Exponential mixing : Lectures from Mumbai

Mark Pollicott

Abstract

We discuss a number of results related to mixing and decay of correlations.

Contents

1 Introduction

The aim of these notes is to describe the rates of mixing for various types of hyperbolic systems, and some applications. In the case of diffeomorphisms exponential mixing is classical. For certain well known examples flows, such as geodesic flows, this is a result of Dolgopyat and Liverani, but significantly harder to prove. However, the hypotheses required for the proofs of these results are very special and the generality of the method is far from being completely understood.

We begin with four definitions of “mix” taken from Google:

1. Combine together

2. Associate with others socially
3. Combine soundtracks into one
4. Be belligerent physically or verbally

In the mathematical sense, mixing is probably closer in spirit to the first definition. However, for dynamical systems we might naturally consider two cases: The discrete case and the continuous case. We first consider both settings in a fairly general sense.

1. Discrete transformations $T : X \to X$ where
   - X is a compact metric space;
   - T is a continuous map;
   - μ is a T-invariant probability measure $\mu(X) = 1$ (i.e., $B \subset X$ is a Borel set and $\mu(B) = \mu(T^{-1}B)$)

   We usually denote $T^n = \underbrace{T \circ \cdots \circ T}_{\times n}$ for $n \geq 0$.

   This allows us to think of this as a $\mathbb{Z}_+$-acton: $\mathbb{Z}_+ \times X \to X$ given by $(n, x) \mapsto T^n x$.

2. Continuous (semi-) flows $\phi_t : Y \to Y$ for $t \in \mathbb{R}$ (or $t \in \mathbb{R}$)
   - Y is a compact metric space;
   - ϕ_t is family of continuous maps such that ϕ_0 is the identity, and $\phi_{s+t} = \phi_s \circ \phi_t$ with $s, t \in \mathbb{R}^+$;
   - μ is a ϕ-invariant probability measure $\mu(Y) = 1$ (i.e., $B \subset X$ is a Borel set and $\mu(B) = \mu(\phi_t^{-1}B)$)

Having introduced the natural setting(s) we can now turn to the dynamical property we want to study. We now recall the definitions of (strong) mixing.

Definition 1.1. *We say μ is mixing for the discrete transformation $T : X \to X$ if for all $f, g \in L^2(X, \mu)$ the correlation function*

$$\rho(n) := \int f \circ T^n g d\mu - \int f d\mu \int g d\mu, n \geq 0$$

satisfies $\rho(n) \to 0$ as $n \to +\infty$.

Similarly, we have an analogous definition for semi-flows.

Definition 1.2. *We say μ is mixing for the (semi)-flow $\phi_t : Y \to Y$ if for all $f, g \in L^2(Y, m)$ the correlation function*

$$\rho(t) := \int f \circ \phi_t g d\mu - \int f d\mu \int g d\mu, t \geq 0$$

satisfies $\rho(t) \to 0$ as $t \to +\infty$.

There is a classical hierarchy of ergodic properties of such systems, with ergodicity the weakest and Bernoulli and K-automorphisms the strongest.

As is well known mixing is somewhere in the middle of the list: being a stronger assumption than ergodicity, i.e., in the discrete case for any function $F \in L^1(X, \mu)$ we have that

$$\frac{1}{N}\sum_{n=0}^{N-1} F(T^n x) \to \int F d\mu, \text{ as } N \to +\infty,$$

and in the continuous case for any function $F \in L^1(Y, \mu)$ we have that

$$\frac{1}{T}\int_0^T F(\phi_t x) \to \int F d\mu, \text{ as } N \to +\infty.$$

2 Discrete case

We begin with a simple example of a one dimensional discrete transformation.

A rather trivial concrete example is the doubling map.

Example 2.1 (Doubling map). *Let $n = 2$ and $x_1 = \frac{1}{2}$. We then specify the map by*

$$T(x) = \begin{cases} 2x & \textit{for } x \in [0, \frac{1}{2}] \\ 2x - 1 & \textit{for } x \in [\frac{1}{2}, 1]. \end{cases}$$

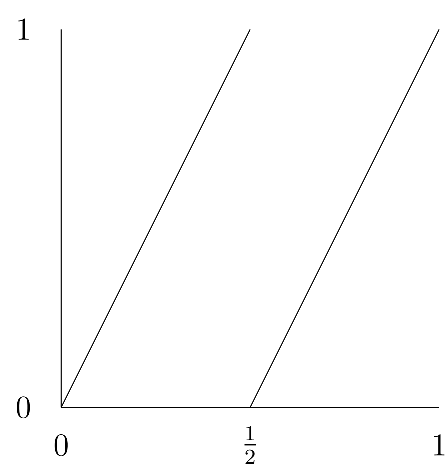

Figure 1: The graph of the doubling map on the unit interval

This transformation preserves the usual Lebesgue measure μ on the unit interval. A little exercise with trigonometric polynomials shows that the correlation function for Lipschitz functions tends to zero like $\rho(n) = O\left(\frac{1}{2^n}\right)$.

We can consider a more general formulation. Given a finite set of points

$$x_0 = 0 < x_1 < \cdots < x_{n-1} < x_n = 1$$

we get a partition of the unit interval , i.e., $[0,1] = \cup_{i=0}^{n-1}[x_i, x_{i+1}]$. We can define a map T on the disjoint union of such intervals by specifying it on each of the intervals $[x_i, x_{i+1}]$. Let us denote by

$$X = \coprod_{i=0}^{n-1}[x_i, x_{i+1}]$$

a disjoint union of the interval and assume $T : X \to X$ satisfies

1. $T : [x_i, x_{i+1}] \to [0,1]$ is C^∞
2. There exists $\beta > 1$ so that for any $x \in X$ we have $|T'(x)| \geq \beta$ (Expanding property)
3. Each image $T([x_i, x_{i+1}])$, for $i = 0, \cdots, n-1$ is a union of intervals from the partition (Markov property);
4. T has a dense orbit (Transitive property).

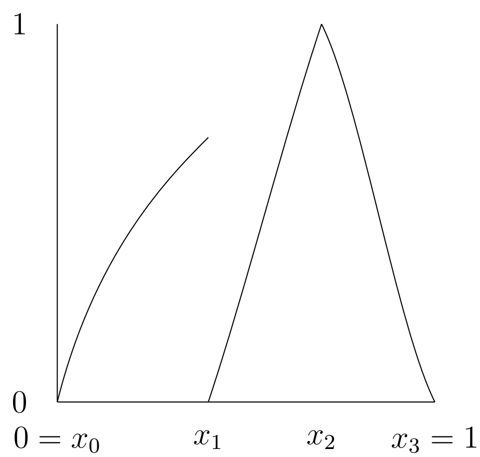

Figure 2: The graph of an expanding Markov map of the unit interval

For such transformations there is a natural T-invariant measure. Moreover, there is an exponential decay of correlations, providing the functions are sufficiently regular.

We begin by recalling the result that shows that there is a natural T-invariant measure.

Theorem 2.2 (Folklore Theorem). *Under the above hypotheses (1)-(4) above there exists a (unique) T-invariant probability measure μ absolutely continuous with respect to Lebesgue measure.*

In this case it is rather difficult to trace Theorem 2.2 back to its origins (hence the name "Folklore"). In a posthumous paper of Bowen [4] from 1979 the Theorem is attributed to Adler. However, in an appendix to Bowen's paper by Adler, he claims that he actually heard it from Flatto and, moreover, he heard it from Benjy Weiss, who in turn claimed it was easily derived from a paper of Sinai from 1968.

The next result shows that we have an example of fast(er) mixing (under suitable regularity assumptions on $f, g : X \to \mathbb{R}$).

Theorem 2.3. *For an interval map $T : X \to X$ as above we have exponential mixing, i.e., There exists $0 < \lambda < 1$ such that for $f, g \in C^\infty(X)$ there exists $B > 0$ such that $|\rho(n)| \leq B\lambda^n$ for $n \geq 0$.*

The usual proof of this result uses operator theory.

3 The Transfer Operator

The proof of Theorem 2.3 is also reassuringly simple and introduces a rather useful and versatile tool: The transfer operator.

