{\Phi_t\}$, for which the ergodic averages converge a.e. for every function $f \in L^1(M)$. We will deal, in the rest of this section, with the following "robustness" question: Assume that a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ converges, in one of the senses specified in Subsection 2.2.4, to a B-regular dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$, then is the system $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ B-regular as well? The following nontrivial example shows that this is not true even if we require strong convergence, in the sense of (2.38). We will see however, in the next theorem, that this is the case if the convergence is sufficiently fast.

Example 3.1.8. Let M be the square $[0, 1) \times [0, 1]$ with the standard Lebesgue measure $dxdy$, and, for every $t \ge 1$, Φ_t^1 be the identity. The ergodic theorem holds for $\{\Phi_t^1\}$. Let $\{\Phi_t^2\}$ be the identity over all of M except on a strip

$$W_t := \{(x, y) \in M | 0 \le a(t) \le x < b(t) \le 1\}, \qquad (3.42)$$

where the functions $a(t)$ and $b(t)$ will be specified later. On W_t, Φ_t^2 reflects the point (x, y) about the line $y = \frac{1}{2}$, namely, on W_t, $\Phi_t^2 : (x, y) \to (x, 1 - y)$. The function f

is chosen to be equal to 1 when $y \geq \frac{1}{2}$ and equal to 0 otherwise, which is clearly in $L^1(M)$. In this example we will see that, for a suitable choice of the functions $a(t)$ and $b(t)$, namely of the size of the strip W_t in (3.42), the average in (3.3) does not exist on a set of positive measure, although $\lim_{t\to\infty} \mu(W_t) = 0$.

Consider an index $j = 0, 1, 2, \ldots$. For every j consider the strips

$$E_{jk} := \{(x, y) \in M | x \in [\frac{k-1}{2^j}, \frac{k}{2^j})\}, \tag{3.43}$$

with $k = 1, 2, \ldots, 2^j$, and set

$$\begin{array}{llll} j = 0, & W_1 = E_{01}, & & \\ j = 1, & W_2 = E_{11}, & W_3 = E_{12}, & \\ j = 2, & W_4 = E_{21}, & W_5 = E_{22}, & W_6 = E_{23}, \quad W_7 = E_{24}, \end{array} \tag{3.44}$$

and so on. It is clear by this definition that $\lim_{t\to\infty} \mu(W_t) = 0$. Assume now, by contradiction, that the ergodic theorem holds for $\{\Phi_t^2\}$ and therefore for almost every $P \in M$

$$\bar{f}(P) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f({}^k\Phi^2 P) \tag{3.45}$$

exists. From (3.39) of Theorem 3.1.7 we also have that

$$\int_M \bar{f}(P) d\mu = \int_M f(P) d\mu = \frac{1}{2}, \tag{3.46}$$

and therefore there exists a set of positive measure U on which $\bar{f}(P)$ is finite and strictly greater than zero. Pick P in $U - \{x = 0\}$ and notice that, denoting the x coordinate of P by x_P, we can choose two integers $l \geq 2$, $n \geq 1$ such that

$$\frac{l-1}{2^n} \leq x_P < \frac{l}{2^n}. \tag{3.47}$$

Let us now study the limit in (3.45). Assume, without loss of generality that P is in the part of M where $f(P) = 1$ (if this is not the case we can always consider another sequence starting with a switch of the two zones of M followed by $\{\Phi_t^2\}$ and our example will go through). In the study of the trajectory, the time can be indexed by j as in (3.44); 2^j instants of time correspond to each j. For each j only one *switch* in the trajectory will occur. This will happen for the m, $1 \leq m \leq 2^j$ such that the set E_{jm} includes P. At that time, the value of the function f will change from 1 to 0 or viceversa. For $j = 0$ the transition will occur at the first step namely *after* no steps and in the sequence $f({}^k\Phi^2 P)$ we will have a number $n(0) = 1$ of ones. In general the transition for the index $j \geq 1$ will occur after

$$n(j) := 2^{j-1} - [2^{j-1} x_P] + [2^j x_P] \tag{3.48}$$

steps, where $[\cdot]$ denotes the integer part. This is so because the transition related to j happens after $[2^j x_P]$ steps of the $j-th$ series plus the steps left out from the previous series, that are indeed $2^{j-1} - [2^{j-1}x_P]$. As a result we will have, in the sequence $f({}^k\Phi^2 P)$ a sequence of $n(0)$ 1's, $n(1)$ 0's, $n(2)$ 1's and so on, where $n(j)$ is given by (3.48). Define now the following sums

$$\gamma(s) := \sum_{j=0}^{s-1} n(j), \tag{3.49}$$

$$\kappa(s) := \sum_{\substack{j=0 \\ j \ \ even}}^{s-1} n(j), \tag{3.50}$$

and notice that $S(s) := \frac{\kappa(s)}{\gamma(s)}$ is a subsequence of the sequence considered in (3.45). Consider the subsequence corresponding to odd s, $S_{odd}(s)$, and observe that

$$S_{odd}(s+1) := \frac{\kappa(s)}{\gamma(s)+n(s)} = \frac{S_{odd}(s)}{1+\frac{n(s)}{\gamma(s)}}. \tag{3.51}$$

Since we have assumed

$$\infty > \lim_{\substack{s\to\infty \\ s \ \ odd}} S(s) = \lim_{\substack{s\to\infty \\ s \ \ odd}} S(s+1) > 0, \tag{3.52}$$

we must have

$$\lim_{\substack{s\to\infty \\ s \ \ odd}} \frac{n(s)}{\gamma(s)} = 0. \tag{3.53}$$

Now notice that, from (3.48) and (3.49), we have

$$\gamma(s) = 1 + \sum_{j=1}^{s-1} 2^{j-1} - [2^{j-1}x_P] + [2^j x_P] = 2^{s-1} - [x_P] + [2^{s-1}x_P] = 2^{s-1} + [2^{s-1}x_P]. \tag{3.54}$$

Using (3.48) and (3.49), the ratio $\frac{n(s)}{\gamma(s)}$ can be written as

$$\frac{n(s)}{\gamma(s)} = \frac{2^{s-1} - [2^{s-1}x_P] + [2^s x_P]}{2^{s-1} + [2^{s-1}x_P]}, \tag{3.55}$$

and, for s large enough,

$$\frac{n(s)}{\gamma(s)} = \frac{\frac{2^{s-1}}{[2^{s-1}x_P]} - 1 + \frac{[2^s x_P]}{[2^{s-1}x_P]}}{\frac{2^{s-1}}{[2^{s-1}x_P]} + 1}. \tag{3.56}$$

Now recall the bound on x_P given by (3.47), so that we can write

$$\left[\frac{2^{s-1}(l-1)}{2^n}\right] \leq [2^{s-1}x_P] < \left[\frac{2^{s-1}l}{2^n}\right] \tag{3.57}$$

and

$$\left[\frac{2^s(l-1)}{2^n}\right] \leq [2^s x_P] < \left[\frac{2^{s-1}l}{2^n}\right]. \tag{3.58}$$

For $s \geq n$, we can omit the integer part on the left and the right hand side of these inequalities. We obtain from (3.57)

$$\frac{(l-1)}{2^n} \leq \frac{[2^{s-1}x_P]}{2^{s-1}} < \frac{l}{2^n}, \tag{3.59}$$

which shows that the denominator in (3.56) is bounded by a bound that depends on x_P. Therefore the only possibility for (3.56) to tend to zero when s tends to infinity is that the numerator in (3.56) tends to zero. Using (3.57), (3.58), with $s \geq n$, we obtain

$$\frac{2(l-1)}{l} < \frac{[2^s x_P]}{[2^{s-1}x_P]} < \frac{2l}{l-1}. \tag{3.60}$$

Using this and (3.59) we obtain

$$\frac{2^{s-1}}{[2^{s-1}x_P]} + \frac{[2^s x_P]}{[2^{s-1}x_P]} > \frac{2^n}{l} + \frac{2(l-1)}{l}. \tag{3.61}$$

Now recall that we chose n and l such that

$$2^n + l > 3, \tag{3.62}$$

which also gives

$$\frac{2^n}{l} + \frac{2(l-1)}{l} > 1 + \frac{1}{l}. \tag{3.63}$$

This, combined with (3.61), gives

$$\frac{2^{s-1}}{[2^{s-1}x_P]} + \frac{[2^s x_P]}{[2^{s-1}x_P]} > 1 + \frac{1}{l}, \tag{3.64}$$

which shows that the numerator of (3.56) is bounded from below by $\frac{1}{l}$, and so is its limit when s goes to infinity, which therefore cannot be zero. This concludes the example.

Theorem 3.1.9. *Assume $\{\Phi_t^1\}$ is a B-regular sequence. Consider another sequence $\{\Phi_t^2\}$ and the sets W_t as in (2.41). Assume*

$$\sum_{t=0}^{\infty} \mu(W_t) < \infty. \tag{3.65}$$

Then $\{\Phi_t^2\}$ is B-regular.

Proof. Assume there exists a set B, with $\mu(B) > 0$, such that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f({}^k\Phi^2 x) > \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f({}^k\Phi^2 x), \qquad \forall x \in B. \tag{3.66}$$

From (3.65) we can choose a $\bar{t}$ such that

$$\sum_{t=\bar{t}}^{\infty} \mu(W_t) < \mu(B). \tag{3.67}$$

From (3.66) for almost every x in B, there exists a $k \geq 0$ such that

$${}^{\bar{t}-1+k}\Phi^2 x \in W_{\bar{t}+k}. \tag{3.68}$$

If this was not the case, then ${}^{\bar{t}+k}\Phi^2_{\bar{t}}\,{}^{\bar{t}-1}\Phi^2 x = {}^{\bar{t}+k}\Phi^1_{\bar{t}}\,{}^{\bar{t}-1}\Phi^2 x$ for each $k \geq 0$, and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f({}^j\Phi^2 x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=\bar{t}}^{n-1} f({}^j\Phi^1_{\bar{t}}\,{}^{\bar{t}-1}\Phi^2 x), \tag{3.69}$$

which exists (a.e.) because of the hypothesis of the B-regularity of $\{\Phi_t^1\}$. Therefore,

$$B \subseteq \bigcup_{k\geq 0} ({}^{\bar{t}-1+k}\Phi^2)^{-1} W_{\bar{t}+k}, \qquad (mod \quad \mu) \tag{3.70}$$

and

$$\mu(B) \leq \sum_{t=\bar{t}}^{\infty} \mu(({}^{t-1}\Phi^2)^{-1} W_t) = \sum_{t=\bar{t}}^{\infty} \mu(W_t), \tag{3.71}$$

which contradicts (3.67). □

We now give another condition for B-regularity of a sequence $\{\Phi_t^2\}$ converging in the strong sense of (2.38) to a B-regular sequence. This condition is given in terms of the topology of the sets W_t in (2.41).

Theorem 3.1.10. *Assume $\{\Phi_t^1\}$ is a B-regular sequence. Consider another sequence $\{\Phi_t^2\}$. Define W_t as in (2.41) and assume $\lim_{t\to\infty}\mu(W_t)=0$. Moreover assume that there exists a $\tilde{t}\geq 0$, such that for every $t>\tilde{t}$, one of the following equivalent conditions holds* [1].

a)
$$W_{t+1}\subseteq \Phi_t^2 W_t;$$

b)
$$W_{t+1}\subseteq \Phi_t^1 W_t;$$

c)
$$\Phi_t^2 \bar{W}_t \cap W_{t+1}=\emptyset;$$

d)
$$\Phi_t^1 \bar{W}_t \cap W_{t+1}=\emptyset.$$

Then, $\{\Phi_t^2\}$ is $B-$regular.

Proof. Given $\tilde{t}\geq 0$, a sequence $\{\Phi_t\}$ is B-regular if and only if $\{\Phi_{\tilde{t}+t}\}$ is B-regular. Moreover, $\{\Phi_t^2\}\to\{\Phi_t^1\}$ if and only if $\{\Phi_{\tilde{t}+t}^2\}\to\{\Phi_{\tilde{t}+t}^1\}$, and the assumptions $a)$ $d)$, hold for $t>\tilde{t}$ for $\{\Phi_t^1\}$, $\{\Phi_t^2\}$ if and only if they hold for $\{\Phi_{\tilde{t}+t}^1\}$, $\{\Phi_{\tilde{t}+t}^2\}$, for $t>0$. In view of these facts, there is no loss of generality in assuming $\tilde{t}=0$.

We first note the equivalence of the conditions $a)$ through $d)$. The equivalence of $a)$ and $c)$ (or $b)$ and $d)$) is an immediate consequence of the invertibility of $\{\Phi_t^2\}$ ($\{\Phi_t^1\}$), for each t. Moreover $c)$ and $d)$ are equivalent since Φ_t^2 and Φ_t^1 are the same on $\bar{W}_t$, for every t, by definition.

Since $\lim_{t\to\infty}\mu(W_t)=0$, for a.e. $x\in M$, there exists a $\bar{t}=\bar{t}(x)>0$ such that $x\notin W_{\bar{t}}$, namely $x\in\bar{W}_t$ or equivalently

$$\Phi_{\bar{t}}^1 x=\Phi_{\bar{t}}^2 x. \tag{3.72}$$

Define V_0 the set of such x's. Applying inductively assumption $c)$, we get that

$$\Phi_{\bar{t}+k}^2\circ\cdots\circ\Phi_{\bar{t}}^2 x=\Phi_{\bar{t}+k}^1\circ\cdots\circ\Phi_{\bar{t}}^1 x,\qquad \forall k\geq 0. \tag{3.73}$$

[1] These relations are meant to hold everywhere except on a set of measure zero. $\bar{A}$ denotes the complement of the set A.

Define now B_1 the set of points $z \in M$ such that $\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} f({}^t\Phi^1 z)$ does not exist. Since $\{\Phi_t^1\}$ is B-regular, we have that $\mu(B_1) = 0$, and, if we define

$$V_1 := V_0 - \bigcup_{t\geq 0} {}^t\Phi^1 B_1, \tag{3.74}$$

we have $\mu(V_1) = 1$. Therefore, we can write $M = V_1 \cup B_2$ for some measurable set B_2, with $\mu(B_2) = 0$. Define now

$$V_2 = M - \bigcup_{t\geq 0} ({}^t\Phi^2)^{-1} B_2, \tag{3.75}$$

that again has full measure. Notice also that, by construction, $V_2 \subseteq V_1 \subseteq V_0$. Consider now $x \in V_2$, since $x \in V_0$, there exists a $\bar{t}$ such that (3.73) holds. Consider

$$y :=^{\bar{t}} \Phi^2 x. \tag{3.76}$$

We have $y \notin B_2$ since $x \in V_2$, therefore $y \in V_1$. Set

$$z := ({}^{\bar{t}}\Phi^1)^{-1} y, \tag{3.77}$$

we have that $z \notin B_1$ since $y \in V_1$. In conclusion, we have

$$\begin{array}{l} \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} f({}^t\Phi^2 x) = \\ \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} f({}^t\Phi^2 \circ ({}^{\bar{t}}\Phi^2)^{-1} \circ^{\bar{t}} \Phi^1 z) = \\ \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} f({}^t\Phi^1 z). \end{array} \tag{3.78}$$

This limit exists, since $z \notin B_1$. Therefore, we have proven that for each x in a set of measure 1, V_2, the limit in (3.78) exists. This concludes the proof of the theorem. □

The conditions added to convergence, in Theorems 3.1.9 and 3.1.10 rule out the situation occurring in the Example 3.1.8, where the set

$$B := \{x \in M | \forall \bar{t} \geq 0, \exists t > \bar{t},^t \Phi^2 x \in W_{t+1}\} \tag{3.79}$$

was a set of positive measure, indeed the whole space M.

We conclude this section discussing the case of periodic systems to which both the Mean Ergodic Theorem and the Pointwise Ergodic Theorem can be extended directly by using the construction of *skew products.* More specifically, if T is the period consider an auxiliary dynamical system $(N, \mathcal{A}, Sh, \nu)$, where N is a set of T atoms representing the time modulo T, $\mathcal{A}$ is the $\sigma-$algebra 2^N of all the possible subsets of N, Sh is a forward shift of times $(mod \quad T)$, and ν is a measure which assigns $\frac{1}{T}$ to each element in N. Given this dynamical system and $(M, \mathcal{S}, \mu, \{\Phi_t\})$, we define the skew-product dynamical system as $(M \times N, \mathcal{A} \times \mathcal{S}, \Psi, \mu' = \mu \times \nu)$, where $\mathcal{A} \times \mathcal{S}$ is the product $\sigma-$algebra namely the $\sigma-$algebra generated by the sets of the form $S \times A$, where S is measurable in $\mathcal{S}$ and A in $\mathcal{A}$. Ψ is the map defined by $\Psi(x,t) = (\Phi_t x, (t+1) \pmod{T})$, and μ' is the product measure namely $\mu'(A \times B) = \mu(A) \times \nu(B)$. Using the properties of $\{\Phi_t\}$ it is easy to show that Ψ is measurable and measure preserving. We can now look at the ergodic averages as averages computed along trajectories of the new dynamical system $(M \times N, \mathcal{A} \times \mathcal{S}, \Psi, \mu')$ and therefore conclude pointwise convergence and mean convergence using the standard results Theorems 3.1.3, and 3.1.5.

3.2 Recurrence properties: Ergodicity and Mixing

Given a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, we will characterize in this section the qualitative features of its trajectories concerned with recurrence properties. More specifically we will try to answer questions like: will a trajectory starting at a point x_0 return arbitrarily close to x_0 (Recurrence)? Will it span the whole space M (Ergodicity)? Will the sets in the $\sigma-$algebra $\mathcal{S}$ get uniformly mixed under sufficient iterations of the map Φ_t (Mixing)?

A simple example shows that recurrence is not always verified for a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$. If M has only two atoms A and B, with measures $\mu(A) = \mu(B) = \frac{1}{2}$, and Φ_t switches the two atoms at time 1 and leaves the positions unchanged, namely it is the identity transformation for the following times, then the trajectory starting at A (or B) will never come back to A (or B). On the other hand, if the transformation were the Identity for all the time that would have been the case. Notice that in this example of sequential system we have convergence, in the strong sense of (2.38), to a system for which recurrence holds. Nevertheless the recurrence property does not hold for this system. The fact that recurrence holds for every system characterized by a single transformation is the content of what can be considered the first result in Ergodic Theory namely the Poincare' Recurrence Theorem.

Definition 3.2.1. *Given a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, a point x in a measurable set $B \in \mathcal{S}$ is said to be recurrent with respect to B if there exists a time $\bar{t} > 0$ such that ${}^{\bar{t}}\Phi x \in B$.*

Theorem 3.2.2. (Poincare' Recurrence Theorem) *Consider a dynamical system characterized by a single transformation Φ, namely of the form $(M, \mathcal{S}, \mu, \Phi)$. For* *every measurable set $B \in \mathcal{S}$, almost every point of B is recurrent with respect to B.*

Proof. Assume there exists a measurable set $A \subseteq B$ of positive measure of nonrecurrent points namely, $\Phi^t A \cap B = \emptyset$, $\forall t > 0$, which implies

$$\Phi^t A \cap A = \emptyset, \quad \forall t > 0. \tag{3.80}$$

This implies that the sets $\Phi^k A$ and $\Phi^j A$ are disjoint for $k \neq j$. In fact, assuming without loss of generality $k > j$, we have $\Phi^k A \cap \Phi^j A = \Phi^j(\Phi^{k-j} A \cap A) = \emptyset$, from (3.80). By countable additivity of the measure μ (see b) Definition 2.1.3), the set $\cup_{t=0}^{\infty} \mu(\Phi^t A)$ has infinite measure, which contradicts the fact tha $\mu(M) = 1$. □

The theorem means that no matter how small we choose a set B in the $\sigma-$algebra $\mathcal{S}$, (almost) all the trajectories starting in B will eventually come back to B. For example, in the Example of Statistical Mechanics 2.2.5, the system will return arbitrarily close to the values of positions and momenta it had at the time instant zero.

We now investigate recurrence for time varying systems under convergence assumptions. We have seen, in the example at the beginning of this section, that convergence alone to a system that satisfies the recurrence property is not sufficient to guarantee that the recurrence property is verified. We can however require convergence to a transformation that moves the points all around the space namely a transformation which is *ergodic* in the sense that will be specified in the next subsection. We can say here that a dynamical system $(M, \mathcal{S}, \mu, \Phi)$ is ergodic if, for each $A \in \mathcal{S}$, with $\mu(A) > 0$, $\mu(\cup_{k=0}^{\infty} \Phi^k A) = 1$. We will see, in the next subsection, that this definition of ergodicity is equivalent to others (see in particular the one given in the Mean Ergodic Theorem for converging sequences of transformations (Theorem 3.1.3)). If we assume that the convergence to an ergodic system is sufficiently fast we can prove the recurrence property for the given system. This will be done in Theorem 3.2.4. We first have the following result.

Lemma 3.2.3 *Let $(M, \mathcal{S}, \mu, \{\Phi_t\})$ be a a dynamical system strongly converging (see (2.38) to an ergodic dynamical system $(M, \mathcal{S}, \mu, \Phi)$. Assume*

$$\sum_{t=0}^{\infty} \mu(W_t) < \infty \tag{3.81}$$

holds with $W_t := \{x \in M | \Phi_t x \neq \Phi x\}$. Then, for each set A, with $0 < \mu(A) \leq 1$,

$$\mu(\cup_{k=1}^{\infty} {}^k\Phi A) = 1, \tag{3.82}$$

holds.

Proof. From condition (3.81), we know that for each $\epsilon > 0$ there exists a $\bar{k} \geq 1$ such that

$$\sum_{k=\bar{k}}^{\infty} \mu(W_k) < \epsilon. \tag{3.83}$$

Consider a set A with $\mu(A) > 0$, and take $\epsilon = \frac{\mu(A)}{2}$ in (3.83), and the corresponding $\bar{k}$. Recall the definition ${}^t\Phi_{\bar{k}} := \Phi_t \circ \cdots \circ \Phi_{\bar{k}}$. Consider the set

$$\tilde{A} :=^{\bar{k}-1} \Phi A - \cup_{j=0}^{\infty} \Phi^{-j} W_{\bar{k}+j}. \tag{3.84}$$

We have

$$\mu(\tilde{A}) \geq \mu(A) - \sum_{k=\bar{k}}^{\infty} \mu(W_k) > \frac{\mu(A)}{2}, \tag{3.85}$$

so that $\tilde{A}$ has positive measure. We also have

$$\Phi^s \tilde{A} =^{\bar{k}+s-1} \Phi_{\bar{k}} \tilde{A}, \quad \forall s \geq 1. \tag{3.86}$$

This is true for $s = 1$ since, if it is not true then there exists an $x \in \tilde{A}$ such that $\Phi x \neq \Phi_{\bar{k}} x$, namely $x \in W_{\bar{k}}$ which is excluded by the definition (3.84). Analogously, for $s \geq 2$, using induction, if $\Phi^s \tilde{A} \neq^{\bar{k}+s-1} \Phi_{\bar{k}} \tilde{A}$, then there exists an $x \in \Phi^{s-1} \tilde{A} =^{\bar{k}+s-2} \Phi_{\bar{k}} \tilde{A}$ with $x \in W_{\bar{k}+s-1}$ which is impossible since $\tilde{A} \cap (\Phi^{s-1})^{-1} W_{\bar{k}+s-1} = \emptyset$.

Now we have

$$\cup_{k=1}^{\infty} {}^k\Phi A \supseteq \cup_{j=0}^{\infty} {}^{\bar{k}+j}\Phi_{\bar{k}} \circ^{\bar{k}-1} \Phi A \supseteq \cup_{j=0}^{\infty} {}^{\bar{k}+j}\Phi_{\bar{k}} \tilde{A} = \cup_{j=1}^{\infty} \Phi^j \tilde{A}, \tag{3.87}$$

where, in the last equality, we have used (3.86). Therefore we have

$$\mu(\cup_{k=1}^{\infty} {}^k\Phi A) \geq \mu(\cup_{j=1}^{\infty} \Phi^j \tilde{A}) = 1, \tag{3.88}$$

where, in the last equality, we used the ergodicity of Φ and the fact that $\tilde{A}$ has positive measure. This proves (3.82). □

It is easy now to prove the result about recurrence for sequences converging sufficiently fast to an ergodic transformation.

