3. Points of view: Dynamical Systems

In the previous arguments the dynamics enters in a very limited way: just to say that the measures considered are invariant. This may seem strange, at least one should take into consideration the fact that measurements are not instantaneous and hence what we see is the result of an action that last for some time, a time in which, given the typical high velocities of the atoms, a lot of things can happen (for example, an atom in the air can have 10^6 collisions in a millisecond). To explore such a point of view let us consider an extremely simple dynamics that has no pretense whatsoever of being physically meaningful: the translation on a circle.

3.1. **Circle rotation.** Let $T_\omega : \mathbb{T}^1 \to \mathbb{T}^1$ be defined by

$$T_\omega(x) = x + \omega \mod 1,$$

where $\omega \in \mathbb{R}$. We are thus interested in the quantity

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x))$$

where, for simplicity, $f \in \mathcal{C}^1(\mathbb{T}^1)$.[1]

Now if $\omega = \frac{p}{q} \in \mathbb{Q}$, it follows

$$T_\omega^q x = x + p \mod 1 = x.$$

Thus, after a time q, the dynamics is exactly identity. Now, let $N = Kq + r$, $0 \leq r < q$, we have

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) &= \frac{1}{N}\sum_{k=0}^{K}\sum_{j=0}^{q-1} f\circ T_\omega^{kq+j}(x) + \frac{1}{N}\sum_{j=0}^{r} f\circ T_\omega^{Kq+j}(x)\\
&= \frac{1}{N}\sum_{k=0}^{K}\sum_{j=0}^{q-1} f\circ T_\omega^{j}(x) + \frac{1}{N}\sum_{j=0}^{r} f\circ T_\omega^{Kq+j}(x)\\
&= \frac{N-r}{N}\frac{1}{q}\sum_{j=0}^{q-1} f\circ T_\omega^{j}(x) + \mathcal{O}(\frac{r}{N}|f|_\infty)
\end{aligned}$$

This means that, if we define the measure $\delta_x(f) := f(x)$, then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) = \frac{1}{q}\sum_{j=0}^{q-1} f\circ T_\omega^{j}(x) = \mu_x(f), \tag{3.1}$$

where the measure $\mu_x := \frac{1}{q}\sum_{j=0}^{q-1}\delta_{T_\omega^j x}$ is simply the average of the delta functions along the periodic trajectory starting at x.

We have thus a situation quite different from the one in the previous chapter: the result of the measurement does depend on the initial point since to each point is associated a different measure μ_x. Clearly such measure are all invariant. We have thus a situation in which there are uncountably many invariant measures and the result of the measurement is described by different invariant measures depending on the initial configuration of the system.

[1] For analogy with the observable of the previous section one may want to consider the case $f = \chi_\Delta$ for some interval Δ. I leave to the reader to check that the following holds also for $f \in BV$, the space of functions of bounded variation.

On the other hand, let us consider the case $\omega \notin \mathbb{Q}$. In this situation it is simple to verify that $\{T_\omega^n x\}$ are all different points.[2] Hence, they must have accumulation points. Let $T_\omega^{n_k} 0$ be a converging subsequence, then for each $\varepsilon > 0$ there exists $n > m$ such that $\text{dist}(T_\omega^n 0 - T_\omega^m 0) \leq \varepsilon$. That is there exists $p \in \mathbb{N}$ such that $\varepsilon \geq |T_\omega^n 0 - T_\omega^m 0 + p| = |(n-m)\omega + p|$, that is, setting $q = n - m$,[3]

$$\left|\omega - \frac{p}{q}\right| \leq \frac{\varepsilon}{q}.$$

In particular, this means $T_\omega^q(x) = T_\alpha(x)$, for some $|\alpha| \leq \varepsilon$. We can thus try to argue similarly to the rational case:

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) &= \frac{1}{N}\sum_{k=0}^{K}\sum_{j=0}^{q-1} f\circ T_\omega^{kq+j}(x) + \frac{1}{N}\sum_{j=0}^{r} f\circ T_\omega^{Kq+j}(x)\\ &= \frac{1}{N}\sum_{j=0}^{q-1}\sum_{k=0}^{K} f\circ T_\alpha^k(T_\omega^j(x)) + \mathcal{O}(\frac{r}{N}|f|_\infty)\end{aligned}$$

