DETERMINISTIC NOISE

CARLANGELO LIVERANI

ABSTRACT. I illustrate how random behaviour might emerge naturally in deterministic systems.
This are super condensed and hastily written notes for the mini-course given at the 14th WORKSHOP ON INTERACTIONS BETWEEN DYNAMICAL SYSTEMS AND PARTIAL DIFFERENTIAL EQUATIONS (JISD2016). Barcelona, July 11 - 15, 2016.
Read at your own risk.

CONTENTS

1. NOISE

The traditional reductionist approach to study nature consists in identify 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.

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 was 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

Date: July 12, 2016.

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 might be large.

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.

The 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 degrees of freedom that we ignore, 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

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

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).

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 (1.1), 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 stochastic behaviour.

2. Chaos

We have seen 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 showed such a behaviour was of zero Lebesgue measure. It is widely believed that such properties holds for a positive measure set of point, but we are far away 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 geodesic flows on manifolds of negative curvature and the automorphisms of a torus.

Yet, to explain the ideas in their simplest form it is better to start with 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 expanding

[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).

circle maps. 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 of the microscopic time. In other words we should be interested in the behaviours 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, if is 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 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. 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. 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_*. 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*

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

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, 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 $h \in \mathbb{V}$,

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

We can then choose a such that $\lambda^{-1}+D+a < 1$, and we have that $\mathcal{L}$ is a contraction on $\mathbb{V}$. Since $\mathcal{L}' \text{Leb} = \text{Leb}$, hence $1 \in \sigma(\mathcal{L})$, 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.[2]

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 here for a quicker, but more detailed, discussion of the present case[3]).

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

$$\dot{\bar{\theta}} = \bar{\omega}. \tag{2.4}$$

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

[3] The latter are personal notes not for diffusion as they are written in a very preliminary form, so read at your own risk.

This is a rather simple example of *averaging*. Something similar holds in general, see [2] for details. We have then seen how a very simple *macroscopic* behaviour arises form an complex *microscopic* behaviour. Remark that (2.4) 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. Let us start doing the former.

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. Say that we can determine that x belongs to some interval I, $|I| = \delta$, $\delta << 1$.[4]

So, we do a measure, 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', $|I_0'| \geq \lambda^{n_0}\delta$. 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 $1-\lambda^{-n_0}$ 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 not 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.

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

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, 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.[5]

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 $\theta_* \in \mathbb{T}$.

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

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

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

$$\left|\zeta_\varepsilon(t) - \sqrt{\varepsilon} \sum_{k=0}^{\lfloor \varepsilon^{-1} t \rfloor - 1} \hat{\omega} \circ f^k(x_0)\right| \leq C_\# \sqrt{\varepsilon}, \tag{3.2}$$

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], standard pairs [3] and spectral methods. The latter, when it works, is the more 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 that it suffices to compute the characteristic function [6], 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 $\phi \in L^1$,

$$\mathcal{L}_\nu \phi(x) = \sum_{f(y)=x} \frac{e^{i\nu\hat{\omega}(y)}}{f'(y)} \phi(y) \tag{3.3}$$

and notice that

$$\mathbb{E}\left(\exp\left[i\xi\sqrt{\varepsilon} \sum_{k=0}^{\lfloor \varepsilon^{-1} t \rfloor - 1} \hat{\omega} \circ f^k\right]\right) = \int_{\mathbb{T}} \mathcal{L}_{\xi\sqrt{\varepsilon}}^{\lfloor \varepsilon^{-1} t \rfloor} \rho. \tag{3.4}$$

[5] Note hovewer that we could have found out that the percentage of discarded systems is different, say $1 - \lambda^{-\alpha n_0}$, 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.

The problem is then reduced to studying the properties of the operator (3.3). 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 3.1. *There exists $\nu_0 > 0$ and continuous functions $C_\nu > 0$ and $\rho_\nu \in (0,1)$ such that, for all $|\nu| \leq \nu_0$, $\mathcal{L}_\nu = e^{\alpha_\nu}\Pi_\nu + Q_\nu$, $\Pi_\nu Q_\nu = Q_\nu \Pi_\nu = 0$, $\|Q_\nu^n\|_{W^{1,1}} \leq C_\nu \rho_\nu^n e^{\alpha_\nu n}$. 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