Theorem 3.2.4. *Let $\{\Phi_t\})$ be a sequence of measure preserving transformations and Φ an ergodic transformation. Assume (3.81) is verified with W_t defined in (2.41). Then, almost every point in a set $B \in S$ is recurrent with respect to B.*

Proof. If there exists a set $A \subseteq B$ such that ${}^k\Phi A \cap B = \emptyset$, for each $k \geq 1$, we have

$$\cup_{k=1}^{\infty} {}^k\Phi A \cap B = \emptyset, \tag{3.89}$$

which is impossible since, from Lemma 3.2.3, we have $\mu(\cup_{k=1}^{\infty} {}^k\Phi A) = 1$. □

We conclude this subsection noticing that for periodic systems $(M, \mathcal{S}, \mu, \{\Phi_t\})$ the recurrence property holds. This can be easily seen by considering the associated time invariant dynamical system whose transformation is the composite transformation associated to $\{\Phi_t\}$ and applying Theorem 3.2.2.

3.2.1 Ergodicity

Given a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, we are interested in how the evolution law $\{\Phi_t\}$ evenly distributes the trajectories in the phase space M in an homogeneous way, and if the trajectory will travel around all of the space M. Also, if we are making measurements on the system with a long-time average procedure (see the introductory part of this section), we would like to know whether the measurement is independent of the initial state of the system. These issues are considered in the definition of *Ergodicity*. We give the following definition for general sequential systems:

Definition 3.2.5. *A dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ is said to be ergodic if, for every $f \in L^1(M)$, the time average,*

$$\bar{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f({}^k\Phi x), \tag{3.90}$$

exists and it is constant almost everywhere.

Notice that it follows immediately from (3.39) of Theorem 3.1.7 that in the case of ergodic systems, the constant $\bar{f}$ has to be equal to the space mean $\int_M f d\mu$ of the function f. Also notice that the definition implies that it does not exist any set A, with measure $0 < \mu(A) < 1$, which is invariant from one instant on, namely such that there exists a $k \geq 1$, with

$$\Phi_t A \subseteq A, \quad \forall t \geq k. \tag{3.91}$$

In fact if this is the case, for any $x \in ({}^{k-1}\Phi)^{-1}A$, the time average for the characteristic function of the set A, χ_A would be

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\chi_A({}^j\Phi x) = \lim_{n\to\infty}\frac{1}{n}(\sum_{j=0}^{k-1}\chi_A({}^j\Phi x) + \sum_{j=k}^{n-1}\chi_A({}^j\Phi_k{}^{k-1}\Phi x))$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{j=k}^{n-1}\chi_A({}^j\Phi_k{}^{k-1}\Phi x) = \lim_{n\to\infty}\frac{n-k}{n} = 1. \tag{3.92}$$

Analogously one can show that if x is in the complement of the set $({}^{k-1}\Phi)^{-1}A$, the time average is equal to zero. Therefore the time average is not constant almost everywhere contradicting ergodicity.

A complete characterization of ergodicity in terms of invariant sets is given in the next theorem for systems described by a single transformation.

Theorem 3.2.6 *Consider a discrete time dynamical system of the form* $(M, \mathcal{S}, \mu, \Phi)$. *The following conditions are equivalent:*

a) For every function $f \in L^1(M)$, *the time average* $\bar{f} = \lim_{n\to\infty}\sum_{k=0}^{n-1} f(\Phi^k x)$ *is constant a.e. and equal to the space average* $\int_M f d\mu$.

b) The only invariant functions ($f = f \circ \Phi$) *in* $L^1(M)$, *are the functions that are constant almost everywhere.*

c) If $A \in \mathcal{S}$ *is an invariant set* ($\Phi A \subseteq A$), *then* $\mu(A) = 0$ *or* $\mu(A) = 1$.

d) For every pair of sets A *and* B *in* $\mathcal{S}$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(\Phi^{-k}A \cap B) = \mu(A)\mu(B). \tag{3.93}$$

e) For any set A, *with* $\mu(A) > 0$,

$$\mu(\cup_{t=0}^{\infty}\Phi^t A) = 1. \tag{3.94}$$

Proof $a \to b$. if there exists an invariant function in $L^1(M)$ which is not constant a.e. then the time averages computed in two points where the values of the function differ, will be different and equal to the values of the function in those points, thus contradicting a. $b \to c$. If A is an invariant set in $\mathcal{S}$, the characteristic function of A, χ_A is Φ-invariant and therefore constant in M. χ_A constant implies χ_A equal to zero or one almost everywhere. Therefore $\mu(A) = 0$ or $\mu(A) = 1$. $c \to a$. If the time average $\bar{f}$ is not constant a.e. then there exists a number α such that the set $A := \{x \in M | \bar{f}(x) < \alpha\}$, is such that $0 < \mu(A) < 1$. However this is impossible since A is an invariant set for Φ. $c \to e$. For every $A \in \mathcal{S}$, with $\mu(A) > 0$, the set $\cup_{n=0}^{\infty} \Phi^n A$, is invariant under Φ, and therefore it has measure 1. $e \to c$. Assume A is an invariant set with $0 < \mu(A) < 1$. Then $\cup_{n=0}^{\infty} \Phi^n A \subseteq A$, and $\mu(\cup_{n=0}^{\infty} \Phi^n A) = \mu(A) < 1$, which contradicts (3.94). $a \to d$. The limit $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(\Phi^k x)\chi_B(x)$ exists a.e. and it is equal to $\mu(A)\chi_B$. Applying the Dominated Convergence Theorem 2.1.6, to the sequence of functions $\frac{1}{n} \sum_{k=0}^{n-1} \chi_A(\Phi^k x)\chi_B(x)$ we get

$$\int_M \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(\Phi^k x)\chi_B(x) d\mu = \mu(A)\mu(B) =$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \int_M \chi_A(\Phi^k x)\chi_B(x) d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\Phi^{-k} A \cap B). \tag{3.95}$$

$d \to c$. If A is invariant, applying d with $B = \bar{A}$, we obtain $\mu(A)\mu(\bar{A}) = 0$, which implies that $\mu(A) = 0$ or $\mu(A) = 1$. □

Notice that the condition (3.93) in d) of the above theorem is equivalent to 'for each A and $B \in \mathcal{S}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\Phi^k A \cap B) = \mu(A)\mu(B). \tag{3.96}$$

We conclude this subsection discussing the case of periodic dynamical systems. We have seen, in the previous section that the ergodic averages exist almost everywhere for these systems. A quick computation, provides an expression for them. Consider f a function in $L^1(M)$, and set $n = kT + l + 1$, with $l = 0, 1, ..., T-1$, $k = 0, 1, 2, ...,$, where T is the period, and $\bar{\Phi} := \Phi_T \circ \cdots \circ \Phi_1$. We can write

$$\begin{aligned} &\tfrac{1}{n} \textstyle\sum_{t=0}^{n-1} f({}^t\Phi x) = \\ &\tfrac{1}{kT+l+1} \left(\textstyle\sum_{j=0}^{k-1} \sum_{i=0}^{T-1} f({}^i\Phi \circ \bar{\Phi}^j x) + \sum_{i=0}^{l} f({}^i\Phi \circ \bar{\Phi}^k x) \right). \end{aligned} \tag{3.97}$$

Relation (3.97) defines T subsequences parametrized by l and indexed by k. For almost every x all of these subsequences have the same limit when k goes to infinity, by the Ergodic Theorem applied to periodic sequences (see the discussion at the end of Section 3.1). Moreover, for almost every $x \in M$ and for every $0 \leq i \leq T-1$, $\lim_{k\to\infty} \frac{1}{k} f({}^i\Phi \circ \bar{\Phi}^k x) = 0$. We have

$$\begin{aligned} &\lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{t=0}^{n-1} f({}^t\Phi x) = \\ &\lim_{k\to\infty} \tfrac{k}{kT+1} \tfrac{1}{k} \left(\textstyle\sum_{j=0}^{k-1} \sum_{i=0}^{T-1} f({}^i\Phi \bar{\Phi}^j x) \right) = \\ &\tfrac{1}{T} \textstyle\sum_{i=0}^{T-1} [f({}^i\Phi)](x), \end{aligned} \tag{3.98}$$

where $[f(^{i}\bar{\Phi})]$ is the time average of the function $f \circ \Phi^i$, $i = 0, ..., T-1$, for the composite transformation $\bar{\Phi}$. Is these averages are constant almost everywhere then also the time average for the sequence is. Therefore ergodicity of $\bar{\Phi}$ implies ergodicity of $\{\Phi_t\}$. The converse is in general not true. In order to see this consider, as a simple example, the usual space M with two atoms A and B, $\mu(A) = \mu(B) = \frac{1}{2}$ and Φ the switch transformation. Obviously $\bar{\Phi} = \Phi \circ \Phi^{-1} = Id$ is not ergodic, every time average of a function $f \in L^1(M)$, for the sequence $\Phi\Phi^{-1}\Phi\Phi^{-1}\Phi\Phi^{-1}\cdots$ is equal to $\frac{1}{2}(f(A) + f(B))$. We formally state the result about periodic systems below.

Theorem 3.2.7 *Consider a periodic dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ and the associated single transformation dynamical system $(M, \mathcal{S}, \mu, \bar{\Phi})$, with $\bar{\Phi} = \Phi_T \circ \cdots \circ \Phi_1$. If the latter system is ergodic then the sequential system is also ergodic.*

3.2.2 Mixing

Consider a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$. If the evolution law $\{\Phi_t\}$ is such that the points in the state space M get homogeneously mixed, we have that for each pair of sets of positive measure A and B, and after a sufficient number of iterations $\bar{n}$, for every $n > \bar{n}$, $\mu(\Phi_n \circ \Phi_{n-1} \circ \cdots \circ \Phi_1 A \cap B) > 0$. This means that from some instant on, there exists always some part of A that is carried in B. This condition is referred to as *partial mixing*. A stronger condition is the one given in the next definition.

Definition 3.2.8. *A dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ is said to be (strongly) mixing if for each pair of measurable sets A and B in $\mathcal{S}$, we have*

$$\lim_{t \to \infty} \mu(^{t}\Phi A \cap B) = \mu(A)\mu(B). \tag{3.99}$$

The mixing condition in the above definition means that for every pair of sets A and B, for sufficiently large time, the percentage of A that is carried into B, namely $\frac{\mu(^{t}\Phi A \cap B)}{\mu(B)}$, is the concentration of A in the whole space, namely $\mu(A)$ since M is assumed to have measure 1.

Mixing is a very well studied property in the case of systems characterized by one single transformation. The following result is useful to relate the mixing properties of a system to the ones of another system strongly convergent to it (see (2.38)).

Theorem 3.2.9. *Let $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ and $(M, \mathcal{S}, \mu, \{\Phi_t^2\})$ be two sequential dynam-*

ical systems such that

$$\sum_{t=1}^{\infty} \mu(W_t) < \infty, \tag{3.100}$$

with $W_t := \{x \in M | \Phi_t^1 x \neq \Phi_t^2 x\}$. *Then,* $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ *is strongly mixing if and only if* $(M, \mathcal{S}, \mu, \{\Phi_t^2\})$ *is strongly mixing.*

Proof We assume that $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ is mixing and prove that $(M, \mathcal{S}, \mu, \{\Phi_t^2\})$ is mixing, the converse follows from symmetry. We also notice that $(M, \mathcal{S}, \mu, \{\Phi_t^1\})$ mixing is equivalent to $(M, \mathcal{S}, \mu, \{\Phi_{t+k}^1\})$ mixing, for each $k \geq 0$.

From condition (3.100), for each $\delta > 0$, we can choose a $\bar{k}$, such that

$$\sum_{t=\bar{k}}^{\infty} \mu(W_t) < \delta. \tag{3.101}$$

Define now the set

$$D := W_{\bar{k}} \cup (\cup_{j=1}^{\infty} (\Phi_{\bar{k}+j-1}^1 \circ \cdots \circ \Phi_{\bar{k}}^1)^{-1} W_{\bar{k}+j}). \tag{3.102}$$

Notice the following two properties of the set D: First

$$\mu(D) \leq \mu(W_{\bar{k}}) + \sum_{j=1}^{\infty} \mu((\Phi_{\bar{k}+j-1}^1 \circ \cdots \circ \Phi_{\bar{k}}^1)^{-1} W_{\bar{k}+j}) = \sum_{t=\bar{k}}^{\infty} \mu(W_t) < \delta. \tag{3.103}$$

Second, defined, for $t \geq \bar{k}$, ${}^t\Phi_{\bar{k}}^1 := \Phi_t^1 \circ \Phi_{t-1}^1 \circ \cdots \circ \Phi_{\bar{k}}^1$ and analogously ${}^t\Phi_{\bar{k}}^2$, for each x in $\bar{D}$, the complement of D, we have

$${}^t\Phi_{\bar{k}}^1 x = {}^t \Phi_{\bar{k}}^2 x, \qquad \forall t \geq \bar{k}. \tag{3.104}$$

This can be proven by induction. First notice that ${}^{\bar{k}}\Phi_{\bar{k}}^1 x = {}^{\bar{k}} \Phi_{\bar{k}}^2 x$, since otherwise x would be in $W_{\bar{k}} \subseteq D$. Then assume ${}^{t-1}\Phi_{\bar{k}}^1 x = {}^{t-1}\Phi_{\bar{k}}^2 x$. If ${}^t\Phi_{\bar{k}}^1 x \neq {}^t\Phi_{\bar{k}}{}^2 x$, then we have ${}^{t-1}\Phi_{\bar{k}}^1 x \in W_t$ which is not possible since otherwise we would have $x \in ({}^{t-1}\Phi_{\bar{k}}^1)^{-1} W_t \subseteq D$ (cfr. (3.102) with $t = \bar{k} + j$).

Consider now two measurable sets A and B as in the definition of strong mixing (3.99). For $t \geq \bar{k}$, we can write

$$\begin{aligned} \mu({}^t\Phi_{\bar{k}}^2({}^{\bar{k}-1}\Phi^2 A \cap \bar{D}) \cap B) &\leq \mu({}^t\Phi_{\bar{k}}^2({}^{\bar{k}-1}\Phi^2 A) \cap B) = \mu({}^t\Phi^2 A \cap B) \\ &= \mu({}^t\Phi_{\bar{k}}^2({}^{\bar{k}-1}\Phi^2 A \cap \bar{D}) \cap B) + \mu({}^t\Phi_{\bar{k}}^2({}^{\bar{k}-1}\Phi^2 A \cap D) \cap B). \end{aligned} \tag{3.105}$$

Noticing that

$$\mu({}^t\Phi_{\bar{k}}^2({}^{\bar{k}-1}\Phi^2 A \cap D) \cap B) < \delta, \tag{3.106}$$

and since ${}^t\Phi^1_{\bar{k}} ={}^t \Phi^2_{\bar{k}}$ on $\bar{D}$, we have

$$\mu({}^t\Phi^1_{\bar{k}}({}^{\bar{k}-1}\Phi^2 A \cap \bar{D}) \cap B) \leq \mu({}^t\Phi^2 A \cap B) \leq \delta + \mu({}^t\Phi^1_{\bar{k}}({}^{\bar{k}-1}\Phi^2 A \cap \bar{D}) \cap B). \quad (3.107)$$

Letting t go to infinity and recalling that, as observed at the beginning of the proof, $\{\Phi^1_{t+\bar{k}}\}$ is mixing, we have

$$\mu({}^{\bar{k}-1}\Phi^2 A \cap \bar{D})\mu(B) \leq \lim_{t\to\infty} \mu({}^t\Phi^2 A \cap B) \leq \delta + \mu({}^{\bar{k}-1}\Phi^2 A \cap \bar{D})\mu(B). \quad (3.108)$$

Noting that $\mu({}^{\bar{k}-1}\Phi^2 A \cap \bar{D}) = \mu({}^{\bar{k}-1}\Phi^2 A) - \mu({}^{\bar{k}-1}\Phi^2 A \cap D)$ and that $\mu({}^{\bar{k}-1}\Phi^2 A) = \mu(A)$ and $\mu({}^{\bar{k}-1}\Phi^2 A \cap D) < \delta$, we finally get

$$\mu(A)\mu(B) - \delta\mu(B) < \lim_{t\to\infty} \mu({}^t\Phi^2 A \cap B) < \delta + \mu(A)\mu(B), \quad (3.109)$$

which, since δ can be chosen arbitrarily small, proves the theorem. □

The following theorem relates the mixing properties of a periodic system to the ones of the associated single transformation system.

Theorem 3.2.10. *Let $(M, \mathcal{S}, \mu, \{\Phi_t\})$ be a periodic system of period T. Define the composite transformation $\bar{\Phi} := \Phi_T \circ \cdots \circ \Phi_1$. Then, $(M, \mathcal{S}, \mu, \{\Phi_t\})$ is strongly mixing, if and only if $(M, \mathcal{S}, \mu, \bar{\Phi})$ is strongly mixing.*

Proof. Assume the periodic system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ is strongly mixing. For any pair of measurable sets A and B and for each $\epsilon > 0$, there exists a $t_\epsilon(A, B)$ such that, for each $t > t_\epsilon(A, B)$,

$$|\mu({}^t\Phi(A) \cap B) - \mu(A)\mu(B)| < \epsilon. \quad (3.110)$$

Therefore, in particular,

$$|\mu(\bar{\Phi}^k(A) \cap B) - \mu(A)\mu(B)| < \epsilon, \quad (3.111)$$

if $k > \frac{t_\epsilon(A,B)}{T}$.

Conversely, assume that $(M, \mathcal{S}, \mu, \bar{\Phi})$ is strongly mixing. This means that, for a fixed ϵ, and for each pair of measurable sets C and D, there exists a $t_\epsilon(C, D)$, such that $\bar{t} > t_\epsilon(C, D)$ implies

$$|\mu(\bar{\Phi}^{\bar{t}}(C) \cap D) - \mu(C)\mu(D)| < \epsilon. \quad (3.112)$$

Now, consider any two measurable sets A and B, and define

$$\bar{t}_\epsilon(A, B) = max(t_\epsilon(A, B), t_\epsilon(A, \Phi_1^{-1}(B)), \cdots, t_\epsilon(A, (\Phi_{n-1} \circ \cdots \circ \Phi_1)^{-1}(B))). \quad (3.113)$$

For each $t > (\tilde{t}_\epsilon(A,B)+1)n$, we can write $t = \bar{t}n + k$, with $\bar{t} > \tilde{t}_\epsilon(A,B)$ and $0 \le k \le n-1$. We have

$$\begin{aligned} |\mu({}^t\Phi(A) \cap B) - \mu(A)\mu(B)| &= |\mu({}^k\Phi \circ \bar{\Phi}^{\bar{t}}(A) \cap B) - \mu(A)\mu(B)| \qquad (3.114) \\ &= |\mu(\bar{\Phi}^{\bar{t}}(A) \cap ({}^k\Phi)^{-1}(B)) - \mu(A)\mu(({}^k\Phi)^{-1}(B))| < \epsilon, \\ & \qquad (3.115) \end{aligned}$$

where, we have used the fact that every transformation is measure preserving, and, in the last inequality, we have used (3.112), with $C = A$ and $D = ({}^k\Phi)^{-1}B$, and $\bar{t} > \tilde{t}_\epsilon(A,B) \ge t_\epsilon(A, ({}^k\Phi)^{-1}(B))$, for each $k = (0,...,n-1)$, from (3.113). □

For periodic systems, mixing is a stronger property than ergodicity. In order to see this notice that, for single transformations, Definition 3.2.7 in property (3.93) of Theorem 3.2.6, implies that a strongly mixing system is ergodic. Then use Theorem 3.2.7 and 3.2.10.

3.3 Entropy

Entropy in dynamical systems theory and ergodic theory was introduced as a measure of the randomness introduced by a transformation. In this respect, it is of obvious interest its study in mixing problems where the aim is to design a transformation or a sequence of transformations which introduces the largest possible amount of disorder in the system.

Assume a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$ is given. A finite, measurable, *partition* (in the following for brevity just partition) of the space M, is a collection α of measurable subsets of M, $\alpha := \{A_i\}_{i\in I}$, where I is a set of indices of finite cardinality, such that

$$A_i \cap A_j = \emptyset, \quad i \ne j, \quad (mod \quad \mu) \qquad (3.116)$$

$$\cup_{i\in I} A_i = M, \quad (mod \quad \mu). \qquad (3.117)$$

Given a partition $\alpha = \{A_i\}_{i\in I}$ and a partition $\beta = \{B_j\}_{j\in J}$, the product partition $\alpha \vee \beta$ is the partition consisting of sets which are intersections of the sets in α with the sets in β. A partial ordering can be defined on the set of all the partitions of M in the following way: Given two partitions α and β, we write $\alpha \le \beta$ if each set in β is a subset of a set in α. β is said to be *finer* than α. In particular, notice that $\alpha \vee \beta$

is finer than bith α and β. Given a partition $\alpha = \{A_i\}_{i\in I}$ and an automorphism Φ, the partition $\Phi\alpha$ is given by

$$\Phi\alpha := \{\Phi A_i\}_{i\in I}. \tag{3.118}$$

It is immediate to verify that

$$\Phi(\alpha \vee \beta) = \Phi\alpha \vee \Phi\beta. \tag{3.119}$$

The definition of entropy involves the following continuous, nonnegative and concave function $z(t)$, defined in $[0,1]$:

$$z(t) = \begin{cases} -t\log(t) & , \quad 0 < t \leq 1 \\ 0 & , \quad t = 0 \end{cases} \tag{3.120}$$

Definition 3.3.1: (Entropy of a partition) *The entropy of a partition $\alpha = \{A_i\}_{i\in I}$, denoted by $h(\alpha)$, is defined as*

$$h(\alpha) := \sum_{i\in I} z(\mu(A_i)). \tag{3.121}$$

Notice that, if N is the cardinality of the partition α, and if every set has measure $\frac{1}{N}$, a simple calculation shows that $h(\alpha) = logN$. Moreover every other (mod μ) partition with the same cardinality has strictly less entropy. This follows using the strict concavity of the function z; we have

$$h(\alpha) = \sum_{i=1}^{N} z(\mu(A_i)) = N\sum_{i=1}^{N} \frac{1}{N} z(\mu(A_i)) \leq Nz(\frac{1}{N}\sum_{i=1}^{N}\mu(A_i)) = Nz(\frac{1}{N}) = logN. \tag{3.122}$$

This also proves that the entropy of a finite partition is always finite. Some other important properties of the entropy of partitions are summarized in the following theorem.

Theorem 3.3.2 *Let $\alpha := \{A_i | i \in I\}$ and $\beta := \{B_j | j \in J\}$ be two arbitrary partitions.*
1) $h(\alpha) \geq 0$ with equality iff $\alpha = \{M\}$, where $\{M\}$, denotes the trivial partition consisting only of the set M;
2) $\alpha \leq \beta \rightarrow h(\alpha) \leq h(\beta)$, namely the entropy is a nondecreasing function of its argument;
3) $h(\alpha\vee\beta) \leq h(\alpha) + h(\beta)$, namely entropy is subadditive with respect to the operation of product.
4) Given an automorphism Φ defined on $(M, \mathcal{S}, \mu)$, then $h(\Phi\alpha) = h(\alpha)$.