Now T_α is a rotation by a very small amount hence

$$f(x)\alpha = \int_x^{T_\alpha x} f(z)dz + \int_x^{T_\alpha x}[f(x) - f(z)]dz = \int_x^{T_\alpha x} f(z)dz + \mathcal{O}(\frac{1}{2}|f|_{\mathcal{C}^1}\alpha^2).$$

Introducing the above fact in the previous formula yields

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) = \frac{1}{N}\sum_{j=0}^{q-1}\sum_{k=0}^{K}\alpha^{-1}\int_{T_\alpha^k(T_\omega^j(x))}^{T_\alpha^{k+1}(T_\omega^j(x))} f(z)dz + \mathcal{O}(\frac{Kq}{2N}|f|_{\mathcal{C}^1}\alpha + \frac{r}{N}|f|_\infty).$$

Since $K\alpha = l + s$, $s < 1$, it follows that the integral goes completely around the circle l times, that is

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) &= \frac{1}{N}\sum_{j=0}^{q-1} l\alpha^{-1}\int_{\mathbb{T}^1} f(z)dz + \mathcal{O}(\frac{Kq}{2N}|f|_{\mathcal{C}^1}\alpha + \frac{r+qs}{N}|f|_\infty)\\ &= \frac{ql\alpha^{-1}}{N}\int_{\mathbb{T}^1} f(z)dz + \mathcal{O}(\frac{Kq}{2N}|f|_{\mathcal{C}^1}\alpha + \frac{r+qs}{N}|f|_\infty)\\ &= \int_{\mathbb{T}^1} f(z)dz + \mathcal{O}(\frac{Kq}{2N}|f|_{\mathcal{C}^1}\alpha + \frac{2r+qs+\alpha^{-1}sq}{N}|f|_\infty).\end{aligned}$$

Thus,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T_\omega^n(x)) = \int_{\mathbb{T}^1} f(z)dz, \tag{3.2}$$

and this time the limit does not depend on the initial condition!

Indeed, the measurement converges toward the average with respect to the Lebesgue measure, which is easily seen to be invariant. In fact, we have gone from an extreme to the other as the following lemma shows.

Lemma 3.1. *The Lebesgue measure is the only invariant probability measure for T_ω, with $\omega \notin \mathbb{Q}$.*

[2] Indeed $T_\omega^n x = T_\omega^m x$ for some $n \neq m$ readily implies $\omega \in \mathbb{Q}$.

[3] Notice that, in this manner we have obtained a weak, and yet non trivial, information on the possibility to approximate irrational numbers. This is a pale example of the many connections between ergodic theory and number theory.

Proof. Suppose ν is an invariant measure, that is $T_\omega^*\nu(f) := \nu(f \circ T_\omega) = \nu(f)$. Next, let $f \in \mathcal{C}^1$,

$$\nu(f) = \frac{1}{N}\sum_{n=0}^{n-1} T_\omega^{*n}\nu(f) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{n-1} T_\omega^{*n}\nu(f) = \lim_{N\to\infty}\nu\left(\frac{1}{N}\sum_{n=0}^{N-1} f\circ T_\omega^n\right).$$

Now, notice that the convergence in (3.2) is, in fact, in the uniform topology, hence

$$\nu(f) = \nu\left(\int_{\mathbb{T}^1} f\right) = \nu(1)\int_{\mathbb{T}^1} f = \int_{\mathbb{T}^1} f.$$

The result follows since $\mathcal{C}^1$ is dense in $\mathcal{C}^0$. □

A system with a unique invariant measure is called *uniquely ergodic*. A moment thought shows that any system with two periodic orbit cannot be uniquely ergodic, this shows how unusual such systems are. Nevertheless, the above considerations show that the existence of a unique "special" invariant measure may yield the phenomenon we are looking for, that is the fact that result of a measurement is given with overwhelming probability by the average computed with respect to such a measure.

To understand a bit better let us look at a still simple but far less trivial example.

