NOISE AND CHAOS

CARLANGELO LIVERANI

Abstract. In the common usage there exists a dichotomy between deterministic and random behaviour. In this lecturers I will discuss in which sense random behaviour can arise in a deterministic system. The emphasis will not be on generalities and philosophy but, rather, on the precise quantitative analysis of simple, but far from trivial, examples.

Contents

1. Noise

The traditional reductionist approach to study nature consists in identifying the phenomenon one is interested in and then consider it as an isolated system. A classical example is provided by Hamiltonian mechanics that describes with remarkable success an incredibly wide array of systems. However, as we look deeper into the phenomena, we realise that, on the one hand, the distance between the fundamental laws that describe a system and the phenomena that we observe keep widening and, on the other hand, at a more attentive scrutiny the very concept of *isolated system* risks to crumble.

Date: September 11, 2016.

This are super condensed and hastily written notes for the mini-course given at the XLI SUMMER SCHOOL ON MATHEMATICAL PHYSICS (RAVELLO - SEPTEMBER 5 - 17, 2016). Ravello, September 12 - 16, 2016.
Read at your own risk.

As a trivial instance of these problems consider *friction*. Friction is not naively part of Hamiltonian mechanics and it is not obvious how to describe it other than by some phenomenological rule. In order to start to have an understanding of how friction might arise (even in the context of Hamiltonian mechanics) one has to realise that the description of the system is incomplete and other degrees of freedom (often a huge number of them) have to be taken into consideration. This has become apparent only with the advent of *statistical mechanics* which posits that the the macroscopic behaviour that we witness is the result of the cumulative effect of an enormous number of degrees of freedom. This counterintuitive fact (that an enormous number of degrees of freedom on a certain scale can give rise to fairly simple cumulative behaviour on a larger scale) is often called *renormalization*. Renormalization is a very vague word that nevertheless inspires a powerful set of technical ideas and tools both in physics and in mathematics.

To better understand the problems with the concept of isolated systems, think of a pendulum. To consider it really isolated one has to worry about the suspension point, that could vibrate if, for example, a car passes near by. Since such vibrations go thru the earth, they will have most likely a frequency of a few hertz. Then one needs to worry about interactions with the air. Even if the air seems still, just talking will produce vibrations that might interfere with the pendulum, such vibrations might be in the order of 1000 Hz. Taking the air out will not help: if you use a cell phone, then you produce electromagnetic oscillations that might interfere with the pendulum, this time of a frequency around 10^{10} Hz; then there is light, this time oscillating around 10^{14} Hz, and so on. Of course, you might argue that all these contributions are small, but what it is worrisome is that they seem to be present at all frequencies, so the cumulative exchange of energy could be enormous.

What is even more worrisome is that even a very small exchange in energy might create a disaster in the perception that the pendulum is isolated. To get acquainted with this problem consider the very concrete example of a pendulum with a vibrating suspension point when the initial condition is close to the unstable fixed point. If the suspension point vibrates with an amplitude of ε^2 at a frequency comparable with the oscillation frequency of the pendulum, then an initial condition at a distance ε from the unstable equilibrium point could gather enough energy in a swing to rotate instead of oscillating at the next passage near the unstable equilibrium. If this happens or not depends on differences on the initial conditions of order ε^2, so it is very hard to predict.

An even more impressive example is given by two billiards ball of radius R is a square table of size L with mass one and kinetic energy K. To simplify matters consider the case in which $2\sqrt{2}R < L < 4R$ so that there exists a length ℓ_0 such that the distance between two collisions of the balls is, at most, ℓ_0. Then a change of the initial condition by ε will create a change of velocities after the next collision proportional to R^{-1}. Thus the change in the trajectories will grow, at least, like $e^{N/R}\varepsilon$, where N is the number of collisions among balls. Note that $N\ell_0 \leq T\sqrt{2K}$, where T is the time. Thus the change in the trajectory due to a small initial perturbation grows, at least, like $e^{T\sqrt{2K}/R^2}\varepsilon$. Thus, if $R = .1$ meters e $\sqrt{2K} = 10$ meters per second (36 Km per hour), then the perturbation of the trajectory grows, at least, like $e^{10^3 T}$. Accordingly, suppose that you observe two identical systems and, at a certain point only on one of them acts, for 10^{-10} seconds a force of size 10^{-90} newtons. This creates a change in velocity of size 10^{-100} meters per second which, after a tenth of a second, will create a difference in the coordinates of the same size of the box. Clearly it is very difficult to imagine that such a system can be isolated.

A standard way of taking into account all the above issues is to add to the system a small random perturbation. Namely, if you have a system of the type

$$\dot{x} = F(x),$$

where $F \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$, you might add to it a noise of the form

$$dx = F(x)dt + \varepsilon\Sigma(x)dB \tag{1.1}$$

where B is a d dimensional Brownian motion and $\Sigma(x)$ is a positive symmetric matrix. We have thus turned a differential equation into a stochastic differential equation, where the noise is supposed to model the (hopefully small) effect of all the degree of freedom that have been ignored.

Yet, note that, in some cases, (1.1) could be a bad model. Consider, for example, the the Hamiltonian system

$$\begin{aligned} dq &= pdt \\ dp &= -V'(q)dt + \varepsilon\sigma dB \end{aligned}$$

where we put the noise only on the second equation because we think of it as a random force acting on the system. Note that, setting $H(q,p) = \frac{1}{2}p^2 + V(q)$, by Ito's formula,

$$dH = \frac{\varepsilon^2\sigma^2}{2}dt + \varepsilon\sigma p dB.$$

Hence,

$$\mathbb{E}(H) = \frac{\varepsilon^2\sigma^2}{2}t.$$

In other words, the systems heats up indefinitely. If this were a good model for the influence of the ignored degrees of freedom, then every system should keep getting hotter and hotter; this is not what we see. The usual fix for this problem is to consider the equations

$$\begin{aligned} dq &= pdt \\ dp &= -V'(q)dt - \gamma p dt + \varepsilon\sigma dB, \end{aligned} \tag{1.2}$$

where we have added a *friction* to the system. The above is called a Langevin equation or an Ornstein-Uhlenbeck process. Such a process does now have an invariant measure. Indeed,[1]

$$\begin{aligned} &\frac{d}{dt}\int e^{-\beta H(q,p)}\mathbb{E}(\varphi(q(t,q,p),p(t,q,p)))\big|_{t=0} \\ &= \int e^{-\beta H(q,p)}\left\{p\partial_q\varphi - [V'(q) + \gamma p]\partial_p\varphi + \frac{\varepsilon^2\sigma^2}{2}\partial_p^2\varphi\right\} \\ &= \int e^{-\beta H(q,p)}\left\{-\beta\gamma p^2 + \gamma + \varepsilon^2\sigma^2\beta^2p^2 - \beta\varepsilon^2\sigma^2\right\}\varphi. \end{aligned}$$

Thus the derivative is zero provided $\gamma = \varepsilon^2\sigma^2\beta$ and we obtain the interpretation that the friction (that can also be interpreted as a drift) is related to the inverse of the temperature and the diffusivity (this is some sort of Einstein relation).

[1] Here $\mathbb{E}$ is the expectation with respect to the Brownian motion and $q(t,q,p), p(t,q,p)$ is the process with initial conditions (q,p).

Note that, possibly with some work, it might be possible to reduced the effect of external factors, hence making ε smaller. It is then clear that, in the study of (1.2), we should be interested only in phenomena that are, in some sense, independent on ε. Indeed, if some behaviour would be present for some level of noise and not for a near by level, this would mean that our model is rather useless for applications.

The study of equations of the type (1.1), (1.2) is a wide, currently active, and interesting branch of mathematics, but will not be our focus. Our focus will be to try to understand in which way a deterministic dynamics can give rise to a behaviour similar to the stochastic one.

2. Chaos

I claimed that a simple system like a pendulum can exhibit the phenomena of strong dependence from initial conditions, colloquially often called *chaos*. However, in such an example the set for which we can show a chaotic behaviour is of zero Lebesgue measure. It is widely believed that such properties holds for a positive measure set of point, but we are very far from a proof of such a fact. By KAM theory a positive measure of trajectory have instead a regular behaviour. Thus we expect, in general, realistic systems to have a mixture of regular and chaotic motion. Unfortunately, we have no idea how to treat such systems. It is then natural to start the study from simpler systems in which one of the possibilities is absent. Here we will concentrate on systems for which all the trajectories have a strong dependence from the initial conditions. This are called *uniformly hyperbolic systems*. Examples of paramount importance are, e.g., geodesic flows on manifolds of negative curvature, the automorphisms of a two dimensional torus and the billiard balls systems we previously discussed.

Yet, the study of the above mentioned systems entails quite a bit of technical difficulties that cloud the main issues. To explain in their simplest form the ideas I want to put forward in this course it is convenient to consider the simplest possible example: smooth expanding maps of a circle. We will therefore consider this seemingly ridiculously simple model: the *macroscopic* degree of freedom is $\theta \in \mathbb{T}$ and does nothing. The *microscopic* dynamics is given by an expanding circle map. The influence of the microscopic variable on the macroscopic one is small. In mathematical terms, such systems are described by maps $F_\varepsilon \in \mathcal{C}^r(\mathbb{T}^2, \mathbb{T}^2)$, $r > 1$, defined as

$$\begin{aligned} &F_\varepsilon(x,\theta) = (f(x,\theta), \theta + \varepsilon\omega(x,\theta)) \\ &\partial_x f \geq \lambda > 1; \quad \|\omega\|_{\mathcal{C}^r} = 1. \end{aligned} \tag{2.1}$$

Given some initial condition $(x,\theta) = (x_0,\theta_0)$, the time evolution of the system is described by $(x_n,\theta_n) = F_\varepsilon^n(x_0,\theta_0)$.

