Chapter 7

Quantitative Statistical Properties, a class of 1-d examples

Given a Dynamical System it is in general very hard to study its ergodic properties, especially if the goal is to have a *quantitative* understanding. To make clear what is meant by a *quantitative understanding* and which type of obstacles may prevent it, I devote this chapter to the study of a simple, but highly non-trivial, class of examples: one dimensional smooth expanding maps.

7.1 The problem

Recall from Examples 6.4.1 that a one dimensional smooth expanding map is a map $T \in \mathcal{C}^2(\mathbb{T}^1, \mathbb{T}^1)$ such that $|DT| \geq \lambda > 1$.

We know already that such maps have a unique absolutely continuous invariant measure (see sections 6.4.1, 6.5.1 Expanding maps).

We would like first to understand other invariant measures in order to have a clearer picture of which measurable Dynamical Systems can be associated to the topological Dynamical System $(\mathbb{T}^1, T)$. This is still at the qualitative level. In addition, we would like to have tools to actually compute such invariant measures with a given precision, and this is a first quantitative issue.

Next, we would like to study statistical properties more in depth. To this end we will restrict to the case $(\mathbb{T}^1, T, \mu)$, where μ is the measure absolutely continuous with respect to Lebesgue. The type of questions we would like to address are

If we make repeated finite time and precision measurements, what do we observe?

Remember that a measurement is represented by the evaluation of a function. The fact that the measurement has a finite precision correspond to the fact that the function has some uniform regularity (otherwise we could identify the point with an arbitrary precision). The fact that the measure is made for finite time means that we are able only to measure finite times averages. In other words we would like to understand the behavior of

$$\sum_{k=0}^{N-1} f \circ T^k$$

for large, but finite, N.

7.2 Invariant measures

Let $\mathcal{M}$ be the set of probability (Borel) measures on $\mathbb{T}^1$. We can then consider the new Dynamical System $(\mathcal{M}, T')$, where $T'\mu(f) = \mu \circ T$ for all $f \in \mathcal{C}^0(\mathbb{T}^1, \mathbb{R})$. The invariant measures are the fixed points of T', let us call them $\text{Fix}(T')$. If $\mu \in \text{Fix}(T')$ then for each $h \in L^\infty(\mathbb{T}^1, \mu)$, $h \geq 0$, $\mu(h) = 1$, we can consider the new probability measure defined by $\mu_h(f) = \mu(hf)$, for all $f \in \mathcal{C}^0(\mathbb{T}^1, \mathbb{R})$. Note that

$$|T'\mu_h(f)| = |\mu(hf \circ T)| \leq |h|_{L^\infty(\mu)} \mu(|f| \circ T) = |h|_{L^\infty(\mu)} \mu(|f|).$$

Hence $T'\mu_h$ is absolutely continuous with respect to μ and $\frac{dT'\mu_h}{d\mu} \in L^\infty(\mu)$. We can then define the operator $\mathcal{L}_\mu : L^\infty(\mathbb{T}^1, \mu) \to L^\infty(\mathbb{T}^1, \mu)$ by $\mathcal{L}_\mu h := \frac{dT'\mu_h}{d\mu}$.

Let $\{I_i\}$ be a partition in interval of $\mathbb{T}^1$ such that $T|_{I_i}$ is invertible, $T(I_i) = \mathbb{T}^1$ and $\cup_i I_i = \mathbb{T}^1$. Call S_i the inverse of the i-th branch of T. Then, setting $\rho_i := \frac{dT'\mu_{\mathbb{1}_{I_i}}}{d\mu}$

$$\begin{aligned} T'\mu_h(f) &= \sum_i \mu(h\mathbb{1}_{I_i} f \circ T) = \sum_i \mu(\mathbb{1}_{I_i}(h \circ S_i f) \circ T) \\ &= \mu\left(\left[\sum_i \rho_i h \circ S_i\right] f\right). \end{aligned}$$

Thus, setting $\rho = \sum_i \rho_i \circ T \mathbb{1}_{I_i}$ we have

$$\frac{dT'\mu_h}{d\mu} = \sum_i (\rho h) \circ S_i =: \mathcal{L}_\rho(h).$$

It follows that $\mathcal{L}_\rho(1) = 1$ and, for each $h \in L^\infty(\mu)$, $\mu(\mathcal{L}_\rho(h)) = T'\mu_h(1) = \mu(h)$.

Problem 7.1 *Compute ρ and $\mathcal{L}_\rho$, in the case in which μ is the unique invariant measure absolutely continuous with respect to Lebesgue.*

The relevant fact is that one has the following (partial) converse.

Lemma 7.2.1 *For $\rho \in \mathcal{C}^0$, $\rho \geq 0$, let $\mathcal{L}_\rho(h)(x) := \sum_{y \in T^{-1}x} \rho(y)h(y)$. If there exists $\lambda \in \mathbb{R}$, $h \in \mathcal{C}^0$, $h > 0$, such that $\mathcal{L}_\rho h = \lambda h$, then there exists a measure $\mu \in \mathcal{M}$ such that $\mu(\mathcal{L}_\rho f) = \lambda \mu(f)$ for all $f \in \mathcal{C}^0$ and there exists an invariant measure absolutely continuous with respect to μ.*

Proof. By continuity there exists $\gamma > 0$ such that $h \geq \gamma > 0$. Thus

$$|\mathcal{L}_\rho^n f| \leq \gamma^{-1} |f|_\infty \mathcal{L}_\rho^n h = \lambda^n \gamma^{-1} |f|_\infty.$$

Hence, calling m the Lebesgue measure $\frac{1}{n}\sum_{k=0}^{n-1} \lambda^{-k} (\mathcal{L}_\rho')^k m$ is a weakly compact sequence. Accordingly the same arguments used in Krylov-Bogoliubov Theorem 6.4.2 imply that there exists a measure μ such that $\lambda^{-1}\mathcal{L}_\rho' \mu = \mu$.

Next, define $\nu(f) := \mu(hf)$. Clearly ν is a measure absolutely continuous with respect to μ, in addition

$$\nu(f \circ T) = \lambda^{-1}(\mathcal{L}_\rho' \mu)(hf \circ T) = \lambda^{-1}\mu(f \mathcal{L}_\rho h) = \mu(fh) = \nu(f).$$

□

7.3 Absolutely continuous invariant measure: revisited

We have already seen that there exists a unique invariant measure with respect to Lebesgue. Here we study this issue by a slightly different technique. Although the main idea is always to study the spectrum of the transfer operator, it is interesting to see how this can be achieved in many different ways, each way having its own advantages and disadvantages. Consider the transfer operator

$$\mathcal{L}h(x) := \sum_{y\in T^{-1}x} |D_yT|^{-1}h(y) \tag{7.3.1}$$

Problem 7.2 *Show that if $d\mu = hdm$, where m is the Lebesgue measure, then $\mu(f\circ T) = m(f\mathcal{L}h)$.*

Problem 7.3 *Show that, for each $n \in \mathbb{N}$,*

$$\mathcal{L}^n h(x) := \sum_{y\in T^{-n}x} |D_yT^n|^{-1}h(y)$$

Notice that, since DT cannot be zero, then its sign is constant. We limit ourselves, for simplicity, to the case $DT \geq \lambda$.

Problem 7.4 *Show that*

$$\begin{aligned}\frac{d}{dx}\mathcal{L}^n h(x) &= \sum_{y\in T^{-1}x} (D_yT)^{-2}h'(y) - D_y^2T(D_yT)^{-3}h(y)\\ &= \mathcal{L}((DT)^{-1}h') - \mathcal{L}(D^2T(DT)^{-2}h)\end{aligned}$$

7.3.1 A functional analytic setting

Let us consider first the Sobolev space $W^{1,1}$ and the space L^1.[1] Then, for each $h \in L^1(\mathbb{T}^1, m)$,

$$\int_{\mathbb{T}^1} |\mathcal{L}h|dm \leq \int_{\mathbb{T}^1} 1\cdot\mathcal{L}|h|dm = \int_{\mathbb{T}^1} 1\circ T|h|dm = \int_{\mathbb{T}^1} |h|dm \tag{7.3.2}$$

that is $\mathcal{L}$ is a bounded operator on L^1 and its norm is bounded by one.

[1] For an open set $U \subset \mathbb{R}$, the spaces $W^{p,q}(U)$ are the completion of $\mathcal{C}^\infty(U,\mathbb{C})$ with respect to the norms $\left[|f|^q_{L^q} + |f'|^q_{L^q} + \cdots + |f^{(p)}|^q_{L^q}\right]^{\frac{1}{q}}$. Note that they are all Banach spaces by construction but the $W^{p,2}$ are also Hilbert spaces (**Exercise**: write the scalar product).

In addition, remembering Exercise 7.2,

$$\int_{\mathbb{T}^1} |\frac{d}{dx}\mathcal{L}h|dm \leq \lambda^{-1}|h'|_{L^1} + D|h|_{L^1}, \tag{7.3.3}$$

where $D := \sup D^2T(DT)^{-2}$.

Problem 7.5 *Iterate the (7.3.2), (7.3.3) and prove, for all $n \in \mathbb{N}$,*

$$|\mathcal{L}^n h|_{L^1} \leq |h|_{L^1}$$
$$|\mathcal{L}^n h|_{W^{1,1}} \leq \lambda^{-n}|h|_{W^{1,1}} + B|h|_{L^1}$$

where $B = 1 + (1-\lambda^{-1})^{-1}D$.

Since $W_{1,1}$ controls the L^∞ norm,[2] then we have that there exists $C > 0$ such that $|\mathcal{L}^n 1|_\infty < C$ for each $n \in \mathbb{N}$.