3.2. **Circle doubling.** Let us consider the map $T : \mathbb{T}^1 \to \mathbb{T}^1$ defined by

$$T(x) = 2x \mod 1. \tag{3.3}$$

First of all let us see that the system is not uniquely ergodic, indeed $T(0) = 0$ and $T(\frac{1}{3}) = \frac{2}{3}$ and $T(\frac{2}{3}) = \frac{1}{3}$, thus we a fixed point and a period two orbit, hence δ_0 and $\frac{1}{2}\delta_{\frac{1}{3}} + \frac{1}{2}\delta_{\frac{2}{3}}$ are both invariant measures for T. Of course, many other such measures can be constructed. In addition,

$$\int_{\mathbb{T}^1} \varphi\circ T = \int_0^{\frac{1}{2}} \varphi(2x)dx + \int_{\frac{1}{2}}^1 \varphi(2x-1)dx = \int_0^1 \varphi. \tag{3.4}$$

That is the Lebesgue measure is another invariant measure. It seems reasonable to think that the Lebesgue measure is more relevant, from the physical point of view, than the above measures concentrated on periodic orbits. It is then natural to wonder if it is the only one supported on sets of positive Lebesgue measure, that is in the class of measures absolutely continuous with respect to the Lebesgue measure. To answer such a question a considerable detour is needed.

A far reaching and very effective strategy to investigate invariant measures is to lift the dynamics on the measures. Namely, one can first define a dynamics on $\mathcal{C}^0(\mathbb{T}^1, \mathbb{R})$ by $T_*\varphi := \varphi \circ T$ and then on the Borel measures[4] $\mathcal{C}^0(\mathbb{T}^1, \mathbb{R})'$ by $T^*\mu(\varphi) := \mu(T_*\varphi)$.[5] In this language (3.4) reads $T^*m = m$. Next, consider the

[4]The Borel measures can be identified with the dual of $\mathcal{C}^0(\mathbb{T}^1, \mathbb{R})$ thanks to the Riesz representation theorem.

[5]Another, more general, way to define a dynamics on measures is to define $T^*\mu(A) := \mu(T^{-1}A)$ for each measurable set A. It is an easy exercise to verify that the two definition coincide if T is a continuous map.

measure $d\mu = hdm$, then

$$T^*\mu(\varphi) = \int_{\mathbb{T}^1} h(x)\varphi(T(x))dx = \int_0^1 \frac{1}{2}h\left(\frac{x}{2}\right)\varphi(x)dx + \int_0^1 \frac{1}{2}h\left(\frac{x+1}{2}\right)\varphi(x)dx$$
$$=: \int_{\mathbb{T}^1} \mathcal{L}h\varphi.$$

That is $\frac{dT^*\mu}{dm} = \mathcal{L}h$, where

$$\mathcal{L}h(x) = \frac{1}{2}h\left(\frac{x}{2}\right) + \frac{1}{2}h\left(\frac{x+1}{2}\right). \tag{3.5}$$

The operator $\mathcal{L}$, often called *transfer operator* or *Ruelle-Perron-Frobenius operator* describes the evolution of the densities of the measures. In the above terms (3.4) reads $\mathcal{L}1 = 1$. More in general, if $d\mu = hdm$ and $T^*\mu = \mu$, then $\mathcal{L}h = h$ and vice versa, if $h \in L^1(\mathbb{T}^1, m)$. Thus, not surprisingly, the answer to our questions on the invariant measures absolutely continuous with respect to Lebesgue can be obtained by studying the operator $\mathcal{L}$. First of all,

$$\int_{\mathbb{T}^1} |\mathcal{L}h| \leq \int_{\mathbb{T}^1} \mathcal{L}|h| \cdot 1 = \int_{\mathbb{T}^1} |h| \cdot 1 \circ T = \int_{\mathbb{T}^1} |h|. \tag{3.6}$$