As mentioned, for $\varepsilon = 0$, θ is a constant of motion. The study of the system (2.1) for $\varepsilon \neq 0$ has proven rather non trivial and is far from being completed. Accordingly, here we will just take it as a motivation that points us in a specific direction of research.

As we have explained, there should be a scale separation between the macroscopic and the microscopic variables. Here the scale separation is in time and is given by ε, hence the proper way of thinking is that the *macroscopic time* is ε slower than the microscopic time. In other words we should be interested in the behaviour of the variable $\theta_\varepsilon \in \mathcal{C}^0([0,T],\mathbb{T})$ defined by

$$\theta_\varepsilon(t) = \theta_{\lfloor \varepsilon^{-1}t\rfloor} + (\varepsilon^{-1}t - \lfloor \varepsilon^{-1}t\rfloor)(\theta_{\lfloor \varepsilon^{-1}t\rfloor+1} - \theta_{\lfloor \varepsilon^{-1}t\rfloor}),$$

and, first, we should ask ourselves:

Does θ_ε has some limiting behaviour for $\varepsilon \to 0$?

To further simplify the problem, let us start with the case $\partial_\theta\omega = \partial_\theta f = 0$. This is called a *skew product*. In such a simple situation

$$\left|\theta_\varepsilon(t) - \varepsilon \sum_{k=0}^{\lfloor \varepsilon^{-1}t\rfloor-1} \omega\circ f^k(x_0)\right| \leq C_\# \varepsilon. \tag{2.2}$$

Thus our variable is described by an ergodic average. By Birkhoff ergodic theorem, see A.4, the $\lim_{\varepsilon\to 0}\varepsilon\sum_{k=0}^{\lfloor \varepsilon^{-1}t\rfloor-1}\omega\circ f^k(x_0)$ exists for almost every point with respect to any invariant measure of f, but what are such invariant measures?

2.1. **Invariant measures.** Deterministic system often have a lot of invariant measures (see Appendix A for more informations on this point). In particular, to any periodic orbit is associated an invariant measure. Given such plentiful possibilities, we need a criteria to select the invariant measures that we think might be physically relevant. A common choice is to consider measures that can be obtained by push forward of measures absolutely continuous with respect to Lebesgue.[2] That is, let $d\mu = h(x)dx$, $h \in L^1(\mathbb{T}^1, \text{Leb})$ and define $f_*\mu(\varphi) = \mu(\varphi\circ f)$. Note that if μ is a probability measure (i.e., $h \geq 0$ and $\mu(1) = 1$), then also $f_*\mu$ is a probability measure. Then

$$\frac{1}{n}\sum_{k=0}^{n-1} f_*^k\mu$$

is a weakly compact set, hence it has accumulation points. On can easily see that such accumulation points are invariant measures for f, that is fixed points for f_* (see Appendix A for details). We would then like to study such fixed points.

A simple change of variables shows that $\frac{d(f_*\mu)}{d\,\text{Leb}} = \mathcal{L}h$ where

$$\mathcal{L}h(x) = \sum_{f(y)=x} \frac{h(y)}{f'(y)}.$$

The operator $\mathcal{L}$ is called a (Ruelle) transfer operator. Of course an operator, to be properly defined, must have a well specified domain. Since

$$\int |\mathcal{L}h(x)|dx \leq \int \mathcal{L}|h|(x)dx = \int |h(x)|dx$$

it follows that $\mathcal{L}$ is a contraction on $L^1(\mathbb{T}, \text{Leb})$. However, the spectrum of $\mathcal{L}$ on L^1 turns out to be the full unit disk, not a very useful fact.

Following Lasota-Yorke, we look then at the action of $\mathcal{L}$ on $W^{1,1}$.

$$\frac{d}{dx}\mathcal{L}h = \mathcal{L}\left(\frac{h}{f'}\right) - \mathcal{L}\left(h\frac{f''}{(f')^2}\right).$$

The above implies the so called *Lasota-Yorke inequalities*

$$\begin{aligned} &\|\mathcal{L}h\|_{L^1} \leq \|h\|_{L^1} \\ &\|(\mathcal{L}h)'\|_{L^1} \leq \lambda^{-1}\|h'\|_{L^1} + D\|h\|_{L^1}. \end{aligned} \tag{2.3}$$

[2] This are often called *physical measures*.

Such inequalities imply that $\mathcal{L}$, when acting on $W^{1,1}$, has a spectral gap. To give an idea of the why, let us consider the simple case in which $D = \|\frac{f''}{(f')^2}\|_{L^\infty}$ is small, more precisely $\lambda^{-1} + D < 1$.

Note that, if $\text{Leb}(h) = 0$, then $\text{Leb}(\mathcal{L}h) = 0$, hence the space $\mathbb{V} = \{h \in L^1 : \text{Leb}(h) = 0\}$ is invariant under $\mathcal{L}$. Also, if $h \in \mathbb{V}$, then, by the mean value theorem and since in one dimension $W^{1,1} \subset \mathcal{C}^0$, there must exists x_* such that $h(x_*) = 0$, thus

$$\|h\|_{L^1} = \int_{\mathbb{T}} |h(x)| = \int_{\mathbb{T}} \int_{x_*}^{x} |h'(y)| \leq \|h'\|_{L^1}.$$

Next, let us define the norm $\|h\|_{W^{1,1}} = \|h'\|_{L^1} + a\|h\|_{L^1}$ for some $a > 0$ to be chosen shortly. Then, for $h \in \mathbb{V}$,

$$\|\mathcal{L}h\|_{W^{1,1}} \leq \lambda^{-1}\|h'\|_{L^1} + (D+a)\|h\|_{L^1} \leq (\lambda^{-1}+D+a)\|h'\|_{L^1} \leq (\lambda^{-1}+D+a)\|h\|_{W^{1,1}}.$$

We can then choose a such that $\lambda^{-1} + D + a < 1$, and we have that $\mathcal{L}$ is a strict contraction on $\mathbb{V}$. Since $\mathcal{L}' \text{Leb} = \text{Leb}$, $1 \in \sigma(\mathcal{L})$ and we have that there exists $h_* \in L^1$ such that $\mathcal{L}h = h_* \text{Leb}(h) + Qh$, where $\|Q\|_{W^{1,1}} < 1$ and $\text{Leb}\, Q = Qh_* = 0$. We have just proven that $h_*(x)dx$ is the only invariant measure of f absolutely continuous with respect to Lebesgue.[3]

In fact, the above spectral decomposition, and hence the uniqueness of the invariant measure absolutely continuous with respect to Lebesgue, holds in much higher generality, in particular for each $f \in \mathcal{C}^2$ such that $|f'| \geq \lambda > 1$ (see [1] for an exhaustive discussion or have a look Appendix C, in particular Theorem C.1, for a quicker discussion sufficient for the present case).

2.2. **Back to our problem.** By the results of the previous section it follows that, for lebesgue almost all x,

$$\lim_{\varepsilon \to 0} \theta_\varepsilon(t) = \bar{\theta}(t) = t \int \omega(x) h_*(x) dx =: t\bar{\omega}.$$

That is, the limit satisfies the differential equation

$$\frac{d}{dt}\bar{\theta} = \bar{\omega}. \tag{2.4}$$

This is a rather simple example of *averaging*. In general, any map $f_\theta(x) = f(x, \theta)$ has a unique invariant physical measure μ_θ with density h_θ and we can define $\bar{\omega}(\theta) = \mu_\theta(\omega(\cdot, \theta))$ and $\bar{\omega} = \omega - \bar{\omega}$. The it is possible to prove that the limit satisfies, see [2] for details,

$$\frac{d}{dt}\bar{\theta} = \bar{\omega}(\bar{\theta}). \tag{2.5}$$

We have then seen how a very simple *macroscopic* behaviour arises form an complex *microscopic* behaviour. Remark that (2.5) looks like the equation of an isolated system, although it describes the evolution of a degree of freedom in contact with another (microscopic) degree of freedom whose effect has been averaged out.

How can we detect that in reality the system is not isolated? To do that we have to look at it a bit more closely or for a longer time. In the next section we will do the former.

---

[3] To make the argument precise use that $W^{1,1}$ is dense in L^1.

3. Noise form determinism

To look more closely means, for example, to consider the variable

$$\zeta_\varepsilon = \frac{1}{\sqrt{\varepsilon}}(\theta_\varepsilon(t) - \bar{\theta}(t))$$

and ask if it has some limiting behaviour when $\varepsilon \to 0$. In order to answer to such a question it is necessary first to discuss which initial conditions are physically reasonable.

3.1. **Initial Conditions.**

Physically to fix an initial condition is equivalent to preparing the system in some state. Let us consider, for example, the problem of preparing a bunch of systems in the "same state". What can we do?

One possibility is to take one system as the reference system. Start with a lot of systems, make a measure, and discard all the systems that give a value different form the reference one. For simplicity, let us consider the system (2.1) and assume that we can make measures only on the variable θ. Clearly, by consecutive measures we can get some information also on the variable x, but not very precise.[4] Say that we can determine that x belongs to some interval I, $|I| = \delta$, $\delta < 1$.[5]

So, we do a measure and we determine that, for the reference system, $x \in I_0$ and we discard all the systems for which $x \notin I_0$. We wait a fixed time, say t_0 (which corresponds to the microscopic time $n_0 = \varepsilon^{-1} t_0$), and repeat the measurement. What will happen?

Due to the expansivity of the map, after time t_0, before the measure we will be able to say only that x belongs to some interval I_0', $|I_0'| \geq \lambda^{n_0}\delta$. Most likely it will be $\lambda^{n_0}\delta > 1$, that is we have o idea of where x might be. We perform the measure and again we are going to discard the systems that differ from the reference one. How many systems we discard? That depends on how the initial systems were distributed. Suppose we discard a percentage $1 - C\lambda^{-n_0}$ of systems, that would mean that originally the systems were distributed not so differently from Lebesgue.