Proof. 1). The fact that $h(\alpha) \geq 0$ is obvious from the definition. Also, it is obvious that if $\alpha = \{M\}$, $h(\alpha = 0)$. Conversely, if $h(\alpha) = 0$, it means that all the terms in the sum in (3.121) are zero. and this is possible, according to (3.120), only if $\alpha = \{M\}$. 2). Let us derive a general expression for $h(\alpha \vee \beta)$.

$$
\begin{aligned}
h(\alpha \vee \beta) &= -\sum_{i,j} \mu(A_i \cap B_j) log \mu(A_i \cap B_j) \\
&= -\sum_{i,j} \mu(A_i \cap B_j) log \frac{\mu(A_i \cap B_j)\mu(A_i)}{\mu(A_i)} \\
&= -\sum_{i,j} \mu(A_i \cap B_j) log \mu(A_i) - \sum_{i,j} \mu(A_i \cap B_j) log(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}) \\
&= -\sum_{i} \mu(A_i) log(\mu(A_i)) - \sum_{i,j} \mu(A_i \cap B_j) log(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}) \\
&= h(\alpha) + \sum_i \mu(A_i) z(\mu(B_j/A_i)),
\end{aligned}
\tag{3.123}
$$

where $\mu(B_j/A_i) := \frac{\mu(B_j \cap A_i)}{\mu(A_i)}$. Now notice that the second term in the right hand side of (3.123) is nonnegative, and that, if $\beta \geq \alpha$ then $\alpha \vee \beta = \beta$, and $h(\alpha \vee \beta) = h(\beta)$ in (3.123). Therefore 2) is proven. 3). Consider again formula (3.123) that we rewrite below

$$
h(\alpha \vee \beta) = h(\alpha) + \sum_j \sum_i \mu(A_i) z(\mu(B_j/A_i)),
\tag{3.124}
$$

and recall that since z is a concave function, we have $\sum_i \lambda_i z(x_i) \leq z(\sum_i \lambda_i x_i)$ for each finite combination of points $x_i \in [0,1]$ and coefficients $\lambda_i \in [0,1]$, with $\sum_i \lambda_i = 1$. Applying this in (3.124), we obtain

$$
\begin{aligned}
&\sum_j \sum_i \mu(A_i) z(\mu(B_j/A_i)) \leq \sum_j z(\sum_i \mu(A_i) \mu(B_j/A_i)) \\
&= \sum_j z(\sum_i \mu(A_i) \frac{\mu(A_i \cap B_j)}{\mu(B_j)}) = \sum_j z(\mu(B_j)) = h(\beta),
\end{aligned}
\tag{3.125}
$$

and using this in (3.124), we obtain 3). 4). This is an immediate consequence of the fact that Φ is measure preserving. □

Definition 3.3.3 (Entropy of a sequence of automorphisms with respect to a partition) *Given a partition α and a sequence of automorphisms $\{\Phi_t\}$, the entropy of $\{\Phi_t\}$ with respect to α, $h(\alpha, \{\Phi_t\})$, is given by the following limit (if it exists)*

$$
\begin{aligned}
& h(\alpha, \{\Phi_t\}) \\
:= \; & \lim_{k\to+\infty} \frac{1}{k} h(\alpha \vee (\Phi_1)^{-1}\alpha \vee \cdots \vee (\Phi_{k-1} \circ \cdots \circ \Phi_1)^{-1}\alpha) \qquad (3.126) \\
= \; & \lim_{k\to+\infty} \frac{1}{k} h(\alpha \vee (\Phi_1)\alpha \vee (\Phi_2 \circ \Phi_1)\alpha \vee \cdots \vee (\Phi_{k-1} \circ \cdots \circ \Phi_1)\alpha),
\end{aligned}
$$

where the equality of the two limits easily follows from 4) of Theorem 3.3.2 and (3.119).

The following theorem shows that the above limit exists for every partition α, for every periodic dynamical system.

Theorem 3.3.4. *Consider a periodic dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, of period T. For every partition α, the limit (3.126) exists*

Proof. Let α be a partition and consider the sequence of positive numbers

$$h_k = h(\alpha \vee \Phi_1^{-1}\alpha \vee \cdots \vee (\Phi_{k-1} \circ \cdots \circ \Phi_1)^{-1}\alpha), \qquad (3.127)$$

We have to prove that the limit

$$\lim_{k\to\infty} \frac{h_k}{k} \qquad (3.128)$$

exists. We notice that h_1,h_2,... are finite, since we are working with finite partitions. It is easily seen from properties 3) and 4) of Theorem 3.3.2, that for any $j \geq 0$, $0 \leq l \leq T-1$, we have

$$h_{jT+l} \leq h_{jT} + h_l, \qquad (3.129)$$

where we set $h_0 := 0$. Now consider the following subsequences of the sequence in (3.127) parametrized by l,

$$h_{jT+l}, \quad l = 0, 1, ..., T-1.$$

We prove that each subsequence $\frac{h_{jT+l}}{jT+l}$ has a limit, when j tends to infinity, which is the same for each l. Let us first prove that $\frac{h_{jT}}{jT}$ has a limit $\frac{L}{T}$. The proof relies on the fact that h_{jT} is a subadditive and nondecreasing sequence. Denote by $\tilde{h}_j := h_{jT}$. Since by subadditivity we have $\tilde{h}_j \leq j\tilde{h}_1$ and we also have that $\tilde{h}_1$ is finite, we can conclude that $\frac{\tilde{h}_j}{j} \leq \tilde{h}_1$ is bounded. As a consequence the $L := \liminf_{j\to\infty} \frac{\tilde{h}_j}{j}$ is finite.

Now, given some $\epsilon > 0$ we can choose a $\bar{j}$ sufficiently large such that $\frac{\tilde{h}_{\bar{j}}}{\bar{j}} < L + \epsilon$. For every j define $i(j)$ the least integer which is greater than or equal $\frac{j}{\bar{j}}$. We have

$$(i(j) - 1)\bar{j} < j \leq i(j)\bar{j}, \tag{3.130}$$

and

$$\frac{\tilde{h}_j}{j} \leq \frac{\tilde{h}_{i(j)\bar{j}}}{j} < \frac{\tilde{h}_{i(j)\bar{j}}}{(i(j)-1)\bar{j}} \leq \frac{i(j)\tilde{h}_{\bar{j}}}{(i(j)-1)\bar{j}} < \frac{i(j)}{i(j)-1}(L+\epsilon). \tag{3.131}$$

Letting j go to infinity in (3.131), we obtain that

$$L = \liminf_{j\to\infty} \frac{\tilde{h}_j}{j} \leq \limsup_{j\to\infty} \frac{\tilde{h}_j}{j} \leq L + \epsilon, \tag{3.132}$$

and since ϵ is arbitrarily small we have that lim sup = lim inf so that the limit exists.

To prove the convergence of $\frac{h_{jT+l}}{jT+l}$, for each $l = 1, ..., T-1$, notice that, from (3.129) and 2) and 3) of Theorem 3.3.2, we have

$$\frac{h_{jT}}{jT+l} \leq \frac{h_{jT+l}}{jT+l} \leq \frac{h_{jT}}{jT+l} + \frac{h_l}{jT+l}, \tag{3.133}$$

and letting j go to infinity, we have that all subsequences $\frac{h_{jT+l}}{jT+l}$ tend to the same limit $\frac{L}{T}$. Therefore, also the sequence $\frac{h_k}{k}$ tends to the same limit $\frac{L}{T}$ when k goes to infinity. □

It is not difficult to manufacture examples where the limit in (3.126) does not exist. Consider a single transformation Φ and a partition α. From the above theorem, it follows that $h(\alpha, \Phi)$ exists. Also, from the proof, it follows that it is finite. Consider the same partition and a sequence $\{\Phi_t\}$ composed by alternating of Φ and the identity map Id. Define $k(n)$ the number of times the map Φ appears in the sequence $\{\Phi_t\}$, in the first n steps. We have

$$\frac{1}{n}h(\vee_{j=0}^{n-1j}\Phi\alpha) = \frac{k(n)}{n}\frac{1}{k(n)}h(\vee_{j=0}^{k(n)}\Phi^j\alpha). \tag{3.134}$$

Letting $n \to \infty$ the last factor in the right hand side of the above expression tends to $h(\alpha, \Phi)$ but the factor $\frac{k(n)}{n}$, the density of maps Φ in n steps, may not have limit.

Definition 3.3.5 (Entropy of a sequence of automorphisms) *If the limit in (3.126) exists for every partition α for a sequence $\{\Phi_t\}$, then the entropy of the dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, $h(\{\Phi_t\})$, is given by*

$$h(\{\Phi_t\}) := sup_\alpha h(\alpha, \{\Phi_t\}). \tag{3.135}$$

Remark 3.3.6 Assume, for simplicity, that the partition α has just two elements A and A^c, that can represent the location of two different fluids at the beginning of the mixing process. After the first iteration in the sequence $\{\Phi_t\}$, Φ_1, some of the fluid in A will go into A^c. This part will be $\Phi(A) \cap A^c$. Similarly, some of the fluid in A^c, $\Phi(A^c) \cap A$, will go into the part that was originally occupied by A. These sets will be found in the partition $\alpha \vee \Phi_1\alpha$. At each step of the sequence, the various regions of the partition move and overlap different areas. If the mixing is good more and more regions overlap, the size of the intersection regions becomes small, the partition $\vee_{k=0}^{n-1}{}^k\Phi\alpha$ finer and, from 2) of Theorem 3.3.2 the entropy will be larger. In this sense, the entropy of a sequence of automorphisms, with respect to a partition measures the *asymptotic randomness* introduced by the sequence in the partition. The entropy of a sequence, introduced in Definition 3.3.5, being the supremum over all of the partitions, measures the asymptotic randomness that a sequence is *able* to produce in the phase space. We will come back to the issue of entropy as a measure of the randomness, in more quantitative terms in subsection 3.3.3.

3.3.1 Methods for the computation of entropy

There exist many methods for the computation of the entropy of a dynamical system. In this subsection we will only mention the fundamental Kolmogorov-Sinai theorem on generators and we will deal more in detail with Pesin's entropy formula which will be one of the major tools used in the following. All of the methods proposed in the literature deal with systems characterized by a single transformation. In this case, as well as in the more general periodic case, the limit (3.126) in the definition of entropy is guaranteed to exist and therefore the definition of entropy (Definition 3.3.5) is well posed. The following result is a tool to reduce the computation of the entropy for the periodic case to the one for the single transformation case.

Lemma 3.3.7. *Let $(M, \mathcal{S}, \mu, \{\Phi_t\})$ be a periodic sequence of measure preserving automorphisms of period T, and consider the associated composite automorphism $\bar{\Phi} := \Phi_T \circ \Phi_{T-1} \circ \cdots \circ \Phi_1$ and the dynamical system $(M, \mathcal{S}, \mu, \bar{\Phi})$. Then we have*

$$h(\bar{\Phi}) = Th(\{\Phi_t\}). \tag{3.136}$$

Proof. First we prove that:

$$h(\bar{\Phi}) \leq Th(\{\Phi_t\}). \tag{3.137}$$

Consider, for a given partition α, the sequence

$$\frac{\bar{h}_j(\alpha)}{T(j-1)+1} := \frac{h(\alpha \vee (\Phi_1)^{-1}\alpha \vee (\Phi_2 \circ \Phi_1)^{-1}\alpha \vee \cdots \vee (\Phi_T \circ \Phi_{T-1} \circ \cdots \circ \Phi_1)^{-j+1}\alpha)}{T(j-1)+1} \tag{3.138}$$

and notice that this is a subsequence of the sequence

$$\frac{h_k(\alpha)}{k} := \frac{h(\alpha \vee \Phi_1^{-1}\alpha \vee (\Phi_2 \circ \Phi_1)^{-1}\alpha \vee \cdots \vee (\Phi_{k-1} \circ \cdots \circ \Phi_1)^{-1}\alpha)}{k} \tag{3.139}$$

considered in (3.126). Therefore, using Theorem 3.3.4, the sequences (3.139) and (3.138) have the same limit, when k goes to infinity, which is by definition $h(\alpha, \{\Phi_t\})$. Moreover, it is

$$h(\alpha, \{\Phi_t\}) = \lim_{j \to +\infty} \frac{\bar{h}_j}{T(j-1)+1} = \lim_{j \to +\infty} \frac{\bar{h}_j}{Tj}. \tag{3.140}$$

Consider the sequence

$$\frac{h'_j(\alpha)}{j} := \frac{h(\alpha \vee (\Phi_T \circ \Phi_{T-1} \circ \cdots \circ \Phi_1)^{-1}\alpha \vee \cdots \vee (\Phi_T \circ \Phi_{T-1} \circ \cdots \circ \Phi_1)^{-j+1}\alpha)}{j}, \tag{3.141}$$

which, when j goes to infinity, tends to $h(\alpha, \bar{\Phi})$, by the definition of $\bar{\Phi}$. From 2) of Theorem 3.3.2, we get

$$\frac{h'_j(\alpha)}{j} \leq \frac{\bar{h}_j(\alpha)}{j}, \tag{3.142}$$

and using this and (3.140), we obtain

$$\frac{1}{T} h(\alpha, \bar{\Phi}) = \frac{1}{T} \lim_{j \to +\infty} \frac{h'_j}{j} \leq \lim_{j \to +\infty} \frac{\bar{h}_j}{Tj} = h(\alpha, \{\Phi_t\}). \tag{3.143}$$

Taking the supremum over all of the partitions α of the terms of (3.143), we obtain (3.137).

We now prove that

$$h(\bar{\Phi}) \geq T h(\{\Phi_t\}). \tag{3.144}$$

Choose a partition α. We show that there exists a partition $\tilde{\alpha}$, such that

$$h(\tilde{\alpha}, \bar{\Phi}) = T h(\alpha, \{\Phi_t\}). \tag{3.145}$$

Pick

$$\tilde{\alpha} := \alpha \vee (\Phi_1)^{-1}\alpha \vee \cdots \vee (\Phi_{T-1} \circ \cdots \circ \Phi_1)^{-1}\alpha. \tag{3.146}$$

We have

$$h(\tilde{\alpha}, \bar{\Phi}) = T \lim_{k\to+\infty} \frac{1}{Tk} h(\alpha \vee \Phi_1^{-1}\alpha \vee \cdots \vee (\Phi_{T-1} \circ \cdots \circ \Phi_1)^{-1}\alpha \vee (\Phi_T \circ \cdots \circ \Phi_1)^{-1}\alpha$$
$$\vee (\Phi_T \circ \cdots \circ \Phi_1)^{-1} \circ \Phi_1^{-1}\alpha \vee \cdots \vee (\Phi_T \circ \cdots \circ \Phi_1)^{-k+1} \circ (\Phi_{T-1} \circ \cdots \circ \Phi_1)^{-1}\alpha); \tag{3.147}$$

The limit on the right hand side of (3.147) is $h(\alpha, \{\Phi_t\})$, therefore, we have

$$h(\tilde{\alpha}, \bar{\Phi}) = Th(\alpha, \{\Phi_t\}). \tag{3.148}$$

Therefore, for each partition α, there exists a partition $\tilde{\alpha}$ such that (3.148) holds. Therefore we have

$$h(\bar{\Phi}) = \sup_\alpha h(\alpha, \bar{\Phi}) \geq T \sup_\alpha h(\alpha, \{\Phi_t\}) = Th(\{\Phi_t\}), \tag{3.149}$$

which is the inequality in (3.144). □

We have already mentioned, in this chapter, that the ergodic properties of the sequences $\{\Phi_t\}$ and $\{\Phi_{t+k}\}$, for any integer $k > 0$ are the same (see the proof of Theorem 3.1.10). In particular, if the ergodic theorem holds for one it also holds for the other. If ergodicity and/or mixing are verified for one they also are for the other and so on. This is quite natural since ergodic theory is the study of *asymptotic* average properties of dynamical systems that are the same for systems that coincide from one instant on. The situation is the same for what entropy is concerned. In particular, if a sequence coincides with another sequence from one instant on then the entropy limit 3.126) exists for the first sequence, for every partition, if and only if it exists for the second and the two sequences have the same entropy. In particular, from Theorem 3.3.4, the limit exists for a sequence which is periodic from a certain instant on and its value is the same as for the associated periodic sequence. We make this observation more precise and prove it in the next proposition.

Proposition 3.3.8. *Consider a dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$, and a dynamical system $(M, \mathcal{S}, \mu, \{\tilde{\Phi}_t\})$ and assume that there exists an integer $k \geq 0$ such that $\Phi_{t+k} = \tilde{\Phi}_t$ for each $t \geq 1$. Then the entropy limit (3.126) exists, for every partition α, for $\{\Phi_t\}$ if and only if it exists for $\{\tilde{\Phi}_t\}$*

Proof. Given the partition α, consider the partition $\tilde{\alpha} := (\Phi_k \circ \cdots \circ \Phi_1)\alpha$, we have, for $n \geq k+1$

$$\frac{1}{n} h(\vee_{j=0}^{n-k-1} {}^j\tilde{\Phi}\tilde{\alpha}) = \frac{1}{n} h(\vee_{j=0}^{n-1-k} {}^{k+j}\Phi_{k+1}\tilde{\alpha}) \tag{3.150}$$
$$\leq \frac{1}{n} h(\vee_{j=0}^{n-1} {}^j\Phi\alpha) \leq \frac{1}{n} h(\vee_{j=0}^{k-1} {}^j\Phi\alpha) + \frac{1}{n} h(\vee_{j=0}^{n-k-1} {}^j\tilde{\Phi}\tilde{\alpha}), \tag{3.151}$$

where we have used 2) and 3) of Theorem 3.3.2. From this expression, letting n go to infinity, we see that if the entropy limit (3.126) exists for $\{\tilde{\Phi}_t\}$ and $\tilde{\alpha}$ then it also exists for $\{\Phi_t\}$ and α, and conversely. Moreover

$$h(\{\Phi_t\}, \alpha) = h(\{\tilde{\Phi}_t\}, \tilde{\alpha}). \tag{3.152}$$

The above computation shows that for each partition α there exists a partition $\tilde{\alpha}$ such that (3.152) holds. The converse is also easily verified, the relation between the two partitions being $\tilde{\alpha} := (\Phi_k \circ \cdots \circ \Phi_1)\alpha$. Therefore we have

$$h(\{\Phi_t\}) = h(\{\tilde{\Phi}_t\}). \tag{3.153}$$

□

Remark 3.3.9. Lemma 3.3.7 in conjunction with Proposition 3.3.8, can be used to prove an interesting property of entropy of dynamical systems described by a single transformation, which is the classical case dealt with in the literature. Consider a dynamical system $(M, \mathcal{S}, \mu, \Phi)$ described by a single transformation Φ, and assume Φ is the composition of T automorphisms, sometimes referred to as *factors*, namely $\Phi = \Phi_T \circ \cdots \circ \Phi_1$. From Lemma 3.3.7, the entropy of Φ is T times the entropy of the sequence $\Phi_1, \Phi_2, \cdots := \{\Phi_t\}$. If we now consider the transformation $\bar{\Phi} := \Phi_k \circ \Phi_{k-1} \circ \cdots \circ \Phi_1 \circ \Phi_T \circ \cdots \circ \Phi_{k+1}$, for some k, $1 \leq k < T$ its entropy is equal to T times the entropy of the periodic sequence of transformations $\Phi_{k+1}, \Phi_{k+2}, \cdots, \Phi_T, \Phi_1, \cdots \Phi_k, \Phi_{k+1} \cdots := \{\bar{\Phi}_t\}$[2]. The entropy of this sequence of transformation is the same as the one of $\{\Phi_t\}$, by Proposition 3.3.8, because the two sequences coincide from one instant on, in particular after $T-k$ steps. It follows that the entropy of Φ and $\bar{\Phi}$ are the same. In a more concise way, we can say $h(\Phi_2 \circ \Phi_1) = h(\Phi_1 \circ \Phi_2)$, for any two automorphisms Φ_1 and Φ_2. Notice that this is that same as the well known property in linear algebra $Trace(AB) = Trace(BA)$, for two matrices A and B, and in fact we will see, in the rest of this subsection and in Chapter 5, that there exists an intimate relation between the entropy of a transformation and the properties, in particular the trace, of the corresponding Jacobian matrix.

Now, we briefly discuss and state, without proof, the Kolmogorov-Sinai theorem on generators. Consider $(M, \mathcal{S}, \mu, \Phi)$. Recall that given a set of subsets af M, $\mathcal{A} := (A_1, A_2, ...)$, $\mathcal{A}$ is said to be a generator of $\mathcal{S}$ if $\mathcal{S}$ is the smallest $\sigma-$algebra which contains the sets in $\mathcal{A}$. If α is a partition, it also identifies a set of subsets of M. α

[2]Notice we have reversed the indexes because a sequence of transformations is, in our convention, read from left to right while in the corresponding composite transformation, the maps appear from right to left

is said a *generator for* $(M, \mathcal{S}, \mu, \Phi)$, if the set of sets in the partition $\vee_{n=-\infty}^{+\infty} \Phi^n \alpha$ is said to be a generator for $\mathcal{S}$.

Theorem 3.3.10. (Kolmogorov-Sinai theorem on generators) *Assume α is a generator for $(M, \mathcal{S}, \mu, \Phi)$, then*

$$h(\Phi) = h(\alpha, \Phi). \tag{3.154}$$

Theorem 3.3.10 is one of the major tools for the computation of the entropy of an automorphism. The strategy is to choose a partition that can be proved to be a generator and then compute the entropy of the automorphism with respect to this partition.

To study Pesin's entropy formula, we have to introduce some further assumptions on the system under study, which are typical of the branch of ergodic theory known as *Smooth Ergodic Theory.*

We shall assume in the following of this section that M is a C^∞ differentiable manifold which is compact. We shall also assume that M is a surface, namely it has dimension 2. Although this assumption is not really necessary for the results that we will state, it will simplify results and notations and it will be good enough for the following development. M is a topological space, and as such it has defined on it a topology, which generate the Borel $\sigma-$ algebra $\mathcal{B}$. At any point $x \in M$, the tangent space is endowed with a norm $|| \cdot ||$. μ is any measure on $\mathcal{B}$, and Φ is an automorphisms which preserves the measure μ. We shall moreover assume that Φ is a C^2 diffeomorphism, defined for almost every $x \in M$ and with derivatives up to the second order uniformly bounded in M. These assumptions are sufficient to show the following theorem.

Theorem 3.3.11. (Oseledec's Multiplicative Ergodic Theorem) *For almost every point x in M, the tangent space T_xM can be decomposed in (at most) two subspaces $E_+(x)$ and $E_-(x)$, such that for each $v_1 \in E_+(x)$ the following limit exists and, it is independent of v_1*

$$\chi_1(x) = \lim_{n\to\infty} \frac{1}{n} log||D\Phi^n(x)v_1||, \tag{3.155}$$

and for every $v_2 \in E_-(x)$ the following limit exists and, it is independent of v_2

$$\chi_2(x) = \lim_{n\to\infty} \frac{1}{n} log||D\Phi^n(x)v_2||. \tag{3.156}$$

Moreover, the above decomposition of T_xM is unique and the functions $\chi_1(x)$ and $\chi_2(x)$ are measurable functions.

The functions χ_1 and χ_2 are called the *Lyapunov exponents* of Φ. The concept of Lyapunov exponent is a generalization of the concept of eigenvalue of a matrix to which it reduces if the Jacobian $D\Phi(x)$ is constant at each point x. The Lyapunov exponents measure the amplification of a vector $v_{1/2}$ in the tangent space under the iterates of the differential $D\Phi$. Moreover, the Lyapunov exponents measure the rate at which two trajectories starting close to each other diverge asymptotically. The easiest way of seeing this is noticing that in the limits (3.155) -(3.156) the vectors v_1 and v_2 are fixed and the limit only depends on how much $D\Phi^n(x)$ grows with n. Assume, for simplicity, that M is a subset of the Euclidean space R^2. If we perform a one term Taylor expansion of $\Phi^n(x_2) - \Phi^n(x_1)$, for two close points x_1 and x_2, we have

$$\Phi^n(x_2) - \Phi^n(x_1) \approx D\Phi^n(x_1)(x_2 - x_1), \tag{3.157}$$

which shows that the more $D\Phi^n(x_1)$ is large the more the distance between $\Phi^n(x_2)$ and $\Phi^n(x_1)$ is large. More specifically, if $x_2 - x_1$ has a component in the direction of the vector v_1 and χ^+ is positive, then the component will be amplified. Therefore, the Lyapunov exponents measure the rate of stretching of two trajectories starting close by. In the case in which Φ preserves the Lebesgue measure in R^2, $|det(D\Phi^n)| = 1$, and $\chi_1(x) = -\chi_2(x)$ for almost every x. The most important use of the concept of Lyapunov exponents is in Pesin's entropy formula which is given in the next theorem. In it we assume to be verified all of the assumptions above plus an extra assumption concerning the measure μ.

Theorem 3.3.12. (Pesin entropy formula) *Assume the measure μ in $(M, \mathcal{S}, \mu, \Phi)$ is absolute continuous with respect to the Lebesgue measure on M* [3]. *Define*

$$\chi(x) := \sum_{\chi_i \geq 0} \chi_i(x) dim E_i(x). \tag{3.158}$$

Then, the entropy of the dynamical system $h(\Phi)$ is given by the following formula

$$h(\Phi) = \int_M \chi(x) d\mu \tag{3.159}$$

In the special case of a volume preserving transformation Φ, there exists only one nonnegative Lyapunov exponent and the formula (3.159) reduces to

$$h(\Phi) = \int_M \chi_1(x) d\mu, \tag{3.160}$$

which is the one that we will use in the following.

[3] Recall the definitions in Subsection 2.1.2.

3.3.2 Entropy as a measure of disorder in dynamical systems

Other than giving a method for the computation of entropy in sufficiently smooth dynamical systems, Pesin entropy formula (3.160) provides an interpretation of entropy as a measure of the disorder introduced by a transformation in the state space. We have seen that the Lyapunov exponents measure the rate of stretching induced by a transformation. Therefore the integrals (3.159), (3.160) are the sums of the rate of stretching over the whole state space M, or (recall $\mu(M)=1$) the average rate of stretching introduced by the transformation in the state space. In fluid dynamics, the more the stretch is the more bounds between different parts of the fluid are broken, the better is the mixing. In Chapter 5 we will consider a prototypical problem of maximization of entropy and we will see that there exists an intimate relation between the shear stress as defined in fluid dynamics and entropy [4].