In other words $\mathcal{L}$ is a contraction (but not a strict one) in $L^1(\mathbb{T}^1, m)$. Yet, the experience has shown that studying $\mathcal{L}$ as a bounded operator on $L^1(\mathbb{T}^1, m)$ it is not a very rewarding activity. It turns out to be much more helpful to restrict it to smoother functions. In fact, let $h \in \mathcal{C}^1(\mathbb{T}^1, \mathbb{R})$, then differentiating (3.5) yields

$$(\mathcal{L}h)' = \frac{1}{2}\mathcal{L}(h')$$

But the above equality implies

$$|(\mathcal{L}h)'|_{L^1} \leq \frac{1}{2}|h'|_{L^1}. \tag{3.7}$$

It is then clear that it is convenient to study $\mathcal{L}$ as an operator on the Sobolev space $W_{1,1}$. In fact, let $V_0 := \{h \in W_{1,1} \; : \; \int h = 0\}$, then $\mathcal{L}V_0 \subset V_0$. Moreover, notice that if $h \in V_0$, then $|h|_{L^1} \leq |h'|_{L^1}$. Accordingly, for each $h \in V_0$, holds

$$|\mathcal{L}^n h|_{L^1} \leq |(\mathcal{L}^n h)'|_{L^1} \leq 2^{-n}|h'|_{L^1}.$$

That is

$$|\mathcal{L}^n h|_{W_{1,1}} \leq 2^{-n+1}|h'|_{L^1} \leq 2^{-n+1}|h|_{W_{1,1}}.$$

Clearly this shows that there are no other absolutely continuous invariant measures.

Lemma 3.2. *If an invariant measure is absolutely continuous with respect to Lebesgue, then it is Lebesgue. That is, the dynamical system* $(T, \mathbb{T}^1, m)$ *is* ergodic.

Proof. Let $h \in L^1$ be such that $\mathcal{L}h = h$, $\int h = 1$, then, for each $\varepsilon > 0$ there exists $h_\varepsilon \in W_{1,1,}$, $\int h_\varepsilon = 1$, such that $|h - h_\varepsilon|_{L^1} \leq \varepsilon$. But then

$$|\mathcal{L}^n h_\varepsilon - h_\varepsilon|_{L^1} \leq |\mathcal{L}^n(h_\varepsilon - h)|_{L^1} + |h_\varepsilon - h|_{L^1} \leq 2\varepsilon,$$

and, since $h_\varepsilon - 1 \in V_0$,

$$|h-1|_{L^1} \leq \varepsilon + |h_\varepsilon - 1|_{L^1} \leq 3\varepsilon + |\mathcal{L}^n(h_\varepsilon - 1)|_{L^1} \leq 3\varepsilon + 2^{-n+1}|h'_\varepsilon|_{L^1} \leq 4\varepsilon,$$

where we have chosen n large enough. By the arbitrariness of ε it follows $h = 1$. □

Since the spectral radius of $\mathcal{L}|_{V_0}$ is $\frac{1}{2}$, the spectrum of $\mathcal{L}$ on $W_{1,1}$ consists of the simple eigenvalue one while the rest of the spectrum is contained in the disk of radius $\frac{1}{2}$.

It will be interesting, for the following to notice that the system under consideration enjoys a stronger property than ergodicity.

Lemma 3.3. *For each $h, f \in L^2(\mathbb{T}^1, m)$, $\int h = 0$, holds*

$$\lim_{n\to\infty} |\mathcal{L}^n h|_{L^1} = 0 \quad \text{and} \quad \lim_{n\to\infty} \int h f \circ T^n = 0.$$

Proof. For each $\varepsilon > 0$ choose $h_\varepsilon \in W_{1,1}$ such that $\int h_\varepsilon = 0$ and $|h - h_\varepsilon|_{L^2} \le \varepsilon$, then

$$|\mathcal{L}^n h|_{L^1} \le |\mathcal{L}^n h_\varepsilon|_{L^1} + |h - h_\varepsilon|_{L^1} \le |\mathcal{L}^n h_\varepsilon|_{L^1} + \varepsilon$$

and

$$\int h f \circ T^n = \int \mathcal{L}^n h_\varepsilon f + \mathcal{O}(\varepsilon |f \circ T^n|_{L^2}) = \mathcal{O}\left(2^{-n+1}|h_\varepsilon|_{W_{1,1}}|f|_{L^2} + \varepsilon|f|_{L^2}\right).$$