Now we can repeat again the measure. Note that now we are considering systems that had the same behaviour for some time. We can then ask ourselves if this means that they will have similar behaviour in the future. That would mean that, next time, we will discard a smaller percentage of systems. If you consider the previous discuss you will see that this is unlikely. If the original systems were distributed not so differently from Lebesgue, then you would expect to keep a percentage λ^{-n_0} of systems every time. In other words, there is no way to determine the variable x with a precisione larger than δ. Asking for a big grant to built a better measurement apparatus will not help you much, you will just decrease a bit the value of δ.

What can we then use as an initial condition? Well, if we have done the experiment, and we have seen that every time we keep a percentage λ^{-n_0} of systems,

---

[4] Such a measure would correspond to measuring the *istantaneous* velocity with which θ changes.

[5] Note that an error in measurements is inevitable, here we are just saying that it is not too much larger than ε.

then we can assume that the same will happen in the future and this is tantamount to assume that the variable x is a random variable distributed according to a probability distribution absolutely continuos with respect to Lebesgue.[6]

From now on this will be our standing assumption. That is, we consider the system (2.1) with random initial conditions such that, for each $\varphi \in \mathcal{C}^0$,

$$\mathbb{E}(\varphi(x_0, \theta_0)) = \int_{\mathbb{T}} \varphi(x, \theta_*)\rho(x)dx \tag{3.1}$$

where $\rho \in W^{1,1}$ and $\theta_* \in \mathbb{T}$.

Remark 3.1. *Technically these initial conditions are a special case of the measures called* standard pairs *introduced by Dolgopyat and that are a basic tool to investigate the statistical properties of systems with some chaoticity.*

3.2. **Central Limit Theorem.** Having explained that we consider (x_0, θ_0) to be random variables, it follows that $\zeta_\varepsilon(t)$ is a random variable as well. It is then natural to try to compute its distribution. It is well known that, to do so, it suffices to compute the characteristic function [6], that is[7]

$$\Phi(\xi) = \mathbb{E}\left(e^{i\xi\zeta_\varepsilon(t)}\right). \tag{3.2}$$

For simplicity, let us consider again the case of a skew product (i.e. $\partial_\theta \omega = \partial_\theta f = 0$). Then

$$\left|\zeta_\varepsilon(t) - \sqrt{\varepsilon}\sum_{k=0}^{\lfloor \varepsilon^{-1}t\rfloor - 1} \hat{\omega} \circ f^k(x_0)\right| \leq C_\# \sqrt{\varepsilon}, \tag{3.3}$$

where $\hat{\omega} = \omega - \bar{\omega}$. So, up to a precision of order ε, our problem is equivalent to the one of studying the characteristic function of the sum.

To study such a sum several approaches are available: martingale approximations [5], reduction, via standard pairs, to a martingale problem [3] (but see [2] for a didactical presentatiton) and spectral methods. The latter, when it works, is the most powerful, yet it needs stronger hypotheses and hence it has a smaller range of applicability. However, for the current presentation is the simplest one to apply and it will then be our method of choice.

The basic idea is to compute directly the characteristic function (3.5), that is

$$\mathbb{E}\left(\exp\left[i\xi\sqrt{\varepsilon}\sum_{k=0}^{\lfloor \varepsilon^{-1}t\rfloor - 1} \hat{\omega} \circ f^k\right]\right).$$

To this end we define the transfer operator, for each $\phi \in L^1$,

$$\mathcal{L}_\nu \phi(x) = \sum_{f(y)=x} \frac{e^{i\nu\hat{\omega}(y)}}{f'(y)}\phi(y) \tag{3.4}$$

[6] Note hovewer that we could have found out that the percentage of discarded systems is different, say $1 - C\lambda^{-\alpha n_0}$, for some $\alpha \neq 1$, and this would mean that our systems are originally distributed according to another measure, a measure singular with respect to Lebesgue.

[7] If in doubt, see the end of the section to see why this is true in the special case of average of smooth functions.

and notice that

$$\mathbb{E}\left(\exp\left[i\xi\sqrt{\varepsilon}\sum_{k=0}^{\lfloor\varepsilon^{-1}t\rfloor-1}\hat{\omega}\circ f^k\right]\right)=\int_{\mathbb{T}}\mathcal{L}_{\xi\sqrt{\varepsilon}}^{\lfloor\varepsilon^{-1}t\rfloor}\rho. \tag{3.5}$$

The problem is then reduced to studying the properties of the operator (3.4). To do so we note that $\mathcal{L}_0=\mathcal{L}$. Since $\mathcal{L}$ has a spectral gap on $W^{1,1}$, it makes sense to try to apply perturbation theory.

Lemma 3.2. *There exists $\nu_0>0$ and continuous functions $C_\nu>0$ and $\rho_\nu\in(0,1)$ such that, for all $|\nu|\le\nu_0$, $\mathcal{L}_\nu=e^{\alpha_\nu}\Pi_\nu+Q_\nu$, $\Pi_\nu Q_\nu=Q_\nu\Pi_\nu=0$, $\|Q_\nu^n\|_{W^{1,1}}\le C_\nu\rho_\nu^n e^{\alpha_\nu n}$. Also $\Pi_\nu(g)=h_\nu\ell_\nu(g)$, $\ell_\nu(h_\nu)=1$, $\ell_\nu(h'_\nu)=0$. In addition, everything is analytic in ν.*

Proof. Note that

$$\|\mathcal{L}_\nu h-\mathcal{L}h\|_{W^{1,1}}\le C_\#|\nu|\|\hat{\omega}\|_{\mathcal{C}^1}\|h\|_{W^{1,1}}.$$

By standard perturbation theory, see [4], it follows that there exists $\nu_0>0$ such that, for all $|\nu|\le\nu_0$ the operator $\mathcal{L}_\nu$ has a maximal simple eigenvalue and a spectral gap. Let e^{α_ν} be the leading eigenvalue, and Π_ν be the associated eigenprojection, again by standard perturbation theory they are analytic in ν. Hence also $\bar{h}_\nu=\Pi_\nu h_*$, $\bar{\ell}_\nu=\operatorname{Leb}\Pi_\nu$ and $\beta_\nu=\bar{\ell}_\nu(\bar{h}_\nu)$ are analytic functions of ν. Moreover, $\bar{h}_0=h_*$ and $\bar{\ell}_0=\operatorname{Leb}$, thus we have $\beta_0=1$ and, provided ν_0 is small enough, $\beta_\nu\neq0$, hence $\Pi_\nu=\bar{h}_\nu\otimes\tilde{\ell}_\nu$ where $\tilde{\ell}_\nu=\beta_\nu^{-1}\bar{\ell}_\nu$. Note however that there is some freedom: $\Pi_\nu=h_\nu\otimes\ell_\nu$ where $\ell_\nu=\gamma_\nu^{-1}\tilde{\ell}_\nu$ and $h_\nu=\gamma_\nu\bar{h}_\nu$ for any arbitrary non zero function γ_ν. We can thus impose the condition

$$0=\ell_\nu(h'_\nu)=\ell_\nu(\gamma'_\nu\bar{h}_\nu+\gamma_\nu\bar{h}'_\nu)=\gamma_\nu^{-1}\gamma'_\nu+\tilde{\ell}_\nu(\bar{h}'_\nu).$$

The above equation yields the choice

$$\gamma_\nu=\exp\left[-\int_0^\nu\tilde{\ell}_{\nu'}(\bar{h}_{\nu'})d\nu'\right].$$

We have thus seen that there are analytic $h_\nu\in W^{1,1}$ and $\ell_\nu\in(W^{1,1})'$, normalised so that $\ell_\nu(h_\nu)=1$, $\ell_\nu(h'_\nu)=0$ and $\Pi_\nu=h_\nu\otimes\ell_\nu$ are analytic in ν. Also

$$\mathcal{L}_\nu h_\nu=e^{\alpha_\nu}h_\nu, \tag{3.6}$$

and $\alpha_0=1$, $h_0=h_*$ and $\ell_0=\operatorname{Leb}$. □

Lemma 3.3. *Fot all $|\nu|\le\nu_0$, the function α_ν satisfies $\alpha_0=\alpha'_0=0$ and $|\alpha_\nu|_{\mathcal{C}^3}\le C_\#$, $\alpha''_0\le0$. Finally, $\alpha''_0=0$ iff there exists $g\in\mathcal{C}^0$ such that $\hat{\omega}=g-g\circ f$; that is, only if $\hat{\omega}$ is a $\mathcal{C}^0$*-coboundary.

Proof. As we mentioned α_ν is analytic in ν by standard perturbation theory, hence the bound on the $\mathcal{C}^3$ norm.

We can differentiate (3.6) obtaining

$$\mathcal{L}'_\nu h_\nu+\mathcal{L}_\nu h'_\nu=\alpha'_\nu e^{\alpha_\nu}h_\nu+e^{\alpha_\nu}h'_\nu. \tag{3.7}$$

Applying ℓ_ν yields

$$\frac{d\alpha_\nu}{d\nu}=i\ell_\nu(\hat{\omega}h_\nu)=:i\mu_\nu(\hat{\omega}). \tag{3.8}$$

Thus $\alpha_0' = 0$. Differentiating again yields

$$\frac{d^2\alpha_\nu}{d\nu^2} = i\ell_\nu'(\hat\omega h_\nu) + i\ell_\nu(\hat\omega h_\nu') = i\ell_\nu'(\omega_\nu h_\nu) + i\ell_\nu(\omega_\nu h_\nu'). \tag{3.9}$$

where $\omega_\nu = \hat\omega - \mu_\nu(\hat\omega)$. On the other hand, from (3.7) and (3.8) we have

$$(\mathbb{1}e^{\alpha_\nu} - \mathcal{L}_\nu)h_\nu' = i\mathcal{L}_\nu(\omega_\nu h_\nu),$$

Since, by construction, $\Pi_\nu h_\nu' = 0$, the above equation can be studied in the space $\mathbb{V}_\nu = (\mathbb{1} - \Pi_\nu)W^{1,1}$ in which $\mathbb{1}e^{\alpha_\nu} - \mathcal{L}_\nu$ is invertible.