Another interpretation of entropy can be given by the use of the fundamental Shannon-McMillan-Breiman theorem that we will discuss in the rest of this subsection. We have already seen, in Remark 3.3.6, how the rate at which the sets in the partition $\vee_{k=0}^{n-1}\Phi^k\alpha$ get smaller with n is a measure of the disorder introduced by the transformation Φ. We will see below that these sets decrease exponentially with the entropy $h(\Phi,\alpha)$.

Given a partition α, the function $I_\alpha(x) = -\sum_{A\in\alpha} log\mu(A)\chi_A(x)$ is called *information function.* It measures the information that we gain when we know to which element of α belongs the point x. For example, if the set A is very small, then when $x \in A$ the value of the information function will be very large when, on the other hand, if $\alpha = \{M\}$ there is no uncertainty about which element of the partition α the point x belongs to, and the information function is identically zero. Notice that $h(\alpha) = \int_M I_\alpha(x)d\mu$, namely it is the average uncertainty associated with the partition α. We have seen, in Theorem 3.3.4, that $\lim_{n\to\infty}\frac{1}{n}\int_M I_{\alpha\vee\Phi_1\alpha\vee\cdots\vee^{n-1}\Phi\alpha}(x)d\mu$ exists for every periodic sequence of transformations and it is by definition $h(\alpha,\{\Phi_i\})$. Here we consider systems described by a single transformation, $(M,\mathcal{S},\mu,\Phi)$ and we assume the system to be ergodic. Under these assumptions, it is possible to prove a result stronger than Theorem 3.3.4, namely that the function $\frac{1}{n}I_{\alpha\vee\Phi\alpha\vee\cdots\vee\Phi^{n-1}\alpha}(x)$ converges almost everywhere to $h(\alpha,\Phi)$.

Theorem 3.3.12. (Shannon-McMillan-Breiman Theorem) *Consider an ergodic*

[4]Notice that this discussion is for single transformation dynamical systems and can be extended to periodic systems via Lemma 3.3.7.

dynamical system $(M, \mathcal{S}, \mu, \Phi)$. Then, for every finite measurable partition α,

$$\lim_{n\to\infty} \frac{1}{n} I_{\alpha\vee\Phi\alpha\vee\cdots\vee\Phi^{n-1}\alpha}(x) = h(\alpha, \Phi) \qquad a.e.x \in M. \tag{3.161}$$

A simple consequence for the sizes of the sets in the partition $\alpha \vee \Phi\alpha \cdots \Phi^{n-1}\alpha$ is the result below.

Corollary 3.3.13. (Entropy Equipartition Theorem) *Under the assumptions of Theorem 3.3.12, for every $\epsilon > 0$, there exists an n_ϵ such that the sets in $\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{n-1}\alpha$, can be divided into two sets $\mathcal{B}$ and $\mathcal{G}$, such that $\mu(\cup_{B\in\mathcal{B}}) < \epsilon)$ and for each set $G \in \mathcal{G}$*

$$e^{-n(h(\alpha,\Phi)+\epsilon)} < \mu(G) < e^{-n(h(\alpha,\Phi)-\epsilon)} \tag{3.162}$$

Proof. From Theorem 3.3.12 we have convergence almost everywhere of $I_{\vee_{k=0}^{n-1}\Phi^k\alpha}$to $h(\alpha, \Phi)$ and therefore convergence in measure. This gives that for any $\epsilon > 0$, there exists an n_ϵ such that

$$\mu(\{x \in M : |\frac{1}{n} I_{\vee_{k=0}^{n-1}\Phi^k\alpha} - h(\alpha, \Phi)| \geq \epsilon\}) < \epsilon. \tag{3.163}$$

Notice $I_{\vee_{k=0}^{n-1}\Phi^k\alpha}$ is constant on every set of $\vee_{k=0}^{n-1}\Phi^k\alpha$ and define $\mathcal{B}$ the sets B in $\vee_{k=0}^{n-1}\Phi^k\alpha$ such that, if $x \in B$, $|\frac{1}{n} I_{\vee_{k=0}^{n-1}\Phi^k\alpha} - h(\alpha, \Phi)| \geq \epsilon$. Obviously, if $n > n_\epsilon$, the union of all the sets in $\mathcal{B}$ has measure less that ϵ by (3.163). In the other sets, which belongs to $\mathcal{G}$, we have

$$|I_{\vee_{k=0}^{n-1}\Phi^k\alpha}(x) - h(\alpha, \Phi)| < \epsilon, \tag{3.164}$$

and evaluating (3.164) for $x \in G$, a set in $\mathcal{G}$, we have, recalling the definition of the information function $I_{\vee_{k=0}^{n-1}\Phi^k\alpha} := -\sum_{G\in\vee_{k=0}^{n-1}\Phi^k\alpha} log\mu(G)\chi_G(x)$,

$$-\frac{1}{n} log\mu(G) - h(\alpha, \Phi)| < \epsilon, \tag{3.165}$$

which immediately gives (3.162). □

3.4 Notes and References

The aim of the introduction to Ergodic Theory we have presented in this chapter has been twofold. On one hand we have introduced the basic concepts of the theory with particular attention to the ones that will be used in the following in dealing with problems of mixing for fluids. On the other hand we have introduced some new results and a new point of view for discrete time sequential systems namely for systems described by possibly different transformations at different times.

Ergodic theory is a very rich mathematical theory combining concepts of measure theory, functional analysis, dynamical systems and other fields of mathematics. There are many general introductory books on the subject such as [42], [32], [17], [26], [40].

A comprehensive book on the ergodic theorems is [20] although they are treated very well also in [32]. The extensions to the case of sequences as well as the Example 3.1.8, were first presented in [9]. The extension te sequences of Poincare' recurrence theorem (Theorem 3.2.4) is also given in [9]. The robustness result for mixing sequences (Theorem 3.2.9) was first presented in [11] along with the correspondence between mixing periodic sequences and mixing maps of Theorem 3.2.10. The entropy theory for periodic sequences presented in Section 3.3, is taken from [12]. An introduction to Pesin formula of Theorem 3.3.12 and Smooth Ergodic Theory can be found in [34] as well as in the supplement of [17]. A more advanced treatment dealing with the case of non-smooth maps can be found in [18].

The proofs of Kolmogorov-Sinai theorem on generators, Theorem 3.3.10, and Shannon-McMillan-Breiman Theorem 3.3.12 can be found in [32] (Theorem 5.3.1 and Theorem 6.2.3 , respectively). Osedelec's multiplicative ergodic Theorem 3.3.11 was first proved in [28] and then generalized in different directions in [39] [25]. A proof for the two-dimensional case dealt with in Theorem 3.3.11, which assumes the system to be ergodic is given in [34] (Theorem 2.1). Pesin entropy formula (Theorem 3.3.12) was proved in [33] and generalized in [18] to the case of maps with singularities.

Chapter 4

Some Facts from Fluid Dynamics

In this brief chapter, we collect some facts from fluid dynamics that will be needed in the following. The purpose of the chapter is to provide a link between tha mathematical concepts of ergodic and dynamical systems theory described in the two previous chapters and the physical phenomenon of fluid mixing.

4.1 Eulerian and Lagrangian point of view

The motion of a fluid is usually described adopting one of two different but complementary points of view: the Eulerian and the Lagrangian. In the Eulerian point of view the motion is described by a vector field of three functions of spatial coordinates and time, $u(x, y, z, t)$, $v(x, y, z, t)$, $w(x, y, z, t)$, each one specifying the x, y and z component of the velocity $\vec{V}$ in the region of space where the fluid motion occurs. In the Lagrangian point of view, the motion is described in a system of coordinates that translate with the motion of the particle that is under study. While a Lagrangian point of view is used in the proof of the conservation laws that regulate the fluid motion, the Eulerian quantities are used for a global description.

4.2 Kinematics

Here we recall some quantities and concepts used in fluid kinematics in order to describe the fluid motion. These quantities are defined in terms of the velocity components u, v and w.

Given an infinitesimal cube of fluid centered at the point x, y, z, the *vorticity* vector $\vec{\omega}$ measures the angular velocity experienced by the element of fluid in a plane perpendicular to the x, y and z direction. Mathematically vorticity is defined by $\vec{\omega} = (\omega_x, \omega_y, \omega_z)$, with $\omega_x := \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}$, $\omega_y := \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}$, $\omega_z := \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$, or in more compact notation $\vec{\omega} = \nabla \times \vec{V}$. The average angular velocity in a plane perpendicular to the $x(y, z)$ axis is $\frac{1}{2}\omega_x(\frac{1}{2}\omega_y, \frac{1}{2}\omega_z)$ and analogously for ω_y and ω_z. The *shear tensor* τ measures the rate at which the sides of a square get closer. If we denote $u_1 := u$, $u_2 := v$, $u_3 := w$, $x_1 := x$, $x_2 := y$, $x_3 := z$, we can write in compact notation

$$\tau_{ij} := \frac{1}{2}(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}), \quad i, j = 1, 2, 3. \tag{4.1}$$

For example, the quantity τ_{12} measures the angular velocity at which the two sides parallel to the x and y axis of cubic element get closer, and analogously for the other components of the tensor. The line integral of the velocity field $\vec{V}$ along a closed curve $\mathcal{L}$ in the fluid is called *circulation* Γ around the curve $\mathcal{L}$. The circulation Γ around the curve $\mathcal{L}$ is equal, by Stokes theorem, to the surface integral of the vorticity on a surface S enclosed by $\mathcal{L}$. In formulas

$$\Gamma := \int_{\mathcal{L}} \vec{V} d\vec{s} = \int_S \vec{\omega} \cdot \vec{n} dS. \tag{4.2}$$

If we follow the motion of a particle in a fluid under the velocity field $\vec{V}(x, y, z, t) := (u(x, y, z, t), v(x, y, z, t), w(x, y, z, t))$, the line which gives the trajectory of this particle is called *pathline*. Lines in the region where the fluid motion occurs that are tangent, at each time to the velocity field $\vec{V}$ are called *streamlines*. In an *unsteady flow*, since the velocity field depends explicitly on time, the streamlines will change with time and will not coincide in general with the pathlines. In *steady flows* , instead, streamlines will be constant in time and will coincide with pathlines.

4.3 Governing equations

A blob of fluid as observed flowing under the action of the velocity field $\vec{V}(x, y, z, t)$ will change its shape but not its mass. This is the *principle of conservation of mass* which is expressed mathematically by the statement

$$\frac{\partial \rho}{\partial t} + \sum_{k=1}^{3} \frac{\partial(\rho u_k)}{\partial x_k} = \frac{\partial \rho}{\partial t} + div(\rho \vec{V}) = 0, \tag{4.3}$$

where $\rho = \rho(x_1, x_2, x_3, t)$ is the density of the fluid as a function of position and time. If we assume, as it will be done from now on, that the fluid is incompressible, namely

$\rho = const$, then the equation of conservation of mass reduces to

$$\sum_{k=1}^{3} \frac{\partial u_k}{x_k} = div\vec{V} = 0. \tag{4.4}$$

In the case of a two dimensional flow, $u_3 = constant$ and equation (4.4) reduces to

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} = 0. \tag{4.5}$$

This can be also expressed saying that there exists a function $\psi(x_1, x_2, t)$ such that

$$u_1 = \frac{\partial \psi(x_1, x_2, t)}{\partial x_2} \tag{4.6}$$

$$u_2 = -\frac{\partial \psi(x_1, x_2, t)}{\partial x_1}, \tag{4.7}$$

since replacing (4.6), (4.7) into (4.5) gives an identity. Notice that the system $\dot{x}_1 = u_1$, $\dot{x}_2 = u_2$, under the conditions (4.6), (4.7) is an Hamiltonian system where the role of the Hamiltonian function is played by the function ψ (cfr. end of Subsection 2.2.1). The lines $\psi(x, y, t) = constant$ are, at each instant, the streamlines of the flow. They vary with the time t. If ψ does not depend on t, and the flow is steady, the streamlines are fixed, and they are actually invariant manifolds for the trajectories of the system. If $x_1(0)$ and $x_2(0)$ are such that $\psi(x_1(0), x_2(0)) = c$, then we have using (4.6) and (4.7),

$$\frac{d}{dt}\psi(x_1(t), x_2(t)) = \frac{\partial \psi}{\partial x_1} u_1 + \frac{\partial \psi}{\partial x_2} u_2 = 0. \tag{4.8}$$

Therefore $\psi(x_1(t), x_2(t)) = c$ along a trajectory and every trajectory moves along a streamline. This also shows, by uniqueness of solution of differential equations, that no trajectory will cross a streamline. Therefore if we have two fluids on two different sides of a streamline in a *steady* two dimensional flow, they will never get mixed and this implies that in order to have mixing, we have to induce, in the fluid an *unsteady* motion. This observation has motivated the study of ergodic theory for sequences of transformations as presented in the previous chapter.

Another fundamental equation is the conservation of momentum that is essentially Newton's second law applied to an infinitesimal element of fluid. This equation is a vectorial equation consisting of three scalar equations given, for $j = 1, 2, 3$, by

$$\rho \frac{\partial u_j}{\partial t} + \rho \sum_{k=1}^{3} u_k \frac{\partial u_j}{\partial x_k} = \sum_{k=1}^{3} \frac{\partial \sigma_{kj}}{\partial x_k} + \rho f_j. \tag{4.9}$$

In (4.9) the right hand side is the $\vec{F}$ part of Newton's law and the left hand side is the $m\vec{a}$ part (everything is divided by the volume). $\sigma_{kj} := \sigma_{kj}(x, y, z, t), k, j = 1, 2, 3$ is the 2-tensor measuring the shear stress acting on a infinitesimal cubic element of fluid under the action of the velocity field. More precisely, σ_{kj} is the force per unit area applied in the direction of the axis x_j to the face of the infinitesimal cubic element centered at x, y, z and perpendicular to the axis x_k. f_j is the *body force* (e.g. gravitational, electromagnetical) per unit mass in the direction j. Notice that the whole equation (4.9) has dimensions of a force over volume. For *Newtonian fluids* that are incompressible, the shear stress tensor σ is related to the tensor rate of shear τ and the pressure p by the relation

$$\sigma_{kj} := -p\delta_{kj} + \nu\tau_{kj} = -p\delta_{kj} + \nu(\frac{\partial u_k}{\partial x_j} + \frac{\partial u_j}{\partial x_k}), \tag{4.10}$$

where δ_{kj} is the Kroneker δ, namely $\delta_{kj} = 0$ if $k \neq j$ and $\delta_{kj} = 1$, if $k = j$. ν is a constant called *dynamic viscosity* (or just *viscosity*) which measures the rate at which the shear stress (force) affects the rate of shear (kinematic quantity). Using (4.10) into (4.9), one gets the Navier-Stokes equations for an incompressible, Newtonian, fluid with constant viscosity:

$$\rho\frac{\partial u_j}{\partial t} + \rho\sum_{k=1}^{3} u_k\frac{\partial u_j}{\partial x_k} = -\frac{\partial p}{\partial x_j} + \nu\sum_{k=1}^{3}\frac{\partial^2 u_j}{\partial x_k^2} + \rho f_j. \tag{4.11}$$

These are, along with the continuity equation (4.4), the basic equations that a vector field $\vec{V}$, describing the motion of a fluid, has to satisfy.

4.4 Fluid motion as a dynamical system

Assume a fluid motion occurs in a bounded region of the space M that we assume to have normalized area or volume, namely Lebesgue measure, equal to one. Every subset of M is the region occupied by some fluid and can be seen as a set in a $\sigma-$ algebra $\mathcal{S}$, with the Lebesgue measure μ given by its area (or volume). The fluid evolves according to the laws described in this Chapter. In particular, assuming constant density ρ, its velocity field $\vec{V}$ satisfies the continuity equation (4.4) and it is therefore divergence free. This implies by Liouville's Theorem 2.2.3, that the flow associated with $\vec{V}$, which is denoted by Φ_t, preserves the Lebesgue measure. So does every discretization of Φ_t, $\Phi := \Phi_{\Delta t}$, so that we have a well defined continuous time dynamical system $(M, \mathcal{S}, \mu, \Phi_t)$ or discrete time dynamical system $(M, \mathcal{S}, \mu, \Phi)$. If the law that regulates the motion changes abruptly at some given instants of time $t_1, t_2, t_3, ...$, every velocity field has to satisfy the governing equations (4.4), (4.11), after discretization at times $t_1, t_2, t_3, \ldots$ the motion will be described by a sequential dynamical system $(M, \mathcal{S}, \mu, \{\Phi_t\})$.

4.5 Mixing

Mixing of fluid in general involves several phenomena which are related to the specific situation at hand. The properties of fluids change enormously from one case to the other. Assume a fluid A has to be mixed with a fluid B. A and B can be *miscible* meaning that we can draw an arbitrarily thin line of A inside B, or *immiscible* meaning that, after stretching A inside B for a while A will break. The latter mechanisms is called *break-up.* If the region occupied by A is sufficiently thin inside B, spontaneous *diffusion* will complete the mixing process. Chemical reactions can happen changing drastically the composition of the substances involved in the process. Moreover, the fluids to be mixed can have different densities or viscosities.

A mathematical theory of control of mixing has to abstract from the physical aspects related to specific situations and focus on the phenomenon that is more easily controlled, namely *mechanical mixing.* In Fig. 4.1 the basic mechanisms involved in mechanical mixing namely *stretching* and *folding* of material lines is described. Two blobs of fluid, black and white, are stretched and folded until the whole fluid will become grey. The dynamics is similar to the one of an horseshoe map, which is the canonical example of chaos in dynamical systems.

In this work we adopt the dynamical system, measure theoretic, model described in the previous section, for the motion of a fluid in a *bounded region* M of $\mathbf{R}^n$. This model is a good description of the real process if we assume that 1) the fluids to be mixed have similar physical properties; 2) They are perfectly miscible; 3) no chemical reactions happen; 4) There is no diffusion and no breakup (however these processes increase mixing, so that the theory is conservative in this respect). Under these assumptions the problem will be how to shape the velocity profiles, namely the transformations $\{\Phi_t\}$ in the measure theoretic model, in order to achieve good mixing. Mathematical precise meaning of what is meant by 'good mixing' is given by the ergodic theoretic concepts of Mixing and Entropy. In particular $\{\Phi_t\}$ has to be chosen mixing and with as large as possible entropy.

4.6 Notes and References

Standard textbooks in Fluid Mechanics include [4], [21], [27], [8].

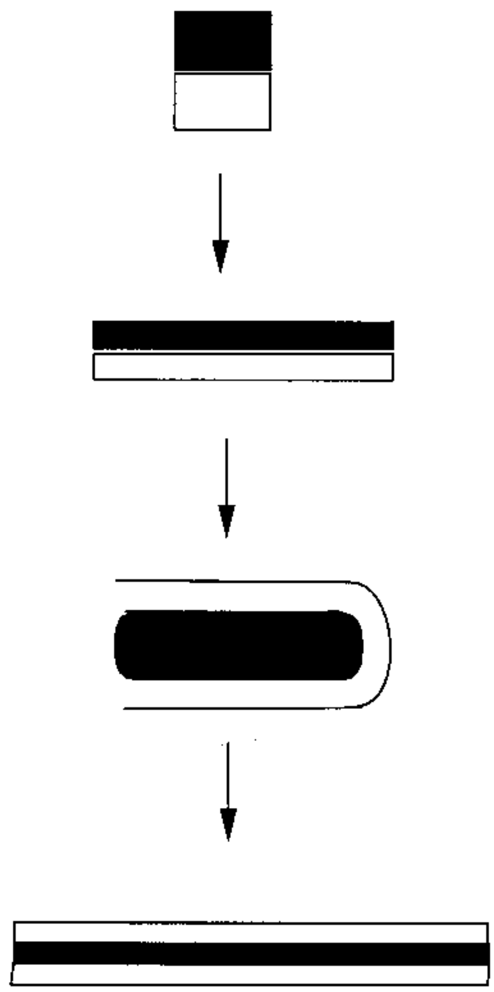

Figure 4.1: Basic Stretching and Folding Mixing Mechanism

Chapter 5

The prototypical flow

The *prototypical flow* is a simplified two-dimensional model which describes the basic mechanisms of *stretching* and *folding* that give raise to mechanical mixing (cfr. Fig. 4.1). The flow is assumed to be a periodic sequence of purely horizontal (H) and vertical (V) velocity profiles on a two-dimensional torus (see Fig. 5.1). The mathematical model assumes that the fluids to be mixed have similar properties (in particular the same viscosity ν). More specifically the flow is assumed to be described in local coordinates on the torus (x, y), $0 \leq x < 1$, $0 \leq y < 1$, and on a certain interval of time, by

$$\begin{aligned} \dot{x} &= h(y), \qquad (mod \quad 1) \\ \dot{y} &= 0, \end{aligned} \tag{5.1}$$

for the purely horizontal (H) flow, or as

$$\begin{aligned} \dot{x} &= 0, \\ \dot{y} &= v(x), \qquad (mod \quad 1) \end{aligned} \tag{5.2}$$

for the purely vertical (V) flow. In the above equations, h and v are functions which we assume C^2 except on an at most countable set of points. Therefore we allow discontinuous flows. In particular, if $h(0) \neq \lim_{y \to 1} h(y)$, h is not continuous, since the flow is on a torus. Analogously for v. Notice also that the velocity profiles (5.1), (5.2) automatically satisfy (a.e.) the continuity equation (4.5) and the Navier-Stokes equations (4.11) stated in the previous chapter.

It is easily seen why a flow composed of alternating velocity profiles (5.1) (5.2) describes a *stretching and folding* mechanism. The action (5.1) (or (5.2) induce some stretching in the flow since, in general, there exists some gradient of the velocity with

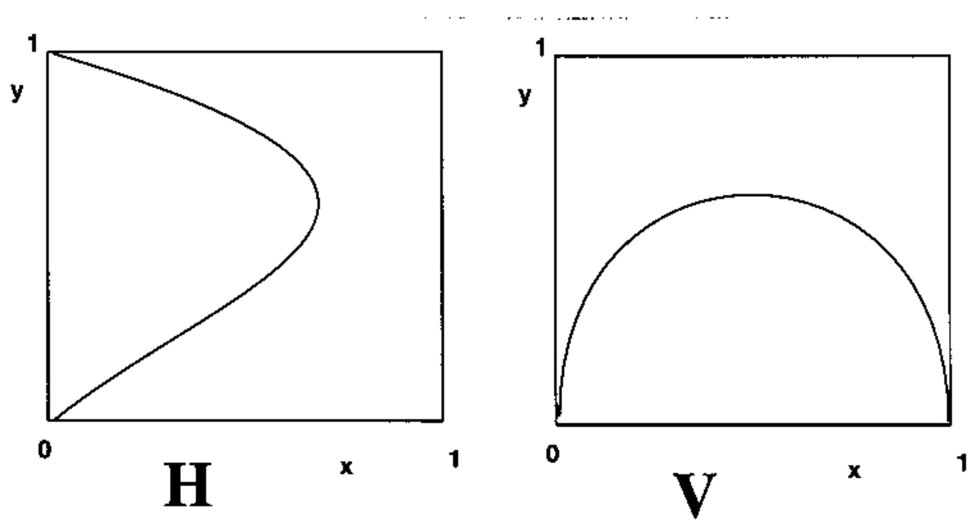

Figure 5.1: Purely horizontal and vertical velocity profiles on a torus

y (or x). The fact that an H (or V) action is followed by a V (or H) action induces a bending in the trajectories of the flow which gives the folding mechanism.

The problem that we consider consists of finding a sequence of actions H and V as specified in (5.1) (5.2), that mixes best. More specifically, we will search for a periodic sequence of transformations that achieves the maximum ergodic theoretic entropy and it is mixing in the ergodic theoretic sense. The criterion of maximization of entropy is motivated by the fact that this is a measure of the randomness induced by a transformation. We will see that once the entropy criterion is satisfied, the optimal sequence will also be 'mixing'.

The control design problem that we consider will not only be on the choice of the sequence to be applied to the flow, but also on the shape of the functions h and v. Therefore, the horizontal and vertical action in (5.1) (5.2) are allowed to change each time. Therefore, for example, if V_1 is the velocity profile in (5.2) with $v = v_1$, and H_1 the one in (5.1) with $h = h_1$, and V_2 the one in (5.2) with $v = v_2$, for some admissible functions v_1, h_1, v_2, a possible sequence will be a periodic one with period 6 equal to $V_1^3 \circ H_1 \circ V_2^2$ (for notational convenience, we will not deal with sequences with subsequent compositions of different vertical (or horizontal) maps of the type $V_i \circ V_{i+1}$ (or $H_i \circ H_{i+1}$)).