The result follows by choosing first ε small and then n large. □

We have finally collect enough knowledge on the statistical properties of the map to be able to tackle the problem of the measure previously discussed for the rotations. To start with, let $f \in W_{1,1}$, $m(f) > 0$, and consider the set

$$A_\delta^+ := \left\{ x \in \mathbb{T}^1 \ : \ \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \ge (1+\delta)m(f) \right\}. \tag{3.8}$$

As before we would like to estimate the measure (in this case the *Lebesgue* measure) of the set. We use the same strategy already used: for $\lambda > 0$,

$$m(A_\delta^+) = m(\{x \ : \ e^{\lambda \sum_{k=0}^{n-1} f\circ T^k(x) - (1+\delta)m(f)} \ge 1\}) \le m(e^{\lambda \sum_{k=0}^{n-1} f\circ T^k - (1+\delta)m(f)}).$$

Now set $g := f - (1+\delta)m(f)$, then

$$m(A_\delta^+) \le m\left(\prod_{k=0}^{n-1} e^{\lambda g} \circ T^k\right) = m(\mathcal{L}_\lambda^n 1)$$

where we have defined the operator $\mathcal{L}_\lambda h := \mathcal{L}(e^{\lambda g} h)$. Since $W_{1,1}$ is an algebra it is easy to see that $\mathcal{L}_\lambda$ is a well defined operator on $W_{1,1}$. The basic idea to accomplish the wanted estimate is to show that the spectral radius of $\mathcal{L}_\lambda$ is strictly smaller than one. To prove such a fact we will try to apply perturbation theory viewing $\mathcal{L}_\lambda$ as a perturbation of $\mathcal{L}$.

Lemma 3.4. *For each $0 < \lambda \le |g|_\infty^{-1}$ holds true*

$$\|\mathcal{L} - \mathcal{L}_\lambda\|_{W_{1,1,}} \le 2e\lambda |g|_{W_{1,1,}}.$$

Proof. Remembering (3.6) and (3.7), just compute

$$\begin{aligned}
\|(\mathcal{L} - \mathcal{L}_\lambda)h\|_{W_{1,1,}} &\le |(1 - e^{\lambda g})h|_{L^1} + \frac{1}{2}|(1-e^{\lambda g})' h + |(1-e^{\lambda g})h'|_{L^1} \\
&\le e\lambda |g|_\infty |h|_{L^1} + \frac{e}{2}\lambda |g|_\infty |h'|_{L^1} + \frac{e}{2}\lambda |g'|_{L^1}|h|_\infty \\
&\le 2e\lambda |g|_{W_{1,1,}} |h|_{W_{1,1,}}.
\end{aligned}$$

□

At this point let us recall a bit of perturbation theory. For each $z \in \mathbb{C}$ let $R_\lambda(z) := (z\mathbf{Id} - \mathcal{L}_\lambda)^{-1}$, if the operator is defined, and set $R(z) := R_0(z)$. Also, given any operator A let $\sigma(A)$ be its spectrum.

Lemma 3.5. *If $z \notin \sigma(\mathcal{L})$ and $\|(\mathcal{L} - \mathcal{L}_\lambda)R(z)\|_{W_{1,1,}} < 1$, then $z \notin \sigma(\mathcal{L}_\lambda)$ and*

$$R_\lambda(z) = R(z)(\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))^{-1}.$$

Proof. Since $\|(\mathcal{L} - \mathcal{L}_\lambda)R(z)\|_{W_{1,1}} < 1$, the operator $(\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))^{-1}$ is well defined and, more precisely,

$$(\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))^{-1} = \sum_{n=0}^{\infty} [(\mathcal{L}_\lambda - \mathcal{L})R(z)]^n .$$

Moreover,

$$(z\mathbf{Id} - \mathcal{L}_\lambda)R(z)(\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))^{-1} = (\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))(\mathbf{Id} + (\mathcal{L} - \mathcal{L}_\lambda)R(z))^{-1} = \mathbf{Id}.$$