Setting $\hat{\mathcal{L}}_\nu := e^{-\alpha_\nu}\mathcal{L}_\nu$and $\hat Q_\nu := e^{-\alpha_\nu}Q_\nu$ we have

$$h_\nu' = i(\mathbb{1} - \hat Q_\nu)^{-1}\hat{\mathcal{L}}_\nu(\omega_\nu h_\nu). \tag{3.10}$$

Doing similar considerations on the equation $\ell_\nu(\mathcal{L}_\nu) = \alpha_\nu \ell_\nu(g)$, we obtain

$$\begin{aligned}
\alpha_\nu'' &= -\ell_\nu(\omega_\nu(\mathbb{1} - \hat Q_\nu)^{-1}(\mathbb{1} + \hat Q_\nu)(\omega_\nu h_\nu)) \\
&= -\sum_{n=1}^{\infty}\ell_\nu(\omega_\nu \hat{\mathcal{L}}_\nu^n(\mathbb{1} + \hat{\mathcal{L}}_\nu)(\omega_\nu h_\nu)) \\
&= -\mu_\nu(\omega_\nu^2) - 2\sum_{n=1}^{\infty}\ell_\nu(\omega_\nu \hat{\mathcal{L}}_\nu^n(\omega_\nu h_\nu)).
\end{aligned} \tag{3.11}$$

Finally, notice that

$$\ell_\nu(\omega_\nu \hat{\mathcal{L}}_\nu^n(\omega_\nu h_\nu)) = \ell_\nu(\hat{\mathcal{L}}_\nu^n(\omega_\nu \circ f^n \omega_\nu h_\nu)) = \mu_\nu(\omega_\nu \circ f^n \omega_\nu)$$

and

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\mu_\nu\left(\left[\sum_{k=0}^{n-1}\omega_\nu \circ f^k\right]^2\right) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k,j=0}^{n-1}\mu_\nu(\omega_\nu \circ f^k \omega_\nu \circ f^j) \\
&= \mu_\nu(\omega_\nu^2) + \lim_{n\to\infty}\frac{2}{n}\sum_{k=1}^{n-1}(n-k)\mu_\nu(\omega_\nu \circ f^k\omega_\nu) \\
&= \mu_\nu(\omega_\nu^2) + 2\sum_{k=1}^{\infty}\mu_\nu(\omega_\nu \circ f^k\omega_\nu).
\end{aligned} \tag{3.12}$$

The above two facts and equation (3.11) yield[8]

$$-\alpha_0'' = \lim_{n\to\infty}\frac{1}{n}\mu_0\left(\left[\sum_{k=0}^{n-1}\omega_0 \circ f^k\right]^2\right) \geq 0. \tag{3.13}$$

Finally, note that the computations in (3.12) imply that, if $\alpha_0'' = 0$,

$$\mu_0\left(\left[\sum_{k=0}^{n-1}\omega_0 \circ f^k\right]^2\right) = -2\sum_{k=1}^{n-1}k\,\mathrm{Leb}(\hat\omega \circ f^k\hat\omega),$$

and the last quantity, by the decay of correlations, is uniformly bounded. Accordingly, $\sum_{k=0}^{n-1}\omega_0 \circ f^k$ is uniformly bounded, and hence weakly compact, in L^2. We

[8] This, together with equation (3.11), is a simple instance of the so called *Green-Kubo formula*.

can then extract a converging subsequence, let $g \in L^2$ be its limit. Then, for each $\phi \in W^{1,1}$,[9]

$$\begin{aligned}\text{Leb}\left(\phi(g\circ f - g)\right) &= \lim_{j\to\infty}\sum_{k=0}^{n_j-1}\text{Leb}\left(\phi[\hat\omega\circ f^{k+1} - \hat\omega\circ f^k]\right)\\ &= -\text{Leb}(\phi\hat\omega) + \lim_{j\to\infty}\text{Leb}(\hat\omega\mathcal{L}^{n_j}\phi) = -\text{Leb}(\phi\hat\omega).\end{aligned}$$

Since $W^{1,1}$ is dense in L^2, it follows that $g - g\circ f = \hat\omega$. The only problem left is to show that g is regular. Note that, it holds as well $g \in L^2(\mu_0)$ and without loss of generality, we can assume $\mu_0(g) = 0$. Then, multiplying by h_0 and applying $\mathcal{L}$

$$\mathcal{L}\hat\omega h_0 = \mathcal{L}gh_0 - \mathcal{L}(g\circ f h_0) = \mathcal{L}(gh_0) - gh_0 = (Q_0 - \mathbb{1})(gh_0).$$

That is $gh_0 = -(\mathbb{1} - Q_0)^{-1}\mathcal{L}\hat\omega h_0 \in W^{1,1} \subset \mathcal{C}^0$. On the other hand note that, since $h_0 \geq 0$, if there exists $\bar x \in \mathbb{T}$ such that $0 = h_0(\bar x)$ then for all $n \in \mathbb{N}$ we have $0 = \mathcal{L}^n h_0(\bar x) = \sum_{y\in f^{-n}(\bar x)} \frac{h_0(y)}{(f^n)'(y)}$. Thus h_0 must be zero on all the pre-images of $\bar x$, but this would imply that $h_0 \equiv 0$. Hence it must be $h_0 > 0$ and then $g \in \mathcal{C}^0$ as claimed. □

We can now collect all our work: suppose we would like to do a measure represented by the function $\psi_{\varepsilon,z}(\zeta_\varepsilon) = \psi((\zeta_\varepsilon - z)\varepsilon^{-\alpha})$, where $\psi \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}_+)$ has support in the interval $[-1, 1]$ and $\alpha \in [0, 1/2)$. This essentially means that we want to know what is probability to find the variable ζ_ε in a given interval of size $2\varepsilon^\alpha$ centred at z. Hence, using (3.3) and (3.5) we want, and can, compute[10]

$$\begin{aligned}\mathbb{E}(\psi_{\varepsilon,z}(\zeta_\varepsilon)) &= \mathbb{E}\left(\frac{1}{2\pi}\int \hat\psi_{\varepsilon,z}(\xi)e^{i\xi\zeta_\varepsilon}\right) = \frac{1}{2\pi}\int \hat\psi_{\varepsilon,z}(\xi)\mathbb{E}(e^{i\xi\zeta_\varepsilon})\\ &= \frac{1}{2\pi}\int \hat\psi_{\varepsilon,z}(\xi)\int_{\mathbb{T}} \mathcal{L}_{\xi\sqrt{\varepsilon}}^{\lfloor\varepsilon^{-1}t\rfloor}\rho + \mathcal{O}(\varepsilon^{\frac12-\alpha})\\ &= \frac{1}{2\pi}\int_{\sqrt{\varepsilon}|\xi|\leq\nu_0} \varepsilon^\alpha e^{i\xi z}\hat\psi(\xi\varepsilon^\alpha)\int_{\mathbb{T}} \mathcal{L}_{\xi\sqrt{\varepsilon}}^{\lfloor\varepsilon^{-1}t\rfloor}\rho + \mathcal{O}(\varepsilon^{\frac12-\alpha})\\ &\quad + \mathcal{O}\left(\int_{|\eta|\geq\nu_0\varepsilon^{\alpha-1/2}} |\hat\psi(\xi)|\right).\end{aligned}$$

Thus, if we set $\beta = \min\{2\alpha, \frac12 - \alpha\} > 0$ and $\sigma^2 = -\alpha_0''$, we have

$$\begin{aligned}\mathbb{E}(\psi_{\varepsilon,z}(\zeta_\varepsilon)) &= \frac{1}{2\pi}\int_{\sqrt{\varepsilon}|\xi|\leq\nu_0} \varepsilon^\alpha e^{i\xi z}\hat\psi(\xi\varepsilon^\alpha)e^{-\frac{t}{2}\sigma^2\xi^2 + \mathcal{O}(\sqrt{\varepsilon}\xi^3)}d\xi + \mathcal{O}(\varepsilon^\beta)\\ &= \frac{\hat\psi(0)}{2\pi}\int_{\mathbb{R}} \varepsilon^\alpha e^{i\xi z}e^{-\frac{t}{2}\sigma^2\xi^2}d\xi + \mathcal{O}(\varepsilon^\beta)\\ &= \text{Leb}(\psi_{\varepsilon,z})\frac{e^{-\frac{z^2}{2\sigma^2 t}}}{\sigma\sqrt{2\pi t}} + \mathcal{O}(\varepsilon^\beta) = \int_{\mathbb{R}} \psi_{\varepsilon,z}(x)\frac{e^{-\frac{x^2}{2\sigma^2 t}}}{\sigma\sqrt{2\pi t}} + \mathcal{O}(\varepsilon^\beta).\end{aligned}$$

Of course, since $\text{Leb}(\psi_{\varepsilon,z}) = \mathcal{O}(\varepsilon^\alpha)$ the above formula is useful only if $\beta > \alpha$, thus we can explore the distribution only till intervals of size $\varepsilon^{\frac14}$. To have informations

[9] Here we are using that the composition with f is a continuous operator in L^2, indeed

$$\|\varphi\circ f\|_{L^2} = \text{Leb}(|\varphi|^2\mathcal{L}1) \leq C_\#\|\varphi\|_{L^2}.$$

[10] Remember that $\hat\psi(\xi) = \int e^{-i\xi x}\psi(x)dx$ and $\psi(x) = \frac{1}{2\pi}\int e^{1\xi x}\hat\psi(\xi)d\xi$.

on smaller scales one must investigate the operators $\mathcal{L}_\nu$ for values of ν beyond the perturbative regime. This is indeed possible, but outside the scopes of the present note.