5.1 Formulation of the optimal flow planning problem

Discretizing (5.1), (5.2), we obtain the map of the type H (purely horizontal)

$$\begin{aligned} x_{t+1} &= x_t + h(y_t) \quad (mod \quad 1), \\ y_{t+1} &= y_t, \end{aligned} \tag{5.3}$$

and the map of the type V (purely vertical)

$$\begin{aligned} x_{t+1} &= x_t, \\ y_{t+1} &= y_t + v(x_t) \quad (mod \quad 1). \end{aligned} \tag{5.4}$$

The flow is a sequence of purely horizontal (H) and vertical (V) flows, whose velocity profiles are determined by the shapes of the functions h and v.

The problem that we consider in this chapter consists of finding a periodic sequence of actions H and V that achieves the maximum ergodic theoretic entropy and it is mixing in the ergodic theoretic sense. We will see that once the entropy criterion is satisfied, the optimal sequence will also be 'mixing'. In addition, using Theorem 3.2.9, we can say that every sequence that converges to the maximum entropy sequence, in a sufficiently fast manner, is also mixing.

The constraint in the entropy optimization problem is a bound on the first derivatives of these functions. In particular, we shall assume

$$|\frac{dh}{dy}| \leq a, \tag{5.5}$$

$$|\frac{dv}{dx}| \leq b. \tag{5.6}$$

This constraint regularizes the problem, since the expression of the maximal entropy which we will derive here tends to infinity when the parameters a and b tend to infinity. The quantity

$$|\nu \frac{dh}{dy}|_{(x_0, y_0)}, \tag{5.7}$$

where ν is the viscosity of the fluid, represents the shear stress (see definitions in Section 4.2) induced on the fluid by the velocity field in (5.1) at the point (x_0, y_0). An analogous physical interpretation holds for $|\nu \frac{dv}{dx}|$. Therefore the entropy optimization problem is posed with a constraint on the possible shear stress on the fluid.

We will restrict ourselves to periodic sequences. However all of the results that we will present can be extended to sequences that are equal to periodic sequences after a finite interval of time. In fact, it has been proven in Proposition 3.3.8 that the definition of entropy is well posed in this case as well as in the periodic case (see the limit (3.126)), and the entropy of the sequence is the same as the corresponding periodic one. Also, the ergodic theoretic mixing properties of the two sequences can be proven to be the same by Theorem 3.2.9. We have seen in Chapter 3 that, for periodic sequences of transformations, many concepts developed in standard ergodic theory, which deals with single transformations, can be naturally extended and studied in terms of the corresponding composite single map.

We shall denote the set of periodic sequences composed by H and V by $\mathcal{P}$ and the set of measure preserving automorphisms obtained by composition of the maps H and V of length T, by $\mathcal{P}_S(T)$. The result of Lemma 3.3.7 allows us to compare entropy of periodic sequences by comparing the entropy of the corresponding composite automorphisms. Gathering the above definitions and properties, we can summarize the problem to be solved in the next sections as the one of finding (subject to (5.5) (5.6))

$$\max_{\{\Phi_t\}\in\mathcal{P}} h(\{\Phi_t\}), \tag{5.8}$$

or equivalently as finding

$$\max_{T} \max_{\Phi\in\mathcal{P}_S(T)} \frac{h(\Phi)}{T}. \tag{5.9}$$

5.2 Maximum Stretch Maps

So far we have introduced two concepts dealing with the stretching experienced by a trajectory of a particle undergoing motion. In ergodic and dynamical systems theory, we have seen that Lyapunov exponents measure the rate at which two trajectories starting close by diverge asymptotically, under the action of the velocity field (see Subsection 3.3.1). In fluid dynamics, the rate of shear tensor τ or alternatively the shear stress tensor σ, (that are related for Newtonian fluids through the viscosity ν (see (4.10)) measure the stretch undergone by the fluid, or equivalently, the velocity gradient. There should be a relation between the velocity gradient and how much two close by trajectories diverge. Therefore, in our problem, in order to increase the Lyapunov exponents and therefore the entropy (see Theorem 3.3.12), it seems opportune to use at each step all of the rate of stretch allowed by the bounds (5.5), (5.6). This is indeed the case, and it is made precise by the following theorem. This theorem allows us to restrict the search for the optimal sequence to sequences composed of maps with maximum stretch, defined as the transformations $|\bar{H}|$ and

$|\bar{V}|$ in (5.3) and (5.4) respectively, with $h(y) = ay$ and $v(x) = bx$, and a and b given in (5.5), (5.6). We first prove a simple lemma about the entropy of diffeomorphisms with constant differential.

Lemma 5.2.1. *Let $(M, \mathcal{S}, \mu, \Phi_A)$ a dynamical system, and assume Φ_A has constant differential A defined almost everywhere on the smooth manifold M. Assume μ is the volume Lebesgue measure on M. The entropy of $(M, \mathcal{S}, \mu, \Phi_A)$ is given by*

$$h(\Phi_A) = \log|\lambda_{max}(A)|, \tag{5.10}$$

where $\lambda_{max}(A)$ is the eigenvalue of A with maximum magnitude.

Proof. From Pesin's formula (3.159), the entropy of $(M, \mathcal{S}, \mu, \Phi_A)$ is given by

$$h(\Phi_A) = \int_M \chi(x, y) d\mu, \tag{5.11}$$

where $\chi(x, y)$ is the sum of positive Lyapunov exponents as defined in formulas (3.155) (3.156). In our case $D_{(x,y)}\Phi_A^n = A^n$ and $detA = 1$ because Φ_A preserves the Lebesgue measure. Assume that A has two eigenvalues with magnitude equal to 1. In that case, the limit in (3.155), (3.156), for every v, is zero, and so is the value of the entropy in (5.11). If A has two different eigenvalues λ_{min} and λ_{max} with $|\lambda_{min}| = \frac{1}{|\lambda_{max}|} < 1$, we compute the Lyapunov exponent in (3.155), using a decomposition of v as $v = \alpha_1 v_{min} + \alpha_2 v_{max}$ where v_{min} and v_{max} are the eigenvectors corresponding to λ_{min} and λ_{max}, respectively. In particular we have

$$\lim_{n\to+\infty} \frac{1}{n} \log ||D_{(x,y)}\Phi_A^n v|| = \lim_{n\to+\infty} \frac{1}{n} \log ||A^n v||$$
$$= \lim_{n\to+\infty} \frac{1}{n}((\log(|\lambda_{max}|^n) + \log(||\alpha_2 v_{max}||))) = \log|\lambda_{max}|.$$

Plugging this into (5.11), we obtain (5.10). □

Theorem 5.2.2. *Consider a periodic sequence of purely horizontal and vertical transformations on the torus of period T. Define the associated composite transformation of the form $S := H_1^{n_1} \circ V_2^{n_2} \circ \cdots \circ H_{s-1}^{n_s-1} \circ H_s^{n_s}$, with $n_1 + \cdots + n_s = T$. The entropy of this sequence is less than or equal to the entropy of the sequence with composite transformation $\bar{S} := |\bar{H}|^{n_1} \circ |\bar{V}|^{n_2} \circ \cdots \circ |\bar{H}|^{n_s-1} \circ |\bar{H}|^{n_s}$.*

Proof. In the following, we denote by $h(S)$ and by $h(\bar{S})$ the entropy of the *sequence* characterized by S and $\bar{S}$, respectively. Observe that the differential $D\bar{S}$ is a constant matrix, defined almost everywhere, and therefore it easily follows, from Lemma 3.3.7 and Lemma 5.2.1 above, that $h(\bar{S}) = \frac{1}{T}log\rho(D\bar{S})$, where $\rho := |\lambda_{max}|$ denotes the spectral radius of a matrix. Also, from the fact that Φ_A preserves the area, Lemma 3.3.7 and (3.155) and (3.160) we have $h(S) = \frac{1}{T}\int_M \chi_S(x)d\mu$, where $\chi_S(x)$ (the positive Lyapunov exponent) is given by

$$\chi_S(x) := \lim_{m\to\infty} \frac{1}{m} log||DS_m DS_{m-1}\cdots DS_1 v||, \tag{5.12}$$

for almost every vector $v \in R^2$. In the notations in (5.12) the subscripts indicate that the differentials are computed at m (possibly) different points. We know, from a linear algebra result (see e.g. [15][Lemma 5.6.10, pg. 297]) that for each $\epsilon > 0$ there exists a norm $||\cdot||_{ch}$ such that

$$||D\bar{S}||_{ch} \le \rho(D\bar{S}) + \epsilon. \tag{5.13}$$

Considering the 1-norm, we notice that, for every m,

$$||DS_m DS_{m-1}\cdots DS_1||_1 \le ||(D\bar{S})^m||_1. \tag{5.14}$$

This follows because every element of $D\bar{S}$ is positive and it is larger than or equal to the corresponding element in DS, wherever this element is evaluated. This follows from the constraints (5.5) (5.6).
Notice now that, for each m,

$$\begin{aligned} ||DS_m DS_{m-1}\cdots DS_1||_{ch} &\le C_1||DS_m DS_{m-1}\cdots DS_1||_1 \\ &\le C_1||(D\bar{S})^m||_1 \le C_2C_1||(D\bar{S})^m||_{ch}, \end{aligned} \tag{5.15}$$

for some constants C_1, C_2. The first inequality follows from equivalence of norms, the second one from (5.14) and the third one again from equivalence of norms. Set $C := C_1C_2$, we have, from (5.15), for every m and every $v \in R^2$ (here v is assumed with norm 1, without loss of generality)

$$\begin{aligned} \frac{1}{m}log||DS_m DS_{m-1}\cdots DS_1 v||_{ch} &\le \frac{1}{m}log||DS_m DS_{m-1}\cdots DS_1||_{ch} \\ &\le \frac{1}{m}log(C||(D\bar{S})^m||_{ch}) \le \frac{1}{m}logC + log||D\bar{S}||_{ch}, \end{aligned} \tag{5.16}$$

where, in the last inequality, we used the submultiplicative property of the norm $||\cdot||_{ch}$. Recalling now (5.13) and taking the limit in (5.16) when m tends to infinity, we obtain, for a.e. $x \in M$,

$$\chi_S(x) \le log(\rho(D\bar{S}) + \epsilon) \tag{5.17}$$

and since ϵ is arbitrarily small, we obtain

$$\chi_S(x) \le log(\rho(D\bar{S})); \tag{5.18}$$

Integrating and dividing by T both sides we obtain

$$h(S) \leq h(\bar{S}). \tag{5.19}$$

□

5.3 Derivation of the sequence with maximum entropy

From now on, we will deal with the problem in terms of the maximum stretch maps $|\bar{H}|$, $|\bar{V}|$ introduced in the previous section, that we will denote, with some abuse of notation simply as H and V. We will also denote with the same notation, H and V, the matrices representing the differentials of the maps H and V, since no confusion should occurr. The maps H and V therefore are given from now on by

$$\begin{aligned} x(t+1) &= x(t) + ay(t) \quad (mod\ 1), \\ y(t+1) &= y(t)), \end{aligned} \tag{5.20}$$

and

$$\begin{aligned} x(t+1) &= x(t), \\ y(t+1) &= y(t) + ax(t) \quad (mod\ 1), \end{aligned} \tag{5.21}$$

respectively. Both of these maps are defined on the unit square $M = \{(x, y) \mid 0 \leq x < 1,\ 0 \leq y < 1\}$, and any time a point exits M it is projected back to it by the *mod* 1 operation. These maps are the discretizations of the flow

$$\begin{aligned} \dot{x} &= ay \quad (mod\ 1), \\ \dot{y} &= y, \end{aligned} \tag{5.22}$$

and

$$\begin{aligned} \dot{x} &= x, \\ \dot{y} &= ax \quad (mod\ 1), \end{aligned} \tag{5.23}$$

respectively. Notice that for each H−type of motion the total kinetic energy of the flow E is given by

$$E = \frac{1}{2} w \int_M \rho a^2 y^2 dx dy = \frac{1}{6} a^2, \tag{5.24}$$

where w is the thickness of the layer of fluid we are considering. Therefore, the kinetic energy per horizontal step is fixed once we have fixed a. The same is true for a vertical motion once we have fixed b. Therefore the whole problem can be interpreted as a maximum entropy problem with constant kinetic energy per step.

In the following we shall denote by $\mathcal{S}(T)$ the set of matrices obtained from $T-$products of the matrices H and/or V. There is an obvious one to one correspondence between the elements in $\mathcal{S}(T)$ and the sequences of T maps H and V. In view of Lemma 5.2.1, for any fixed value of the period T, we look for

$$\max_{S\in\mathcal{S}(T)} |\lambda_{max}(S)|. \tag{5.25}$$

Notice that the eigenvalues of matrices with determinant equal to 1, such as the ones in $\mathcal{S}(T)$, can be computed by

$$\lambda_{1,2} = \frac{Tr(S) \pm \sqrt{(Tr(S))^2 - 4}}{2}. \tag{5.26}$$

Notice also that it can be easily shown (see formula (5.35) below) that for any element S in $\mathcal{S}(T)$, we have $Tr(S) \geq 2$. Therefore, we have, from formula (5.26) that the maximum eigenvalue is an increasing real invertible function of the trace. In view of this observation, we can restate our problem as

$$\max_{T} \max_{S\in\mathcal{S}(T)} Tr(S). \tag{5.27}$$

And for any fixed T, we have to solve the following problem

$$\max_{S\in\mathcal{S}(T)} Tr(S). \tag{5.28}$$

It will turn out that the solution of (5.28) is the same for every T, therefore when (5.28) is solved, it will be immediate to solve (5.27), and therefore the original problem (5.8).

In the next subsections, we first tackle the problem in the special case $a = b$. This will be done in the next Subsections 5.3.1 and 5.3.2. The extension to the case $a \neq b$ will be done in Subsection 5.3.3.

5.3.1 Auxiliary Results

Assume

$$H = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \tag{5.29}$$

and

$$V = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}. \tag{5.30}$$

We develop here few algebraic relations for the $T-$products of the matrices H and V in (5.29) and (5.30). We notice that

$$H = V^T, \tag{5.31}$$

and, also,

$$H = PVP, \tag{5.32}$$

where P is the permutation matrix $P := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Notice that, using (5.32) and the fact that the trace does not change with similarity transformations, we have the following property, where S_i, $i = 1, ..., n$ are matrices H or V, and $\bar{S}_i$, $i = 1, ..., n$ is H if S_i is V and viceversa.

$$Tr(S_1S_2 \cdots S_n) = Tr(PS_1S_2 \cdots S_nP) = Tr(PS_1PPS_2PP \cdots PPS_nP) = Tr(\bar{S}_1\bar{S}_2 \cdots \bar{S}_n). \tag{5.33}$$

In view of the above relation, in maximizing the trace in (5.27), we can restrict our attention to the elements of $\mathcal{S}(T)$ such that the first matrix from the left is H. Also, from the property of the trace

$$Tr(AB) = Tr(BA), \tag{5.34}$$

when we consider elements of $\mathcal{S}(T)$ other than H^T, we can always restrict our attention to matrix products such that V is the last matrix on the right. Therefore the product which maximizes the trace, for any n, is in the subset of $\mathcal{S}(T)$ given by H^T along with the products of the form $H^{n_1}V^{n_2} \cdots V^{n_s}$, for a certain $s \geq 2$, with $n_1 + n_2 + \cdots + n_s = T$. The formula for the trace of this matrix can be computed explicitly. We have

Lemma 5.3.1.

$$Tr(H^{n_1}V^{n_2} \cdots V^{n_s}) =$$

$$2 + \Big(\sum_{\substack{i_1 < i_2 \\ i_1 \not\sim i_2}} n_{i_1}n_{i_2}\Big)a^2 + \Big(\sum_{\substack{i_1 < i_2 < i_3 < i_4 \\ i_1 \not\sim i_2 \\ i_2 \not\sim i_3 \\ i_3 \not\sim i_4}} n_{i_1}n_{i_2}n_{i_3}n_{i_4}\Big)a^4 + \cdots + n_1n_2 \cdots n_s a^s, \tag{5.35}$$

where $\not\sim$ denotes the relation of two integers to be neither both odd nor both even.

Proof. First, we notice that

$$H^k = \begin{pmatrix} 1 & ka \\ 0 & 1 \end{pmatrix}, \tag{5.36}$$

and analogously

$$V^k = \begin{pmatrix} 1 & 0 \\ ka & 1 \end{pmatrix}, \tag{5.37}$$

for each integer k. Therefore, we can write

$$Tr(H^{n_1}V^{n_2}\cdots V^{n_s}) = Tr((I+N_1)(I+N_2)\cdots(I+N_s)), \tag{5.38}$$

where the matrix N_i, $i = 1, ..., s$, is given by

$$N_i = \begin{pmatrix} 0 & n_i a \\ 0 & 0 \end{pmatrix}, \tag{5.39}$$

if i is odd, and by

$$N_i = \begin{pmatrix} 0 & 0 \\ n_i a & 0 \end{pmatrix}, \tag{5.40}$$

if i is even.

Formula (5.38) is easily seen (by induction on s) to give

$$Tr(H^{n_1}V^{n_2}\cdots V^{n_s}) = Tr(I + \sum_{k=1}^{s} \bar{\sum} N_{i_1}\cdots N_{i_k}), \tag{5.41}$$

where the sum $\bar{\sum}$ is taken over all the $\binom{s}{k}$ combinations of the indices $i_1, ..., i_k$ in $1, ..., s$ with $i_1 < \cdots < i_s$. Now notice that from the definitions (5.39) and (5.40) $N_iN_j = 0$ if i and j are both odd or both even. Therefore we can consider in the sum $\bar{\sum}$ in (5.41) only products where neighboring matrices correspond to indexes that are not both even and not both odd. If k is odd in the term $N_{i_1}\cdots N_{i_k}$, because of the previous observation, either i_1 and i_k are both odd or they are both even. In one case the product matrix takes the form $\begin{pmatrix} 0 & \star \\ 0 & 0 \end{pmatrix}$, in the other, it takes the form $\begin{pmatrix} 0 & 0 \\ \star & 0 \end{pmatrix}$. In both cases the trace of the resulting matrix is zero. Finally, for k even, it is easily seen that the trace of $N_{i_1}\cdots N_{i_k}$ is given by $n_{i_1}\cdots n_{i_k}a^k$. These observations, along with the linearity of the trace, show that we can write (5.38) as (5.35). $\square$

A consequence of (5.35) is that the only elements of $\mathcal{S}(T)$ that have zero entropy are H^T and V^T. Also,

$$\frac{\partial TrS}{\partial a} > 0 \text{ for } a > 0$$

implies that the entropy of a particular periodic sequence is increasing with the modulus of a. We will not perform a direct maximization of the coefficients of a in (5.35), to solve our optimization problem (5.28), but we shall proceed along a different route.

Given a matrix A, A^σ denotes the antitranspose of the matrix A, namely the matrix obtained by reflecting the elements of A along the secondary diagonal. It is easily seen that $(A^n)^\sigma = (A^\sigma)^n$. The following two matrices will be important in the sequel

$$HV = \begin{pmatrix} (1+a^2) & a \\ a & 1 \end{pmatrix}, \tag{5.42}$$

$$VH = \begin{pmatrix} 1 & a \\ a & (1+a^2) \end{pmatrix}. \tag{5.43}$$

It is clear by inspection that

$$HV = (VH)^\sigma, \tag{5.44}$$

and we shall denote, for a given integer $k \geq 0$,

$$(HV)^k = ((VH)^k)^\sigma = \begin{pmatrix} u & v \\ v & z \end{pmatrix}. \tag{5.45}$$

The following result will be very useful.

Lemma 5.3.2. *Consider $(HV)^k$ with $k \geq 0$, as in (5.45). Then we have*

$$u = z + va \tag{5.46}$$

Proof. The proof is by induction on k. (5.46) is true when $k = 0$. Assume that for $k-1$ we have

$$(HV)^{k-1} = \begin{pmatrix} u' & v' \\ v' & z' \end{pmatrix}, \tag{5.47}$$

with

$$u' = z' + av'. \tag{5.48}$$

Using (5.42), we obtain

$$(HV)^k = (HV)^{k-1}(HV) = \begin{pmatrix} ((1+a^2)u' + v'a) & (u'a + v') \\ (v' + z'a + v'a^2) & (z' + av') \end{pmatrix}, \tag{5.49}$$

and using (5.48), we have

$$(HV)^k := \begin{pmatrix} u & v \\ v & z \end{pmatrix} = \begin{pmatrix} ((1+a^2)u' + v'a) & (u'a + v') \\ (u'a + v') & u' \end{pmatrix}, \tag{5.50}$$

from which it is immediate to verify that (5.46) holds. □

5.3.2 Solution of the problem for equal shear flows

In this subsection, we will derive the solution of (5.28) for the case $a = b$. We will see that the periodic sequence which maximizes entropy, in the set of periodic sequences constructed with H and V in (5.29), (5.30) is the alternate sequence $H \circ V \circ H \circ V \cdot \cdots$. Moreover, we will see that this is actually an absolute maximum in the set of all the periodic sequences constructed with H and V in (5.29), (5.30). The extension to the case $a \neq b$ will be given in the next subsection. We will first work with $T = n_1 + n_2 + ... + n_s$ even, as the case in which T odd easily follows from the solution of the problem with T even. We first assume $a \geq 1$

Case $a \geq 1$

Lemma 5.3.3. *Assume in (5.29), (5.30) $a \geq 1$. Consider a matrix*

$$S := H^{n_1} V^{n_2} \cdots V^{n_s}, \tag{5.51}$$

in $\mathcal{S}(T)$, and assume that there are some exponents n_i, $i \in \{1, ..., s\}$, such that $n_i \geq 3$. Then, there exists a matrix

$$\bar{S} := H^{\bar{n}_1} V^{\bar{n}_2} \cdots \bar{V}^{\bar{n}_{\bar{s}}}, \tag{5.52}$$

with $\bar{n}_i \leq 2$, for all $i = 1, ..., \bar{s}$, in $\mathcal{S}(T)$ such that

$$Tr(S) \leq Tr(\bar{S}). \tag{5.53}$$

Proof. Assume that, for a certain j, $n_j \geq 3$ in (5.51). Assume also, without loss of generality, that n_j is an exponent of H in S, otherwise we can use the property in (5.33) and (5.34), to generate another matrix (with H and V as first and last element, respectively), with the same trace, such that this holds, and the following computations will go through the same way. We have

$$Tr(S) = Tr(H^{n_1} V^{n_2} \dots V^{n_{j-1}} H^{n_j} V^{n_{j+1}} \dots V^{n_s}) = Tr(H^{n_j} V^{n_{j+1}} \dots V^{n_s} H^{n_1} \dots V^{n_{j-1}}), \tag{5.54}$$

where we have used the property of the trace in (5.34). Now define

$$L := \begin{pmatrix} \lambda & \mu \\ \nu & \gamma \end{pmatrix} = (H^{n_j-3}V^{n_{j+1}} \cdots V^{n_s}H^{n_1} \cdots V^{n_{j-1}}), \tag{5.55}$$

and notice that, since $a \geq 1$, L is composed of nonnegative elements. We write

$$Tr(S) = Tr(H^3L) = Tr(HLH^2), \tag{5.56}$$

and notice that the product in the matrix L ends with a matrix V. Now we show, by direct computation that $Tr(HLH^2) \leq Tr(HLHV)$. We use the fact that L is a matrix of nonnegative elements and the relations in (5.36), (5.42). We have

$$\begin{aligned} Tr(S) = Tr(HLH^2) &= Tr\left(\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \lambda & \mu \\ \nu & \gamma \end{pmatrix}\begin{pmatrix} 1 & 2a \\ 0 & 1 \end{pmatrix}\right) \\ &= \lambda + 3a\nu + \gamma, \end{aligned} \tag{5.57}$$

$$\begin{aligned} Tr(HLHV) &= Tr\left(\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \lambda & \mu \\ \nu & \gamma \end{pmatrix}\begin{pmatrix} (1+a^2) & a \\ a & 1 \end{pmatrix}\right) \\ &= (1+a^2)\lambda + (1+a^2)\gamma + (2a+a^3)\nu + a\mu. \end{aligned} \tag{5.58}$$

Comparing (5.58) and (5.57), and recalling that $a \geq 1$, we have that $Tr(HLHV) \geq Tr(HLH^2)$. Therefore we have constructed a matrix in $\mathcal{S}(n)$, which has trace at least as large as the original one and such that the exponent $n_j \geq 3$ has been changed to $n_j - 2$. Proceeding iteratively this way we end up with a matrix $\bar{S}$ as in (5.52), whose exponents $\bar{n}_i$, $i = 1, \ldots, \bar{s}$ are all 1 or 2, which has trace greater or equal than the trace of the original matrix. □

Lemma 5.3.4. *Assume $a \geq 1$. For every matrix $\bar{S}$ with exponents ≤ 2 as in (5.52), the following holds*

$$Tr(\bar{S}) \leq Tr((HV)^{\frac{T}{2}}), \tag{5.59}$$

with equality if and only if $\bar{S} = (HV)^{\frac{T}{2}}$.