Using such a fact we can obtain similar inequalities in the Hilbert spaces L^2 and $W^{1,2}$. Indeed

$$\|\mathcal{L}^n h\|_{L^2}^2 = \int_{\mathbb{T}^1} h(\mathcal{L}^n h)\circ T^n \leq \|h\|_{L^2}\left[\int_{\mathbb{T}^1}(\mathcal{L}^n h)^2 \circ T^n\right]^{\frac{1}{2}} = \|h\|_{L^2}$$
$$\left[\int_{\mathbb{T}^1}(\mathcal{L}^n h)^2\mathcal{L}^n 1\right]^{\frac{1}{2}} \leq C^{\frac{1}{2}}\|h\|_{L^2}\|\mathcal{L}^n h\|_{L^2}$$

Which implies $\|\mathcal{L}^n h\|_{L^2} \leq C^{\frac{1}{2}}\|h\|_{L^2}$ for each $n \in \mathbb{N}$. Hence,

$$\|\frac{d}{dx}\mathcal{L}^n h\|_{L^2} \leq \lambda^{-n}C^{\frac{1}{2}}\|h'\|_{L^2} + D_n\|h\|_{L^2}.$$

Iterating as before we have, for all $n \in \mathbb{N}$,

$$\begin{aligned} |\mathcal{L}^n h|_{L^2} &\leq C|h|_{L^2} \\ |\mathcal{L}^n h|_{W^{1,2}} &\leq A\lambda^{-n}|h|_{W^{1,2}} + B|h|_{L^2}, \end{aligned} \tag{7.3.4}$$

for some appropriate constants A, B, C depending only on the map T.

To prove the existence of an invariant measure absolutely continuous with respect to Lebesgue we can try to mimic the Krylov-Bogolubov approach, but to do so we need a compactness result to substitute the weak compactness of the unit ball of the dual of a Banach space. This takes us in a very interesting detour in some fact of functional analysis.

[2] If $f \in \mathcal{C}^\infty$, then the mean value theorem asserts $\int h = h(\xi)$ for some ξ. Then $h(x) = h(\xi) + \int_\xi^x h'(z)dz$. Thus $|h|_\infty \leq |h|_{L^1} + |h'|_{L^1} = |h|_{W^{1,1}}$. The result extends then to all elements of $W^{1,1}$ by a standard approximation argument.

7.3.2 Deeper in Functional analysis

Since we are on a circle it is a good idea to use Fourier series. For each function $h \in \mathcal{C}^\infty(\mathbb{T}, \mathbb{C})$ let h_k be its Fourier coefficients and define

$$(\mathbb{A}_n h)(x) = \sum_{|k| \leq m} h_k e^{2\pi i k x} \tag{7.3.5}$$

Clearly, for all $m > 0$,

$$\begin{aligned}|h - \mathbb{A}_m|_{L^2}^2 &= \sum_{|k|>m} |h_k|^2 = \sum_{|k|>m} |h_k|^2 |k|^{-2}|k|^2 \leq m^{-2} \sum_{|k|>m} |(h')_k|^2 \\ &\leq m^{-2}|h'|_{L^2}^2 \leq m^{-2}|h|_{W^{1,2}}^2.\end{aligned} \tag{7.3.6}$$

Using the above fact we can prove.

Lemma 7.3.1 *The unit ball of $W^{1,2}$ is (sequentially) compact in L^2.*

PROOF. Consider a sequence $\{h_m\} \subset W^{1,2}$, $|h_m|_{W^{1,2}} \leq 1$. Since $\mathbb{A}_l$ are all finite rank operators, $\{\mathbb{A}_l h_m\}$ for l fixed are contained in a bounded finite dimensional (hence compact) set, thus there exists a converging subsequence for all l while (7.3.6) shows that the sequences for fixed m are all convergent. Using the usual diagonalization trick we can then extract a converging subsequence. □

Consider now $h_n := \frac{1}{n}\sum_{k=0}^{n-1} \mathcal{L}^k 1$. By the above lemma $\{h_n\}$ is relatively compact and thus we can extract a subsequence $\{h_{n_j}\}$ converging in L^2. Let h_* be the limit. Note that $\int h_n = 1$ for all $n \in \mathbb{N}$, thus $h_* \not\equiv 0$ and $\int h_* = 1$.

Problem 7.6 *Show that $\mathcal{L}h_* = h_*$, that is $d\mu := h_* dm$ is an invariant measure absolutely continuous with respect to Lebesgue and with L^2 density.*

Of course, at this point it is natural to ask if μ is the only measure with such a property or there exist others. To answer such a question we need some more facts.

7.3.3 Even deeper in Functional analysis

Since we have to do it, let us do in the following general setting.

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

i. $|h| \leq \|h\|$ 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 $\|\mathbb{A}_\varepsilon\| \leq C$ and $|h - \mathbb{A}_\varepsilon h| \leq \varepsilon \|h\|$ for all $h \in \mathbb{B}$.[3]

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

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

c. $\|\mathcal{L}^n h\| \leq A\lambda^{-n}\|h\| + B|h|$ 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 7.3.2 *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}.*[4]

We can now prove our main result.

[3]In fact, this last property can be weakened to: The unit ball $\{h \in \mathbb{B} : \|h\| \leq 1\}$ 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 D.0.3), then one can ask the $\mathbb{A}_\varepsilon$ to be compact instead than finite rank making easier their construction in concrete cases.

[4]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).

Proof of Theorem 7.3.2. 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

$$\|\mathcal{L}^n h - \mathcal{L}_{n,\varepsilon} h\| \leq A\lambda^{-n}\|(\mathbb{1}-\mathbb{A}_\varepsilon)h\| + B|(\mathbb{1}-\mathbb{A}_\varepsilon)h| \leq A(1+C)\lambda^{-n}\|h\| + B\varepsilon\|h\|.$$

By choosing $\varepsilon = \lambda^{-n}$ we have that there exists $C_1 > 0$ such that

$$\|\mathcal{L}^n - \mathcal{L}_{n,\varepsilon}\| \leq C_1 \lambda^{-n}.$$

For each $z \in \mathbb{C}$ we can now write

$$\mathbb{1} - z\mathcal{L} = (\mathbb{1} - z(\mathcal{L} - \mathcal{L}_{n,\varepsilon})) - z\mathcal{L}_{n,\varepsilon}.$$

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.[5] 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 D.0.3 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. □

7.3.4 The harvest

We are finally in the position to use all the above result to gain a deep understanding of the properties of the Dynamical Systems under consideration.

Problem 7.7 *Show that Theorem 7.3.2 implies that there exists $\sigma \in (0,1)$, $\{\theta_k\}_{k=1}^p$ and $L > 0$ such that*

$$\mathcal{L} = \sum_{k=1}^{p} e^{i\theta_k} \Pi_{\theta_k} + R$$

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

where Π_{θ_k} *and* R *are operators on* $W^{1,2}$ *such that* $\Pi_{\theta_k}\Pi_{\theta_j} = \delta_{jk}\Pi_{\theta_k}$ *and* $R\Pi_{\theta_k} = \Pi_{\theta_k}R = 0$*. Moreover* $|R^n| \leq L\sigma^n$*.(Hint: Read section 6 of the Third Chapter of [Kat66] and recall that the operator is power bounded to exclude Jordan blocks.)*

The above implies that

$$\Pi_\theta := \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} e^{-i\theta k}\mathcal{L}^k = \begin{cases}\Pi_{\theta_i} & \text{iff } \theta = \theta_j\\ 0 & \text{otherwise.}\end{cases} \tag{7.3.7}$$

Problem 7.8 *Using equations (7.3.4) show that, for each* $h \in L^2$

$$\|\Pi_\theta h\|_{W^{1,2}} \leq C\|h\|_{L^2}.$$

(Hint: prove it first for $h \in W^{1,2}$ *and then do a density argument).*

Next, note that Exercise 7.6 implies that $h_* = \Pi_0 1 \neq 0$, that is one is in the spectrum on $\mathcal{L}$, this means that the spectral radius of $\mathcal{L}$ is one.

Accordingly, if $\Pi_\theta h = h$ we have $h \in W^{1,2} \subset \mathcal{C}^0$ and[6]

$$|h| = |\Pi_\theta h| \leq \lim_{j\to\infty}\frac{1}{n_j}\sum_{k=0}^{n_j-1}\mathcal{L}^k|h| = \Pi_0|h| \leq |h|_\infty h_*.$$

This means that all the eigenvectors of the peripheral spectrum are of the form $h = gh_*$ with $g \in \mathcal{C}^0$. Thus, if h_i is an $W^{1,2}$ orthonormal a base of the eigenspace associated to an eigenvalue θ, then the eigenprojector must have the form

$$\Pi_\theta h = \sum_i h_i \int \ell_i \cdot h,$$

with $\ell_i \in L^2$ and $\int \ell_i h_j = \delta_{ij}$. Hence $\Pi_\theta\mathcal{L} = e^{i\theta}\Pi_\theta$ implies

$$e^{i\theta}\sum_k h_k \int \ell_k \cdot h = \sum_k h_k \int \ell_k \cdot \mathcal{L}h = \sum_k h_k \int \ell_k \circ T \cdot h.$$

[6] Remember that exercise 7.8 implies that the sequence in (7.3.7) converges in L^2, accordingly there exists a subsequence that converges almost everywhere with respect to Lebesgue.