4. But, really, where does probability comes from?

In the last lecture we have seen how random behaviour can arise from a deterministic one. Only some of you might feel that I have been cheating: after all a system starts form a certain initial condition and does not care if we know it or not! So, the probability has been introduced as a representation of our ignorance and why should the system worry about we do or do not know?

To further the discussion along this lines would lead us to argue about the relation between the frequentist interpretation of probability and the Bayesian view of probability. Such a discussion could easily go on indefinitely without getting anywhere. Therefore I'd like to take a different point of view and ask: is it possible to obtain an almost everywhere results? That is: random behaviour can occur for almost all initial conditions? Of course, this does not solve completely the problem: *almost all* implies a probability, and it remains open the issue of which reference probability we should consider. But at least it would eliminate the average with respect to the initial conditions, which is rather unsatisfactory.[11]

Only, if you fix the initial condition then $\zeta_\varepsilon(t)$ will be some path, there is no randomness, so, what can we say? If you think a bit you will see that the same problem occurs for the Brownian motion itself: if you look only at one realisation, how do you deicide that the motion is random? There is no probability over there! This is a problem that experimentalists know very well, they often have at disposal only one system, hence how to compute averages?[12] The usual answer is to look at the system at different time intervals and consider such measures and their relations. For example, for Brownian motion $B(t)$, the increments should be distributed according to a Gaussian and should be independent. One can then choose a time interval h and different times $\{t_i\}_{i=1}^N$, $t_{i+1} - t_i \geq h$ and study the quantities

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \varphi(B(t_i+h) - B(t_i)) \\
&\frac{1}{N}\sum_{i=1}^N \varphi(B(t_{i+1}+h) - B(t_{i+1}))g(B(t_i+h) - B(t_i)).
\end{aligned}
\tag{4.1}
$$

Then, by Birkhoff ergodic theorem, the first quantity, for $N \to \infty$, should converge to the average of φ with respect to a Gaussian, and the second should converge to the product of the averages of φ and g, for almost all the trajectories.

If an experimentalist would measure the quantities (4.1) and find out the above behaviour, then she would be rather satisfied that she is observing a genuine Brownian motion. It is then natural to ask: in the model we are discussing what happens

[11] If you want to make a theory that explains how to boil eggs, you should be weary of one that tells you that doing such and such the average egg will be properly boiled: it could be that half of the eggs are burned and the other half frozen!

[12] This is particularly true, e.g., in cosmology as we have only one universe at our disposal from which to draw data.

to the analogous of (4.1), that is to

$$\begin{aligned}(4.2)\qquad &\frac{1}{N}\sum_{i=1}^{N}\varphi(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i))\\ &\frac{1}{N}\sum_{i=1}^{N}\varphi(\zeta_\varepsilon(t_{i+1}+h)-\zeta_\varepsilon(t_{i+1}))g(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i)).\end{aligned}$$

Let us analyse the first quantity, the second being similar. We have seen in the previous lecture that the convergence of the random variables is implied by the convergence of the characteristic function. Hence we would like to show that, Lebesgue almost surely,

$$\begin{aligned}(4.3)\qquad \lim_{N\to\infty}\lim_{\varepsilon\to 0}\frac{1}{N}\sum_{i=1}^{N}\exp\left[i\xi(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i))\right] &= \mathbb{E}(e^{i\xi(\zeta_\varepsilon(h)-\zeta_\varepsilon(0))})\\ &= \exp\left[-\frac{\xi^2\sigma^2 h}{2}\right].\end{aligned}$$

To this end, let us start computing

$$\begin{aligned}&\lim_{\varepsilon\to 0}\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\exp\left[i\xi(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i))\right]-\exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right|^2\right]\\ &=\lim_{\varepsilon\to 0}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i=j}^{N}\mathbb{E}\left[\exp\left[i\xi\left\{(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_j))-(\zeta_\varepsilon(t_j+h)-\zeta_\varepsilon(t_i))\right\}\right]\right]\\ &\quad-\exp\left[-\xi^2\sigma^2 h\right].\end{aligned}$$

Recalling (3.3), we have, setting $\Delta_{i,j}=(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_j))-(\zeta_\varepsilon(t_j+h)-\zeta_\varepsilon(t_i))$,

$$\left|\Delta_{i,j}-\sqrt{\varepsilon}\sum_{k=\lfloor\varepsilon^{-1}t_i\rfloor}^{\lfloor\varepsilon^{-1}(t_i+h)\rfloor-1}\hat\omega\circ f^k(x_0)+\sqrt{\varepsilon}\sum_{k=\lfloor\varepsilon^{-1}t_j\rfloor}^{\lfloor\varepsilon^{-1}(t_j+h)\rfloor-1}\hat\omega\circ f^k(x_0)\right|\le C_\#\sqrt{\varepsilon}.$$

We can then introduce again the transfer operators

$$\mathcal{L}h(x)=\sum_{f(y)=x}\frac{1}{f'(y)}h(y);\qquad \mathcal{L}_\nu h(x)=\sum_{f(y)=x}\frac{e^{i\nu\hat\omega(y)}}{f'(y)}h(y),$$

and write, if $i>j$,

$$\begin{aligned}\mathbb{E}\left[e^{i\xi\Delta_{i,j}}\right]&=\mathbb{E}\left[\mathcal{L}_{\xi\sqrt{\varepsilon}}^{h/\varepsilon}\mathcal{L}^{[t_i-t_j-h]/\varepsilon}\mathcal{L}_{-\xi\sqrt{\varepsilon}}^{h/\varepsilon}\mathcal{L}^{t_i/\varepsilon}\rho\right]+\mathcal{O}(\sqrt{\varepsilon})\\ &=\exp\left[-\xi^2\sigma^2 h\right]+\mathcal{O}(\sqrt{\varepsilon}+e^{-c_\# h/\varepsilon})\end{aligned}$$

while, if $i=j$, then $\mathbb{E}\left[i\xi\Delta_{i,j}\right]=1+\mathcal{O}(\sqrt{\varepsilon})$. Thus, for $h\ll\varepsilon$,

$$\begin{aligned}&\lim_{\varepsilon\to 0}\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\exp\left[i\xi(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i))\right]-\exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right|^2\right]\\ &=\lim_{\varepsilon\to 0}\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^{N}\left\{1-\exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right\}\right]=\mathcal{O}\left(\frac{h}{N}\right).\end{aligned}$$

Then, by Chebyshev's inequality, setting $S_N = \frac{1}{N}\sum_{i=1}^N \exp\left[i\xi(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_i))\right]$, we have

$$\mathbb{P}\left[\left\{\left|S_N - \exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right| \geq \delta\right\}\right] \leq \frac{C_\# h}{\delta^2 N}.$$

On the other hand

$$|S_{N+m} - S_N| \leq C_\# \frac{m}{N}.$$

Hence, for $k \in \mathbb{N}$ and $j, m \leq 2^{k/2}$,

$$|S_{2^k + j2^{k/2}+m} - S_{2^k+jk}| \leq C_\# 2^{-k/2}.$$

Accordingly, for $N \geq C_\# \ln \delta^{-1}$,

$$\begin{aligned}
&\mathbb{P}\left[\left\{\sup_{n\geq N}\left|S_n - \exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right| \geq 2\delta\right\}\right] \\
&\leq \sum_{k\geq \ln_2 N}\sum_{j=0}^{2^{k/2}-1}\mathbb{P}\left[\left\{\left|S_{2^k+j2^{k/2}} - \exp\left[-\frac{\xi^2\sigma^2 h}{2}\right]\right| \geq \delta\right\}\right] \\
&\leq \sum_{k\geq \ln_2 N} 2^{-k/2}\delta^{-2} \leq \frac{C_\# h}{\sqrt{N}\delta^2},
\end{aligned}$$

which proves equation (4.3).

Remark 4.1. *Note that if we choose $t_{i+1} - t_i = 2h$, then the time average takes place in a time $T = 2hN$. Also it is interesting to choose $\delta = h^{1+\alpha/2}$, otherwise the error is larger than the difference form the exponential and one. Then the error in the equation above reads*

$$\frac{C_\#}{\sqrt{T}h^{\frac{1}{2}+\alpha}}.$$

Accordingly, to obtain a small error, it is necessary to take an average for a macroscopic time much larger than $h^{-\frac{1}{2}}$. Note however that a better estimate could be obtained by estimating higher moments.

Appendix A. Invariant measures and Von Neuman Theorem

First of all we need a useful characterization of invariance.

Lemma A.1. *Given a compact metric space X and map T continuous apart from a compact set K,*[13] *a Borel measure μ, such that $\mu(K) = 0$, is invariant if and only if $\mu(f \circ T) = \mu(f)$ for each $f \in \mathcal{C}^0(X)$.*

Proof. To prove that the invariance of the measure implies the invariance for continuous functions is obvious since each such function can be approximate uniformly by simple functions–that is, sum of characteristic functions of measurable sets–for which the invariance it is immediate.[14] The converse implication is not so obvious.

The first thing to remember is that the Borel measures, on a compact metric space, are regular [RS80]. This means that for each measurable set A the following holds[15]

[13] This means that, if $C \subset X$ is closed, then $T^{-1}C \cup K$ is closed as well.

[14] This is essentially the definition of integral.