Sketch proof of Theorem 2.3. Since μ is an absolutely continuous probability measure we can consider the Radon-Nikodym derivative $\frac{d\mu}{dx} = \rho$. We can then denote $\omega(x) = \frac{\rho(x)}{|T'(x)|\rho(Tx)}$ and observe that

$$\sum_{Ty=x} \omega(y) = 1$$

(which is an easy exercise using the change of variable formula, cf. [5] for a nice account of this).

Let $C^1(X)$ denote the space of C^1 functions on X with the norm

$$\|f\| := \|f\|_\infty + \|f'\|_\infty \text{ for } f \in C^1(X).$$

In particular, $C^1(X)$ is a Banach space with the norm $\|\cdot\|$. We can now associate a linear operator $\mathcal{L} : C^1(X) \to C^1(X)$ where we write

$$\mathcal{L}f(x) = \sum_{Ty=x} \omega(y)f(y).$$

This is what we call a *transfer operator*.

The good news is that we have the useful identity

$$\rho(n) = \int f \circ T^n g d\mu = \int f(\mathcal{L}^n g) d\mu, \text{for} \quad n \geq 0,$$

which relates the transfer operator to the correlation function. The even better news is that $\mathcal{L}$ has the useful properties:

1. $\mathcal{L}\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the constant function taking the value 1; and
2. the operator $\mathcal{L} : C^1(X)/\mathbb{C} \to C^1(X)/\mathbb{C}$ on the quotient space $C^1(X)/\mathbb{C}$ (with associated norm $\|\cdot\|_{C^1(X)/\mathbb{C}}$) has spectral radius λ, say, which is strictly smaller than 1.

Thus we can combine these simple results to deduce that

$$|\rho(n)| \leq \| \int f(\mathcal{L}^n g)\|_\infty \|f\|_\infty \|\mathcal{L}^n\|_{C^1(X)/\mathbb{C}} \|g\|_{C^1} = O\left((\lambda + \epsilon)^n\right)$$

for any $\epsilon > 0$ (by using the spectral radius theorem, which gives that $\lambda = \lim_{n \to +\infty} \|\mathcal{L}^n\|_{C^1(X)/\mathbb{C}}^{\frac{1}{n}}$). $\square$

Remark 3.1. Of course, at the heart of the proof of Theorem 2.3 are the properties 1 and 2. To show the "spectral gap" for $\mathcal{L}$ is pure functional analysis. It is also easy using two basic ingredients:

a) The unit ball in $C^1(X)$ is compact in the C^0 topology (i.e., the classical Arzela-Ascoli Theorem); and

b) There exists $C > 0$ and $0 < \alpha < 1$ such that for all $n \geq 0$ and all $f \in C^1(X)$:

$$\|\mathcal{L}^n f\|_{C^1} \leq C\|f\|_\infty + \alpha^n \|f\|_{C^1}$$

(by the chain rule for differentiation).

The inequality in b) is one version of a result which has been (re)-discovered many times. It began as the Doeblin-Fortet inequality from 1937 [6], reappeared as the Ionescu-Tulcea-Marinescu inequality from 1942 [10], and finally appeared as the Lasota-Yorke inequality from 1975 [12]. However, for brevity, we will refer to it as the Doeblin-Fortet inequality.

Vincent Doeblin (1915-1940) was a talented young French-German mathematician (son of the famous novelist Alfred Doeblin) who died during the German invasion of France.

Remark 3.2. The transfer operator could have made an earlier appearance in these notes in the proof of Theorem 2.2 on the existence of the absolutely continuous T-invariant measure μ (see [5]). In particular, μ is a fixed point for the dual transfer operator, i.e., $\mathcal{L}^*\mu = \mu$.

Another explicit (almost) example is the continued fraction transformation.

Example 3.3 (Continued Fraction Transformation)**.** *We can consider a partition of* $[0, 1]$ *into countably many intervals* $[\frac{1}{n+1}, \frac{1}{n}]$*, plus the extra point* 0*. The Continued fraction transformation (or Gauss map)* $T : [0, 1] \to [0, 1]$ *is defined by*

$$T(x) = \begin{cases} \frac{1}{x} - n & \textit{if } \frac{1}{n+1} \leq x \leq \frac{1}{n} \\ 0 & \textit{if } x = 0. \end{cases}$$

This has an explicit T-invariant probability measure μ:

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{dx}{1 + x}, \qquad \textit{whenever } B \subset [0, 1] \textit{ is a Borel set.}$$

called the Gauss measure.

If we consider sufficiently smooth functions f then again the correlation funcion $\rho(n)$ tends to zero with a bound $O(\alpha^n)$ for some $0 < \alpha < 1$, i.e., there is exponential decay of correlations. However, the value of $\alpha = 0 \cdot 303663\ldots$ in not known explicitly, but up to a high degree of numerical precision, starting from the work of Wirsing. Earlier, it was only established by Kuzmin and Lévy that there was a slower subexponential (in fact, stretched exponential) decay. This particular problem has its roots in the famous correspondence of Gauss and Laplace from 1812.

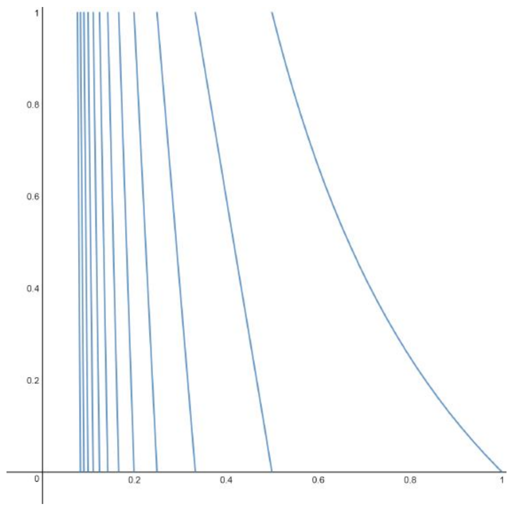

Figure 3: The graph of the Gauss map on the unit interval

Remark 3.4. There are well known results of a similar flavour for diffeomorphisms. We can assume that $T : M \to M$ is a transitive Anosov diffeomorphism and μ is a T-invariant absolutely continuous probability measure (if it exists, and the SRB measure if it doesn't). If we let $f, g : M \to \mathbb{R}$ be smooth functions then again the corresponding correlation function $\rho(n)$ tends to zero exponentially fast. Moreover, the standard approach to proving this is (at least morally) to go via the case of expanding maps. The bridge between the two is the use of Markov Partitions for M.

4 Continuous case

Let us now move onto the case of flows, rather than getting too bogged down in the discrete case. We begin with the most famous example of a continuous transformation: the Geodesic Flow.

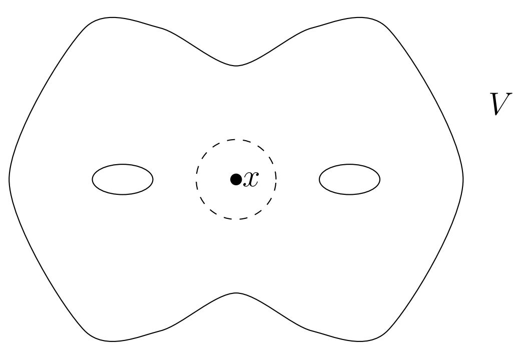

Figure 4: A surface V of negative curvature.

Let V be a compact oriented C^∞ surface with $\kappa(x) < 0$ where κ is the Gaussian curvature given by

$$\kappa(x) = \lim_{r \to 0} \frac{\pi r^2 - \mathrm{Vol}(B(x,r))}{\pi r^4/24}.$$

The geodesic flow takes place not on the surface V, but on the three dimensional unit tangent bundle.

Definition 4.1. *Let $SV = \{v \in TV : \|v\| = 1\}$ be the unit tangent bundle. We define the geodesic flow $\phi_t : SV \to SV$ by associating to $v \in SV$ the geodesic $\gamma_v : \mathbb{R} \to V$ with $\gamma_v'(o) = v$. We then define $\phi_t(v) = \dot\gamma_v(0)$.*

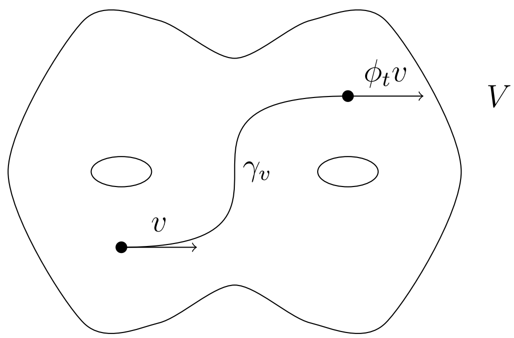

Figure 5: The geodesic flow on the surface V

The first basic result on geodesic flows is the following.