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

By standard perturbation theory, see [4], it follows that there exists $\nu_0 > 0$ such that, for all $|\nu| \leq \nu_0$ 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 $\bar{h}_\nu = \Pi_\nu h_*$, $\bar{\ell}_\nu = \mathrm{Leb}\,\Pi_\nu$ and $\beta_\nu = \bar{\ell}_\nu(\bar{h}_\nu)$ are analytic functions of ν. Moreover, $\bar{h}_0 = h_*$ and $\bar{\ell}_0 = \mathrm{Leb}$, thus we have $\beta_0 = 1$ and, provided ν_0 is small enough, $\beta_\nu \neq 0$, 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 $\ell_\nu \in (W^{1,1})'$, the eigenvalue of the dual operator and normalised so that $\ell_\nu(h_\nu) = 1$, such that $\ell_\nu(h_\nu) = 1$, $\ell_\nu(h'_\nu) = 0$ and $\Pi_\nu = h_\nu \otimes \ell_\nu$ are analytic in ν. Also

$$\mathcal{L}_\nu h_\nu = e^{\alpha_\nu} h_\nu, \tag{3.5}$$

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

Lemma 3.2. *Fot all $|\nu| \leq \nu_0$, the function α_ν satisfies $\alpha_0 = \alpha'_0 = 0$ and $|\alpha_\nu|_{\mathcal{C}^3} \leq C_\#$, $\alpha''_0 \leq 0$. 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 (3.5) obtaining

$$\mathcal{L}'_\nu h_\nu + \mathcal{L}_\nu h'_\nu = \alpha'_\nu e^{\alpha_\nu} h_\nu + e^{\alpha_\nu} h'_\nu. \tag{3.6}$$

Applying ℓ_ν yields

$$\frac{d\alpha_\nu}{d\nu} = i\ell_\nu(\hat{\omega}h_\nu) =: i\mu_\nu(\hat{\omega}). \tag{3.7}$$

Thus $\alpha'_0 = 0$. Differentiating again yields

$$\frac{d^2\alpha_\nu}{d\nu^2} = i\ell'_\nu(\hat{\omega}h_\nu) + i\ell_\nu(\hat{\omega}h'_\nu) = i\ell'_\nu(\omega_\nu h_\nu) + i\ell_\nu(\omega_\nu h'_\nu). \tag{3.8}$$

where $\omega_\nu = \hat{\omega} - \mu_\nu(\hat{\omega})$. On the other hand, from (3.6) and (3.7) 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

$$h'_\nu = i(\mathbb{1} - \hat{Q}_\nu)^{-1}\hat{\mathcal{L}}_\nu(\omega_\nu h_\nu). \tag{3.9}$$

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

$$\begin{aligned}
\alpha''_\nu &= -\ell_\nu(\omega_\nu(\mathbb{1} - \hat{Q}_\nu)^{-1}(\mathbb{1} + \hat{Q}_\nu)(\omega_\nu h_\nu)) \\
&= -\sum_{n=1}^{\infty} \ell_\nu(\omega_\nu \hat{\mathcal{L}}_\nu^n(\mathbb{1} + \hat{\mathcal{L}}_\nu)(\omega_\nu h_\nu)) \\
&= -\mu_\nu(\omega_\nu^2) - 2\sum_{n=1}^{\infty} \ell_\nu(\omega_\nu \hat{\mathcal{L}}_\nu^n(\omega_\nu h_\nu)).
\end{aligned} \tag{3.10}$$

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

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\mu_\nu\left(\left[\sum_{k=0}^{n-1} \omega_\nu \circ f^k\right]^2\right) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k,j=0}^{n-1} \mu_\nu(\omega_\nu \circ f^k \omega_\nu \circ f^j) \\
&= \mu_\nu(\omega_\nu^2) + \lim_{n\to\infty} \frac{2}{n}\sum_{k=1}^{n-1}(n-k)\mu_\nu(\omega_\nu \circ f^k \omega_\nu) \\
&= \mu_\nu(\omega_\nu^2) + 2\sum_{k=1}^{\infty} \mu_\nu(\omega_\nu \circ f^k \omega_\nu).
\end{aligned} \tag{3.11}$$

The above two facts and equation (3.10) yield[6]

$$-\alpha''_0 = \lim_{n\to\infty} \frac{1}{n}\mu_0\left(\left[\sum_{k=0}^{n-1} \omega_0 \circ f^k\right]^2\right) \geq 0. \tag{3.12}$$

Finally, note that the computations in (3.11) imply that, if $\alpha''_0 = 0$,

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

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 can then extract a converging subsequence, let $g \in L^2$ be its limit. Then, for each $\phi \in W^{1,1}$,[7]

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

[6] This, together with equation (3.10), is a simple instance of the so called *Green-Kubo formula*.

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

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

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 $\mathcal{L}$

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

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 line 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 interval of size $2\varepsilon^\alpha$ centred at z. Hence, using (3.2) and (3.4) we want, and can, compute[8]

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

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

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

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 till intervals of size $\varepsilon^{\frac14}$. To have informations 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 notes.

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

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

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.[9]

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? 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, 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, the 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 to the analogous of (4.1), that is to

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

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,

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

[9] 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!

To this end, let us start computing

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

Recalling (3.2), we have, setting $\Delta_{i,j}=(\zeta_\varepsilon(t_i+h)-\zeta_\varepsilon(t_j))-(\zeta_\varepsilon(t_j+h)-\zeta_\varepsilon(t_i))$,

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

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 $i>j$,

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

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

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

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

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

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

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

which proves equation (4.3).

References

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

[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).

[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`