[15] This is rather clear if one thinks of the Carathéodory construction starting from the open sets.

$$\mu(A) = \inf_{\substack{G \supset A \\ G = \mathring{G}}} \mu(G) = \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(C). \tag{A.1}$$

Next, remember that for each closed set A and open set $G \supset A$, there exists $f \in \mathcal{C}^0(X)$ such that $f(X) \subset [0,1]$, $f|_{G^c} = 0$ and $f|_A = 1$ (this is Urysohn Lemma for Normal spaces [Roy88]). Hence, setting $B_A := \{f \in \mathcal{C}^{(0)}(X) \mid f \geq \chi_A\}$,

$$\mu(A) \leq \inf_{f \in B_A} \mu(f) \leq \inf_{\substack{G \supset A \\ G = \mathring{G}}} \mu(G) = \mu(A). \tag{A.2}$$

Accordingly, for each A closed, we have

$$\mu(T^{-1}A) \leq \inf_{f \in B_A} \mu(f \circ T) = \inf_{f \in B_A} \mu(f) = \mu(A).$$

In addition, using again the regularity of the measure, for each A Borel holds[16]

$$\begin{aligned}\mu(T^{-1}A) &= \inf_{\substack{U \supset K \\ U = \mathring{U}}} \mu(T^{-1}A \backslash U) \leq \inf_{\substack{U \supset K \\ U = \mathring{U}}} \sup_{\substack{C \subset T^{-1}A \backslash U \\ C = \overline{C}}} \mu(T^{-1}(TC)) \\ &\leq \inf_{\substack{U \supset K \\ U = \mathring{U}}} \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(T^{-1}C) \leq \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(C) = \mu(A).\end{aligned}$$

Applying the same argument to the complement A^c of A it follow that it must be $\mu(T^{-1}A) = \mu(A)$ for each Borel set. □

Proposition A.2 (Krylov–Bogoluvov). *If X is a metric compact space and $T : X \to X$ is continuous, then there exists at least one invariant (Borel) measure.*

Proof. Consider any Borel probability measure ν and define the following sequence of measures $\{\nu_n\}_{n \in \mathbb{N}}$:[17] for each Borel set A

$$\nu_n(A) = \nu(T^{-n}A).$$

The reader can easily see that $\nu_n \in \mathcal{M}^1(X)$, the sets of the probability measures. Indeed, since $T^{-1}X = X$, $\nu_n(X) = 1$ for each $n \in \mathbb{N}$. Next, define

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \nu_i.$$

Again $\mu_n(X) = 1$, so the sequence $\{\mu_i\}_{i=1}^{\infty}$ is contained in a weakly compact set (the unit ball) and therefore admits a weakly convergent subsequence $\{\mu_{n_i}\}_{i=1}^{\infty}$; let μ be the weak limit.[18] We claim that μ is T invariant. Since μ is a Borel measure it

[16] Note that, by hypothesis, if C is compact and $C \cap K = \emptyset$, then TC is compact.

[17] Intuitively, if we chose a point $x \in X$ at random, according to the measure ν and we ask what is the probability that $T^n x \in A$, this is exactly $\nu(T^{-n}A)$. Hence, our procedure to produce the point $T^n x$ is equivalent to picking a point at random according to the evolved measure ν_n.

[18] This depends on the Riesz-Markov Representation Theorem [RS80] that states that $\mathcal{M}(X)$ is exactly the dual of the Banach space $\mathcal{C}^0(X)$. Since the weak convergence of measures in this case correspond exactly to the weak-* topology [RS80], the result follows from the Banach-Alaoglu theorem stating that the unit ball of the dual of a Banach space is compact in the weak-* topology.

suffices to verify that for each $f \in \mathcal{C}^0(X)$ holds $\mu(f \circ T) = \mu(f)$ (see Lemma A.1). Let f be a continuous function, then by the weak convergence we have[19]

$$\begin{aligned}\mu(f \circ T) &= \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} \nu_i(f \circ T) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} \nu(f \circ T^{i+1}) \\ &= \lim_{j\to\infty} \frac{1}{n_j} \left\{ \sum_{i=0}^{n_j-1} \nu_i(f) + \nu(f \circ T^{n_j}) - \nu(f) \right\} = \mu(f).\end{aligned}$$

□

Theorem A.3 (Von Neumann)**.** *Let (X, T, μ) be a Dynamical System, then for each $f \in L^2(X, \mu)$ the ergodic average converges in $L^2(X, \mu)$.*

Proof. We have already seen that it can be useful to lift the dynamics at the level of the algebra of function or at the level of measures. This game assumes different guises according to how one plays it, here is another very interesting version.

Let us define $U : L^2(X, \mu) \to L^2(X, \mu)$ as

$$Uf := f \circ T.$$

Then, by the invariance of the measure, it follows $\|Uf\|_2 = \|f\|_2$, so U is an L^2 contraction (actually, and L^2-isometry). If T is invertible, the same argument applied to the inverse shows that U is indeed unitary, otherwise we must content ourselves with

$$\|U^* f\|_2^2 = \langle UU^* f, f\rangle \le \|UU^* f\|_2 \|f\|_2 = \|U^* f\|_2 \|f\|_2,$$

that is $\|U^*\|_2 \le 1$ (also U^* is and L^2 contraction).

Next, consider $V_1 = \{f \in L^2 \mid Uf = f\}$ and $V_2 = \text{Rank}(\mathbb{1} - U)$. First of all, note that if $f \in V_1$, then

$$\|U^* f - f\|_2^2 = \|U^* f\|_2^2 - \langle f, U^* f\rangle - \langle U^* f, f\rangle + \|f\|_2^2 \le 0.$$

Thus, $f \in V_1^* := \{f \in L^2 \mid U^* f = f\}$. The same argument applied to $f \in V_1^*$ shows that $V_1 = V_1^*$. To continue, consider $f \in V_1$ and $h \in L^2$, then

$$\langle f, h - Uh\rangle = \langle f - U^* f, h\rangle = 0.$$

This implies that $V_1^\perp = \overline{V_2}$, hence $V_1 \oplus \overline{V_2} = L^2$. Finally, if $g \in V_2$, then there exists $h \in L^2$ such that $g = h - Uh$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} U^i g = \lim_{n\to\infty} \frac{1}{n}(h - U^n h) = 0.$$

On the other hand if $f \in V_1$ then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} U^i f = f$. The only function on which we do not still have control are the g belonging to the closure of V_2 but not in V_2. In such a case there exists $\{g_k\} \subset V_2$ with $\lim_{k\to\infty} g_k = g$. Thus,

$$\|\frac{1}{n} \sum_{i=0}^{n-1} U^i g\|_2 \le \|\frac{1}{n} \sum_{i=0}^{n-1} U^i g_k\|_2 + \|g - g_k\|_2 \le \|\frac{1}{n} \sum_{i=0}^{n-1} U^i g_k\|_2 + \frac{\varepsilon}{2},$$

[19] Note that it is essential that we can check invariance only on continuous functions: if we would have to check it with respect to all bounded measurable functions we would need that μ_n converges in a stronger sense (*strong convergence*) and this may not be true. Note as well that this is the only point where the continuity of T is used: to insure that $f \circ T$ is continuous and hence that $\mu_{n_j}(f \circ T) \to \mu(f \circ T)$.

provided we choose k large enough. Then, by choosing n sufficiently large we obtain

$$\|\frac{1}{n}\sum_{i=0}^{n-1} U^i g\|_2 \leq \varepsilon.$$

We have just proven that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} U^i = P$$

where P is the orthogonal projection on V_1. □

Next, we state, without proof, a slightly more refined, and more useful, result.

Theorem A.4 (Birkhoff). *Let (X, T, μ) be a dynamical system, then for each $f \in L^1(X, \mu)$*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$$

exists for almost every point $x \in X$. In addition, setting

$$f_+(x) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x),$$

it holds true

$$\int_X f_+ d\mu = \int_X f d\mu.$$

Finally, next Lemma is part of the usual magic of the ergodic theory.

Lemma A.5. *Let (X, T, μ) be an invertible dynamical system and, for some $f \in L^1(\mu)$, let f_+ be the ergodic average with resect to T and f_- the one with respect to f_-. Then f_+ and f_- coincide μ-almost everywhere.*

Proof. Let

$$\mathcal{A}_+ = \{x \in \mathbb{T}^2 \mid f_+(x) > f_-(x)\};$$

by definition $\mathcal{A}_+$ is an invariant set, hence

$$\int_{\mathcal{A}_+} [f_+(x) - f_-(x)]\, d\mu(x) = \int_{\mathcal{A}_+} f(x) d\mu(x) - \int_{\mathcal{A}_+} f(x) d\mu(x) = 0$$

which implies $\mu(\mathcal{A}_+) = 0$ and $f_+ \leq f_-$ μ-almost everywhere. The same argument, this time applied to the set $\mathcal{A}_- = \{x \in \mathbb{T}^2 \mid f_-(x) > f_+(x)\}$, implies the converse inequality. □

Appendix B. Ergodicity

Definition 1. *A measurable set A is invariant for T if $T^{-1}A \subset A$.*

A dynamical system (X, T, μ) is ergodic if each invariant set has measure zero or one.

The definition for continuous dynamical systems being exactly the same.

Note that if A is invariant then $\mu(A \backslash T^{-1}A) = \mu(A) - \mu(T^{-1}A) = 0$, moreover $\Lambda = \cap_{n=0}^{\infty} T^{-n}A \subset A$ is invariant as well. In addition, by definition, $\Lambda = T\Lambda$, which implies $\Lambda = T^{-1}\Lambda$ and $\mu(A\backslash\Lambda) = 0$. This means that, if A is invariant, then it always contains a set Λ invariant in the stronger (maybe more natural) sense that $T\Lambda = T^{-1}\Lambda = \Lambda$. Moreover, Λ is of full measure in A. Our definition of invariance is motivated by its greater flexibility and the fact that, from a measure theoretical point of view, zero measure sets can be discarded.

In essence, if a system is ergodic then most trajectories explore all the available space. In fact, for any A of positive measure, define $A_b = \cup_{n\in\mathbb{N}\cup\{0\}} T^{-n}A$ (this are the points that eventually end up in A), since $A_b \supset A$, $\mu(A_b) > 0$. Since $T^{-1}A_b \subset A_b$, by ergodicity follows $\mu(A_b) = 1$. Thus, the points that never enter in A (that is, the points in A_b^c) have zero measure. Actually, if the system has more structure (topology) more is true (see Problem 1).

The reader should be aware that there are many equivalent definitions of ergodicity, see Problems 5, 7, 8.

A stronger, and often more useful property of a dynamical systems is mixing.