Lemma 4.2. *There is a (unique) ϕ-invariant probability measure m equivalent to volume (called the Liouville measure). Moreover, m is mixing.*

This result is harder to prove than in the discrete case. However, even more challenging is the following result.

Question 4.3. *Given C^∞ functions $F, G : SV \to \mathbb{R}$, how fast does the correlation function $\rho(t) \to 0$ as $t \to +\infty$?*

One of the first settings where the question was shown to have an affirmative answer (with an exponentially fast decay) was in the special case that V has constant curvature.

Example 4.4 (Constant curvature geodesic flows)**.** *If we assume $\kappa = -1$ then $\phi_t : SV \to SV$ mixes exponentially quickly (i.e., there exists $\alpha > 0$ such that for any f, g there exists $C > 0$ with $|\rho(t)| \le Ce^{-\alpha t}$, $t \ge 0$).*

In this setting this can be seen by writing $V = \mathbb{D}^2/\Gamma$ where

$$\mathbb{D}^2 = \{z = x + iy \in \mathbb{C} : |z| < 1\}$$

with Riemann metric

$$ds^2 = 4\frac{dx^2 + dy^2}{1 - (x^2 + y^2)}$$

Figure 6: The Poincaré disk (tessellated by hyperbolic triangles).

and covering transformations

$$\Gamma < \mathrm{Isom}_0(\mathbb{D}^2) = PSL(2, \mathbb{R}).$$

Using unitary representation theory one can decompose $L^2(SV) = \oplus_\alpha H_\alpha$ *and write*

$$\rho_{f,g}(n) = \sum_\alpha \rho_{f_\alpha, g_\alpha}(t).$$

One can show each correlation function $\rho_{f_\alpha, g_\alpha}(t) \to 0$ *with explicit bounds. (This is usually referred to as Decay of Matrix Coefficients). The regularity of the function* $f, g : SV \to \mathbb{R}$ *guarantees the the exponential decay for the terms translates into exponential decay for the summation.*

Remark 4.5. One of the interesting consequences of this approach is that one gets explicit information on α, which is related to the smallest non-zero eigenvalue of the Laplacian on V.

5 Exponential Decay of correlations for flows

5.1 Overview

This brings us to a central question.

Question 5.1. *What happens in the more general case that* V *has variable curvature?*

Unfortunately, there is no representation theory available in variable curvature, but we can try to use the transfer operators that served so well in the discrete case. We need to replace the flow by an invertible hyperbolic map, the expanding map by a semi-flow and the single transfer operator by a family of transfer operators. This is how Dolgopyat proved exponential decay of correlations [7].

Theorem 5.2 (Dolgopyat)**.** *If $\phi_t : SV \to SV$ is the geodesic flow for a compact surface of (variable) negative curvature, m is the Liouville probability measure, and $f, g \in C^\infty(SV)$ then the flow mixes exponentially fast.*

Remark 5.3. In contrast to the case of constant negative curvature, using this method there are typically no explicit estimates on the speed of mixing $\alpha > 0$.

Remark 5.4. Even after more than 20 years the method of Dolgopyat is essentially the only approach to proving exponential mixing for (semi-)flows beyond the setting of constant negative curvature geodesic flows. However, it has been extended to a number of settings:

1. m can be generalized to Gibbs measures for Hölder potentials [7];

2. V can be extended to higher dimensions (with negative sectional curvatures) with the Liouville measure [13]; However, in higher dimensions for other measures the problem of exponential mixing is still fairly open (although for polynomial mixing it is known to be true).

3. The result can be extended to contact Anosov flows, Teichmüller flows, etc. [1].

Remark 5.5. There is a connection with geodesic flows (for $\kappa = -1$) and expanding maps of the interval and the work of Bowen–Series. The surface V gives rise to a fundamental domain F and the geodesics which make up the sides of F extend to give points $\{x_i\}$ on the boundary (i.e., the unit circle). These points partition the boundary into arcs $[x_i, x_{i+1}]$, say. We then define a transformation T on the disjoint union of these arcs by $T|[x_i, x_{i+1}] = g$, which is one of the Möbius maps occurring as a side pairing. There is some ambiguity in the choices, but this is resolved by using some consistent choice.

We define $r : X := \coprod_i [x_i, x_{i+1}] \to \mathbb{R}$ to represent the time a geodesic takes to cross F.

Example 5.6 (Classical(non)-example: Modular surface)**.** *Let $V = \mathbb{H}^2/PSL(2,\mathbb{Z})$ be the modular surface. This is not compact, but the associated surface is non-compact. The underlying dynamics in this case is the classical continued fraction transformation $T(x) = 1/x - [1/x]$. The associated function can be taken to be $r(x) = -\log|T'(x)| = -2\log x$*

The basic approach is to use the mixing for a simpler model and then translate this into a result for the geodesic flow.

5.2 The simplified model: Suspension semi-flow

Let $T : X \to X$ be a piecewise $C^{1+\alpha}$ expanding Markov map of the interval and let $r : X \to \mathbb{R}$ be a piecewise C^1 function. We then define

$$Y = \{(x, u) \in X \times \mathbb{R} : 0 \le u \le r(x)\}/(x, r(x)) \sim (Tx, 0)$$

(i.e., where we identify points at the top of the graph with certain points at the bottom of the graph). We also define the suspension semi-flow $\phi_t : Y \to Y$ (locally) by $\phi_t(x, u) = (x, u+t)$, subject to the identification.

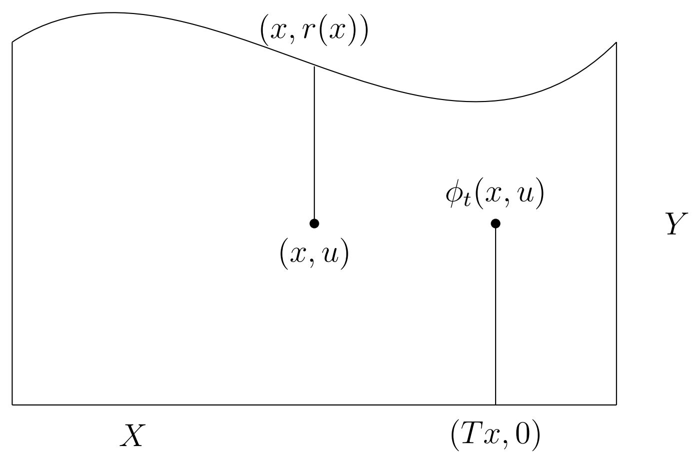

Figure 7: The suspension semi-flow (on the area under the graph).

Finally, given the T-invariant probability μ we can define a ϕ-invariant probability measure m defined by

$$dm = \frac{d\mu \times dt}{\int d\mu}$$

where the denominator is there to give the correct normalization.

Given $F, G : Y \to \mathbb{R}$ we recall that

$$\rho(t) = \int F \circ \phi_t G dm - \int F dm \int G dm.$$

To prove exponential mixing, we want to apply the Paley-Wiener Theorem [16].

Consider the Laplace transform

$$\widehat{\rho}(s) = \int_0^\infty e^{-st} \rho(t) dt$$

which is easily seen to converge to an analytic function on $Re(s) > 0$.

Theorem 5.7 (Paley-Wiener)**.** *If $\widehat{\rho}(s)$ extends analytically to $R(s) > -\epsilon_0$ ($\epsilon_0 > 0$) and there is an L^1-condition on the restrictions $t \mapsto \widehat{\rho}(\sigma + it)$ then for any $0 < \epsilon < \epsilon_0$ we have that there exists $D > 0$ such that*

$$|\rho(t)| \leq De^{-\epsilon t}, \quad t \geq 0.$$

The theorem is probably more familiar in the context of Fourier transforms, but it is is more convenient for us to formulate it this way.

Remark 5.8. R. Paley (1907-1933) was an English mathematician who died aged 26 while skiing near Banff, and is buried in the graveyard there.