That is $e^{i\theta}\ell_k = \ell_k \circ T$. But then if we set $f_k := \bar{\ell}_k h_* \in L^2$, we have

$$\mathcal{L} f_k = e^{i\theta}\mathcal{L}(\bar{\ell}_k \circ T h_*) = e^{i\theta}\bar{\ell}_k \mathcal{L} h_* = e^{i\theta}\bar{\ell}_k h_* = e^{i\theta} f_k$$

By the above facts, this implies $\Pi_\theta f_k = f_k \in W^{1,2}$, that is $\ell_k \in \mathcal{C}^0$. But then for each $p \in \mathbb{N}$ we can set $h_p := \bar{\ell}_k^p h_*$ obtaining

$$\mathcal{L} h_p = e^{ip\theta} h_p.$$

Since the the peripheral spectrum consists of finitely many eigenvalues it follows that there must exist $p \in \mathbb{N}$ such that $p\theta = \theta \mod 2\pi$, that is the spectrum on the unit circle must be the union of finitely many cyclic groups. In turn this implies that there exists $\bar{p} \in \mathbb{N}$ such that $\bar{p}\theta = 0 \mod 2\pi$, hence $\ell_k^{\bar{p}} = \ell_k^{\bar{p}} \circ T$. But this implies that if we define the sets $A_L := \{x \in \mathbb{T} \;:\; |\ell_k^{\bar{p}}| \leq L\}$, $L \in \mathbb{R}$, they are all invariant. So if χ_L is the characteristic function of the set A_L, then $\chi_L \circ T = \chi_L$ and $\mathcal{L}(\chi_L h_*) = \chi_L h_*$. We can thus produce a lot of eigenvalues of $\mathcal{L}$, but we know that such eigenvalues form a finite dimensional space. The only possibility is that only finitely many of the A_L are different. This is like saying that ℓ_k takes only finitely many values. But $\ell_k^{\bar{p}}$ is a continuous function, so it must be constant. Hence ℓ_k can assume only $\bar{p}$ different values, thus, again by continuity, must be constant. Finally this implies $\theta = 0$.

The conclusion is that one is the only eigenvalue on the unit circle and that the associated eigenprojector has rank one. So one is a simple eigenvalue and h_* is the only invariant density for the map.

7.3.5 conclusions

If we have any probability measure ν absolutely continuous with respect to Lebesgue and with density $h \in W^{1,2}$, then setting $d\mu = h_* dm$, for each $\varphi \in W^{1,2}$ we have

$$|\mu(\varphi \circ T^n) - \nu(\varphi \circ T^n)| = \left| \int \varphi \mathcal{L}^n (h - h_*) \right| \leq \|\varphi\|_{1,2} C \sigma^n \|h - h_*\|_{1,2}$$

where σ is the largest eigenvalue of modulus smaller than one (or λ^{-1} is no such eigenvalue exist).

Remark 7.3.3 *The above means that the evolution of the present* chaotic *system, if seen at the level of the absolutely continuous measures, becomes simply a dynamics with an uniformly attracting fixed point, the simplest dynamics of all!*

7.4 General transfer operators

In the previous sections we have been very successful in studying the measure absolutely continuous with respect to Lebesgue. We have seen in §7.2 (crf. Lemma 7.2.1) that to study other invariant measures one has to analyze more general transfer operators. Here we will restrict ourselves to studying

$$\mathcal{L}_g h := \mathcal{L}(e^g h)$$

where $\mathcal{L}$ is the usual transfer operator. This are called *transfer operators with weight* and g is sometime called the *potential*. We will consider first the case of $g : \mathbb{T}^1 \to \mathbb{C}$ and specialize to real potential later on.

For convenience, and also for didactical purposes, we will use the Banach spaces $\mathcal{C}^1$ and $\mathcal{C}^0$. Hence, form now on, we will assume $T \in \mathcal{C}^2(\mathbb{T}^1, \mathbb{T}^1)$ and $g \in \mathcal{C}^1(\mathbb{T}^1, \mathbb{C})$.

The first step is to compute the powers of $\mathcal{L}_g$ and study how they behave with respect to derivation.

Problem 7.9 *Show that, for each $n \in \mathbb{N}$, holds true*

$$\mathcal{L}_g^n h = \mathcal{L}^n \left[e^{g_n} h\right],$$

where $g_n = \sum_{k=0}^{n-1} g \circ T^k$.

Problem 7.10 *Show that for each $n \in \mathbb{N}$ and $h \in \mathcal{C}^1$ holds true*

$$\frac{d}{dx}\mathcal{L}_g^n h = \mathcal{L}_g^n \left[\frac{h'}{(T^n)'} - \frac{(T^n)''}{[(T^n)']^2} h + \frac{(g_n)'}{(T^n)'} h\right]$$

Note that $|\mathcal{L}_g^n h|_\infty \leq |h|_\infty \mathcal{L}_{\Re(g)}^n 1$. In addition,[7]

$$\left|\frac{(T^n)''(y)}{[(T^n)'(y)]^2}\right| = \left|\frac{\frac{d}{dy}\prod_{k=0}^{n-1} T'(T^k y)}{[(T^n)'(y)]^2}\right|$$
$$\leq \sum_{k=0}^{n-1}\left|\frac{T''(T^k y)}{(T^{n-k})'(T^k y)}\right| \leq \sum_{k=0}^{n-1}|T''|_\infty \lambda^{-n+k+1} \leq \frac{|T''|_\infty}{1-\lambda^{-1}}.$$

Analogously,

$$\left|\frac{(g_n)'}{(T^n)'}\right| \leq \frac{|g'|_\infty}{1-\lambda^{-1}}.$$

The above inequalities imply

$$\left|\frac{d}{dx}\mathcal{L}_g^n h\right| \leq \lambda^{-n}\mathcal{L}_{\Re(g)}^n |h'| + B\mathcal{L}_{\Re(g)}^n |h|. \tag{7.4.8}$$

Which, taking the sup over x, yields

$$\left|\frac{d}{dx}\mathcal{L}_g^n h\right|_\infty \leq \lambda^{-n}|h'|_\infty \mathcal{L}_{\Re(g)}^n 1 + B_*|h|_\infty \mathcal{L}_{\Re(g)}^n 1,$$

Note that the above inequality implies that the spectral radius is bounded by $\rho = \lim_{n\to\infty}\|\mathcal{L}_{\Re(g)}^n 1\|_{\mathcal{C}^0}^{\frac{1}{n}}$ while the essential spectral radius is bounded by $\lambda^{-1}\rho$. The reader should notice that for positive potentials the above bounds are essentially sharp while for non positive, or complex, potential typically there will be cancellations that induce a smaller spectral radius. To control exactly such cancellations is, in general, a very hard problem.

7.4.1 Real potential

In this section we will restrict to the case of $g \in \mathcal{C}^1(\mathbb{T}^1, \mathbb{R})$, i.e. real potentials.

If we define the cone $\mathcal{C}_a := \{h \in \mathcal{C}^1 \;:\; h > 0 \; |h'(x)| \leq ah(x)\}$, then equation (7.4.8), for $h > 0$, implies that, for each $\sigma \in (0, \lambda^{-1})$, $\mathcal{L}_g\mathcal{C}_a \subset \mathcal{C}_{\sigma_a}$ provided $a \geq B(\sigma - \lambda^{-1})^{-1}$.[8] We can then apply the theory of Appendix A to conclude the following.

[7]The quantity estimated here is usually called *distortion*. In fact, it measure how much the maps distorts intervals.

[8]Note that this cone is almost the same than the one in Example 6.5.1, more precisely is its infinitesimal version.

Lemma 7.4.1 *For each real potential $g \in \mathcal{C}^1(\mathbb{T}^1, \mathbb{R})$, the transfer operator $\mathcal{L}_g$ has the Perron-Frobenius property, i.e. it has a simple strictly positive maximal eigenvalue and all the other eigenvalues are strictly smaller in modulus. In particular, the maximal eigenvalue of $\mathcal{L}_{\tau g}$, $\tau \in \mathbb{R}$, is analytic in τ.*[9]

7.4.2 Variational principle

. . . .

7.5 Limit Theorems

Given $f \in \mathcal{C}^1$, $n \in \mathbb{N}$ and $a \in \mathbb{R}_+$ let

$$A_{a,n}(f) := \left\{ x \in \mathbb{T}^1 \ : \ \left| \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) - \mu(f) \right| \geq a \right\}. \tag{7.5.9}$$

By the ergodic theorem $\lim_{n\to\infty} \mu(A_{a,n}(f)) = 0$. A natural question is:

Question 3 *How large is $m(A_{a,n})$?*

Note that we can write $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) - \mu(f) = \frac{1}{n}\sum_{k=0}^{n-1} \hat{f} \circ T^k(x)$ where $\hat{f} := f - \mu(f)$. So we can reduce the question to the study of zero average function. A more refined question could be.

Question 4 *Does it exists a sequence $\{c_n\}$ such that*

$$\frac{1}{c_n} \sum_{k=0}^{n-1} \hat{f} \circ T^k(x)$$

converges in some sense to a non zero finite object?

[9]This follows from the fact that the maximal eigenvalue must always be simple and the results in Appendix C.4.

7.5.1 Large deviations. Upper bound

Note that it suffices to study the set

$$A^+_{a,n}(f) := \left\{ x \in \mathbb{T}^1 \;:\; \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) - \mu(f+a) \geq 0 \right\}.$$

since $A_{a,n}(f) = A^+_{a,n}(f) \cap A^+_{a,n}(-f)$. On the other hand, setting $\hat{f} := f - \mu(f)$, for each $\lambda \geq 0$ we have

$$\begin{aligned} m(A^+_{a,n}(f)) &= m(\{x \;:\; e^{\lambda \sum_{k=0}^{n-1}(\hat{f}\circ T^k(x) - a)} \geq 1\}) \leq e^{-n\lambda a} m(e^{\lambda \sum_{k=0}^{n-1} \hat{f}\circ T^k}) \\ &= e^{-n\lambda a} m(e^{\lambda \sum_{k=0}^{n-1} \hat{f}\circ T^k}). \end{aligned}$$

Accordingly,

$$m(A^+_{a,n}(f)) \leq e^{-n\lambda a} m(\mathcal{L}^n_\lambda 1) \tag{7.5.10}$$

where we have defined the operator $\mathcal{L}_\lambda g := \mathcal{L}(e^{\lambda \hat{f}} g)$, $\mathcal{L}$ being the Transfer operator of the map T.