□

In addition notice that, since $\mathcal{L}_\lambda$ is an analytic function of λ, so is $R_\lambda(z)$. Now, calling $\Pi(h) := m(h)$ the projector on the eigenvalue 1 of $\mathcal{L}$, the previous results and an easy computation imply

$$R(z) = (z - 1)^{-1}\Pi + Q(z)$$

where $Q(z) = z^{-1} \sum_{n=0}^{\infty} z^{-n}\mathcal{L}^n(\mathbf{Id} - \Pi)$. Thus if $\Gamma := \{z \in \mathbb{C} \;:\; |z - 1| = \frac{1}{4}\}$ then there exists a constant $c > 0$ such that

$$\sup_{z\in\Gamma} \|(\mathcal{L} - \mathcal{L}_\lambda)R(z)\|_{W_{1,1}} < C\lambda|g|_{W_{1,1}}.$$

This means that that, for λ small enough, $\Gamma \cap \sigma(\mathcal{L}_\lambda) = \emptyset$ and, calling Π_λ the eigenprojector associated at the portion of $\sigma(\mathcal{L}_\lambda)$ in the interior of Γ, holds

$$\Pi_\lambda = \frac{1}{2\pi i} \int_\Gamma R_\lambda(z)dz. \tag{3.9}$$

Clearly $\Pi_0 = \Pi$, in addition the dimension of the range of Π_λ is an analytic function and hence must be constant, that is Π_λ is a rank one projector. Accordingly, there exists $h_\lambda \in W_{1,1}$ and $\ell_\lambda \in W'_{1,1}$, both analytic in λ, such that $\Pi_\lambda(h) = h_\lambda\ell_\lambda(h)$. Obviously

$$\mathcal{L}_\lambda h_\lambda = \alpha_\lambda h_\lambda, \tag{3.10}$$

and $\alpha_0 = 1$, $h_0 = 1$. Notice that h_λ and ℓ_λ are not uniquely defined: by $\Pi_\lambda^2 = \Pi_\lambda$ follows $\ell_\lambda(h_\lambda) = 1$ but one normalization can be chosen freely, let us choose $m(h_\lambda) = 1$. All the above discussion is summarize by the following Lemma.

Lemma 3.6. *There exists a constant $C_1, C_2 > 0$ and $0 < \rho < \frac{3}{4}$ such that, for $\lambda \leq C_1|g|_{W_{1,1}}^{-1}$, $\mathcal{L}_\lambda = \alpha_\lambda\Pi_\lambda + Q_\lambda$, $\Pi_\lambda Q_\lambda = Q_\lambda\Pi_\lambda = 0$, $\|Q_\lambda^n\|_{W_{1,1}} \leq C_2\rho^n$. Moreover everything is analytic in λ.*

In view of the above fact we can differentiate (3.10) obtaining

$$\mathcal{L}'_\lambda h_\lambda + \mathcal{L}_\lambda h'_\lambda = \alpha'_\lambda h_\lambda + \alpha_\lambda h'_\lambda \;; \quad m(h'_\lambda) = 0. \tag{3.11}$$

Integrating with respect to m and setting $\lambda = 0$ yields

$$\alpha'_0 = m(\mathcal{L}g) = m(g) = -\delta m(f) < 0.$$

This means that we can choose λ such that the norm of $\mathcal{L}_\lambda^n$ is strictly smaller than one, yet to know how small we can take it, it is necessary to investigate the second derivative of α_λ. Taking the derivative of (3.11), integrating with respect to m and setting $\lambda = 0$ yields

$$\alpha_0'' = m(g^2) + 2m(gh_0').$$

On the other hand, (3.11) implies

$$(\mathbf{Id} - \mathcal{L})h_0' = \mathcal{L}(g - m(g)) = \mathcal{L}(\mathbf{Id} - \Pi)g$$

and, setting $\hat{\mathcal{L}} := \mathcal{L}(\mathbf{Id} - \Pi)$,

$$h_0' = (\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}g. \tag{3.12}$$

Hence

$$\alpha_0'' = m(g^2) + 2m(g(\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}g). \tag{3.13}$$

Since $\alpha_\lambda = 1 - \alpha_0'\lambda + \frac{1}{2}\alpha_0''\lambda^2 + \mathcal{O}(\lambda^3)$, it follows that the situation is drastically different if α_0'' is positive or negative. Looking at (3.13) the sign is far from evident, yet a careful analysis shows that the sign is often positive.