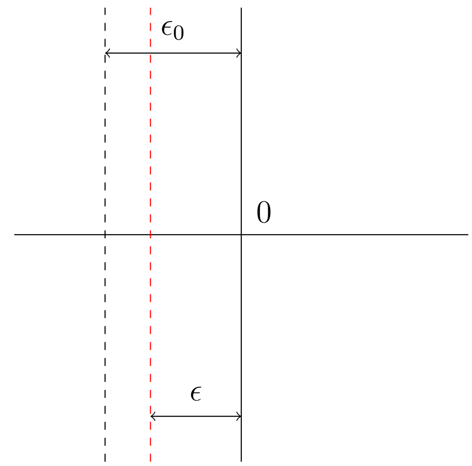

Figure 8: The domain of the Laplace transform $\widehat{\rho}(s)$.

We need to relate the suspension semi-flow to the Paley-Wiener Theorem.

Fix $s \in \mathbb{C}$. For the suspension semi-flow we first associate to suitable $F, G : Y \to \mathbb{R}$ functions $f_s, g_s : X \to \mathbb{R}$ defined by

$$f_s(x) = \int_0^{r(x)} F(x,u)e^{-su}du$$

and

$$g_s(x) = \int_0^{r(x)} G(x,u)e^{-su}du.$$

Assume without loss of generality, and for greater convenience, that $\int F dm = \int G dm = 0$.

The next part of the argument is very well explained in the paper of Naud [14]. However, to work in greater generality one needs to follow the rest of the argument from the articles of Baladi—Vallée [3] when X is one dimensional, or of Avila—Gouëzel—Yoccoz [1] when X is higher dimensional. We need some condition on r to ensure (exponential) mixing.

Remark 5.9. For the (non)-example of a semi-flow where r takes a constant value the semi-flow is not even mixing.

Consider the (formal) identity

$$F(\phi_t(x,u)) = \sum_{n=0}^{\infty} \int_0^{r(T^n x)} F(T^n x, v)\delta(u + t - v - r^n(x))dv$$

where $\delta(\cdot)$ represents the Dirac delta function on the real line and

$$r^n(x) = \sum_{j=0}^{n-1} r(T^j x) \text{ for } n \geq 1.$$

We can then rewrite

$$\begin{aligned}
\widehat{\rho}(s) &= \int_0^\infty e^{-st}\rho(t)dt \\
&= \int_0^\infty e^{-st}\left(\int_Y F(\phi_t(x,u))G(x,u)\frac{d\mu(x)du}{\int r d\mu}\right)dt \\
&= \frac{1}{\int r d\mu}\sum_{n=0}^\infty \int_X e^{-sr^n(x)}\left(\int_0^{r(x)} e^{-sr(x)}F(x,u)du\right)\left(\int_0^{r(T^n x)} e^{sr(x)}G(T^n x,u)du\right)d\mu(x) \\
&= \frac{1}{\int r d\mu}\sum_{n=0}^\infty \int_X e^{-sr^n(x)} f_s(x) g_{-s}(T^n x) d\mu(x).
\end{aligned}$$

Whereas for expanding interval maps we used a single transfer operator, for suspension semi-flows we need to consider a family of transfer operators $\mathcal{L}_s$, in light of the need to accommodate the dependence on $s \in \mathbb{C}$.

Definition 5.10. *For each $s \in \mathbb{C}$ we associate a linear operator $\mathcal{L}_s : C^1(X) \to C^1(X)$ defined by*

$$\mathcal{L}_s f(x) = \sum_{Ty=x} e^{-sr(y)} f(y).$$

Thus $\mathcal{L}_s f(x) = \mathcal{L}(e^{-sr}f)$ and we are looking at a *family of complex operators.* Moreover, we can write the expression for $\widehat{\rho}(s)$ more succinctly as:

$$\begin{aligned}
\widehat{\rho}(s) &= \frac{1}{\int r d\mu}\sum_{n=0}^\infty \int \left(\mathcal{L}(f_s g_{-s}\circ T^n e^{-sr^n})\right)d\mu(x) \\
&= \frac{1}{\int r d\mu}\int g_{-s}(1-\mathcal{L}_s)^{-1} f_s d\mu
\end{aligned} \tag{5.1}$$

Thus the analytic extension of $\widehat{\rho}(s)$ will come from the spectral properties of the transfer operator.

Lemma 5.11 ("Claim 1")**.** *There exists $\epsilon > 0$ and $|Im(s)|$ sufficiently small such that the spectral radius of $\mathcal{L}_s$ is strictly smaller than 1, i.e.,*

$$\operatorname{spr}\left(\mathcal{L}_s : C^1([0,1]) \to C^1([0,1])\right) < 1$$

Establishing this result is the core of the proof. Then we have the following chain of deductions:

(I) Establishing uniform bounds on the spectral radius of $\mathcal{L}_s$ (Lemma 5.11);

(II) gives analytic extension of the complex function $\widehat{\rho}(s)$ (by equation (5.1)); and

(III) then applying the Paley-Wiener theorem (Theorem 5.7) to get estimates on the rates of mixing.

5.3 A couple of preliminaries

We next need to establish some basic properties of transfer operators. Let $\sigma > 0$. We begin with a classical result in the case of a real positive transfer operator due to Ruelle.

Theorem 5.12 (Ruelle Operator theorem). *Consider $\mathcal{L}_\sigma : C^1(X) \to C^1(X)$, for $\sigma \in \mathbb{R}$.*

1. *There exists a simple positive eigenvalue $\lambda_\sigma > 0$ of maximum modulus;*
2. *The associated eigenfunction h_σ is strictly positive; and*
3. *The associated eigenmeasure ν_σ is positive.*

In particular, we can write

$$\mathcal{L}_\sigma h_\sigma = \lambda_\sigma h_\sigma \in C^1(X)$$

and

$$\mathcal{L}_\sigma^* \nu_\sigma = \lambda_\sigma \nu_\sigma \in C^1(X).$$

Remark 5.13. This is completely analogous to the Perron-Frobenius theorem for positive matrices, and for this reason the result is sometimes called the Ruelle-Perron-Frobenius Theorem.

As we have seen above, in order to establish the necessary properties of $\mathcal{L}_s$ we need to consider the more general case of a complex transfer operator, and in particular establish Lemma 5.11.

Definition 5.14. *Let $s = \sigma + it$. It is convenient to make a simplification by first replacing $\mathcal{L}_s$ by $\widetilde{\mathcal{L}}_s : C^1(X) \to C^1(X)$ where*

$$\widetilde{\mathcal{L}}_s(f) = \frac{1}{\lambda_\sigma}\frac{1}{h_\sigma}\mathcal{L}(h_\sigma f).$$

In particular, $\widetilde{\mathcal{L}}_\sigma 1 = 1$ and we can think of $\widetilde{\mathcal{L}}_\sigma$ as being a normalized version of $\mathcal{L}_\sigma$.

For the next ingredient in establishing Lemma 5.11 we need some bounds on iterates of the operator(s). One can compare the second inequality below for a family of operators with the, perhaps, better known version for a single operator in the sketch proof for interval maps in Remark 3.1.

Lemma 5.15 (Doeblin-Fortet Lemma (for families of operators)). *Let $\sigma_0 < \sigma_1$. There exists $C > 0$ and $0 < \theta < 1$ such that for all $s = \sigma + it$ with $\sigma_0 < \sigma < \sigma_1$ and $t \in \mathbb{R}$:*

1. *$\|\widetilde{\mathcal{L}}_s^n f\|_\infty \le \|f\|_\infty$ (uniform contraction), for $n \ge 1$; and*
2. *$\|(\widetilde{\mathcal{L}}_s^n f)'\|_\infty \le C(|t|\|f\|_\infty + \theta^n \|f'\|_\infty)$, for $n \ge 1$.*

The first inequality is a direct consequence of the simplification above. The second inequality follows by simple calculus and the product rule for differentiation.

It simplifies the notation a little if we make the norm depend on the value $|t|$.