By Lemma 7.4.1 $\mathcal{L}_\lambda$ has a maximal eigenvalue α_λ depending analytically on λ. Hence by the same argument used in Lemma 7.2.1 there exists $c \in \mathbb{R}$ such that

$$m(A^+_{a,n}(f)) \leq e^{-n(\lambda a - \ln \alpha_\lambda) + c}.$$

Since λ has been chosen arbitrarily we have obtained

$$m(A^+_{a,n}(f)) \leq e^{-n\tilde{I}(a) + c} \tag{7.5.11}$$

where $\tilde{I}(a) := \sup_{\lambda \in \mathbb{R}^+}\{\lambda a - \ln \alpha_\lambda\}$. The problem is then reduced to studying the function $I(a)$ which is commonly called *rate function*. Note that $\tilde{I}$ is not necessarily finite. Indeed if $a > \|\hat{f}\|_\infty$, then clearly $m(A^+_{a,n}(f)) = 0$.

To better understand the rate function it is helpful to make a little digression into convex analysis.

Recall that a function $f : \mathbb{R}^d \to \mathbb{R}^d$ is convex if for each $x, y \in \mathbb{R}^d$ and $t \in [0,1]$ we have $f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$ (if the inequality is everywhere strict, then the function is *stricly convex*.

Problem 7.11 *Show that if $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$, then f is convex iff $\frac{\partial^2 f}{\partial x^2}$ is a positive matrix.*[10] *Give a condition for strict convexity.*

Problem 7.12 *If a function $f : D \subset \mathbb{R}^d \to \mathbb{R}$, D convex,*[11] *is convex and bounded, then it is continuous.*

Given a function $f : \mathbb{R}^d \to \mathbb{R}$ let us define its *Legendre transform* as

$$f^*(x) = \sup_{y \in \mathbb{R}^d} \{\langle x, y\rangle - f(y)\} \tag{7.5.12}$$

Remark that f^* can take the value $+\infty$.

Problem 7.13 *Prove that f^* is convex.*

Problem 7.14 *Prove that $f^{**} \leq f$.*

Problem 7.15 *Prove that is $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ is strictly convex, then the function $h(y) := \frac{\partial f}{\partial y}(y)$ is invertible and f^* is strictly convex. Moreover, calling g the inverse function of h, we have*

$$f^*(x) = \langle x, g(x)\rangle - f \circ g(x).$$

Problem 7.16 *Show that if $f \in \mathcal{C}^2$ is strictly convex, then $f^{**} = f$.*

Problem 7.17 *Show that, for each $x, y \in \mathbb{R}^d$, $\langle x, y\rangle \leq f^*(x) + f(y)$, (Young inequality).*

From the above discussion it follows that the rate function is defined very similarly to the Legendre transform of the logarithm of the maximal eigenvalue, which is commonly called *pressure of $\hat{f}$*. In fact, setting $I(a) = \max_{\lambda \in \mathbb{R}}(\lambda a - \ln \alpha_\lambda)$ we will se that, for $a \geq 0$, $I(a) = \tilde{I}(a)$. Unfortunately, to see that the rate function is exactly a Legendre transform takes some work. Let us start by studying the function α_λ.

Lemma 7.5.1 *There exists continuous functions $C_\lambda > 0$ and $\rho_\lambda \in (0, 1)$ such that, for $\lambda \leq 0$, $\mathcal{L}_\lambda = \alpha_\lambda \Pi_\lambda + Q_\lambda$, $\Pi_\lambda Q_\lambda = Q_\lambda \Pi_\lambda = 0$, $\|Q_\lambda^n\|_{\mathcal{C}^1} \leq C_\lambda \rho_\lambda^n \alpha_\lambda^n$. Also $\Pi_\lambda(g) = h_\lambda \ell_\lambda(g)$, $\ell_\lambda(h_\lambda) = 1$, $\ell_\lambda(h'_\lambda) = 0$. In addition, $\mu_\lambda(\cdot) := \ell_\lambda(h_\lambda \cdot)$ is an invariant probability measure. Moreover everything is analytic in λ.*

---

[10] A matrix $A \in GL(\mathbb{R}, d)$ is called *positive* if $A^T = A$ and $\langle v, Av\rangle \geq 0$ for each $v \in \mathbb{R}^d$.

[11] A set D is convex if, for all $x, y \in D$ and $t \in [0, 1]$, olds true $ty + (1 - t)x \in D$.

PROOF. As we have seen, there exists $h_\lambda \in \mathcal{C}^1$ and a measure ℓ_λ, both analytic in λ, such that the projection on the maximal eigenvalue of $\mathcal{L}_\lambda$ reads $\Pi_\lambda(h) = h_\lambda \ell_\lambda(h)$. Obviously

$$\mathcal{L}_\lambda h_\lambda = \alpha_\lambda h_\lambda, \tag{7.5.13}$$

and $\alpha_0 = 1$, $h_0 = h$ and $\ell_0 = m$. 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.

Problem 7.18 *Show that the normalization of ℓ_λ, h_λ can be chosen so that $\ell_\lambda(h'_\lambda) = 0$.*

□

Lemma 7.5.2 *The functions α_λ and $\ln \alpha_\lambda$ are convex. Moreover,*

$$\left| \frac{d}{d\lambda} \ln \alpha_\lambda \right| \leq |\hat{f}|_\infty.$$

PROOF. Note that

$$\frac{d^2}{d\lambda^2} \ln \alpha_\lambda = \frac{\alpha''_\lambda \alpha_\lambda - (\alpha'_\lambda)^2}{\alpha_\lambda^2}, \tag{7.5.14}$$

thus the convexity of $\ln \alpha_\lambda$ implies the convexity of α_λ.

In view of the above fact we can differentiate (7.5.13) obtaining

$$\mathcal{L}'_\lambda h_\lambda + \mathcal{L}_\lambda h'_\lambda = \alpha'_\lambda h_\lambda + \alpha_\lambda h'_\lambda. \tag{7.5.15}$$

Applying ℓ_λ yields

$$\frac{d\alpha_\lambda}{d\lambda} = \alpha_\lambda \ell_\lambda(\hat{f} h_\lambda)) = \alpha_\lambda \mu_\lambda(\hat{f}). \tag{7.5.16}$$

Thus $\alpha'_0 = 0$. Note that, as claimed,

$$\left| \frac{d}{d\lambda} \ln \alpha_\lambda \right| \leq |\mu_\lambda(\hat{f})| \leq |\hat{f}|_\infty.$$

Differentiating again yields

$$\frac{d^2 \alpha_\lambda}{d\lambda^2} = \alpha_\lambda \mu_\lambda(\hat{f})^2 + \alpha_\lambda \ell'_\lambda(\hat{f} g h_\lambda) + \alpha_\lambda \ell_\lambda(\hat{f} h'_\lambda). \tag{7.5.17}$$

On the other hand, from (7.5.15) we have

$$(\mathbb{1}\alpha_\lambda - \mathcal{L}_\lambda)h'_\lambda = \mathcal{L}_\lambda(f_\lambda h_\lambda),$$

where $f_\lambda = \hat{f} - \mu_\lambda(\hat{f})$. Since, by construction, $\Pi_\lambda h'_\lambda = \Pi_\lambda(f_\lambda h_\lambda) = 0$, the above equation can be studied in the space $\mathbb{V}_\lambda = (\mathbb{1} - \Pi_\lambda)\mathcal{C}^1$ in which $\mathbb{1}\alpha_\lambda - \mathcal{L}_\lambda$ is invertible.

Setting $\hat{\mathcal{L}}_\lambda := \alpha_\lambda^{-1}\mathcal{L}_\lambda$, we have

$$h'_\lambda = (\mathbb{1} - \hat{\mathcal{L}}_\lambda)^{-1}\hat{\mathcal{L}}_\lambda(f_\lambda h_\lambda). \tag{7.5.18}$$

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

$$\begin{aligned}\alpha''_\lambda &= \alpha_\lambda\mu_\lambda(\hat{f})^2 + \alpha_\lambda\ell_\lambda(f_\lambda(\mathbb{1} - \hat{\mathcal{L}}_\lambda)^{-1}(\mathbb{1} + \hat{\mathcal{L}}_\lambda)(f_\lambda h_\lambda))\\ &= \alpha_\lambda\mu_\lambda(\hat{f})^2 + \alpha_\lambda\sum_{n=1}^{\infty}\ell_\lambda(f_\lambda\hat{\mathcal{L}}_\lambda^n(\mathbb{1} + \hat{\mathcal{L}}_\lambda)(f_\lambda h_\lambda))\\ &= \frac{(\alpha'_\lambda)^2}{\alpha_\lambda} + \left[\mu_\lambda(f_\lambda^2) + 2\sum_{n=1}^{\infty}\ell_\lambda(f_\lambda\hat{\mathcal{L}}_\lambda^n(f_\lambda h_\lambda))\right]\alpha_\lambda.\end{aligned} \tag{7.5.19}$$