Lemma 3.7. *Setting $\hat{f} := f - m(f)$, either $\alpha_0'' \geq C > 0$, with C independent on δ, or, for each periodic orbit $\{x_i = T^i x_0\}_{i=0}^{n-1}$, it holds true $\sum_{i=0}^{n-1} \hat{f}(x_i) = 0$.*

Proof. First of all $(\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}g = (\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}\hat{f} \in V_0$, thus

$$\alpha_0'' = m(\hat{f}^2) + \delta^2 m(f)^2 + 2m(\hat{f}(\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}\hat{f}) \geq m(\hat{f}^2) + 2m(\hat{f}(\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}\hat{f}).$$

Next consider the following

$$\begin{aligned}
0 \leq m\left(\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\hat{f}\circ T^k\right]^2\right) &= \frac{1}{n}\sum_{k,j=0}^{n-1} m(\hat{f}\circ T^k \hat{f}\circ T^j) \\
&= m(\hat{f}^2) + \frac{2}{n}\sum_{j=0}^{n-1}\sum_{k=j+1}^{n-1} m(\hat{f}\hat{f}\circ T^{k-j}) = m(\hat{f}^2) + \frac{2}{n}\sum_{k=1}^{n-1}\sum_{j=1}^{k} m(\hat{f}\hat{\mathcal{L}}^j\hat{f}) \\
&= m(\hat{f}^2) + 2\sum_{j=1}^{n-1}\frac{n-j}{n} m(\hat{f}\hat{\mathcal{L}}^j\hat{f}).
\end{aligned}$$

Accordingly,

$$\begin{aligned}
0 \leq \sigma^2 := \lim_{n\to\infty} m\left(\left[\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\hat{f}\circ T^k\right]^2\right) &= m(\hat{f}^2) + 2\sum_{j=1}^{\infty} m(\hat{f}\hat{\mathcal{L}}^j\hat{f}) \\
&= m(\hat{f}^2) + 2m(\hat{f}(\mathbf{Id} - \hat{\mathcal{L}})^{-1}\hat{\mathcal{L}}\hat{f}).
\end{aligned}$$

Clearly, if $\sigma > 0$ the lemma is proven, thus we need only to analyze the case $\sigma = 0$.

If $\sigma = 0$, we have

$$\begin{aligned} m\left(\left[\sum_{k=0}^{n-1}\hat f\circ T^k\right]^2\right) &= n\left[m(\hat f^2)+2\sum_{j=1}^{n-1}\frac{n-j}{n}m(\hat f\hat{\mathcal L}^j\hat f)\right]\\ &= -2n\sum_{j=n}^{\infty}m(\hat f\hat{\mathcal L}^j\hat f)-2\sum_{j=1}^{n-1}jm(\hat f\hat{\mathcal L}^j\hat f)\\ &\leq C_3\left[n2^{-n}+\sum_{j=0}^{\infty}j2^{-j}\right]|\hat f|_{L^1}\|\hat f\|_{W_{1,1}}\leq C_4|\hat f|_{L^1}\|\hat f\|_{W_{1,1}} \end{aligned}$$

Accordingly, the sequence $\sum_{k=0}^{n-1}\hat f\circ T^k$ is bounded in L^2 and hence weakly compact. Let $\sum_{k=0}^{n_j-1}\hat f\circ T^k$ a weakly convergent subsequence, that is there exists $\phi\in L^2$ such that for each $\varphi\in L^2$ holds

$$\lim_{j\to\infty}\int\varphi\sum_{k=0}^{n_j-1}\hat f\circ T^k=\int\varphi\phi.$$