Modifying the norm. Let $s = \sigma + it$. We can change the norm on $C^1(X)$ to be

$$\|f\|_t \colon = \|f\|_\infty + \frac{1}{t}\|f'\|_\infty.$$

Remark 5.16. A simple consequence of the above version of the Doeblin-Fortet inequality is the following:

$$\begin{aligned}\|\widetilde{\mathcal{L}}_s^n f\|_t &= \|\widetilde{L}_s^n f\|_\infty + \frac{1}{|t|}\|(\widetilde{\mathcal{L}}_s^n f)'\|_\infty \\ &\le \|f\|_\infty + \frac{1}{|t|}(C|t|\|f\|_\infty + \theta^n\|f'\|_\infty) \\ &\le (C+1)\|f\|_\infty + \frac{\|f'\|_\infty}{|t|} \\ &\le (C+1)\|f\|_t,\end{aligned}$$

where we use the notation $\|f\|_t$ as introduced above.

A slightly stronger (and more applicable) version of Lemma 5.11 is the following.

Lemma 5.17 ("Claim 2")**.** *There exists $C > 0$, $\beta > 0$ such that for all σ sufficiently close to 1 and all $|t|$ sufficiently large we have*

$$\|\widetilde{\mathcal{L}}_s^{n(t)} f\|_t \le \frac{\|f\|_t}{|t|^\beta}, \forall f \in C^1(X)$$

where $n(t) := [C \log |t|]$.

This fundamental result is due to Dolgopyat [7].

Remark 5.18. To see that Claim 2 implies Claim 1 (i.e, Lemma 5.17 implies Lemma 5.11) we proceed as follows. The spectral radius of $\mathcal{L}_s$ satisfies

$$\begin{aligned}\limsup_{n\to+\infty} \|\mathcal{L}_s^n\|_t^{1/n} &\le \lambda_\sigma \limsup_{n\to+\infty} \|\widetilde{\mathcal{L}}_s^n\|_t^{1/n} \\ &\le \lambda_\sigma \limsup_{n\to+\infty} \|\widetilde{\mathcal{L}}_s^{n[C\log|t|]}\|_t^{1/(n[C\log|t|])} \\ &\le \lambda_\sigma \exp(-\beta/C) < 1\end{aligned}$$

provided σ is close to 1, by perturbation theory.

Moreover, Claim 2 (Lemma 5.17) follows from the following (apparently weaker) lemma describing contraction of the operator in the L^2-norm.

Lemma 5.19 ("Claim 3")**.** *There exists $C > 0$ and $\exists \beta > 0$ such that $\forall |t|$ sufficiently large and σ sufficiently close to 1:*

$$\left(\int_X \left|\widetilde{\mathcal{L}}_s^{n(t)} f\right|^2 d\mu_\sigma\right)^{1/2} \le \frac{1}{|t|^\beta}\|f\|_t, \quad \text{for } f \in C^1(X),$$

where $n(t) = [C \log |t|]$.

Continuing to work through the implications of these claims, let us next explain why Claim 3 (Lemma 5.19) implies Claim 2 (Lemma 5.17).

Proof (that Claim 3 implies Claim 2). There are 2 steps.

Step 1 (uniform contraction). Assuming Claim 3 we can first show a bound on the supremum norm of the operator:

$$\Big\| \underbrace{\widetilde{\mathcal{L}}_s^{2n} f}_{\widetilde{\mathcal{L}}_s^n(\widetilde{\mathcal{L}}_s^n f)} \Big\|_\infty \le \int \left| \widetilde{L}_s^n f \right| d\mu_\sigma + O\left(\alpha^n \|\widetilde{\mathcal{L}}_s^n f\|_t\right) \qquad \text{(using the spectral gap)} \tag{1}$$

$$\le \underbrace{\left(\int \left| \widetilde{\mathcal{L}}_s^n f \right|^2 d\mu_\sigma \right)^{1/2}}_{\text{Cauchy—Schwarz}} + \underbrace{O\left(\alpha^n \|f\|_t\right)}_{\text{Doeblin—Fortet}} \tag{2}$$

$$\le \underbrace{\left(\frac{1}{|t|^\beta} \|f\|_t \right)}_{\text{(by assumption)}} + \underbrace{O\left(\frac{1}{|t|} C |\log \alpha| \|f\|_t \right)}_{\text{since } n=[C|\log t|]} \tag{3}$$

By choosing $C > \frac{\beta}{|\log \alpha|}$ and β slightly smaller we have this is bounded above by $\frac{\|f\|_t}{|t|^\beta}$. This completes the sketch of the proof of Step 1. $\square$

Step 2 (norm contraction). It remains to bound the derivative of the operator acting on functions. The additional hypothesis needed comes from the non-integrability of the horocycle foliations

$$\frac{1}{|t|} \Big\| \underbrace{(\mathcal{L}_s^{2n} f)'}_{=\mathcal{L}_s^n(\mathcal{L}_s^n f)} \Big\|_\infty \le C \cdot \|\mathcal{L}_s f\|_\infty + \frac{\theta^n}{|t|} \|(\mathcal{L}_s f)'\|_\infty \tag{4}$$

(by using the Doeblin—Fortet inequality)

$$\le \frac{C}{|t|^\beta} \|f\|_t + \frac{\theta^n}{|t|} (C|t| + 1) \|f\|_t \tag{5}$$

(by using Step 1, and the Corollary to the Doeblin—Fortet inequality)

$$\le \frac{\|f\|_t}{|t|^\beta} + \frac{\|f\|_t}{|t|^{1 + C|\log \theta|}} \tag{6}$$

(since $n = [C \log |t|]$).

By choosing β slightly smaller, if necessary, we can assume that $C > \frac{\beta - 1}{|\log \theta|}$ and we can choose $|t|$ larger, if necessary, so that we can assume the we can bound (6) by $\frac{1}{|t|^\beta} R \|f\|_t$. This completes the sketch of the proof of Step 2. $\square$

Step 3 (L^2 contraction) It remains to prove (or at least, in our case, to sketch the proof of) claim 3 (Lemma 5.19). In order to do this we now need to use some additional hypothesis (which is implied by the geodesic flow).

Additional hypothesis: There does not exist a function $u \colon X \to \mathbb{R}$ such that

1. u is C^1 on each interval $[x_i, x_{i+1}]$
2. $\psi\colon = r + \underbrace{u \circ T - u}_{\text{coboundary}}$ is constant on each $[x_i, x_{i+1}]$.

This hypothesis is simple to state and is equivalent to what is *actually* used in the proof. We will actually make use of the following:

Technical version of hypothesis: There exists $\varepsilon > 0$ such that $T\colon X \to X$ has inverse branches $T_i, T_j\colon X \to X$ (i.e. locally $T \circ T_i =$ identity and $T \circ T_j =$ identity) and $R_{ij}\colon = r(T_i x) - r(T_j x)$ satisfies $|R'_{ij}(x)| \geq \varepsilon$. [1]

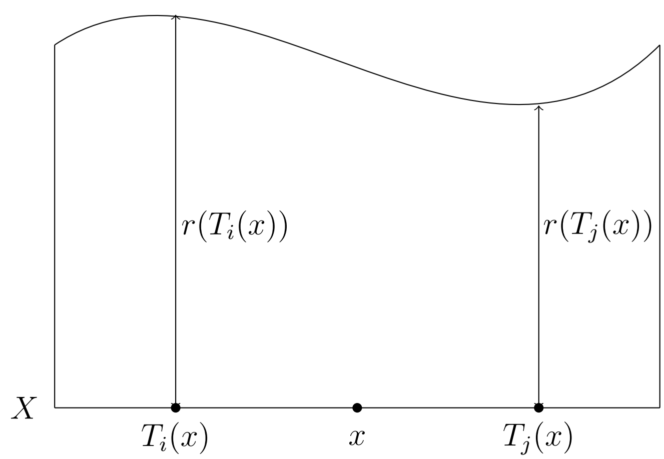

Figure 9: The difference between the values of r evaluated at the pre-images $T_i(x)$ and $T_j(x)$ of x under T determine the (locally defined) function $R_{ij}(x)$

The basic idea of the proof of Claim 3 (Lemma 5.19) if the following. We can write that

$$\mathcal{L}_s f(x) = \sum_{y\colon Ty=x} \left(\omega(y) f(y) e^{-\sigma r(y)}\right) e^{-itr(y)}$$

and thus taking the absolute value we have that

$$|\mathcal{L}_s f(x)| = \left| \sum_{y\colon Ty=x} \left(f(y) e^{-\sigma r(y)}\right) e^{-itr(y)} \right|.$$

However, in Claim 3 (Lemma 5.19) it is the integral in x of this function which we need to bound. But, by the technical hypothesis, for every x we have that the contribution of

[1] Actually, one needs a slightly weaker assumption: There exists $\varepsilon > 0$ and (an arbitrary large) n such that $T^n\colon X \to X$ has inverse branches $T_i, T_j\colon X \to X$ (i.e. locally $T^n \circ T_i =$ identity and $T^n \circ T_j =$ identity) and $R_{ij}\colon = r(T_i x) - r(T_j x)$ satisfies $|R'_{ij}(x)| \geq \varepsilon$.

two pre-images $y = T_i x$ and $y' = T_j x$, say, satisfy that $|R'_{ij}(x)| \geq \varepsilon$. In particular, for any given sufficiently large $|t|$ there is a fixed proportion of the μ_σ measure of X for which locally $|t|R_{ij}(\cdot)$ lies in an interval $[\pi/2, 3\pi/2]$ (mod 1), say.