Proof. The result is easily verified by comparing the traces of (5.36), (5.37) and (5.42), if $T = 2$. Therefore we have to prove the result only for $T \geq 4$. Assume by (5.34) and (5.33) that H^2 is at the first position from the left. Since T is even this is not the only square term and there exists at least another square factor in the product. There are two cases: 1) The closest square element to H^2 on the right is

V^2; 2) The closest square element H^2 on the right is H^2. Let us consider these two cases separately.

Case 1) The matrix $\bar{S}$ has the form

$$\bar{S} = H^2(VH)^k V^2 L, \tag{5.60}$$

for a suitable $k \geq 0$. L is a general matrix with nonnegative elements obtained by some multiplications of H and V, for which we will use the same notation with greek letters as in (5.55). L starts with H and ends with V. We claim that

$$Tr(\bar{S}) = Tr(H^2(VH)^k V^2 L) < Tr(L(HV)^{k+2}). \tag{5.61}$$

Recalling the definitions of $(HV)^k$ of (5.45) and the fact that $(VH)^k$ is the antitranspose of $(HV)^k$ and (5.36) and (5.37), we have

$$\begin{aligned} Tr(H^2(VH)^k V^2 L) &= Tr\left(\begin{pmatrix} 1 & 2a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} z & v \\ v & u \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2a & 1 \end{pmatrix}\begin{pmatrix} \lambda & \mu \\ \nu & \gamma \end{pmatrix}\right) \\ &= (z + 4av + 4a^2u)\lambda + (v + 2au)\nu + (v + 2au)\mu + u\gamma, \end{aligned} \tag{5.62}$$

and

$$\begin{aligned} Tr(L(HV)^{k+2}) &= Tr\left(\begin{pmatrix} \lambda & \mu \\ \nu & \gamma \end{pmatrix}\begin{pmatrix} u & v \\ v & z \end{pmatrix}\begin{pmatrix} (1+a^2) & a \\ a & 1 \end{pmatrix}^2\right) \\ &= ((1 + 3a^2 + a^4)u + (2a + a^3)v)\lambda \\ &+ ((1 + 3a^2 + a^4)v + (2a + a^3)z)\mu \\ &+ ((2a + a^3)u + (1 + a^2)v)\nu \\ &+ ((2a + a^3)v + (1 + a^2)z)\gamma. \end{aligned} \tag{5.63}$$

Using the expression of z obtained from (5.46) in (5.62) and (5.63), we obtain

$$Tr(H^2(VH)^k V^2 L) = ((1 + 4a^2)u + 3av)\lambda + (v + 2au)\nu + (v + 2au)\mu + u\gamma, \tag{5.64}$$

and

$$\begin{aligned} Tr(L(HV)^{k+2}) &= ((1 + 3a^2 + a^4)u + (2a + a^3)v)\lambda \\ &+ ((1 + a^2)v + (2a + a^3)u)\mu \\ &+ ((2a + a^3)u + (1 + a^2)v)\nu \\ &+ (av + (1 + a^2)u)\gamma. \end{aligned} \tag{5.65}$$

Comparing the coefficients of λ, μ, ν and γ in (5.64) and (5.65), recalling that $a \geq 1$, it is easily seen that every coefficient in (5.65) is greater or equal than the corresponding coefficient in (5.64). Moreover notice that the element γ in L and the element u in HV are always greater or equal to zero. They are actually greater or equal to one since they come from products of matrices H and V and $a \geq 1$. Moreover v is also greater than or equal to zero. With these facts in mind, it is easy to verify that the element containing γ in (5.65) is *strictly* greater than the element containing γ in (5.64), which proves the claim in (5.61).

Case 2) The matrix $\bar{S}$ has the form

$$\bar{S} = H^2(VH)^k VH^2 L, \tag{5.66}$$

for a suitable $k \geq 0$. Here also L is a general matrix with nonnegative elements, obtained by some multiplications of H and V. Notice also that, in this case, L starts with V and ends with V. We claim that

$$Tr(\bar{S}) = Tr(H^2(VH)^k VH^2 L) < Tr(HL(HV)^{k+2}). \tag{5.67}$$

As in the Case 1), we compute explicitly the traces of these two matrices and compare the coefficients of λ, μ, ν, and γ, with the help of (5.46). We obtain

$$\begin{aligned} Tr(H^2(VH)^k VH^2 L) &= (z + 3av + 2a^2u)\lambda \\ &+ (2az + (1 + 6a^2)v + 2a(1 + 2a^2)u)\nu \\ &+ (v + au)\mu + (2av + (1 + 2a^2)u)\gamma, \end{aligned} \tag{5.68}$$

and

$$\begin{aligned} Tr(HL(HV)^{k+2}) &= ((1 + 3a^2 + a^4)u + (2a + a^3)v)\lambda \\ &+ ((1 + 3a^2 + a^4)v + (2a + a^3)z)\mu \\ &+ ((3a + 4a^3 + a^5)u + (1 + 3a^2 + a^4)v)\nu \\ &+ ((3a + 4a^3 + a^5)v + (1 + 3a^2 + a^4)z)\gamma. \end{aligned} \tag{5.69}$$

Replacing z as given in (5.46) in (5.68) and (5.69), we obtain

$$\begin{aligned} Tr(H^2(VH)^k VH^2 L) &= ((1 + 2a^2)u + 2av)\lambda + (v + au)\mu \\ &+ ((4a + 4a^3)u + (1 + 4a^2)v)\nu + (2av + (1 + 2a^2)u)\gamma, \end{aligned} \tag{5.70}$$

and

$$\begin{aligned} Tr(HL(HV)^{k+2}) &= ((1+3a^2+a^4)u+(2a+a^3)v)\lambda \\ &+ ((1+a^2)v+(2a+a^3)u)\mu \\ &+ ((3a+4a^3+a^5)u+(1+3a^2+a^4)v)\nu \\ &+ ((2a+a^3)v+(1+3a^2+a^4)u)\gamma. \end{aligned} \tag{5.71}$$

Comparing the coefficients of (5.70) and (5.71), and recalling that $a \geq 1$ along with a closer look at the term containing γ as in Case 1) gives (5.67).

The result of the lemma follows by iterating the above procedure, in particular eliminating possible square powers in the matrix L in $L(HV)^{k+2}$ and $HL(HV)^{k+2}$.
□

Lemmas 5.3.3 and 5.3.4 solve the subproblem (5.28) when $a = b \geq 1$ as the case of T odd can be solved from the analysis developed above: the sequence of period 3, ...HVV..., for example, has the same entropy as the sequence of period 6, ...$HVVHVV$..., and all the sequences of period 6 have entropy strictly less than ...$HVHVHV$.... The only periodic sequence that have the same entropy of $H \circ V \circ H \circ V \cdots$, is the sequence $V \circ H \circ V \circ H \cdots$, and this is easily seen by the fact that $Tr(HV) = Tr(VH)$. We now use a continuity argument to extend the result to the case $a < 1$.

Case $a < 1$

Given an even T, we have to prove that $Tr(HV)^{\frac{T}{2}} > Tr(S)$, where S is any matrix different from $(HV)^{\frac{T}{2}}$ of the form $S = H^{n_1}V^{n_2}\cdots V^{n_s}$, with $n_1+n_2+\cdots+n_s = T$, for a certain $s < T$. The traces of S and $(HV)^{\frac{T}{2}}$ can be expressed as polynomials in a by (5.35). From (5.35), we notice that $Tr((HV)^{\frac{T}{2}})$ is a monic polynomial of degree T with integer coefficients. $Tr(S)$ is a polynomial with integer coefficients and of degree s strictly less than T. The difference

$$Tr((HV)^{\frac{T}{2}}) - Tr(S) := p(a) \tag{5.72}$$

is a monic polynomial of degree T with integer coefficients. So is $\bar{p}(a)$ which is defined by the factorization $p(a) := a^k\bar{p}(a)$, where $k \geq 0$, is the number of roots at zero of $p(a)$.

From Lemma 5.3.3 and Lemma 5.3.4, $p(a) > 0$ if $a \geq 1$. For $a < 1$ we argue by contradiction. Assume that $p(a) \leq 0$ for some a in $(-1,1)$. Then, by continuity, there must exist at least one root of $\bar{p}(a)$ different from zero which must necessarily be in $(-1,1)$. Let us denote these roots by $a_1, ..., a_r$ with $r > 0$. We have, since $\bar{p}(a)$

is monic, $\prod_{i=1}^{r} |a_i| = |c|$, where c is the constant term in $\bar{p}(a)$. But this cannot be since c is an integer and all of the a_i are in $(-1, 1)$.

We summarize below the results achieved so far for the case of identical shear flows.

Theorem 5.3.5. *Consider $a > 0$ in (5.29) (5.30). The sequence of period 2, $H \circ V$ is the one which maximizes entropy among the periodic ones composed of H and V. Moreover every other periodic sequence, other than $V \circ H \circ V \circ H \cdots$ gives strictly less entropy*

5.3.3 Solution for different linear shear flows

In this case, the matrices H and V are defined as

$$H := \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \qquad V := \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}. \tag{5.73}$$

For the matrices H and V defined in (5.73), we have

$$H = P^{-1} V P, \tag{5.74}$$

where

$$P := \begin{pmatrix} 0 & 1 \\ b/a & 0 \end{pmatrix}$$

and thus property (5.33) is preserved. This, along with (5.34), allows us to consider again only products of the form $H^{n_1} V^{n_2} \cdots V^{n_s}$. The formula for the trace of this matrix can be computed explicitly by simply modifying the proof of Lemma 5.3.1. We have

$$Tr(H^{n_1} V^{n_2} \cdots V^{n_s}) =$$

$$2 + \Big(\sum_{\substack{i_1 < i_2 \\ i_1 \not\sim i_2}} n_{i_1} n_{i_2}\Big) ab + \Big(\sum_{\substack{i_1 < i_2 < i_3 < i_4 \\ i_1 \not\sim i_2 \\ i_2 \not\sim i_3 \\ i_3 \not\sim i_4}} n_{i_1} n_{i_2} n_{i_3} n_{i_4}\Big) a^2 b^2 + \cdots + n_1 n_2 \cdots n_s a^{s/2} b^{s/2}. \tag{5.75}$$

Since a and b are of the same sign, we are allowed to let $c^2 = ab$. For given a, b

the sequence with maximal entropy is the one for which $n_1, ..., n_s$ maximize

$$2+\Big(\sum_{\substack{i_1<i_2\\ i_1 \not\sim i_2}} n_{i_1}n_{i_2}\Big)c^2+\Big(\sum_{\substack{i_1<i_2<i_3<i_4\\ i_1\not\sim i_2\\ i_2\not\sim i_3\\ i_3\not\sim i_4}} n_{i_1}n_{i_2}n_{i_3}n_{i_4}\Big)c^4+\cdots+n_1n_2\cdots n_sc^s. \quad (5.76)$$

We have already shown that $n_1 = 1$, $n_2 = 1$ maximizes the above expression for a fixed c.

5.4 Properties of the optimal sequence

We start by recalling the result derived in the two previous sections.

Theorem 5.4.1. *Consider a periodic sequence of purely horizontal and vertical transformations on the torus. This sequence has entropy strictly less than the one of the alternate sequence* $|\bar{H}|\circ|\bar{V}|\circ|\bar{H}|\circ|\bar{V}|\cdots$ *(or* $|\bar{V}|\circ|\bar{H}|\circ|\bar{V}|\circ|\bar{H}|\cdots$

Remark 5.4.2. The fact that the map $|\bar{H}|\circ|\bar{V}|$ has the same entropy as the map $|\bar{V}|\circ|\bar{H}|$ follows from the property of the trace (5.34) but it also follows from the property of entropy discussed in Remark 3.3.9.

Remark 5.4.3. From formula (5.26) and Lemmas 5.2.1 and 3.3.7, one can find an expression for the entropy of the sequence $|\bar{H}|\circ|\bar{V}|\circ|\bar{H}|\circ|\bar{V}|\cdots$, namely the maximum entropy achievable with maps composed of purely horizontal and vertical maps on the torus with the bounds (5.5) (5.6). It is given by

$$h(|\bar{H}|\circ|\bar{V}|\circ|\bar{H}|\circ|\bar{V}|\cdots) = \frac{1}{2}log\frac{(2+ab)+\sqrt{ab(4+ab)}}{2}. \quad (5.77)$$

This also shows that the constraints in (5.5) (5.6) regularize the problem since when a and/or b go to infinity the entropy becomes unbounded.

We now discuss the mixing properties of the optimal sequence. We know from Theorem 3.2.10 that it is sufficient to check if the associated composite transformation, in this case $|\bar{H}|\circ|\bar{V}|$ is mixing. Actually, we know more. We know from Theorem 3.2.9 that if we prove that $|\bar{H}|\circ|\bar{V}|$ is mixing, then every sequence which converges to the alternate sequence $|\bar{H}|\circ|\bar{V}|\circ|\bar{H}|\circ|\bar{V}|\cdots$ is mixing as long as the convergence is sufficiently fast in the sense of (3.100).

In order to check if $|\bar{H}|\circ|\bar{V}|$ is mixing we need to recall a result on automorphisms with singularities on the torus due to Vaienti and Chernov. We state this result without the proof which is too long. The result identifies a class of automorphisms on the two dimensional torus which is mixing in the sense of the Definition 3.2.8.

Theorem 5.4.4. (Vaienti-Chernov) *Consider a dynamical system $(M, \mathcal{S}, \mu, \Phi)$ where M is the two dimensional torus, $\mathcal{S}$ is the Borel σ-algebra induced by the usual topology, and μ the Lebesgue measure. Assume Φ has constant differential $D\Phi$ defined almost everywhere in M. We have $|det(D\Phi)| = 1$ and assume $D\Phi$ has two real eigenvalues λ_{max} and λ_{min}, with $\frac{1}{|\lambda_{min}|} = |\lambda_{max}| > 1$. Denote by $S_+ \subset M$ and $S_- \subset M$ the discontinuity sets of Φ and Φ^{-1}, respectively and by E_u and E_s the eigenvectors of $D\Phi$ corresponding to λ_{max} and λ_{min}. Assume S_+ and S_- consist of segments on M not parallel to the eigenvectors E_u and E_s of $D\Phi$. Define $S_n = \Phi^{-n+1}S_+$ and $S_{-n} = \Phi^{n-1}S_-$. Assume for every m and n greater or equal to one S_m and S_n are lines always intersecting transversally, and the same for S_{-m} and S_{-n}. Assume that the vectors E_u and E_s intersect transversally with S_+, S_-, S_n, S_{-n}. Then, the system $(M, \mathcal{S}, \mu, \Phi)$ is mixing.*

The map $|\bar{H}| \circ |\bar{V}|$, when a and b are positive integers in (5.29) (5.30), has no discontinuities and the mixing property follows from the above theorem, just noticing that the differential has two real eigenvalues with absolute value different from one. In the case in which a and b are not integers the singularity set for $|\bar{H}| \circ |\bar{V}|$ is given by the singularity set for $|\bar{V}|$, $\gamma_V := \{(x, y) \in M | x = 0, 0 \le y < 1\}$ along with the set $|\bar{V}|^{-1}\gamma_H$, where $\gamma_H := \{(x, y) \in M | 0 \le x < 1, y = 0\}$ is the singularity set for $|\bar{H}|$. Noticing that

$$|\bar{V}|^{-1}\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \quad mod \quad 1, \tag{5.78}$$

it is easily seen that $S_+ = \gamma_V \cup |\bar{V}|^{-1}\gamma_H$ is the union of a vertical line γ_V which is parallel to the vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and segments of slope $-b$ namely parallel to the vector $v = \begin{pmatrix} 1 \\ -b \end{pmatrix}$. Call J the differential of $|\bar{H}| \circ |\bar{V}|$. It is immediate to verify that neither $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ nor $\begin{pmatrix} 1 \\ -b \end{pmatrix}$ are eigenvectors of J. If S_m intersects S_n not transversally, for some $m > 0 \neq n > 0$, then the following situation has to happen for $r = -n + 1 \ge 0$ and $s = -m + 1 \ge 0$: There exists a real number α such that

$$J^{-r}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \alpha J^{-s}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.79}$$

or

$$J^{-r}\begin{pmatrix} 1 \\ -b \end{pmatrix} = \alpha J^{-s}\begin{pmatrix} 1 \\ -b \end{pmatrix}, \tag{5.80}$$

or

$$J^{-r}\begin{pmatrix}0\\1\end{pmatrix} = \alpha J^{-s}\begin{pmatrix}1\\-b\end{pmatrix}. \tag{5.81}$$

The first case (5.79) and the second case (5.80) are impossible, otherwise $\begin{pmatrix}0\\1\end{pmatrix}$ or $\begin{pmatrix}1\\-b\end{pmatrix}$ would be eigenvectors of J, which is not true. Finally the case (5.81) also cannot be. In fact, assume this is the case, assume $s > r$, we would have

$$J^{s-r}\begin{pmatrix}0\\1\end{pmatrix} = \alpha\begin{pmatrix}1\\-b\end{pmatrix}, \tag{5.82}$$

which is impossible since the elements of J are all positive and the vector at the right hand side has two components with different sign. If $r > s$ then, it should be

$$\begin{pmatrix}0\\1\end{pmatrix} = \alpha J^{r-s}\begin{pmatrix}1\\-b\end{pmatrix}, \tag{5.83}$$

and notice that $J\begin{pmatrix}1\\-b\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}$ and a simple induction argument shows that the first component in the vector on the right hand side of (5.83) is always different from zero which shows that also (5.83) is impossible.

The conclusion of the above argument is that, for our map, S_n and S_m always intersect transversally. A similar analysis can be carried out for S_{-m} and S_{-n} so that we can conclude that the map $|\bar{H}| \circ |\bar{V}|$ satisfies the hypotheses of Theorem 4.5.4. and it is therefore mixing.

Figure (5.2) compares the mixing behaviour of the sequence with alternate purely horizontal and vertical actions with $h(y) = sin(\pi y)$ and $v(x) = sin(\pi x)$ in (5.3) (5.4) (right column) with the optimal sequence with a and b equal to π, the maximum slope of the functions h and v, (left column). It is clear that for the sequence with maximum entropy the mixing is faster. Simulations for times greater than $T = 4$ (not shown in Figure (5.2)) indicate that the sequence characterized by the sine function will eventually mix homogeneously after numerous iterations. The symmetry of the sine function creates particular patterns in the flow that are not created by the sequence with maximum entropy, and are eventually destroyed.

Figures (5.3) and (5.4) show the mixing behaviour of the protocols $|\bar{H}|^8|\bar{V}|^2$ and $(|\bar{H}||\bar{V}|)^5$. Initial points are placed on a grid in the rectangle $0 \leq x \leq 1$, $0 \leq y \leq 0.5$. We used the values $a = b = 0.28$. Although both protocols mix well, both being mixing in the ergodic theoretic sense, the one with larger entropy creates disorder in a faster manner. In general, the speed of mixing improves dramatically by increasing

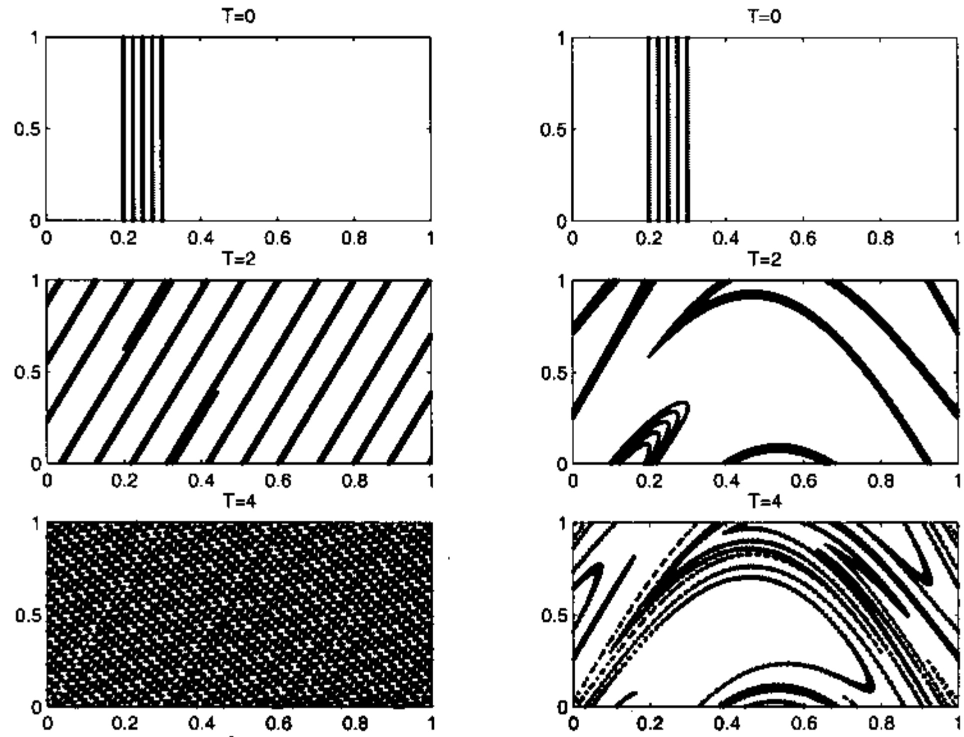

Figure 5.2: Mixing behaviour of the maximum entropy sequence ; $a = b = 3.14$ (left column) versus the HV sequence characterized by the function $sin(\pi x)$. At time $T = 0$ particles are posed in the strip $\{0.2 \leq x \leq 0.3\}$.

the parameters a and b, but for different initial configurations the difference between the two protocols is not always evident. In general the mixing behaviour will be dependent on the initial configuration. From Pesin's formula (5.11) we know that entropy is equal to the stretching rate (given by the Lyapunov exponents) averaged over the whole phase space. In choosing the maximizing entropy we did not refer to any particular configuration of the fluids to be mixed. For some particular configurations, we would need more stretching to break bounds in particular directions or in particular points of the flow. In this sense, the protocol proposed here is the best if we do not consider particular initial configurations, or if we would like to design a mixer that works well for various initial conditions.

5.5 Notes and Reference

The prototypical flow dealt with in this chapter was first proposed in [29] as a simplified example of the basic stretching and folding mechanisms of mechanical mixing. It is called there *the egg-beater flow.* The problem of designing a mixing protocol for this flow is dealt with in [13]. The approach there is to look at the symmetries of the flow and to apply at each step a velocity profile that distroys such symmetries. The quantitative approach based on ergodic theory which has been presented here was first outlined in [10] and the problem was solved for the maximum stretch maps in [12]. The extension to every type of map was given in [11]. Theorem 5.4.4 was proved by Vaienti in [41] and further ergodic properties of the maps considered here were proved by Chernov in [7].