It follows that, for each $\varphi\in W_{1,1}$,

$$\begin{aligned} \int\varphi[\hat f-\phi+\phi\circ T]&=\varphi\hat f-\lim_{j\to\infty}\sum_{k=0}^{n_j-1}\int\varphi\hat f\circ T^k+\int\mathcal L\varphi\hat f\circ T^k\\ &=\lim_{j\to\infty}\int\mathcal L^{n_j}[\varphi-m(\varphi)]\hat f=0 \end{aligned}$$

And, since $W_{1,1}$ is dense in L^2, it follows

$$\hat f=\phi-\phi\circ T. \tag{3.14}$$

A function with the above property is called a *coboundary*, in this case an L^2 coboundary since we know only that $\phi\in L^2$. In fact, this it is not not enough to conclude the Lemma: we need to show, at least, that $\phi\in\mathcal C^0$.

First of all notice that, since for each $\beta\in\mathbb R$ we have $\hat f=\phi+\beta-(\phi+\beta)\circ T$, we can assume without loss of generality $\int\phi=0$. But them

$$\hat{\mathcal L}\hat f=\mathcal L\hat f=\mathcal L\phi-\phi=\hat{\mathcal L}\phi-\phi=-(\mathbf{Id}-\hat{\mathcal L})\phi.$$

Hence

$$\lim_{n\to\infty}\sum_{k=0}^{n-1}\hat{\mathcal L}^k(\mathbf{Id}-\hat{\mathcal L})\phi=\phi-\lim_{n\to\infty}\mathcal L^n\phi=\phi,$$

where we have used Lemma 3.3. Accordingly

$$\phi=(\mathbf{Id}-\hat{\mathcal L})^{-1}\hat{\mathcal L}\hat f\in W_{1,1}$$

That is $\hat f$ is a continuous coboundary. From this the lemma follows immediately. □

Accordingly, unless $\hat f$ it is not a continuous coboundary, the best we can do is to choose $\lambda=c\delta m(f)$, whereby obtaining $\alpha_\lambda=e^{-c\delta^2}$.

We can finally conclude

$$m(A_\delta^+)\leq m(\mathcal L^n_{c\delta m(f)}1)\leq Ce^{-c\delta^2 n}.$$

Since similar arguments hold for the set $A_\delta^- := \{x \in \mathbb{T}^1 \;:\; \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k - (1-\delta)m(f) \leq 0\}$, it follows that we have an exponentially small probability to observe a deviation from the average larger than a δ percentage of the average. In particular we have the following.

Lemma 3.8. *For each $f \in L^1$, holds*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) = m(f), \quad m - a.e.$$

Proof. Notice that, for each $f \in W_{1,1}$, $m(f) > 0$, and for each $m \in \mathbb{N}$, $\delta > 0$, holds

$$\begin{aligned}
&m(\{x \in \mathbb{T}^1 \;:\; \exists n \geq m \;:\; |\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) - m(f)| \geq \delta m(f)\}) \\
&\leq \sum_{n=m}^{\infty} m(\{x \in \mathbb{T}^1 \;:\; |\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) - m(f)| \geq \delta m(f)\}) \\
&\leq \sum_{n=m}^{\infty} Ce^{-c\delta^2 n} \leq C\delta^{-2} e^{-c\delta^2 m}
\end{aligned}$$

Next, if $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \neq m(f)$, then there exists $\delta(x)$ and an infinite sequence $\{n_j(x)\}$ such that $|\frac{1}{n_j(x)}\sum_{k=0}^{n_j(x)-1} f \circ T^k(x) - m(f)| \geq \delta(x)m(f)$. It is easy to see that $\delta(x)$ is a measurable function, thus, for each $\varepsilon > 0$ and $m \in \mathbb{N}$,

$$m(\delta \geq \varepsilon) \leq m(\{x \;:\; \exists n \geq m \;:\; |\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) - m(f)| \geq \delta m(f)\}) \leq C\varepsilon^{-2} e^{-c\varepsilon^2 m}.$$

Since m and ε are arbitrary, it follows $m(\delta > 0) = 0$, hence the lemma for such an f. The result for general f follows by standard approximation arguments. □

At this point a natural question, of clear relevance for the applications, would be to understand a bit better in which way the limit is achieved. This is the content of the next subsection.