Finally, notice that

$$\ell_\lambda(f_\lambda\hat{\mathcal{L}}_\lambda^n(f_\lambda h_\lambda)) = \ell_\lambda(\hat{\mathcal{L}}_\lambda^n(f_\lambda \circ T^n f_\lambda h_\lambda)) = \mu_\lambda(f_\lambda \circ T^n f_\lambda)$$

and

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\mu_\lambda\left(\left[\sum_{k=0}^{n-1} f_\lambda\circ T^k\right]^2\right) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k,j=0}^{n-1}\mu_\lambda(f_\lambda\circ T^k f_\lambda\circ T^j)\\ &= \mu_\lambda(f_\lambda^2) + \lim_{n\to\infty}\frac{2}{n}\sum_{k=1}^{n-1}(n-k)\mu_\lambda(f_\lambda\circ T^k f_\lambda)\\ &= \mu_\lambda(f_\lambda^2) + 2\sum_{k=1}^{\infty}\mu_\lambda(f_\lambda\circ T^k f_\lambda).\end{aligned} \tag{7.5.20}$$

The above two facts and equations (7.5.14), (7.5.19) yield

$$\frac{d^2}{d\lambda^2}\ln\alpha_\lambda = \lim_{n\to\infty}\frac{1}{n}\mu_\lambda\left(\left[\sum_{k=0}^{n-1} f_\lambda\circ T^k\right]^2\right) \geq 0. \tag{7.5.21}$$

□

Note that equation (7.5.16) implies $\alpha_0' = 0$, hence $\alpha_\lambda' \geq 0$ for $\lambda \geq 0$. Since the maximum of $\lambda a - \ln \alpha_\lambda$ is taken either at $\alpha_\lambda a = \alpha_\lambda'$ or at infinity (if $a > \sup_{\lambda > 0} \frac{\alpha_\lambda'}{\alpha_\lambda}$), it follows that

$$\tilde{I}(a) = \sup_{\lambda \geq 0}(\lambda a - \ln \alpha_\lambda) = \sup_{\lambda}(\lambda a - \ln \alpha_\lambda) = I(a)$$

as announced. In fact, more can be said.

Lemma 7.5.3 *Either the rate function I is strictly convex, or there exists $\beta \in \mathbb{R}, \phi \in \mathcal{C}^0$ such that $f - \beta = \phi - \phi \circ T$.*

Proof. By Problem 7.15 it suffices to prove that $\ln \alpha_\lambda$ is strictly convex. On the other hand equations (7.5.14) and (7.5.21) imply that if the second derivative of $\ln \alpha_\lambda$ is zero for some λ, then

$$\begin{aligned}\mu_\lambda\left(\left[\sum_{k=0}^{n-1} f_\lambda \circ T^k\right]^2\right) &= n\left[\mu_\lambda(\hat{f}^2) + 2\sum_{k=1}^{n-1}\frac{n-k}{n}\mu_\lambda(f_\lambda \circ T^k\, f_\lambda)\right] \\ &= -2n\sum_{k=n}^{\infty}\ell_\lambda(f_\lambda \hat{\mathcal{L}}_\lambda^k(f_\lambda\, h_\lambda)) - 2\sum_{k=1}^{n-1}k\ell_\lambda(f_\lambda \hat{\mathcal{L}}_\lambda^k(f_\lambda\, h_\lambda)) - \alpha_\lambda \mu_\lambda(\hat{f})^2 \\ &\leq C(\lambda)\left[n\rho_\lambda^n + \sum_{k=0}^{\infty}k\rho_\lambda^k\right]\end{aligned}$$

Accordingly, the sequence $\sum_{k=0}^{n-1} f_\lambda \circ T^k$ is bounded in $L^2(\mathbb{T}^1, \mu_\lambda)$ and hence weakly compact. Let $\sum_{k=0}^{n_j-1} f_\lambda \circ T^k$ a weakly convergent subsequence,[12] that is there exists $\phi_\lambda \in L^2$ such that for each $\varphi \in L^2$ holds

$$\lim_{j\to\infty}\mu_\lambda(\varphi\sum_{k=0}^{n_j-1} f_\lambda \circ T^k) = \mu_\lambda(\varphi\phi_\lambda).$$

It follows that, for each $\varphi \in \mathcal{C}^1$,

$$\begin{aligned}\mu_\lambda(\varphi[f_\lambda - \phi_\lambda + \phi_\lambda \circ T]) &= \mu_\lambda(\varphi f_\lambda) + \lim_{j\to\infty}\sum_{k=0}^{n_j-1}\mu_\lambda(\varphi f_\lambda \circ T^{k+1} - \varphi f_\lambda \circ T^k) \\ &= \lim_{j\to\infty}\mu_\lambda(\varphi f_\lambda \circ T^{n_j}) = \lim_{j\to\infty}\ell_\lambda(f_\lambda \hat{\mathcal{L}}_\lambda^{n_j}(\varphi h_\lambda)) \\ &= \mu_\lambda(\varphi)\mu_\lambda(f_\lambda) = 0.\end{aligned}$$

[12] Such a subsequence always exists [LL01].

thus, since $\mathcal{C}^1$ is dense in L^2, it follows

$$f_\lambda = \phi_\lambda - \phi_\lambda \circ T\,, \qquad \mu_\lambda - \text{a.s.} \tag{7.5.22}$$

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

First of all notice that, since for each $\beta \in \mathbb{R}$ we have $f_\lambda = \phi_\lambda + \beta - (\phi_\lambda + \beta) \circ T$, we can assume without loss of generality $\mu_\lambda(\phi_\lambda) = 0$. But then

$$\hat{\mathcal{L}}_\lambda(f_\lambda\, h_\lambda) = \hat{\mathcal{L}}_\lambda(\phi_\lambda\, h_\lambda) - \phi_\lambda\, h_\lambda = -(\mathbb{1} - \hat{\mathcal{L}}_\lambda)\phi_\lambda\, h_\lambda.$$

Hence

$$\phi_\lambda = h_\lambda^{-1}(\mathbb{1} - \hat{\mathcal{L}}_\lambda)^{-1}\hat{\mathcal{L}}_\lambda(f_\lambda\, h_\lambda) \in \mathcal{C}^1.$$

□

Remark 7.5.4 *The above result is quite sharp. Indeed, it shows that if I is not strictly convex, then for each invariant measure ν holds $\nu(f) = \beta = \mu(f)$. So it suffices to find two invariant measures for which the average of f differs (for example the average on two periodic orbits) to infer that I is strictly convex.*

Problem 7.19 *Set $\sigma := \alpha''(0)$. Show that, for a small, $I(a) = \frac{a^2}{2\sigma} + \mathcal{O}(a^3)$. Show that if $a > |f|_\infty$, then $I(a) = +\infty$.*

The above discussion allows to conclude

$$m(A^+_{a,n}(f)) \leq m(\mathcal{L}^n_{\lambda\_}h) \leq Ce^{-\frac{a^2}{2\sigma^2}n + \mathcal{O}(a^3 n)}.$$

Since similar arguments hold for the set $A^+_{a,n}(-f)$, it follows that we have an exponentially small probability to observe a deviation from the average. Moreover, the expected size of a deviation is of order $n^{-\frac{1}{2}}$, to see if this is really the case we a lower bound.

7.5.2 Large deviations. Lower bound

Let $I = (\alpha, \beta)$, fix $c \in (0, \frac{\beta - \alpha}{2})$ and let us consider a $\lambda \in \mathbb{R}$ such that $\mu_\lambda(\hat{f}) \in (\alpha + c, \beta - c) = I_c$. Let $S_n = \sum_{k=0}^{n-1} \hat{f} \circ T^k$, then $\mu_\lambda(S_n) =$

$n\mu_\lambda(\hat f)$ and, by (7.5.20)

$$\mu_\lambda\left(\left[\sum_{k=0}^{n-1}\hat f\circ T^k - n\mu_\lambda(\hat f)\right]^2\right)\leq C_\lambda n,$$

where C_λ depends continuously by λ. Thus, setting $A_{n,I}=\{x\in\mathbb{T}^1 : \frac{1}{n}S_n(x)\in I\}$,

$$\begin{aligned}\mu_\lambda(A^c_{n,I})&\leq \mu_\lambda\left(\left\{\left|\sum_{k=0}^{n-1}f_\lambda\circ T^k\right|\geq cn\right\}\right)\\ &\leq c^{-2}n^{-2}\mu_\lambda\left(\left|\sum_{k=0}^{n-1}f_\lambda\circ T^k\right|^2\right)\leq C_\lambda c^{-2}n^{-1}.\end{aligned}$$

It follows that there exists $n_\lambda\in\mathbb{N}$ such that, for all $n\geq n_\lambda$, $\mu_\lambda(A_{n,I})\geq \frac{1}{2}$. We can then write

$$\frac{1}{2}\leq \ell_\lambda(A_{n,I}h_\lambda)\leq C_\# e^{-(n+m)\ln\alpha_\lambda}\ell_\lambda\left(\mathcal{L}^{n+m}_\lambda(\mathbb{1}_{A_{n,I}})\right). \tag{7.5.23}$$

To conclude we must analyze a bit the characteristic function of $A_{n,I}$. First of all, notice that if $|T^nx-T^ny|\leq\varepsilon$ then for each $z\in[x,y]$

$$\begin{aligned}|D_xT^n-D_zT^n|&\leq |D_xT^n|\cdot(e^{\sum_{k=0}^{n-1}|\ln D_{T^kx}T-\ln D_{T^kz}T|}-1)\\ &\leq |D_xT^n|(e^{C_\#\sum_{k=0}^{n-1}\lambda^{-k}\varepsilon}-1)\leq C_\#|D_xT^n|.\end{aligned}$$

By a similar estimate follows $|D_xT^n-D_zT^n|\geq C_\#|D_xT^n|$ as well. Moreover,

$$|S_n(x)-S_n(y)|\leq\sum_{k=0}^{n-1}|f|_{\mathcal{C}^1}C_\#\lambda^{-k}\varepsilon\leq C_\#\varepsilon.$$