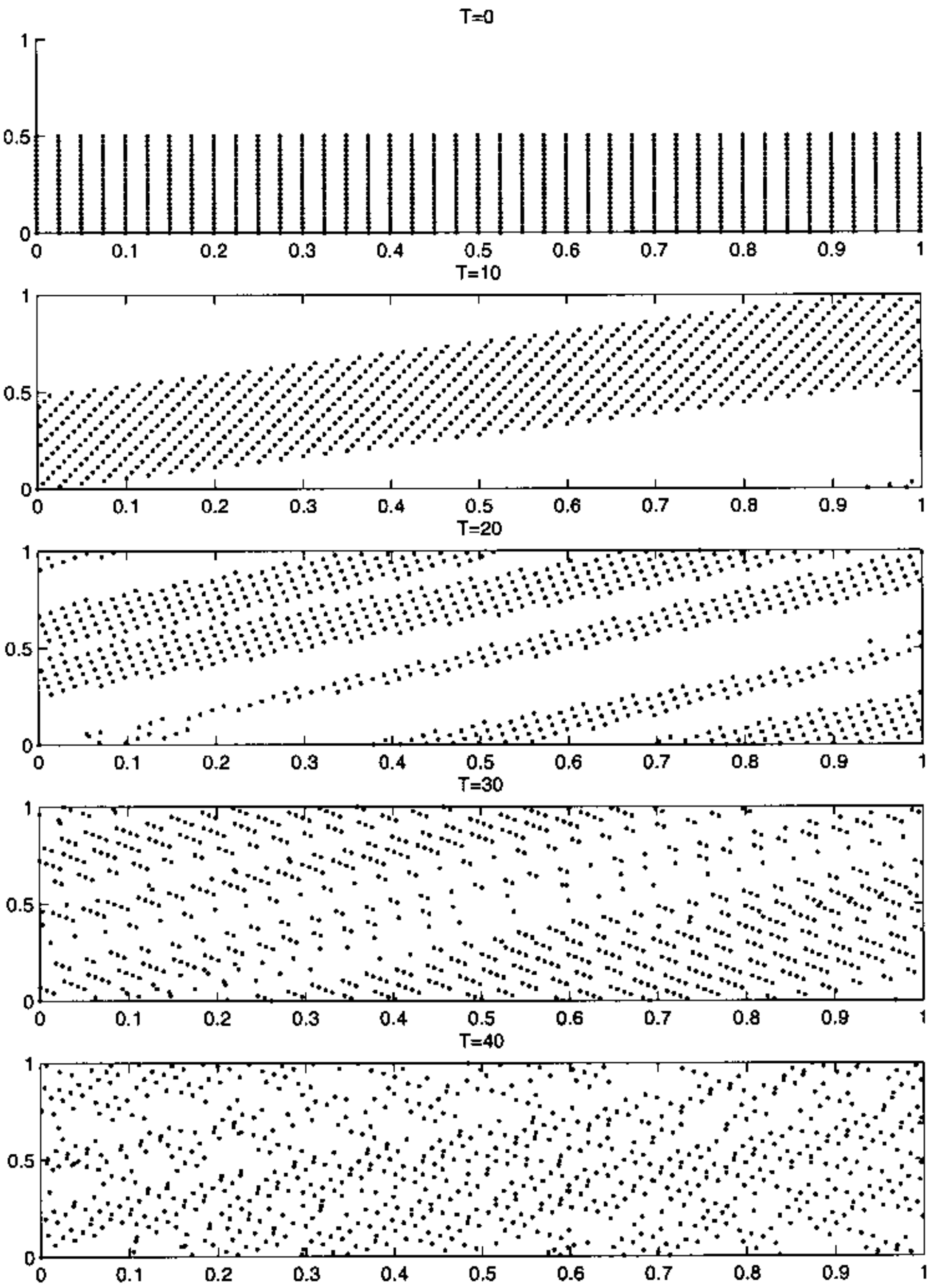

Figure 5.3: Mixing behaviour of the sequence of period 10 $|\bar{H}|^8|\bar{V}|^2$; $a = b = 0.28$.

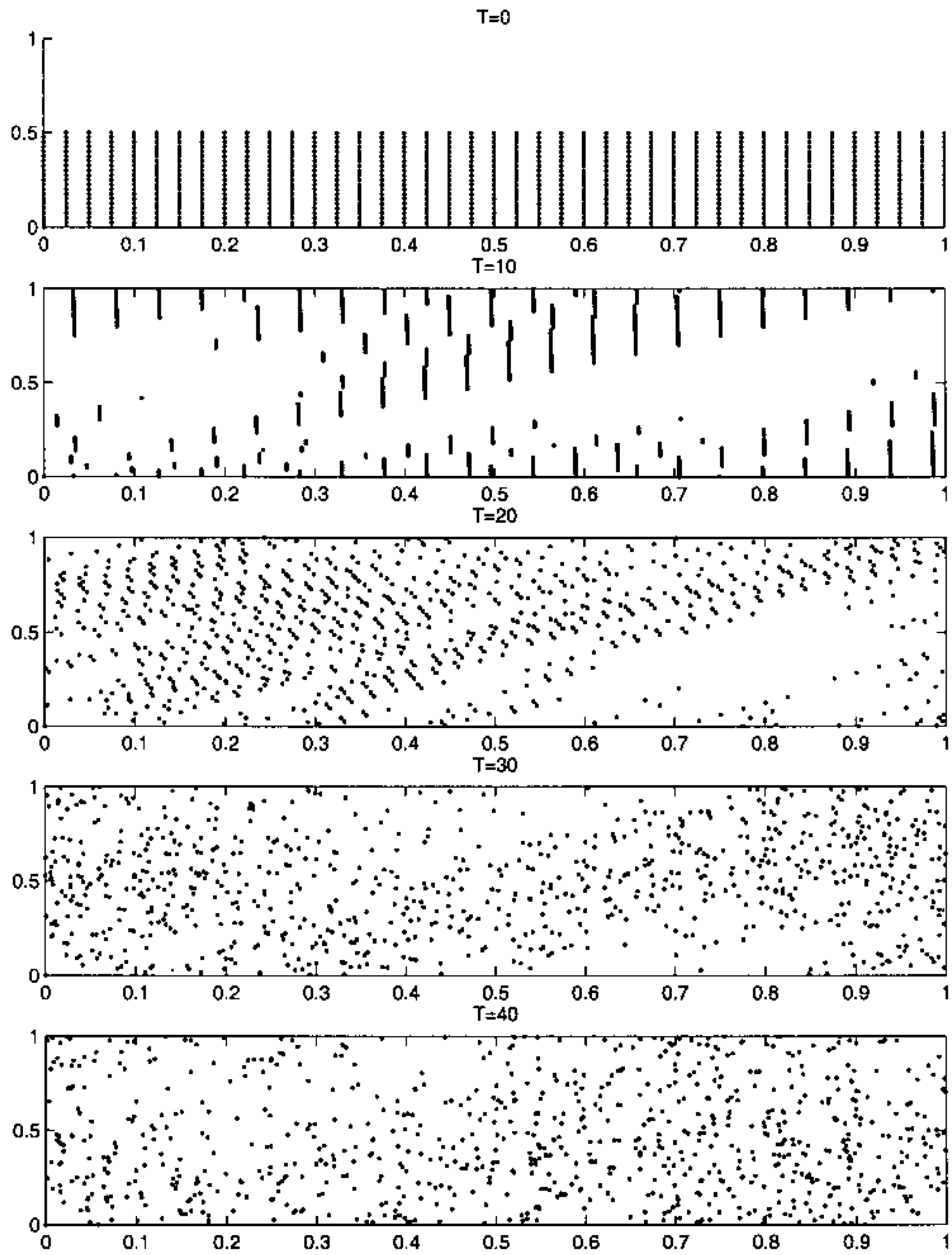

Figure 5.4: Mixing behaviour of the sequence of period 10 $(|\bar{H}||\bar{V}|)^5$; $a = b = 0.28$.

Chapter 6

Other Mixers

The main goal of this chapter is to test the entropy criterion considered in the previous chapter for mixers of physical interest. In these cases the analysis is necessarily more qualitative than the one we have carried out in the previous chapter for the prototypical flow. Computer experiments are necessary for the computation of entropy and the visualization of the flows. We will deal with a particular mixing device called *Partitioned Pipe Mixer.* For this mixer the equations governing the dynamics can be achieved with acceptable accuracy. The basic tool for the computation of entropy will be, once again, the concept of Lyapunov exponents and Pesin's entropy formula (3.159).

6.1 The partitioned pipe mixer

The partitioned pipe mixer consists of a pipe partitioned into a sequence of pairs of semi-circular ducts by rectangular plates placed orthogonally to each-other. We shall refer to a pair of semicircular ducts as an *element* of the mixer. Two elements of the mixer, adjacent to each-other, are shown in Figure 6.1.

The mixer has to be thought of as a sequence, theoretically infinite, of elements. The fluid is forced through the pipe by an axial pressure gradient. In each element, the rectangular plate is fixed and the circular walls of the element rotate. If we assume a no slip condition at the walls, namely that the fluid moves with the walls, this motion causes the mixing process between the streams introduced in the first element in two different semicircular ducts. The walls of two adjacent elements can rotate in the same direction and this will be referred to as the *co-rotating* case or in opposite direction, *counter-rotating* case. The performances of the mixer in these two cases might be completely different, when other parameters are the same, and

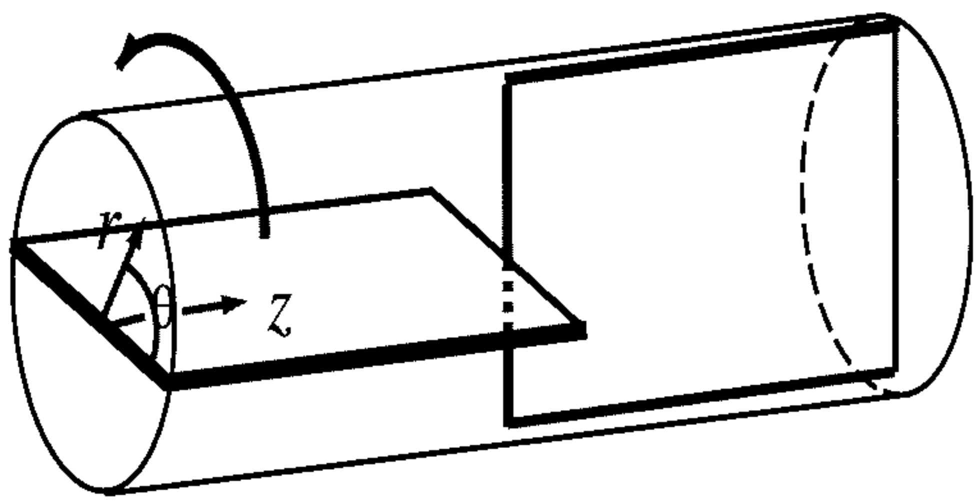

Figure 6.1: Pair of elements of the Partitioned Pipe Mixer

the analysis we will carry out in the following will explain these differences in terms of entropy.

Let us follow the motion of a particle of fluid as it moves through the pipe. In both co-rotating and counter-rotating case the flow is periodic in space, in particular, it is repeated every two elements of the mixer. We record the position of the particle at each cross section at the end of a pair of elements. The superposition of these intersections in a theoretically infinite length mixer gives the *Poincaré map* associated with the flow. This map describes the performance of the mixer.

The quantitative description of the motion of the particle can be carried out in a satisfactory way and approximate expressions of the velocities in cylindrical coordinates can be computed. We have

$$\dot{r} = \beta r(1 - r^{\alpha}) sin 2\theta, \tag{6.1}$$

$$\dot{\theta} = -\beta(2 - (2 + \alpha) r^{\alpha}) sin^2\theta, \tag{6.2}$$

$$\dot{z} = |\frac{16\pi}{\pi^2 - 8} \sum_{k=1}^{3} ((r^{2k-1} - r^2) \frac{sin[(2k-1)\theta]}{(2k-1)[4-(2k-1)^2]})|. \tag{6.3}$$

In the above formulas, $\alpha = \sqrt{\frac{11}{3}} - 1$ and β is a dimensionless parameter, referred to as the *mixing strength* which measures the *cross-sectional* stretching as opposed to *axial* stretching; more specifically, we define

$$\beta := \frac{4 v_R L}{3\alpha < v_z > R}, \tag{6.4}$$

where v_R is the angular velocity of the walls, L is the length of an element, $< v_z >$ is the average velocity in the zeta direction, namely in the direction parallel to the pipe, the average being taken on (any) one cross section and R is the radius of the pipe.

In equations (6.1) (6.2) (6.3), r is the radius, normalized with respect to the radius of the pipe R, z is the axial distance normalized with respect to the length of an element L and time is normalized with respect to $\frac{L}{<v_z>}$. Notice that the direction of the flow is counterclockwise, in both the semicircles of the elements if β is positive.

Equations (6.1), (6.2), (6.3) are valid in the first element. In order to get the flow in the following element, we have to rotate the system of coordinates by 90 degrees and changing the sign of β if the mixer is counter-rotating.

6.2 Describing Maps

We are interested in how the flow maps points that are at the first cross section of a pair of elements to the last cross section. Two cases, will be considered. One case is the so-called *Plug Flow* where, in the equations (6.1) (6.2) (6.3), we set $\dot{z} = 1$, for any point r, θ. This flow is a simplified one, in particular there is no no-slip condition assumed at the boundary with the wall. However, it is directly amenable of an ergodic theoretic analysis, which can be carried out in rigorous terms, since the map from one cross section to the other preserves the Lebesgue measure expressed in cylindrical coordinates as $rd\theta dr$. The second case we will consider is the actual flow (6.1) (6.2) (6.3). In this case, the measure $rdrd\theta$ from one cross section to the other is not preserved, due to the gradient of the velocity in the direction z with r and θ. In particular, if we consider an area at the initial cross section, its points will reach the final cross section at different times. As a consequence, we cannot apply conservation of mass in the form $V_1 A_1 = V_2 A_2$, with $V_1 = V_2$, where V_i, $i = 1, 2$ and $A_i = 1, 2$ are velocities and areas respectively. We cannot guarantee that the area at the final cross section is the same as the one at the initial cross section. It is not rigorous to apply ergodic theoretic criteria to this case, since one of the basic assumptions of ergodic theory, namely that the considered transformation preserves the measure is not verified in this case. However, we will see, in the simulations in Section 6.3, that the heuristics based on ergodic theoretic concepts (in particular entropy) still provide an explanation of the different behaviour of the mixer in different situations.

The next two subsections provide a derivation of the maps to be studied for plug and real flows.

6.2.1 Plug flow

The map Φ from the first cross section to the last cross section of a pair of elements is constructed as follows: a point (r_0, θ_0) is mapped to a point (r_1, θ_1) at the cross section which separates two elements by a map which is given by the flow of the differential equation (6.1) (6.2), $\Phi^t(r_0, \theta_0)$ at time $t = 1$ (recall the velocity in the z direction is assumed to be equal to 1 and the length of an element is also equal to 1). In the second element the flow is the same as in the first element, if we perform a change of coordinates ($R := \theta \rightarrow \theta - \frac{\pi}{2}$). The final point is denoted by (r_2, θ_2) in Figure 6.2. If we are dealing with the counter-rotating mixer equations (6.1)- (6.2) also change sign in the second element. As a consequence the associated flow will be Φ^{-t}. In conclusion the overall map between the first and the last cross section of a pair of elements (namely between (r_0, θ_0) and (r_2, θ_2) in Figure 6.2))

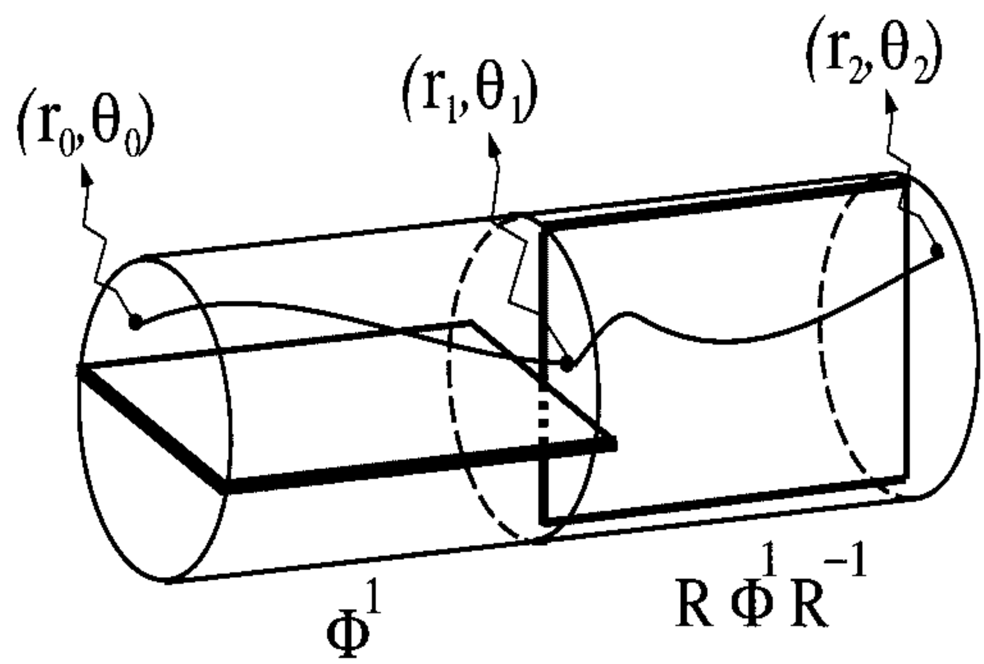

Figure 6.2: Trajectory between initial and final cross section of a pair of elements for the Partitioned Pipe Mixer (co-rotating case)

will be

$$\Phi_{co} := R^{-1} \circ \Phi^1 \circ R \circ \Phi^1, \tag{6.5}$$

in the co-rotating case and

$$\Phi_{counter} := R^{-1} \circ \Phi^{-1} \circ R \circ \Phi^1, \tag{6.6}$$

in the co-rotating. Both of these maps preserve the measure $rdrd\theta$ since they are composed of measure preserving transformations. We are interested in studying the mixing properties of these maps, in particular *Strong Mixing* and *Entropy* as discussed in the previous section. There is almost no way to check analyticly if a map associated to a flow such as (6.5) and (6.6) is mixing in the sense of Definition 3.2.8, and we have to rely on the inspection of the Poincare' maps as done in the next section. For the computation of entropy, we have however enough tools to carry on the analysis in a quantitative way. We will see that bigger entropy will result in a better quality of mixing.

The computation of entropy is based on formula (3.159). In order to use this formula we have to compute the Lyapunov exponents using the definition (3.155). Notice that in this case there is at most one positive Lyapunov exponent since the map preserves the area. Since the map Φ^t is a flow of a differential equation $\dot{x} = f(x)$, the differential $D_x\Phi^t$ satisfies the linear time varying matrix differential *variational equation* (see

$$\frac{d}{dt} D_x\Phi^t = (Df(\Phi^t(x)))D_x\Phi^t, \tag{6.7}$$

with initial condition $D_x\Phi^0 = I$. This was proven in Subsection 2.2.1.

Numerical integration of (6.7) was used in the computation of Lyapunov exponents. Once we have $D_x\Phi^1$, and $D_x\Phi^{-1}$, it is immediate, by the chain rule, and using (6.5) and (6.6), to compute $D_x\Phi_{co}$ and $D_x\Phi_{counter}$, then to compute the Lyapunov exponents by (3.155) and then entropy by (3.159).

6.2.2 Real flow

In the real case, the map from one cross section to another is not simply the flow Φ^t associated to the differential equation (6.1) (6.2). In fact, the time t_f at which the point crosses a given cross section depends on the initial condition because of gradients of the velocity $\dot{z}$ with r and θ. More specifically, the time $t_f(r,\theta)$ at which the point crosses the cross section between the first and the second element of a couple of elements is implicitly defined by

$$\Phi_3^{t_f} := \int_0^{t_f(r,\theta)} f_3(\Phi_1^t(r,\theta), \Phi_2^t(r,\theta))dt = 1. \tag{6.8}$$

Here f_3 denotes the right hand side of (6.3) and $\Phi_1^t(r,\theta)$, $\Phi_2^t(r,\theta)$ are the first and the second components of the flow Φ^t, respectively. As a result, of this dependence of t_f on (r,θ), the map does not preserve the measure $rdrd\theta$ and therefore the Definition of entropy 3.3.5 and Pesin formula (3.159) cannot be rigorously applied in this context with this measure. However, if the limit in (3.155) exists for almost every x, we can still use our interpretation of the entropy integral in 3.159) as a measure of the stretching rate and of the stretched area and therefore as a measure of mixing. Our approach for what the real flow is concerned, is to compute, for a fixed vector v and every $x \in M$, the quantity

$$\tilde{\chi} := \frac{1}{2n} log||(D\Phi^n(x))^T(D\Phi^n(x))||, \tag{6.9}$$

for n very large, where Φ is the map under consideration. This should give an idea of the rate of stretching experienced by the map Φ. Then we compute a *generalized entropy* $\tilde{h}$ as

$$\tilde{h}(\Phi) := \int_M \tilde{\chi} d\mu. \tag{6.10}$$

We now study the form of the map for the real flow. This is given by

$$\Phi_{co} := R^{-1} \circ \Phi^{t_f} \circ R \circ \Phi^{t_f}, \tag{6.11}$$

in the co-rotating, and by

$$\Phi_{counter} := R^{-1} \circ \Phi^{-\bar{t}_f} \circ R \circ \Phi^{t_f}, \tag{6.12}$$

in the counter-rotating case, where $\bar{t}_f$ is defined implicitly, analogously to (6.8) as

$$\Phi_3^{-\bar{t}_f} := \int_0^{\bar{t}_f(r,\theta)} f_3(\Phi_1^{-t}(r,\theta), \Phi_2^{-t}(r,\theta))dt = 1. \tag{6.13}$$

Some more computations are needed in this case in order to get the Jacobian to be used in the computations in (6.9). Assume we want to compute $D_x\Phi_{co}$ (a similar discussion holds for $D_x\Phi_{counter}$). By the chain rule, the problem will be solved if we know how to compute $D_x\Phi^{t_f}$. To this goal, set $x_1 := r$, $x_2 := \theta$, $x := (x_1, x_2)$, and f_i, $i = 1,2,3$, the right hand sides of (6.1), (6.2), (6.3), respectively. We have, for $i,j = 1,2$

$$\begin{aligned} \frac{\partial \Phi_i^{t_f}}{\partial x_j} &= \frac{\partial \Phi_i^{t_f}}{\partial x_j} + \frac{\partial \Phi_i^{t_f}}{\partial t_f}\frac{\partial t_f}{\partial x_j} \\ &= \frac{\partial \Phi_i^{t_f}}{\partial x_j} + f_i(\Phi^{t_f}(x))\frac{\partial t_f}{\partial x_j}. \end{aligned} \tag{6.14}$$

Differentiating (6.8) with respect to x_j, we get

$$\frac{\partial \Phi_3^{t_f}}{\partial x_j} + \frac{\partial \Phi_3^{t_f}}{\partial t_f}\frac{\partial t_f}{\partial x_j} = 0, \tag{6.15}$$

$$\frac{\partial \Phi_3}{\partial x_j} + f_3(\Phi^{t_f}(x))\frac{\partial t_f}{\partial x_j} = 0, \tag{6.16}$$

from which we get

$$\frac{\partial t_f}{\partial x_j} = -\frac{\frac{\partial \Phi_3^{t_f}}{\partial x_j}}{f_3(\Phi^{t_f}(x))}. \tag{6.17}$$

Replacing this into (6.14), we obtain

$$\frac{d\Phi_i^{t_f}}{dx_j} = \frac{\partial \Phi_i}{\partial x_j} - \frac{f_i(\Phi^{t_f}(x))}{f_3(\Phi^{t_f}(x))}\frac{\partial \Phi_3^{t_f}}{\partial x_j}, \tag{6.18}$$

$i, j = 1, 2$. Notice now that all of the terms in (6.18) are available after numerical integration of (6.1), (6.2), (6.3), (6.7). The previous mathematical development and in particular formula (6.18) are the basis of our numerical algorithm for computation of (6.10) for the case of the real flow.

6.3 Numerical experiments

Numerical simulations provided the Poincare' maps associated with the partitioned-pipe mixer both for the case of the plug flow and for the real flow. For the *plug flow*, the results of the simulations are shown in Figures 6.3-6.6, for the cases of $\beta = 2$ and $\beta = 3$. For intermediate values of the mixing strength (see the case $\beta = 2$), the mixing behavior for the partitioned- pipe mixer when two adjacent elements rotate in opposite direction (counter-rotating case) is better than the one for which two adjacent elements rotate in the same direction (co-rotating case). Accordingly the entropy is higher. Results for higher values of β show that both co-rotating and counter-rotating mixers mix well and the difference in entropy values decreases. For the *real mixer* simulation results are shown in the Figures 6.7-6.12, for $\beta = 2, 3, 5$. Also in this case, a higher value of the entropy corresponds to a better mixing. A comparison of the quality of the mixing for different situations with the values of the entropy, confirm hence what was discussed in a theoretic context in Chapters 3 through 5, namely entropy is a good quantifier of the quality of mixing.

In the simulations, points were placed on an $r - \theta$ grid with eight values of r equally spaced between 0 and 1 and seven values of θ equally spaced between 0 and

π, and simulations executed for 300 iterations (namely 300 pairs of elements). The program exploited the symmetry of the flow both in the computation of the Poincare' maps and in the computation of Lyapunov exponents through (6.9). An approximate value of the entropy was achieved via numerical integration.