Definition 2. *A Dynamical System (X, T, μ) is called mixing if for every pairs of measurable sets A, B we have*

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).$$

B.1. **Examples.**

B.1.1. ***Rotations.*** The ergodicity of a rotation depends on ω. If $\omega \in \mathbb{Q}$ then the system is not ergodic. In fact, let $\omega = \frac{p}{q}$ $(p, q \in \mathbb{N})$, then, for each $x \in \mathbb{T}$ $T^q x = x + p \mod 1 = x$, so T^q is just the identity. An alternative way of saying this is to notice that all the points have a periodic trajectory of period q. It is then easy to exhibit an invariant set with measure strictly larger than 0 but strictly less than 1. Consider $[0, \varepsilon]$, then $A = \cup_{i=1}^{q-1} T^{-i}[0, \varepsilon]$ is an invariant set; clearly $\varepsilon \le \mu(A) \le q\varepsilon$, so it suffices to choose $\varepsilon < q^{-1}$.

The case $\omega \notin \mathbb{Q}$ is much more interesting. First of all, for each point $x \in \mathbb{T}$ we have that the closure of the set $\{T^n x\}_{i=0}^{\infty}$ is equal to $\mathbb{T}$, which is to say that the orbits are dense.[20] The proof is based on the fact that there cannot be any periodic orbit. To see this suppose that $x \in \mathbb{T}$ has a periodic orbit, that is there exists $q \in \mathbb{N}$ such that $T^q x = x$. As a consequence there must exist $p \in \mathbb{Z}$ such that $x + p = x + q\omega$ or $\omega \in \mathbb{Q}$ contrary to the hypothesis. Hence, the set $\{T^k 0\}_{k=0}^{\infty}$ must contain infinitely many points and, by compactness, must contain a convergent subsequence k_i. Hence, for each $\varepsilon > 0$, there exists $m > n \in \mathbb{N}$:

$$|T^m 0 - T^n 0| < \varepsilon.$$

Since T preserves the distances, calling $q = m - n$, holds

$$|T^q 0| < \varepsilon.$$

Accordingly, the trajectory of $T^{jq}0$ is a translation by a quantity less than ε, therefore it will get closer than ε to each point in $\mathbb{T}$ (i.e., the orbit is dense). Again by the conservation of the distance, since zero has a dense orbit the same will hold for every other point.

Intuitively, the fact that the orbits are dense implies that there cannot be a non trivial invariant set, henceforth the system is ergodic. Yet, the proof it is not trivial since it is based on the existence of Lebesgue density points [Roy88] (see Problem 11). It is a fact from general measure theory that each measurable set $A \subset \mathbb{R}$ of positive

[20] A system with a dense orbit called *Topologically Transitive*.

Lebesgue measure contains, at least, one point $\bar{x}$ such that for each $\varepsilon \in (0, 1)$ there exists $\delta > 0$:

$$\frac{m(A \cap [\bar{x} - \delta, \bar{x} + \delta])}{2\delta} > 1 - \varepsilon.$$

Hence, given an invariant set A of positive measure and $\varepsilon > 0$, first choose δ such that the interval $I := [\bar{x} - \delta, \bar{x} + \delta]$ has the property $m(I \cap A) > (1 - \varepsilon)m(I)$. Second, we know already that there exists $q, M \in \mathbb{N}$ such that $\{T^{-kq}x\}_{k=1}^{M}$ divides $[0, 1]$ into intervals of length less that $\frac{\varepsilon}{2}\delta$. Hence, given any point $x \in \mathbb{T}$ choose $k \in \mathbb{N}$ such that $m(T^{-kq}I \cap [x - \delta, x + \delta]) > m(I)(1 - \varepsilon)$ so,

$$\begin{aligned} m(A \cap [x - \delta, x + \delta]) &\geq m(A \cap T^{-kq}I) - m(I)\varepsilon \\ &\geq m(A \cap I) - m(I)\varepsilon \geq (1 - 2\varepsilon)2\delta. \end{aligned}$$

Thus, A has density everywhere larger than $1 - 2\varepsilon$, which implies $\mu(A) = 1$ since ε is arbitrary.

The above proof of ergodicity it is not so trivial but it has a definite dynamical flavor (in the sense that it is obtained by studying the evolution of the system). Its structure allows generalizations to contexts whit a less rich algebraic structure. Nevertheless, we must notice that, by taking advantage of the algebraic structure (or rather the group structure) of $\mathbb{T}$, a much simpler and powerful proof is available.

Let $\nu \in \mathcal{M}_T^1$, then define

$$F_n = \int_{\mathbb{T}} e^{2\pi i n x} \nu(dx), \quad n \in \mathbb{N}.$$

A simple computation, using the invariance of ν, yields

$$F_n = e^{2\pi i n \omega} F_n$$

and, if ω is irrational, this implies $F_n = 0$ for all $n \neq 0$, while $F_0 = 1$. Next, consider $f \in \mathcal{C}^{(2)}(\mathbb{T}^1)$ (so that we are sure that the Fourier series converges uniformly), then

$$\nu(f) = \sum_{n=0}^{\infty} \nu(f_n e^{2\pi i n \cdot}) = \sum_{n=0}^{\infty} f_n F_n = f_0 = \int_{\mathbb{T}} f(x) dx.$$

Hence m is the unique invariant measure (unique ergodicity). This is clearly much stronger than ergodicity (see Problem 5)

B.2. **Some easy, and not so easy, Problems.**

(1) A topological Dynamical System (X, T) is called *Topologically transitive*, if it has a dense orbit. Show that if $(\mathbb{T}^d, T, m)$ is ergodic and T is continuous, then the system is topologically transitive.
(2) Give an example of a system with a dense orbit which it is not ergodic.
(3) Give an example of an ergodic system with no dense orbit.
(4) Give an example of a Dynamical Systems which does not have any invariant probability measure.
(5) Show that a Dynamical Systems (X, T, μ) is ergodic if and only if there does not exists any invariant probability measure absolutely continuous with respect to μ, beside μ itself.
(6) Prove that Birkhoff theorem implies Von Neumann theorem.
(7) Prove that if (X, T, μ) is ergodic, then all $f \in L^1(X, \mu)$ such that $f \circ T = f$ are a.e. constant. Prove also the converse.

(8) For each measurable set A, let
$$F_{A,n}(x) = \frac{1}{n}\sum_{i=0}^{n-1}\chi_A(T^i x).$$
be the average number of times x visits A in the time n. Show that there exists $F_A = \lim_{n\to\infty} F_{A,n}$ a.e. and prove that, if the system is ergodic, $F_A = \mu(A)$.

(9) Prove that the Lebesgue measure is invariant for the rotations on $\mathbb{T}$.

(10) Consider a rotation by $\omega \in \mathbb{Q}$, find invariant measures different from Lebesgue.

(11) Prove Lebesgue density theorem: for each measurable set A, $m(A) > 0$, there exists $x \in A$ such that for each $\varepsilon > 0$ exists $\delta > 0$ such that $m(A \cap [x-\delta, x+\delta]) > (1-\varepsilon)2\delta$.

Appendix C. Some Functional analysis

Consider two Banach space $(\mathbb{B}, \|\cdot\|)$ and $(\mathbb{B}_0, |\cdot|)$ such that $\mathbb{B} \subset \mathbb{B}_0$ and

i. $|h| \le \|h\|$ for all $h \in \mathbb{B}$,

ii. if $h \in \mathbb{B}$ and $|h| = 0$, then $h = 0$.

iii. There exists $C > 0:$ for each $\varepsilon > 0$ there exists a finite rank operator $\mathbb{A}_\varepsilon \in L(\mathbb{B}, \mathbb{B})$ such that $\|\mathbb{A}_\varepsilon\| \le C$ and $|h - \mathbb{A}_\varepsilon h| \le \varepsilon\|h\|$ for all $h \in \mathbb{B}$.[21]

In addition consider a bounded operator $\mathcal{L} : \mathbb{B}_0 \to \mathbb{B}_0$, constants $A, B, C \in \mathbb{R}_+$, and $\lambda > 1$, such that

a. $|\mathcal{L}^n| \le C$ for all $n \in \mathbb{N}$,

b. $\mathcal{L}(B) \subset B$

c. $\|\mathcal{L}^n h\| \le A\lambda^{-n}\|h\| + B|h|$ for all $h \in \mathbb{B}$ and $n \in \mathbb{N}$.

In particular $\mathcal{L}$ can be seen as a bounded operator on $\mathbb{B}$.

Theorem C.1. *The spectral radius of the operator $\mathcal{L} \in L(\mathbb{B}, \mathbb{B})$ is bounded by 1 while the essential spectral radius is bounded by λ^{-1}.*[22]

We can now prove our main result.

Proof of Theorem C.1. The first assertion is a trivial consequence of (c), (a) and (i).

The second part is much deeper. Let $\mathcal{L}_{n,\varepsilon} := \mathcal{L}^n\mathbb{A}_\varepsilon$, clearly such an operator is finite rank, in addition
$$\|\mathcal{L}^n h - \mathcal{L}_{n,\varepsilon}h\| \le A\lambda^{-n}\|(\mathbb{1} - \mathbb{A}_\varepsilon)h\| + B|(\mathbb{1} - \mathbb{A}_\varepsilon)h| \le A(1+C)\lambda^{-n}\|h\| + B\varepsilon\|h\|.$$
By choosing $\varepsilon = \lambda^{-n}$ we have that there exists $C_1 > 0$ such that
$$\|\mathcal{L}^n - \mathcal{L}_{n,\varepsilon}\| \le C_1\lambda^{-n}.$$
For each $z \in \mathbb{C}$ we can now write
$$\mathbb{1} - z\mathcal{L} = (\mathbb{1} - z(\mathcal{L} - \mathcal{L}_{n,\varepsilon})) - z\mathcal{L}_{n,\varepsilon}.$$

[21] In fact, this last property can be weakened to: The unit ball $\{h \in \mathbb{B} : \|h\| \le 1\}$ is relatively compact in $\mathbb{B}_0$. We use the present stronger condition since, on the one hand, it is true in all the applications we will be interested in and, on the other hand, drastically simplifies the argument. Note also that, if one uses the Fredholm alternative for compact operators rather than finite rank ones (Theorem C.2), then one can ask the $\mathbb{A}_\varepsilon$ to be compact instead than finite rank making easier their construction in concrete cases.