We can then write $A_{n,I}\supset\cup_l J_l\supset A_{n,I_c}$ where J_l are disjoint intervals such that $|T^nJ_l|\leq\varepsilon$. Choosing ε small enough it follow that the oscillation of S_n on each J_l is smaller than c. Moreover

$$\begin{aligned}\|\mathcal{L}^n\mathbb{1}_{J_l}\|_{BV}&=\sup_{|\varphi|_\infty\leq 1}\int_{J_l}\varphi'\circ T^n\leq\sup_{|\varphi|_\infty\leq 1}\int_{J_l}\frac{d}{dx}\left[(DT^n)^{-1}\varphi\circ T^n\right]+B|J_l|\\ &\leq 2\sup_{x\in J_l}|D_xT^n|^{-1}+B|J_l|\leq C_\#|J_l|.\end{aligned}$$

We can then continue our estimate started in (7.5.23),

$$
\begin{aligned}
\frac{1}{2} &\leq C_\# e^{-(n+m)\ln\alpha_\lambda + n\lambda\beta + mC_\#} \sum_l \ell_\lambda\left(\mathcal{L}^{n+m}(\mathbb{1}_{J_l})\right) \\
&= C_\# e^{-(n+m)\ln\alpha_\lambda + n\lambda\beta + mC_\#} \sum_l \ell_\lambda\left(m(J_l)(1+\mathcal{O}(\rho^m))\right)) \\
&\leq C_\# e^{-n(\ln\alpha_\lambda - \lambda\beta)} m(A_{n,I}),
\end{aligned}
$$

where we have chosen m large but fixed. The above computations imply that, for each $L > 0$,

$$
m(A_{n,I}) \geq C_L e^{-J_L(I)n}
$$

where $J_L(I) = \max_{\{\lambda \leq L \,:\, \mu_\lambda(f) \in I_c\}} \lambda a - \ln\alpha_\lambda$. Note that, if f is not a coboundary and hence $\ln\alpha_\lambda$ is strictly convex, the maximum of $\lambda a - \ln\alpha_\lambda$ is attained at some finite value, hence, for L large enough, $J_L(I) = \sup_{\{\lambda\in\mathbb{R} \,:\, \mu_\lambda(f)\in I_c\}} \lambda a - \ln\alpha_\lambda$. This implies that

$$
m(A^+_{a,n}) \geq C_\# e^{-J(a)n}
$$

where $J(a) = \sup_{\{\lambda \,:\, \mu_\lambda(f) > a\}} \lambda a - \ln\alpha_\lambda$.

The surprising fact is that the upper and lower bound are essentially the same. To see this a little argument is needed.

7.5.3 Large deviations. Conclusions

In fact, it is possible to give a variational characterization of the rate function in the spirit of general Large deviation theory [Var84, Str84, DZ98].

Lemma 7.5.5 *Let $\mathcal{M}_T$ be the set of invariant probability measures invariant with respect to T. Then*

$$
I(a) = -\sup_{\{\nu\in\mathcal{M}_T \,:\, \nu(f)\geq a\}} h_\nu(T) = J(a).
$$

Proof. By section 7.4.2 we have that, for each $\nu \in \mathcal{M}_T$, $\ln\alpha_\lambda = \sup_{\nu\in\mathcal{M}_T}\{h_\nu(T) + \lambda\nu(f)\} = h_{\mu_\lambda}(T) + \lambda\mu_\lambda(f)$. Thus for each $\nu \in \mathcal{M}_T$ such that $\nu(f) \geq a$, we can write

$$
I(a) \leq \max_{\lambda\geq 0}\{\lambda(a - \nu(f)) - h_\nu(T)\} = -h_\nu(T).
$$

On the other and

$$I(a) = \sup_{\lambda \geq 0}\{\lambda(a - \mu_\lambda(f)) - h_{\mu_\lambda}(T)\}.$$

If $a > \sup \mu_\lambda(f)$, then $I(a) = +\infty$, otherwise let λ_* be such that $\mu_{\lambda_*}(f) = a$,[13] then

$$I(a) \geq -h_{\mu_{\lambda_*}}(T) \geq -\sup_{\{\nu \in \mathcal{M}_T \,:\, \nu(f) \geq a\}} h_\nu(T).$$

Finally, since μ_λ and h_{μ_λ} depend smoothly from λ,

$$J(a) = \sup_{\{\lambda \,:\, \mu_\lambda(f) > a\}} \lambda a - \lambda \mu_\lambda(f) - h_{\mu_\lambda}(T) = I(a).$$

□

7.5.4 The Central Limit Theorem

We can now address the second question we have posed. From the above discussion is clear that we must chose $c_n = \sqrt{n}$.

Let $f \in BV$ and set $\hat{f} := f - \mu(f)$, then

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

Let us set $\Psi_n := \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \hat{f} \circ T^k$. We can consider Ψ_n a random variable with distribution $F_n(t) := \mu(\{x \,:\, \Psi_n(x) \leq t\})$. It is well known that, for each continuous function g holds[14]

$$\mu(g(\Psi_n)) = \int_{\mathbb{R}} g(t) dF_n(t)$$

[13] Actually one must show that the sup is a max.

[14] If $g \in \mathcal{C}_0^1$, then

$$\int_{\mathbb{R}} g dF_n = -\int_{\mathbb{R}} F_n(t) g'(t) dt = -\int_{\mathbb{R}} dt \int_{\mathbb{T}^1} dx \chi_{\{z \,:\, \Psi_n(z) \leq t\}}(x) g'(t).$$

Applying Fubini yields

$$\int_{\mathbb{R}} g dF_n = -\int_{\mathbb{T}^1} dx \int_{\mathbb{R}} dt \chi_{\{z \,:\, \Psi_n(z) \leq t\}}(x) g'(t) = -\int_{\mathbb{T}^1} dx \int_{\Psi_n(x)}^{\infty} g'(t) dt = \int_{\mathbb{T}^1} dx g(\Psi_n(x)).$$

where the integral is a Riemann-Stieltjes integral. It is thus clear that if we can control the distribution F_n, we have a very sharp understanding of the probability to have small deviations (of order $\sqrt{n}$) from the limit. From the work in the previous section it follows that there exists $\delta > 0$ such that, for each $|\lambda| \leq \delta\sqrt{n}$,

$$\begin{aligned}\varphi_n(\lambda) := \mu(e^{i\lambda\Psi_n}) = \mu(\mathcal{L}^n_{i\lambda/\sqrt{n}}h) &= (1 - \frac{\sigma^2\lambda^2}{2n} + \mathcal{O}(\lambda^3 n^{-\frac{3}{2}} + \rho^n)\|f\|_{BV})^n \\ &= e^{-\frac{\sigma^2\lambda^2}{2}}(1 + \mathcal{O}(\lambda^3 n^{-\frac{1}{2}} + n\rho^n)\|f\|_{BV}).\end{aligned} \tag{7.5.24}$$

The above quantity is called *characteristic function* of the random variable and determines the distribution (at continuity points) via the formula

$$F_n(b) - F_n(a) = \lim_{\Lambda\to\infty} \frac{1}{2\pi}\int_{-\Lambda}^{\Lambda} \frac{e^{-ia\lambda} - e^{-ib\lambda}}{i\lambda}\varphi_n(\lambda)d\lambda,$$

as can be seen in any basic book of probability theory.[15]

Formula (7.5.24) means in particular that

$$\lim_{n\to\infty} m(e^{\lambda\Psi_n}) = e^{-\frac{\sigma^2\lambda^2}{2}} =: \varphi(\lambda).$$

What can we infer from the above facts? First of all a simple computation shows that

$$g(t) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-it\lambda}\varphi(\lambda)d\lambda = \frac{1}{\sqrt{\pi}\sigma}e^{-\frac{t^2}{2\sigma^2}}$$

a random variable with such a density is called a Gaussian random variable with zero average and variance σ. Accordingly, formula (7.5.24) can be interpreted by saying that there exists a Gaussian random variable G such that

$$\frac{1}{n}\sum_{k=0}^{n-1}\hat{f}\circ T^k \sim \frac{1}{\sqrt{n}}G(1 + \mathcal{O}(n^{-\frac{1}{2}}))$$

[15] In the case when there exists a density, that is an L^1 function f_n such that $F_n(b) - F_n(a) = \int_a^b f_n(t)dt$, then the formula above becomes simply

$$f_n(t) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-it\lambda}\varphi_n(\lambda)d\lambda,$$

and follows trivially by the inversion of the Fourier transform.

in distribution. But what does this means concretely. Actual estimates are made difficult by the fact that the distribution under study not necessarily have a density, thus we are Fourier transforming function that behave quite badly at infinity. To overcome such a problem we can smoothen the quantities involved.