Comparing the arguments of the terms in the summation we see that $|\mathcal{L}_s^n f(x)|$ is uniformly smaller than 1 on a fixed proportion of the measure of X. Moreover, these bounds can be made uniform in $|t|$.

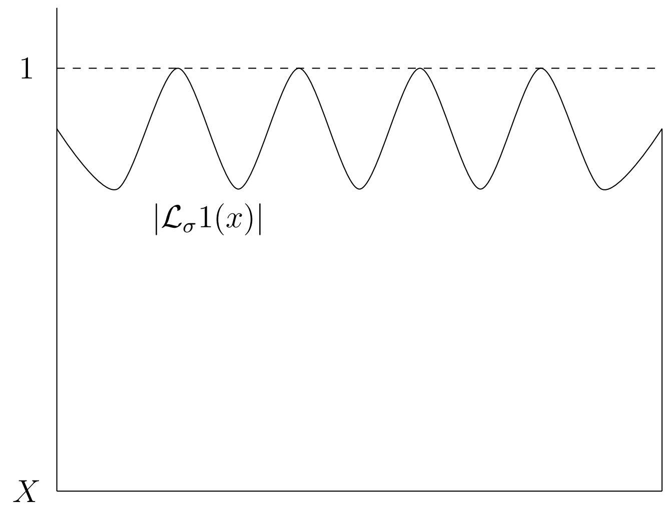

Figure 10: The absolute value of $\mathcal{L}_\sigma 1(x)$ is uniformly smaller than 1 for sets whose μ_σ measure is uniformly bounded away from 0 (independent of t, for $|t|$ sufficiently large)

(In particular, as $|t|$ increases the arguments in the summation change more rapidly but still the proportion of the measure of X for which $|t|R_{ij}(\cdot)$ lies in an interval $[\pi/2, 3\pi/2]$ (mod 1) is uniformly bounded from below).

In summary, by assumption at least two of the terms are "out of phase" since $|R'_{ij}(x)| \geq \varepsilon$. This leads to the bound on $\|\mathcal{L}_s^n\|_{L_1}$.

Remark 5.20. In the case of geodesic flows on n-dimensional manifolds with negative sectional curvature the proof is very similar. The main difference is in the proof of the analogue of Claim 3. The integral over the corresponding n-dimensional space X is approached by integrating over the one-dimensional leaves of a foliation of X. The argument that served us well in one dimension applies along the leaves of the foliation. This explained in the original article of Dolgopyat and in the related article of Avila-Gouëzel-Yoccoz, for Teichmüller flows.

This very simplified account of Dolgopyat's original approach makes use of the original formulation using Markov sections, which was still a standard approach at that time. However, it does suggest the following question:

Question 5.21. *Why can't we just work with the original space SV?*

The short answer is that one can, and that this approach was advanced by Liverani. It leads to slightly stronger results, although it may not be technically simpler. We give a brief description of the method, without entering into any of the technical details.

Consider the perhaps more natural operators:

$$U_t : L^2(SV, m) \to L^2(SV, m)$$
$$U_t : f \mapsto h \circ \varphi_t$$

for $t \geq 0$, say. As is well known, these operators have rather poor spectral properties on $L^2(SV, m)$. However Liverani (and Butterley) introduced a (bigger) space of distributions $\mathcal{B}$ and then defined a family of linear operators by:

$$\mathcal{R}(s) \colon \mathcal{B} \to \mathcal{B} \qquad \mathcal{R}(s)f = \int_0^\infty e^{-st} U_t f dt$$

For such Banach spaces we have the following properties:

1. The spectrum has no eigenvalues in a half plane $Re(s) > -\beta$, say, where β can be made arbitrarily large by choosing an appropriate $\mathcal{B}$; and
2. The Banach space $\mathcal{B}$ is based on functions in the stable direction and distributions in the unstable direction.

In particular, 1. leads one to the same conclusion on $\widehat{\rho}(s)$ that allow the application of the Paley-Wiener theorem. The underlying mechanism to prove these results is a useful reformulation of the approach of Dolgopyat. One might hope that there are many different techniques available, but unfortunately this doesn't appear to be the case.

Confession: The only two ways I know how to get exponential mixing for geodesic flows are:

1. Unitary representations ($\kappa = -1$); and
2. Dolgopyat method ($\kappa < 0$).

6 Other applications of transfer operators

Now that we (or, more precisely, Dolgopyat) has done the hard work — can we use the same machinery to prove other results? That is, can we use the properties established for families of transfer operators in other contexts? The following diagram helps to summarize some of the other applications:

$$\boxed{\text{Property of } \mathcal{L}_s} \implies \begin{cases} \text{Counting Problems} \\ \text{Teichmüller Flows} \\ \text{Skew Products} \\ \text{Algorithms} \\ \text{Exponential Mixing} \end{cases}$$

6.1 Problems that count (closed geodesics)

Let V be a C^∞ compact surface. Let γ denote a closed geodesic (there are a countable infinity of closed geodesics - one in each free homotopy class of V). We can then denote by $\ell(\gamma)$ the length of γ (where the lengths tend to infinity).

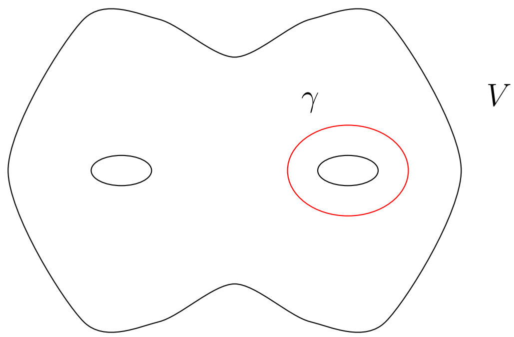

Figure 11: A closed geodesic on the surface V.

We want to consider the growth of the number of closed geodesics and begin with the following definition.

Definition 6.1. *For each $T > 0$, we let*

$$\pi(T) \colon = \#\{\gamma \colon \ell(\gamma) \le T\}$$

count the (finite) number of closed geodesics whose lengths are at most T.

The function $\pi(T)$ is clearly monotone increasing, and tends to infinity as T increases.
As we have already seen in the case of the mixing rates for $\rho(t)$, in the particular case of constant curvature the strongest results have been well known for many years.

Example 6.2 ($\kappa = -1$)**.** *Let V be a surface of constant curvature $\kappa = -1$. Recall that the isometries of $\mathbb{D}^2$ are isomorphic to $PSL(2,\mathbb{R})$ and that $\mathbb{D}^2$ is the Universal Cover of V. Thus we can write $V = PSL(2,\mathbb{R})/\Gamma$ where $\Gamma < PSL(2,\mathbb{R})$ is a discrete subgroup with $\Gamma = \pi_1(V)$. The closed geodesics γ correspond to conjugacy classes $\langle g \rangle$ of elements $g \in \Gamma - \{e\}$.*

In this specific setting the asymptotic formula is due to Huber [9]:

Theorem 6.3 (Huber : $\kappa = -1$)**.** *There is an asymptotic formula*

$$\pi(T) \sim \frac{e^T}{T}, \text{ as } T \to +\infty$$

(i.e., $\lim_{T\to+\infty} \frac{\pi(T)T}{e^T} = 1$). In fact there is an even stronger estimate

$$\pi(T) = \mathrm{li}(e^T)\left(1 + O(e^{-\varepsilon T})\right) \text{ as } T \to +\infty$$

for some $\varepsilon > 0$, where $\mathrm{li}(v) = \int_2^v \frac{du}{\log u}$ (i.e., there exists $C > 0$ with $|\pi(T) - \mathrm{li}(e^T)| \le Ce^{-\varepsilon T}$).