[22] The definition of *essential spectrum* varies a bit from book to book. Here we call essential spectrum the complement, in the spectrum, of the isolated eigenvalues with associated finite dimensional eigenspaces (which is also called the Fredholm spectrum).

Since

$$\|z(\mathcal{L} - \mathcal{L}_{n,\varepsilon})\| \leq |z| C_1 \lambda^{-n} < \frac{1}{2},$$

provided that $|z| \leq \frac{1}{2C_1}\lambda^n$. Thus, given any z in the disk $D_n := \{|z| < \frac{1}{2C_1}\lambda^n\}$ the operator $B(z) := \mathbb{1} - z(\mathcal{L} - \mathcal{L}_{n,\varepsilon})$ is invertible.[23] Hence

$$\mathbb{1} - z\mathcal{L} = \left(\mathbb{1} - z\mathcal{L}_{n,\varepsilon} B(z)^{-1}\right) B(z) =: (\mathbb{1} - F(z))B(z).$$

By applying Fredholm analytic alternative (see Theorem C.2 for the statement and proof in a special case sufficient for the present purposes) to $F(z)$ we have that the operator is either never invertible or not invertible only in finitely many points in the disk D_n. Since for $|z| < 1$ we have $(\mathbb{1} - z\mathcal{L})^{-1} = \sum_{n=0}^{\infty} z^n \mathcal{L}^n$, the first alternative cannot hold hence the Theorem follows. □

Here we give a proof of the Analytic Fredholm alternative in a special case.

Theorem C.2 (Analytic Fredholm theorem–finite rank).[24] *Let D be an open connected subset of $\mathbb{C}$. Let $F : \mathbb{C} \to L(\mathbb{B}, \mathbb{B})$ be an analytic operator-valued function such that $F(z)$ is finite rank for each $z \in D$. Then, one of the following two alternatives holds true*

- *$(\mathbb{1} - F(z))^{-1}$ exists for no $z \in D$*
- *$(\mathbb{1} - F(z))^{-1}$ exists for all $z \in D \backslash S$ where S is a discrete subset of D (i.e. S has no limit points in D). In addition, if $z \in S$, then 1 is an eigenvalue for $F(z)$ and the associated eigenspace has finite multiplicity.*

Proof. First of all notice that, for each $z_0 \in D$ there exists $r > 0$ such that $D_{r(z_0)}(z_0) := \{z \in \mathbb{C} \; : \; |z - z_0| < r(z_0)\} \subset D$, and

$$\sup_{z \in D_{r(z_0)}(z_0)} \|F(z) - F(z_0)\| \leq \frac{1}{2}.$$

Clearly if we can prove the theorem in each such disk we are done.[25] Note that

$$\mathbb{1} - F(z) = \left(\mathbb{1} - F(z_0)(\mathbb{1} - [F(z) - F(z_0)])^{-1}\right)(\mathbb{1} - [F(z) - F(z_0)]).$$

Thus the invertibility of $\mathbb{1} - F(z)$ in $D_r(z_0)$ depends on the invertibility of $\mathbb{1} - F(z_0)(\mathbb{1} - [F(z) - F(z_0)])^{-1}$. Let us set $F_0(z) := F(z_0)(\mathbb{1} - [F(z) - F(z_0)])^{-1}$.

Let us start by looking at the equation

$$(\mathbb{1} - F_0(z))h = 0. \tag{C.1}$$

Clearly if a solution exists, then $h \in \text{Range}(F_0(z)) = \text{Range}(F(z_0)) := \mathbb{V}_0$. Since $\mathbb{V}_0$ is finite dimensional there exists a basis $\{h_i\}_{i=1}^N$ such that $h = \sum_i \alpha_i h_i$. On the

[23] Clearly $B(z)^{-1} = \sum_{n=0}^{\infty} [z(\mathcal{L} - \mathcal{L}_{n,\varepsilon})]^n$.

[24] The present proof is patterned after the proof of the Analytic Fredholm alternative for compact operators (in Hilbert spaces) given in [RS80, Theorem VI.14]. There it is used the fact that compact operators in Hilbert spaces can always be approximated by finite rank ones. In fact the theorem holds also for compact operators in Banach spaces but the proof is a bit more involved.

[25] In fact, consider any connected compact set K contained in D. Let us suppose that for each $z_0 \in K$ we have a disk $D_{r(z_0)}(z_0)$ in the theorem holds. Since the disks $D_{r(z_0)/2}(z_0)$ form a covering for K we can extract a finite cover. If the first alternative holds in one such disk then, by connectness, it must hold on all K. Otherwise each $S \cap D_{r(z_0)/2}(z_0)$, and hence $K \cap S$, contains only finitely many points. The Theorem follows by the arbitrariness of K.

other hand there exists an analytic matrix $G(z)$ such that[26]

$$F_0(z)h = \sum_{ij} G(z)_{ij}\alpha_j h_i.$$

Thus (C.1) is equivalent to

$$(\mathbb{1} - G(z))\alpha = 0,$$

where $\alpha := (\alpha_i)$.

The above equation can be satisfied only if $\det(\mathbb{1} - G(z)) = 0$ but the determinant is analytic hence it is either always zero or zero only at isolated points.[27]

Suppose the determinant different from zero, and consider the equation

$$(\mathbb{1} - F_0(z))h = g.$$

Let us look for a solution of the type $h = \sum_i \alpha_i h_i + g$. Substituting yields

$$\alpha - G(z)\alpha = \beta$$

where $\beta := (\beta_i)$ with $F_0(z)g =: \sum_i \beta_i h_i$. Since the above equation admits a solution, we have $\text{Range}(\mathbb{1} - F_0(z)) = \mathbb{B}$, Thus we have an everywhere defined inverse, hence bounded by the open mapping theorem.

We are thus left with the analysis of the situation $z \in S$ in the second alternative. In such a case, there exists h such that $(\mathbb{1} - F(z))h = 0$, thus one is an eigenvalue. On the other hand, if we apply the above facts to the function $\Phi(\zeta) := \zeta^{-1}F(z)$ analytic in the domain $\{\zeta \neq 0\}$ we note that the first alternative cannot take place since for $|\zeta|$ large enough $\mathbb{1} - \Phi(\zeta)$ is obviously invertible. Hence, the spectrum of $F(z)$ is discrete and can accumulate only at zero. This means that there is a small neighborhood around one in which $F(z)$ has no other eigenvalues, we can thus surround one with a small circle γ and consider the projector

$$\begin{aligned} P :=& \frac{1}{2\pi i}\int_\gamma (\zeta - F(z))^{-1} d\zeta = \frac{1}{2\pi i}\int_\gamma \left[(\zeta - F(z))^{-1} - \zeta^{-1}\right] d\zeta \\ =& \frac{1}{2\pi i}\int_\gamma F(z)\zeta^{-1}(\zeta - F(z))^{-1} d\zeta. \end{aligned}$$

By standard functional calculus it follows that P is a projector and it clearly projects on the eigenspace of the eigenvector one. But the last formula shows that P must project on a subspace of the range of $F(z)$, hence it must be finite dimensional. □

References

[1] V.Baladi, *Positive transfer operators an decay of correlations*, World scientific, Singapore (2000).

---

[26] To see the analyticity notice that we can construct linear functionals $\{\ell_i\}$ on $\mathbb{V}_0$ such that $\ell_i(h_j) = \delta_{ij}$ and then extend them to all $\mathbb{B}$ by the Hahn-Banach theorem. Accordingly, $G(z)_{ij} := \ell_j(F_0(z)h_i)$, which is obviously analytic.

[27] The attentive reader has certainly noticed that this is the turning point of the theorem: the discreteness of S is reduced to the discreteness of the zeroes of an appropriate analytic function: a determinant. A moment thought will immediately explain the effort made by many mathematicians to extend the notion of determinant (that is to define an analytic function whose zeroes coincide with the spectrum of the operator) beyond the realm of matrices (the so called Fredholm determinants).

[2] Jacopo De Simoi, Carlangelo Liverani, *The Martingale approach after Varadhan and Dolpogpyat* . In "Hyperbolic Dynamics, Fluctuations and Large Deviations", Dolgopyat, Pesin, Pollicott, Stoyanov editors, Proceedings of Symposia in Pure Mathematics, **89**, AMS (2015).

[3] Dmitry Dolgopyat. *Averaging and invariant measures. Mosc. Math. J.*, 5(3):537–576, 742, 2005.

[4] Tosio Kato. *Perturbation theory for linear operators.* Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[5] Carlangelo Liverani, *Central Limit Theorem for Deterministic Systems*, International Conference on Dynamical Systems, Montevideo 1995, a tribute to Ricardo Ma e, Pitman Research Notes in Mathemaics Series, 362, editori F.Ledrappier, J.Levovicz, S.Newhouse, (1996).

[Roy88] H. L. Royden. *Real analysis.* Macmillan Publishing Company, New York, third edition, 1988.

[RS80] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, second edition, 1980. Functional analysis.

[6] Varadhan, S. R. S. *Probability theory.* Courant Lecture Notes in Mathematics, **7**. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001. viii+167 pp.

[7] Varadhan, S. R. S. *Stochastic processes.* Courant Lecture Notes in Mathematics, **16**. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2007. x+126 pp.

Carlangelo Liverani, Dipartimento di Matematica, II Università di Roma (Tor Vergata), Via della Ricerca Scientifica, 00133 Roma, Italy.

E-mail address: liverani@mat.uniroma2.it