Let $j \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}_+)$ such that $\int_{\mathbb{R}} j(t)dt = 1$, $j(t) = j(-t)$, and $j(t) = 0$ for all $|t| > 1$, for each $\varepsilon > 0$ defined then $j_\varepsilon(t) := \varepsilon^{-1} j(\varepsilon^{-1} t)$ and

$$F_{n,\varepsilon}(t) := \int_{\mathbb{R}} j_\varepsilon(t-s) F_n(s) ds. \tag{7.5.25}$$

A simple computation shows that, for each $a, b \in \mathbb{R}$, holds

$$F_n(b+\varepsilon) - F_n(a-\varepsilon) \geq F_{n,\varepsilon}(b) - F_{n,\varepsilon}(a) \geq F_n(b-\varepsilon) - F_n(a+\varepsilon)$$

that is: if the measurements have a precision worst than 2ε, then $F_{n,\varepsilon}$ is as good as F_n to describe the resulting statistics. On the other hand calling $\varphi_{n,\varepsilon}$ the characteristic function associated to $F_{n,\varepsilon}$, holds $\varphi_{n,\varepsilon}(\lambda) = \varphi_n(\lambda)\hat{j}(\varepsilon\lambda)$, where $\hat{j}$ is the Fourier transform of j. Since now $F_{n,\varepsilon}$ is the law of a smooth random variable it has a density $f_{n,\varepsilon}$ and

$$f_{n,\varepsilon}(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\lambda t} \varphi_n(\lambda) \hat{j}(\varepsilon\lambda) d\lambda$$

since j is smooth it follows that there exists $C > 0$ such that $|\hat{j}(\lambda)| \leq C(1+\lambda^2)^{-2}$. We can finally use formula (7.5.24) to obtain a quantitative estimate

$$\begin{aligned}
f_{n,\varepsilon}(t) &= \frac{1}{2\pi} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} e^{-i\lambda t} \varphi_n(\lambda) \hat{j}(\varepsilon\lambda) d\lambda + \mathcal{O}(\varepsilon^{-5} n^{-\frac{3}{2}}) \\
&= \frac{1}{2\pi} \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} e^{-i\lambda t} \varphi(\lambda) \hat{j}(\varepsilon\lambda) d\lambda + \mathcal{O}(\varepsilon^{-5} n^{-\frac{3}{2}} + n^{-\frac{1}{2}}) \\
&= g(t) + \mathcal{O}(\varepsilon + \varepsilon^{-5} n^{-\frac{3}{2}} + n^{-\frac{1}{2}}) = g(t) + \mathcal{O}(n^{-\frac{1}{2}})
\end{aligned}$$

provided we choose $n^{-\frac{1}{2}} \geq \varepsilon \geq n^{-5}$. Which, as announced, means that, if the precision of the instrument is compatible with the statistics, the typical fluctuations in measurements are of order $\frac{1}{\sqrt{n}}$ and Gaussian. This is well known by sperimentalists who routinely assume that the result of a measurement is distributed according to a Gaussian.[16]

[16]Note however that our proof holds in a very special case that has little to do

7.6 Perturbation theory

To answer the questions posed at the beginning we need some perturbation theorems. Few such results are available (e.g., see [Kif88], [BY93] or [Bal00] for a review), here we will follow mainly the theory developed in [KL99, GL06] adapted to the special cases at hand.

For simplicity let us work directly with the densities and in the case $d = 1$. Then $\mathcal{L}$ is the transfer operator for the densities. We will start by considering an abstract family of operators $\mathcal{L}_\varepsilon$ satisfying the following properties.

Condition 1 *Consider a family of operators $\mathcal{L}_\varepsilon$ with the following properties*

1. *A uniform Lasota-Yorke inequality:*

$$\|\mathcal{L}_\varepsilon^n h\|_{BV} \leq A\lambda^{-n}\|h\|_{BV} + B|h|_{L^1}, \quad |\mathcal{L}_\varepsilon^n h|_{L^1} \leq C|h|_{L^1} \ ;$$

2. $\int \mathcal{L}h(x)dx = \int h(x)dx$ *;*

3. *For $L : BV \to BV$ define the norm*

$$|||L||| := \sup_{\|h\|_{BV}\leq 1} |Lf|_{L^1},$$

that is the norm of L as an operator from $BV \to L^1$. Then we require that there exists $D > 0$ such that

$$|||\mathcal{L} - \mathcal{L}_\varepsilon||| \leq D\varepsilon.$$

Condition 1-(3) specifies in which sense the family $\mathcal{L}_\varepsilon$ can be considered an approximation of the unperturbed operator $\mathcal{L}$. Notice that the condition is rather weak, in particular the distance between $\mathcal{L}_\varepsilon$ and $\mathcal{L}$ as operators on BV can be always larger than 1. Such a notion of closeness is completely inadequate to apply standard perturbation theory, to get some perturbations results it is then necessary to drastically

with a real experimental setting. To prove the analogous statement for a realistic experiment is a completely different ball game.

restrict the type of perturbations allowed, this is done by Conditions 1-(1,2) which state that all the approximating operators enjoys properties very similar to the limiting one.[17]

To state a precise result consider, for each operator L, the set

$$V_{\delta,r}(L) := \{z \in \mathbb{C} \mid |z| \leq r \text{ or } \operatorname{dist}(z, \sigma(L)) \leq \delta\}.$$

Since the complement of $V_{\delta,r}(L)$ belongs to the resolvent of L it follows that

$$H_{\delta,r}(L) := \sup \left\{ \|(z-L)^{-1}\|_{BV} \mid z \in \mathbb{C} \setminus V_{\delta,r}(L) \right\} < \infty.$$

By $R(z)$ and $R_\varepsilon(z)$ we will mean respectively $(z-\mathcal{L})^{-1}$ and $(z-\mathcal{L}_\varepsilon)^{-1}$.

Theorem 7.6.1 ([KL99]) *Consider a family of operators $\mathcal{L}_\varepsilon : BV \to BV$ satisfying Conditions 1. Let $H_{\delta,r} := H_{\delta,r}(\mathcal{L})$; $V_{\delta,r} := V_{\delta,r}(\mathcal{L})$, $r > \lambda^{-1}$, $\delta > 0$, then, if $\varepsilon \leq \varepsilon_1(\mathcal{L}, r, \delta)$, $\sigma(\mathcal{L}_\varepsilon) \subset V_{\delta,r}(\mathcal{L})$. In addition, if $\varepsilon \leq \varepsilon_0(\mathcal{L}, r, \delta)$, there exists $a > 0$ such that, for each $z \notin V_{\delta,r}$, holds true*

$$|||R(z) - R_\varepsilon(z)||| \leq C\varepsilon^a.$$

Proof.[18] To start with we collect some trivial, but very useful algebraic identities.

[17] Actually only Condition 1-(1) is needed in the following. Condition 1-(2) simply implies that the eigenvalue one is common to all the operators. If 1-(2) is not assumed, then the operator $\mathcal{L}_\varepsilon$ will always have one eigenvalue close to one, but the spectral radius could vary slightly, see [LMD03] for such a situation.

[18] This proof is simpler than the one in [KL99], yet it gives worst bounds, although sufficient for the present purposes.

For each operator $L : BV \to BV$ and $n \in \mathbb{Z}$ holds

$$\frac{1}{z}\sum_{i=0}^{n-1}(z^{-1}L)^i(z-L) + (z^{-1}L)^n = \mathbb{1} \tag{7.6.26}$$

$$R(z)(z-\mathcal{L}_\varepsilon) + \frac{1}{z}\sum_{i=0}^{n-1}(z^{-1}\mathcal{L})^i(\mathcal{L}_\varepsilon - \mathcal{L}) + R(z)(z^{-1}\mathcal{L})^n(\mathcal{L}_\varepsilon - \mathcal{L}) = \mathbb{1} \tag{7.6.27}$$

$$(z-\mathcal{L}_\varepsilon)\left[G_{n,\varepsilon} + (z^{-1}\mathcal{L}_\varepsilon)^n R(z)\right] = \mathbb{1} - (z^{-1}\mathcal{L}_\varepsilon)^n(\mathcal{L}_\varepsilon - \mathcal{L})R(z) \tag{7.6.28}$$

$$\left[G_{n,\varepsilon} + (z^{-1}\mathcal{L}_\varepsilon)^n R(z)\right](z-\mathcal{L}_\varepsilon) = \mathbb{1} - (z^{-1}\mathcal{L}_\varepsilon)^n R(z)(\mathcal{L}_\varepsilon - \mathcal{L}), \tag{7.6.29}$$

where we have set $G_{n,\varepsilon} := \frac{1}{z}\sum_{i=0}^{n-1}(z^{-1}\mathcal{L}_\varepsilon)^i$.

Let us start applying the above formulae. For each $h \in BV$ and $z \notin V_{r,\delta}$ holds

$$\begin{aligned}\|(z^{-1}\mathcal{L}_\varepsilon)^n(\mathcal{L}_\varepsilon - \mathcal{L})R(z)h\|_{BV} &\le (r\lambda)^{-n}A\|(\mathcal{L}_\varepsilon - \mathcal{L})R(z)h\|_{BV} + \frac{B}{r^n}|(\mathcal{L}_\varepsilon - \mathcal{L})R(z)h|_{L^1}\\ &\le [(r\lambda)^{-n}A2C_1 + Br^{-n}D\varepsilon]H_{r,\delta}\|h\|_{BV} < \|h\|_{BV}\end{aligned}$$

Thus $\|(z^{-1}\mathcal{L}_\varepsilon)^n(\mathcal{L}_\varepsilon - \mathcal{L})R(z)\|_{BV} < 1$ and the operator on the right hand side of (7.6.28) can be inverted by the usual Neumann series. Accordingly, $(z-\mathcal{L}_\varepsilon)$ has a well defined right inverse. Analogously,

$$\|(z^{-1}\mathcal{L}_\varepsilon)^n R(z)(\mathcal{L}_\varepsilon - \mathcal{L})h\|_{BV} \le (r\lambda)^{-n}A\|R(z)(\mathcal{L}_\varepsilon - \mathcal{L})h\|_{BV} + Br^{-n}|R(z)(\mathcal{L}_\varepsilon - \mathcal{L})h|_{L^1}.$$

This time to continue we need some informations on the L^1 norm of the resolvent. Let $g \in BV$, then equation (7.6.26) yields

$$\begin{aligned}|R(z)g|_{L^1} &\le \frac{1}{r}\sum_{i=0}^{n-1}|(z^{-1}\mathcal{L})^i g|_{L^1} + \|R(z)(z^{-1}\mathcal{L})^n g\|_{BV}\\ &\le \frac{1}{r^n(1-r)}|g|_{L^1} + H_{\delta,r}A(r\lambda)^{-n}\|g\|_{BV} + H_{\delta,r}Br^{-n}|g|_{L^1}\\ &\le r^{-n}(H_{\delta,r}B + (1-r)^{-1})|g|_{L^1} + H_{\delta,r}A(r\lambda)^{-n}\|g\|_{BV}\end{aligned}$$

Substituting, we have

$$\begin{aligned}&\|(z^{-1}\mathcal{L}_\varepsilon)^n R(z)(\mathcal{L}_\varepsilon - \mathcal{L})h\|_{BV} \le \{(r\lambda)^{-n}AH_{\delta,r}2C_1[1 + Br^{-n}]\\ &+ Br^{-2n}[H_{\delta,r}B + (1-r)^{-1}]D\varepsilon\}\|h\|_{BV} < 1,\end{aligned}$$

again, provided ε is small enough and choosing n appropriately. Hence the operator on the right hand side of (7.6.29) can be inverted, thereby providing a left inverse for $(z-\mathcal{L}_\varepsilon)$. This implies that z does not belong to the spectrum of $\mathcal{L}_\varepsilon$.