The method of proof makes use of the Selberg trace formula.

More generally, we can consider the case that V is a compact surface with a C^∞ Riemannian metric of variable negative curvature.

Theorem 6.4 (Margulis). *($\kappa < 0$): For surfaces of variable negative curvature there exists $h > 0$ such that*

$$\pi(T) \sim \frac{e^{hT}}{hT} \text{ as } T \to +\infty.$$

The proof of Margulis' more general result is dynamical, but it doesn't naturally lead to an error term. However, one can use the transfer operator estimates of Dolgopyat, which served so well in showing exponential mixing, to get these stronger results [15].

Theorem 6.5 (Pollicott-Sharp, after Dolgopyat). *For surfaces of variable negative curvature:*

$$\pi(T) = \mathrm{li}(e^T)(1 + O(e^{-\varepsilon T})), \text{ as } T \to +\infty,$$

for some $\varepsilon > 0$.

The proof parallels that of exponential mixing for geodesic flows. For this problem the associated analytic tool is now a zeta function (rather than the Laplace transform of the correlation function $\rho(t)$). In this setting one of the theorems which gets one from the properties of the analytic domain to the asymptotic formula for the counting function $\pi(T)$ is the Ikehara-Wiener Tauberian Theorem (famous for its use in the proof of the Prime Number Theorem).

This comparison is summarized in the following table.

	Object	Complex Function	Analytic Tool
Exponential mixing	$\rho(t)$	$\widehat{\rho}(s)$	Paley—Wiener
Counting problem	$\pi(T)$	$\underbrace{\zeta(s) = \prod \left(1 - e^{-s\ell(\gamma)}\right)^{-1}}_{\text{zeta function}}$	Ikehara—Wiener

6.2 Multiple mixing for geodesic flows

Let V again be a compact surface of negative curvature $\kappa < 0$. One can also consider the case of multiple mixing for the associated geodesic flow. In particular, we can consider (for simplicity) three smooth functions

$$F, G, H \colon SV \xrightarrow{C^\infty} \mathbb{C}$$

satisfying

$$\int F d\mu = \int G d\mu = \int H d\mu = 0.$$

Given the geodesic flow $g_t \colon SV \to SV$ we can define the multiple correlation function:

$$\rho(t_1, t_2) \colon = \int_{SV} F(g_{t_1+t_2}x) G(g_{t_2}x) H(x) dm \qquad (t_1, t_2 \geq 0)$$

where m is the normalized Liouville measure.

There is a natural analogue of the exponential mixing result of Dolgopyat (Theorem 5.2).

Theorem 6.6 (Multiple Exponential Mixing)**.** *There exists $\varepsilon > 0$ such that for F, G, H there exists $C > 0$ such that*

$$\rho(t_1, t_2) \leq C \exp(-\varepsilon \min\{t_1, t_2\}) \text{ as } t_1, t_2 \to +\infty.$$

The proof is a natural generalization of the proof of exponential decay of correlations in the case of two functions.

Sketch of Proof. We briefly outline the main steps in the proof.

Step 1: We first replace the geodesic flow $g_t \colon SV \to SV$ by a model suspension semi-flow $\varphi_t : Y \to Y$.

Step 2: We can write the Laplace transform of the correlation function (this time in two variables)

$$\widehat{\rho}(s_1, s_2) = \int_0^\infty \int_0^\infty e^{-(s_1 t_1 + s_2 t_2)} \rho(t_1, t_2) dt_1 dt_2 \qquad (s_1, s_2 \in \mathbb{C}),$$

for the suspension semi-flow $\varphi_t : Y \to Y$ in terms of transfer operators:

$$\widehat{\rho}(s_1, s_2) = \int_X f_{s_1} \left(1 - \mathcal{L}_{s_1}\right)^{-1} \left(g_{s_2} (1 - \mathcal{L}_{s_2})^{-1} h_{s_1+s_2}\right) d\mu,$$

where $f_s(x) = \int_0^{r(x)} f(x, u) e^{-su} du$, etc.

Step 3: We can use spectral properties of the transfer operator, in particular Lemma 5.11, to show $\widehat{\rho}(s_1, s_2)$ is analytic for $\Re(s_1), \Re(s_2) > \sigma_0$ for some $\sigma_0 < 0$.

Step 4: Finally, we can apply a version of the Paley-Wiener theorem for two variables to deduce the result. □

Remark 6.7. One would expect that for the Teichmüller geodesic flows the same argument would give multiple exponential mixing.

Remark 6.8. If we assume weaker conditions on the flow then we would anticipate weaker conditions on the behaviour of $\rho(t_1, t_2)$. For example, we consider more general Axiom A or hyperbolic flows for which the period of at least one of the closed orbits is a diophantine number (i.e., badly approximated as a real number). Although in this case we have weaker estimates on the spectra and norm of the family of transfer operators there is a corresponding result with polynomial bounds, i.e., there exists $\varepsilon > 0$ such that for F, G, H there exists $C > 0$ such that

$$\rho(t_1, t_2) \leq C \min\{t_1^{-\epsilon}, t_2^{-\epsilon}\} \text{ as } t_1, t_2 \to +\infty.$$

6.3 Skew Products

One can also use the uniform estimates on the transfer operators which have served us well for flows in a different setting. In the case of flows the estimates are needed to deal with the flow (or neutral) direction. In the case of skew products the neutral direction corresponds to the second component.

To illustrate this, let $T\colon X \to X$ be C^2 expanding interval and let μ be a T-invariant absolutely continuous probability measure. Given a C^1-function $\Theta\colon X \to \mathbb{R}/\mathbb{Z}$ valued on the unit circle, we can consider the skew product defined on the product space $X \times \mathbb{R}/\mathbb{Z}$ by

$$\begin{aligned} \widehat{T}\colon X \times \mathbb{R}/\mathbb{Z} &\to X \times \mathbb{R}/\mathbb{Z} \\ \widehat{T}(x,z) &= (T(x), z + \Theta(x)), \end{aligned}$$

where $x \in X$ and $z \in \mathbb{R}/\mathbb{Z}$.

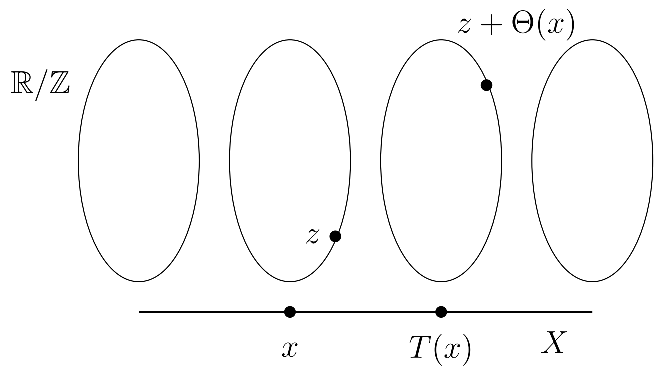

Figure 12: The skew product $\widehat{T}\colon X \times \mathbb{R}/\mathbb{Z}$

We can associate the natural $\widehat{T}$-invariant product measure

$$dm = d\mu \times d\,\mathrm{Haar}.$$

We can then introduce the correlation function for two C^∞ functions $F, G\colon X \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ such that $\int F dm = 0$ and $\int G dm = 0$.

Definition 6.9. *We define the correlation function by*

$$\rho(n) = \int F \circ \widehat{T}^n G d\,m - \int F d\,m \cdot \int G d\,m, \text{ for } n \geq 0.$$

We recall that we say that the measure m is *mixing* with respect to the skew product transformation $\widehat{T}$ if $\rho(n) \to 0$ as $n \to +\infty$. Clearly, we need some extra criteria for m to be even mixing. For example, choosing the function $\Theta \equiv 0$ to be identically zero is a poor choice, since the associated skew product cannot be mixing as any set $X \times B$, where $B \subset \mathbb{R}/\mathbb{Z}$ is Borel. The following is a standard result that guarantees that the measure m is mixing:

Lemma 6.10. *Assume that there do not exist non-trivial C^1 solutions $\chi\colon X \to \mathbb{C}$ to the equation*

$$\chi \circ T = e^{i\theta}\chi. \tag{6.1}$$

Then the measure m is strong mixing for the skew products.