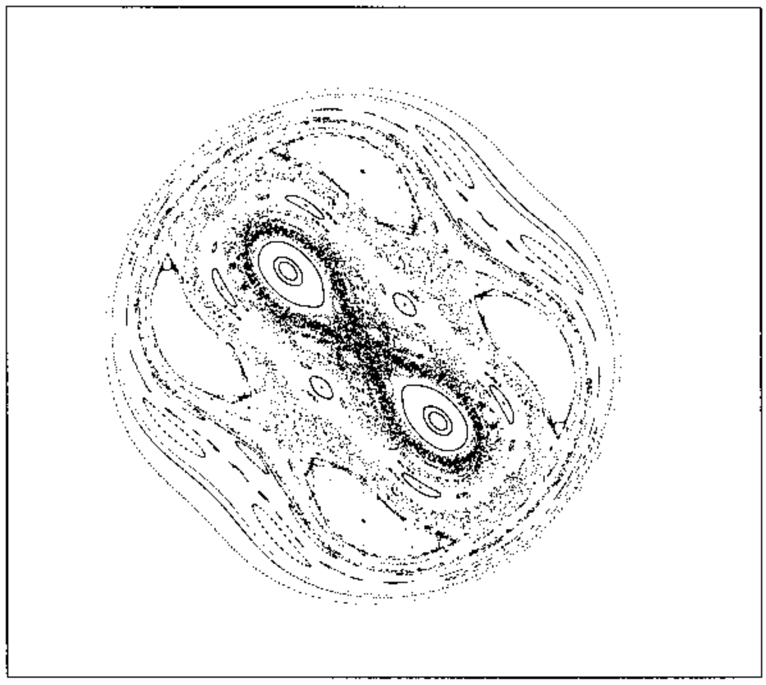

Figure 6.3: (Plug Flow) Poincare' map for co-rotating mixer; $\beta = 2$; Entropy=0.063

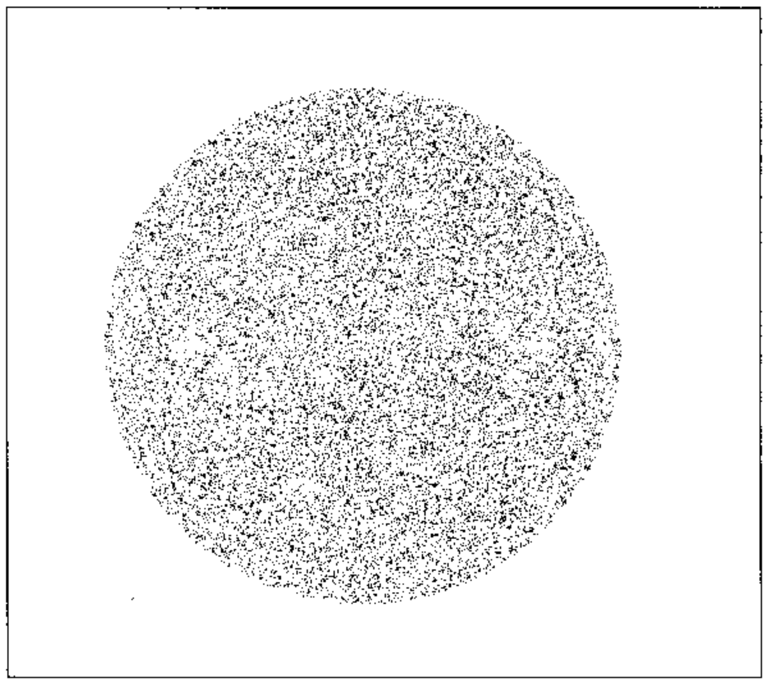

Figure 6.4: (Plug Flow) Poincare' map for counter-rotating mixer; $\beta = 2$; Entropy=0.61

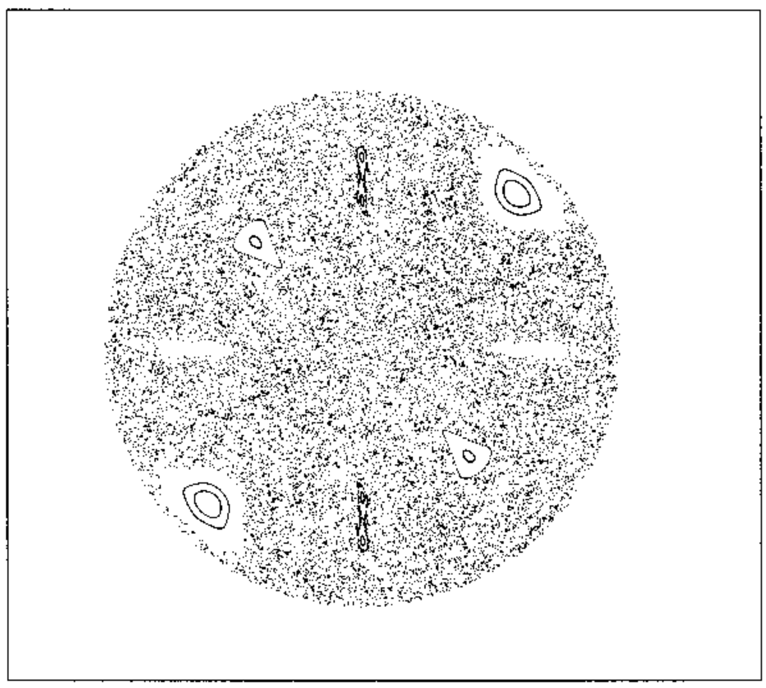

Figure 6.5: (Plug Flow) Poincare' map for co-rotating mixer; $\beta = 3$; Entropy=0.54

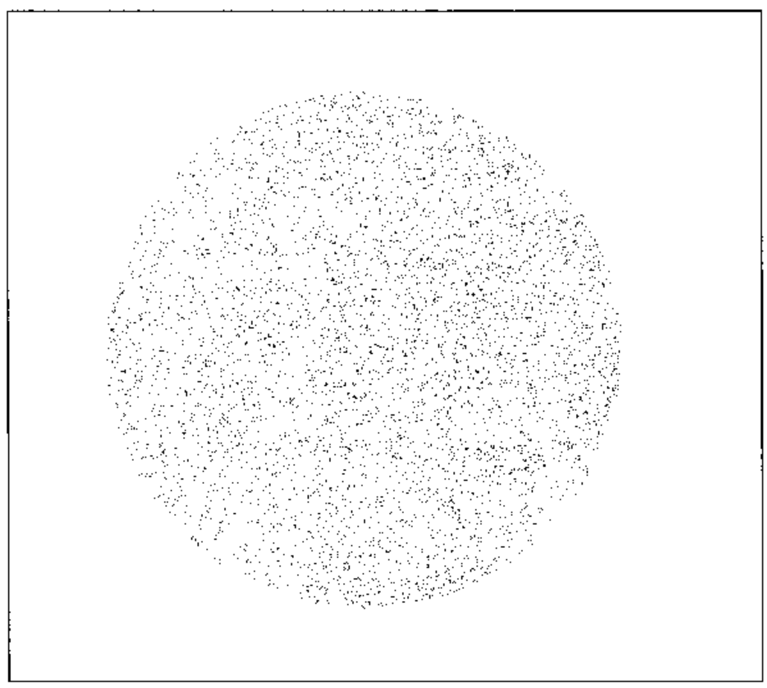

Figure 6.6: (Plug Flow) Poincare' map for counter-rotating mixer; $\beta = 3$; Entropy=1.4415

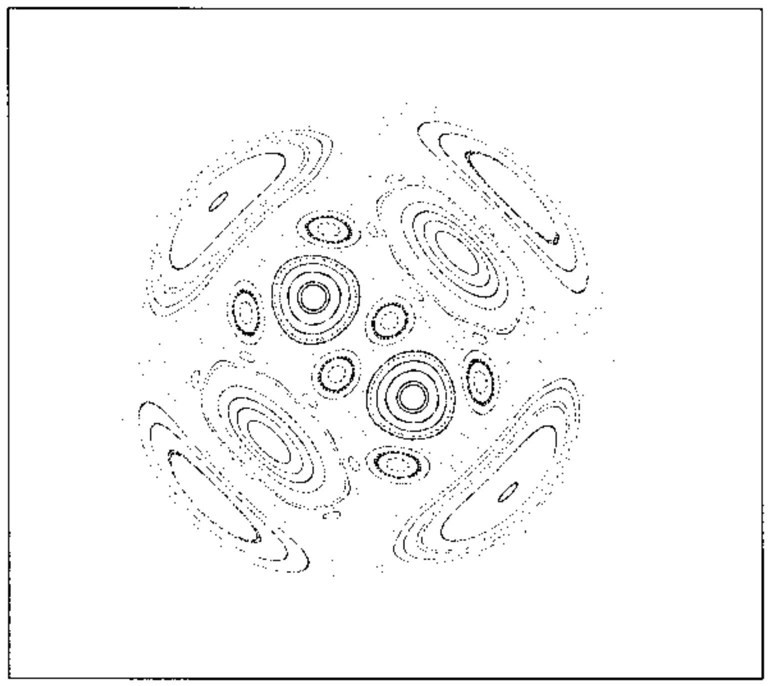

Figure 6.7: (Real Flow) Poincare' map for co-rotating mixer; $\beta = 2$; Entropy=0.026

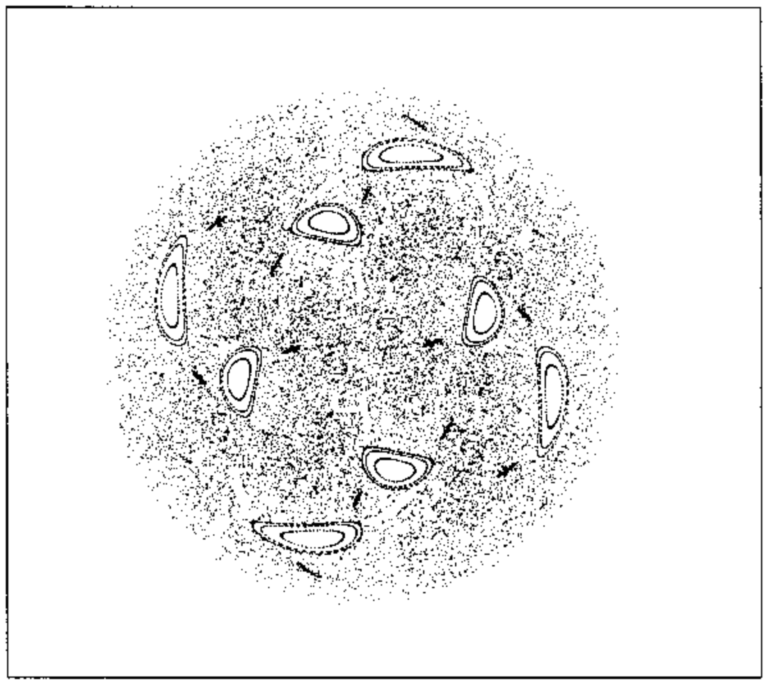

Figure 6.8: (Real Flow) Poincare' map for counter-rotating mixer; $\beta = 2$; Entropy=0.087

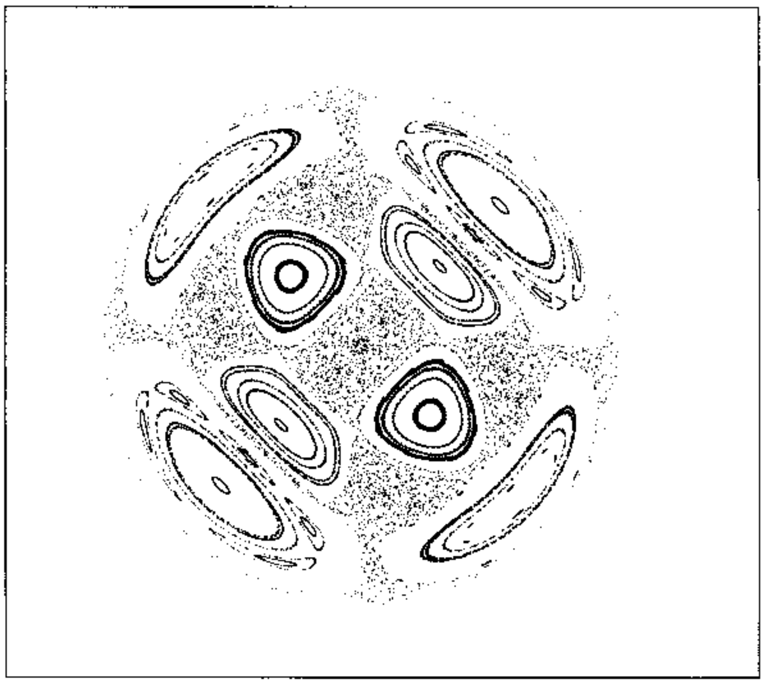

Figure 6.9: (Real Flow) Poincare' map for co-rotating mixer; $\beta = 3$; Entropy=0.042

6.4 Notes and References

The Partitioned Pipe Mixer was studied in [16] and then in [30] where the problem was posed about the different mixing behavior for co-rotating and counter-rotating mixers. The numerical analysis we have carried out in this chapter provides an explanation of the different quality of mixing in these cases in terms of ergodic theoretic entropy. In this context, an analogy can be drawn with the results of the previous chapter where sequences of shear flows were studied on the two dimensional torus which are purely horizontal or purely vertical. It was shown that the alternating sequence of different actions is the one with maximum entropy.

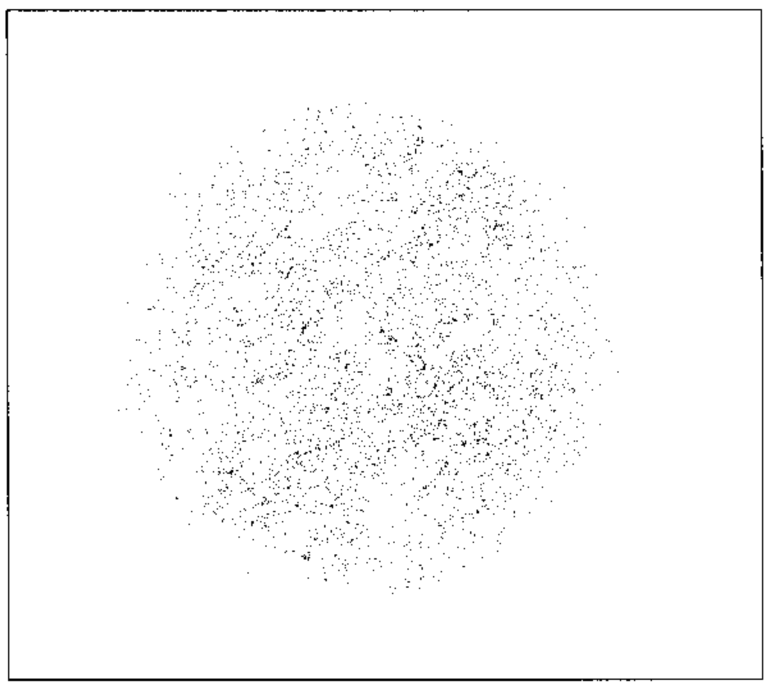

Figure 6.10: (Real Flow) Poincare' map for counter-rotating mixer; $\beta = 3$; Entropy=0.18

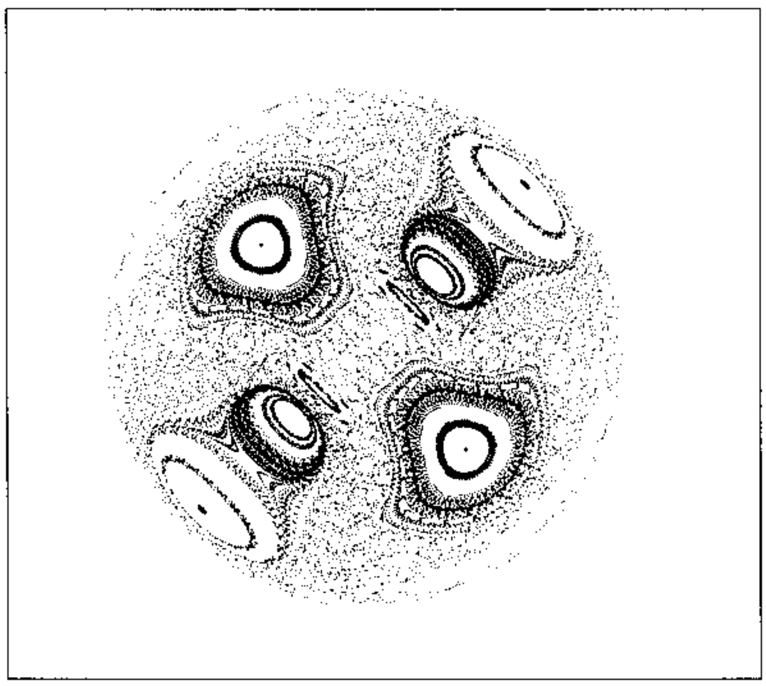

Figure 6.11: (Real Flow) Poincare' map for co-rotating mixer; $\beta = 5$; Entropy=0.049

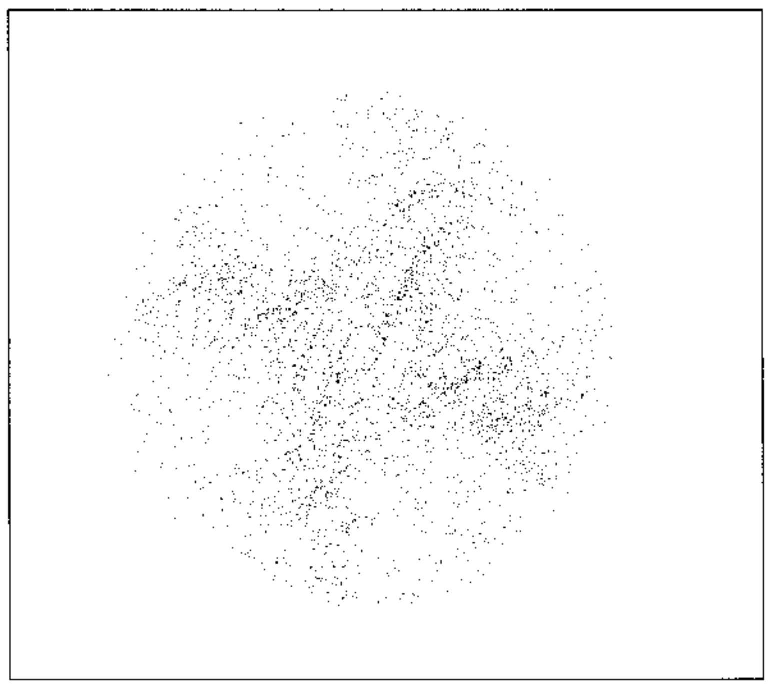

Figure 6.12: (Real Flow) Poincare' map for counter-rotating mixer; $\beta = 5$; Entropy=0.358

Bibliography

[1] H. Aref, Stirring by chaotic advection, *Journal of Fluid Mechanics 143*, pp. 1-21, 1984.

[2] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, 2 nd Edition, Springer Verlag, New York 1989

[3] Arnold V.I. and Avez A., *Ergodic Problems of Classical Mechanics*, W.A. Benjamin, Inc., New York, 1968.

[4] G. K. Batchelor, *An Introduction to Fluid Dynamics*, Cambridge University Press, London, 1967

[5] P. Billingsley, *Ergodic Theory and Information*, John Wiley & Sons, New York, 1965

[6] Bruckner A.M., J.B. Bruckner and B.S. Thomson, *Real Analysis*, Prentice-Hall, Inc., Upple Saddle River, New Jersey, 1997

[7] N. I. Chernov, Ergodic and Statistical Properties of Piecewise Linear Hyperbolic Automorphisms of the 2-Torus, *Journal of Statistical Physics* Vol. 69 pp. 111-133, (1992)(see also the Addendum in *Journal of Statistical Physics* Vol. 71 pp. 341-347.) (1993)

[8] I. G. Currie, *Fundamental Mechanics of Fluids*, 2nd ed., McGraw-Hill, Inc, New York NY, 1993

[9] D. D'Alessandro, I Mezić and M. Dahleh, On the existence of time averages for time-varying dynamical systems, in *Proceedings of the 37 th Conference on Decision and Control*, Tampa FL, Dec. 1998, also submitted to *Journal of Statistical Physics*

[10] D. D'Alessandro, M. Dahleh and I. Mezić, Maximum entropy approach to the control of mixing, *Proceedings of the 1998 American Control Conference*, Albuquerque, NM, June 1997

[11] D. D'Alessandro, M. Dahleh and I. Mezić, Control of fluid mixing using entropy methods, *Proceedings of the 1998 American Control Conference*, Philadelphia, PA, June 1998

[12] D. D'Alessandro, M. Dahleh and I. Mezić, Control of mixing in fluid flows: a maximum entropy approach, to appear in *IEEE Transactions on Automatic Control*

[13] Franjione J.G. and J.M. Ottino, Symmetry concepts for the geometric analysis of mixing flows, *Phil. Trans. R. Soc. Lond.*, Series A, (Vol. 338) pp. 301-323. (1992)

[14] P.R. Halmos, In general a measure preserving transformation is mixing, *Ann. of Math.*, Vol. 45, pp. 786-792, (1944)

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985

[16] Khakhar, D.V., J.G. Franjione, and J.M. Ottino, A case study of chaotic mixing in deterministic flows: the partitioned-pipe mixer, *Chem. Eng. Sci.*, **42**, 2909-26.

[17] Katok A. and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1997

[18] A. B. Katok and J. Strelcyn, *Invariant Manifolds, Entropy and Billiards; Smooth maps with Singularities*, Lecture Notes in Mathematics, Vol. 1222, Springer Verlag, Berlin, Heidelberg, New York, 1986

[19] A. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea, New York, 1950

[20] U. Krengel, *Ergodic Theorems*, de Gruyter, Berlin-New York, 1985

[21] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd edition, Course of Theoretical Physics, Volume 6, Pergamon Press 1987

[22] F.H. Ling, Chaotic mixing in a spatially periodic continuous mixer, *Phys. Fluids A*, Vol. 5, pp. 2147-2160. (1993)

[23] F.H. Ling, The effect of mixing protocol on mixing in discontinuous cavity flows, *Phys. Letters A*, Vol. 177, pp. 331-337. (1993)

[24] Liu M, F.J. Muzzio and R.L.Peskin, Quantification of mixing in aperiodic chaotic flows, *Chaos, Solitons & Fractals*, Vol. 4, pp. 869-893. (1994)

[25] R. Mañé, Lyapunov exponents and stable manifolds for compact transformations, in *Geometric Dynamics*, edited by J. Palis, Lecture Notes in Mathematics, Vol. 1007, pp. 522-577, Springer Verlag, Berlin, Heidelberg, New York, 1983

[26] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, Springer Verlag, Berlin, Heidelberg, New York, 1987

[27] L. M. Milne-Thompson, *Theoretical Hydrodynamics*, 4th ed., The Macmillan Company, New York, N. Y., 1962

[28] V. I. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, *Transactions of the Moscow Mathematical Society*, Vol. 19, pp. 197-221, (1968)

[29] J.M. Ottino, The mixing of fluids, *Scientific American*, pp. 56-67, January 1989.

[30] J.M. Ottino, *The kinematics of mixing: stretching, chaos, and transport*, Cambridge University Press, New York, 1989.

[31] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, McGraw-Hill, 1991

[32] K. Petersen, *Ergodic Theory*, Cambridge University Press, Cambridge, 1983

[33] Y. Pesin, Characteristic Lyapunov exponents and smooth ergodic theory, *Russian Math. Surveys*, Vol. 32-4, pp. 55-114, (1977)

[34] M. Pollicott, *Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds,* Cambridge University Press, Cambridge, 1993

[35] H.S. Robertson, *Statistical Thermophysics*, Prentice-Hall, Englewood Cliff, New Jersey, 1993

[36] V.A. Rokhlin, A general measure-preserving transformation is not mixing, *Dokl. Akad. Nauk SSSR*, Vol. 60, pp. 349-351, (1948)

[37] V. A. Rokhlin, Selected topics from the metric theory of dynamical systems, *Uspehi Mat. Nauk*, Vol.4, pp. 57-128. *Amer. Math. Soc. Transl. Series 2*, Vol.49, pp. 171-240, (1966)

[38] W. Rudin, *Real and Complex Analysis*, Mc-Graw-Hill series in Higher Mathematics, 1987

[39] D. Ruelle, Ergodic theory of differentiable dynamical systems, *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, Vol. 50, pp. 27-58, (1979)

[40] Y. Sinai, *Introduction to Ergodic Theory*, Princeton University Press, Princeton, NJ, 1977

[41] S. Vaienti, Ergodic properties of the discontinuous sawtooth map, *Journal of Statistical Physics* 67 pp. 251-269. (1992)

[42] P. Walters, *An Introduction to Ergodic Theory*, Springer Verlag, Berlin, Heidelberg, New York, 1982