To investigate the second statement note that (7.6.27) implies

$$R(z)-R_\varepsilon(z) = \frac{1}{z}\sum_{i=0}^{n-1}(z^{-1}\mathcal{L})^i(\mathcal{L}_\varepsilon-\mathcal{L})R_\varepsilon(z)-R(z)(z^{-1}\mathcal{L})^n(\mathcal{L}_\varepsilon-\mathcal{L})R_\varepsilon(z).$$

Accordingly, for each $\varphi\in BV$ holds

$$|R(z)\varphi-R_\varepsilon(z)\varphi|_{L^1} \leq \{r^{-n}(1-r)^{-1}\varepsilon+H_{\delta,r}(\lambda r)^{-n}2AC_1+H_{\delta,r}B\varepsilon\}\|R_\varepsilon(z)\varphi\|_{BV}.$$

□

7.6.1 Deterministic stability

The $\mathcal{L}_\varepsilon$ are Perron-Frobenius (Transfer) operators of maps T_ε which are $\mathcal{C}^1$–close to T, that is $d_{\mathcal{C}^1}(T_\varepsilon, T)=\varepsilon$ and such that $d_{\mathcal{C}^2}(T_\varepsilon,T)\leq M$, for some fixed $M>0$. In this case the uniform Lasota-Yorke inequality is trivial. On the other hand, for all $\varphi\in\mathcal{C}^1$ holds

$$\int(\mathcal{L}_\varepsilon f-\mathcal{L}f)\varphi = \int f(\varphi\circ T_\varepsilon - \varphi\circ T).$$

Now let $\Phi(x):=(D_xT)^{-1}\int_{Tx}^{T_\varepsilon x}\varphi(z)dz$, since

$$\Phi'(x) = -(D_xT)^{-1}D_x^2T\Phi(x)+D_xT_\varepsilon(D_xT)^{-1}\varphi(T_\varepsilon x)-\varphi(Tx)$$

follows

$$\int(\mathcal{L}_\varepsilon f-\mathcal{L}f)\varphi = \int f\Phi'+\int f(x)[(D_xT)^{-1}D_x^2T\Phi(x)+(1-D_xT_\varepsilon(D_xT)^{-1})\varphi(T_\varepsilon x)].$$

Given that $|\Phi|_\infty\leq\lambda^{-1}\varepsilon|\varphi|_\infty$ and $|1-D_xT_\varepsilon(D_xT)^{-1}|_\infty\leq\lambda^{-1}\varepsilon$, we have

$$\int(\mathcal{L}_\varepsilon f-\mathcal{L}f)\varphi \leq \|f\|_{BV}\lambda^{-1}|\varphi|_\infty\varepsilon+|f|_{L^1}\lambda^{-1}(B+1)\varepsilon|\varphi|_\infty \leq D\|f\|_{BV}\varepsilon|\varphi|_\infty.$$

By Lebesgue dominate convergence theorem we obtain the above inequality for each $\varphi \in L^\infty$, and taking the sup on such φ yields the wanted inequality.

$$|\mathcal{L}_\varepsilon f - \mathcal{L} f|_{L^1} \le D\|f\|_{BV}\varepsilon.$$

We have thus seen that all the requirements in Condition 1 are satisfied. See [Kel82] for a more general setting including piecewise smooth maps.

7.6.2 Stochastic stability

Next consider a set of maps $\{T_\omega\}$ depending on a parameter $\omega \in \Omega$. In addition assume that Ω is a probability space and consider a measure P on Ω. Consider the process $x_n = T_{\omega_n} \circ \cdots \circ T_{\omega_1} x_0$ where the ω are i.i.d. random variables distributed accordingly to P and let E_μ be the expectation of such process when x_0 is distributed according to μ. Then, calling $\mathcal{L}_\omega$ the transfer operator associated to T_ω, we have

$$E(f(x_{n+1}) \mid x_n) = \mathcal{L}_P f(x_n) := \int_\Omega \mathcal{L}_\omega f(x_n) P(d\omega).$$

Then if

$$|\mathcal{L}_\omega h|_{BV} \le \lambda_\omega^{-1}|h|_{BV} + B_\omega |h|_{L^1}$$

integrating yields

$$|\mathcal{L}_P h|_{BV} \le E(\lambda_\omega^{-1})|h|_{BV} + E(B_\omega)|h|_{L^1}$$

And the operator $\mathcal{L}_P$ satisfy a Lasota-Yorke inequality provided that $E(\lambda^{-1}) < 1$ and $E(B) < \infty$.

In addition, if for some map T and associated transfer operator $\mathcal{L}$,

$$E(|\mathcal{L}_\omega h - \mathcal{L} h|) \le \varepsilon |h|_{BV}$$

then we can apply perturbation theory and obtain stochastic stability.

7.6.3 Computability

If we want to compute the invariant measure and the rate of decay of correlations, we can use the operator P_t defined in (7.3.6) and define $\mathcal{L}_{t,m} = P_t \mathcal{L}^m$. By the estimates in Lemma **??** it follows

$$|\mathcal{L}_{t,m} h|_{BV} \le 4^d \sigma^m |h|_{BV} + B|h|_{L^1}.$$

We can then chose the smallest m so that $4^d\sigma^m = \sigma_1 < 1$. Moreover, we also saw that

$$|\mathcal{L}_{t,m}h - \mathcal{L}h| \le t^{-1}|h|_{BV}.$$

So we are again in the realm of our perturbation theory and we have that the finite dimensional operator $\mathcal{L}_{t,m}$ has spectrum close to the one of the transfer operator. We can then obtain all the info we want by diagonalizing a matrix.

7.6.4 Linear response

Linear response is a theory widely used by physicists. In essence it says the follow: consider a one parameter family of systems T_s and the associated (e.g.) invariant measures μ_s, then, for a given observable f one want to study the response of the system to a small change in s, and, not surprisingly, one expects $\mu_s(f) = \mu_0(f) + s\nu(f) + o(s)$. That is one expects differentiability in s. Yet differentiability is is not ensured by Theorem 7.6.1. Is it possible to ensure conditions under which linear response holds? The answer is yes (for example if holds if the maps are sufficiently smooth and the dependence on the parameter is also smooth in an appropriate sense). To prove it one need a sophistication of Theorem 7.6.1 that can be found in [GL06].

7.6.5 The hyperbolic case

One can wonder is the previous approach can be applied to uniformly hyperbolic systems and partially hyperbolic system. The answer is yes although the work in this direction is still in progress and the price to pay is the need to consider rather unusual functional spaces (space of anysotropic distributions). Just to give a vague idea let us look at a totally trivial example: toral automorphisms.

Then one can consider the norms:

$$\|f\|_{p,q} := \sum_{k\in\mathbb{Z}^{2d}\setminus\{0\}} |f_k| \frac{|k|^p}{1 + |\langle v^s, k\rangle|^{p+q}} + |f_0|,$$

where f_k are the Fourier coefficients of f and v^s is the unit vector in the stable direction. Then

$$\begin{aligned} &\|[\mathcal{L}f\|_{p,q} \le C_1\|f\|_{p,q}, \\ &\|[\mathcal{L}^n f\|_{p,q} \le C_3\mu^n\|f\|_{p,q} + B\|f\|_{p-1,q+1}. \end{aligned} \tag{7.6.30}$$

we have thus the Lasota-Yorke inequality. Moreover on can easily check the relative compactness of $\{\|f\|_{p,q} \leq 1\}$ with respect to the topology induced by the norm $\|\cdot\|_{p-1,q+1}$, hence our previous theory applies almost verbatim.

To have a more precise idea of what can be done, see [GL06, BT07].

Hints to solving the Problems

7.18 Let ℓ_λ, h_λ be analytic. Let us define $z_\lambda = e^{-\int_0^\lambda \ell_\xi(h'_\xi)d\xi}$, define $\hat{h}_\lambda = z_\lambda h_\lambda$ and $\hat{\ell}_\lambda = z_\lambda^{-1}\ell_\lambda$ and check that they are normalized as required.

Notes

Large deviations are taken from Lai-Sang article and Keller book.

The stochastic stability is reasonably well understood (Cowienson) but what about the smooth dependence from a parameter (linear response)? Counterexamples in $d = 1$ but unknown in higher dimensions. The uniformly hyperbolic case is well understood but not much is know on how to apply the present ideas to the partially hyperbolic case and to the case of systems with discontinuities, although a concentrated effort is taking place to extend the theory in such directions.