The C^1 condition on the solution χ is not essential, but is equivalent to the existence of a solution $\chi \in L^\infty(\mu)$.

As before, we can consider a stronger property that $\rho(n)$ tends to zero exponentially fast.

Definition 6.11. *We say that the skew product and the measure m is exponential mixing if there exist constants $C > 0$ and $0 < \lambda < 1$ with $|\rho(n)| \le C\lambda^n$.*

We need a stronger hypothesis to guarantee that the skew-product is exponentially mixing.

Hypothesis 6.12. *Assume there does not exist a C^1 function $u\colon X \to \mathbb{C}$ such that $r + u \circ T - u$ is constant on each $[x_i, x_{i+1}]$.*

We call such a function $u \circ T - u$ a *coboundary.* In fact, the hypothesis above implies that the function r actually satisfies a stronger condition corresponding to that coming from the joint non-integrability condition for the geodesic flow. (This comes from the work of Avila-Gouëzel-Yoccoz [1].)

The next theorem guarantees the exponential mixing.

Theorem 6.13. *Under the above hypothesis on the skew product the measure m mixes exponentially fast.*

If we use the Fourier series expansion in the second co-ordinate to write

$$F(x,z) = \sum_{k\in\mathbb{Z}} a_k(x) e^{2\pi i k z} \text{and } G(x,z) = \sum_{l\in\mathbb{Z}} b_l(x) e^{2\pi i l z}$$

then substituting into the expression for $\rho(n)$, for $n \ge 0$, gives

$$\rho(n) = \sum_{k,l\in\mathbb{Z}} \int a_k(T^n x) e^{ik\Theta^n(x)} b_l(x) d\mu(x) = \sum_{k,l\in\mathbb{Z}} \int a_k(x) (\mathcal{L}^n_{ik\Theta} b_l)(x) d\mu(x)$$

where

$$\Theta^n(x) := \theta(x) + \Theta(Tx) + \cdots + \Theta(T^{n-1}x)$$

and we denote

$$\mathcal{L}_{i\Theta} w(x) = \sum_{Ty=x} g(y) e^{i\Theta(y)} w(y) \text{ for } w \in C^1(X)$$

where $\sum_{Ty=x} g(y) = 1$, with g corresponding to the measure μ.

The proof of exponential mixing comes from uniform estimates in $k \in \mathbb{Z}$ on families of transfer operators, coming from the Fourier modes in the $\mathbb{R}/\mathbb{Z}$. These, in turn, follow from a similar (and equally long) argument to that for suspension flows in §5.

Remark 6.14. The formulation of the discussion above for skew products over expanding maps is primarily for convenience. The same basic argument would work both for higher dimensional C^1 uniformly expanding base maps, and for hyperbolic invertible diffeomorphisms. In the latter case, there are additional complications whereby we need to reduce the diffeomorphism to an expanding map by a Markov partition. This requires the stable foliation to be C^1 and one also has to address the problem of changing the skewing function so that it is defined on the space for the (new) expanding map.

6.4 Skew Products and Flows

It is more challenging to deal with the issue of exponential mixing for skew products of (semi-)flows, which combines the features (and problems) of the last two cases. However, this becomes easier when one realises the correct formulation of the mechanism to use.

Consider the basic model of a suspension semi-flow $\phi_t : Y \to Y$ we considered before in §5 where Y is given by the graph of the function $r : X \to \mathbb{R}$. To this we can add the skewing function $\Theta : X \to \mathbb{R}/\mathbb{Z}$ and then define a (semi-)flow skew product by

$$\begin{aligned} \widehat{\phi}_t &: Y \times \mathbb{R}/\mathbb{Z} \to Y \times \mathbb{R}/\mathbb{Z} \\ \widehat{\phi}_t(x, u, z) &= (x, u + t, z) \end{aligned}$$

subject to the natural identification $(x, r(x), z) = (T(x), 0, z + \Theta(x))$.

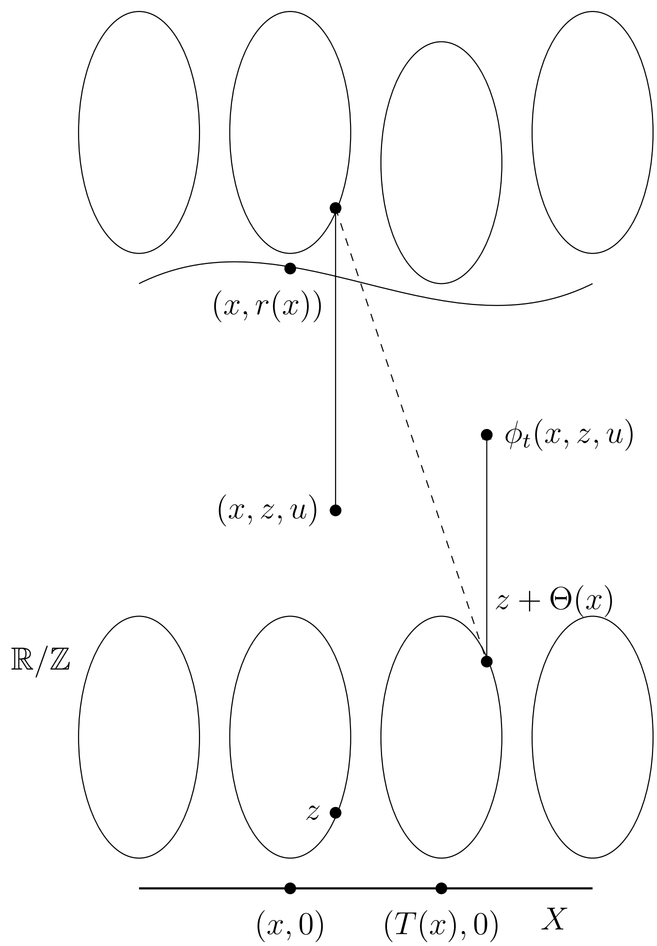

Figure 13: The skew product over the suspension semi-flow $\widehat{\phi}_t : Y \times \mathbb{R}/\mathbb{Z} \to Y \times \mathbb{R}/\mathbb{Z}$

Given C^1 functions $F, G : Y \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ we can write the correlation function as

$$\rho(t) = \int_{Y \times \mathbb{R}/\mathbb{Z}} F \circ \widehat{\phi}_t G d\mu(x) dt dz$$

As in the case of skew products, we can use the Fourier series expansion in the z coordinate to write

$$F(x,u,z) = \sum_{k\in\mathbb{Z}} a_k(x,u)e^{2\pi ikz} \text{and } G(x,u,z) = \sum_{l\in\mathbb{Z}} b_l(x,u)e^{2\pi ilz}$$

and then, as in the case of (semi-)flows in §5, the analysis of this correlation function reduces to the study of the Laplace transform of the individual terms (for each k and l):

$$\begin{aligned}\widehat{\rho}(s) &= \frac{1}{\int r d\mu} \sum_{n=0}^{\infty} \int \mathcal{L}^n(a_s b_{-s} \circ T^n e^{-sr^n} e^{ik\Theta^n}) d\mu(x) \\ &= \frac{1}{\int r d\mu} \int a_{-s}(1-\mathcal{L}_{s,\Theta})^{-1} b_s d\mu\end{aligned}$$

where :

(a) the functions $a_s, b_s : X \to \mathbb{R}$ are defined by

$$a_s(x) = \int_0^{r(x)} a_k(x,u)e^{-su}du$$

and

$$b_s(x) = \int_0^{r(x)} b_l(x,u)e^{-su}du;$$

and

(b) the operator $\mathcal{L}_{s,\Theta} : C^1(X,\mathbb{C}) \to C^1(X,\mathbb{C})$ is defined by

$$\mathcal{L}_{s,k}w(x) = \sum_{Ty=x} g(y)e^{-sr(y)}e^{i\Theta(y)}w(y)$$

for $w \in C^1(X,\mathbb{C})$.

In order to show that $\rho(t)$ tends to zero exponentially fast, we want to show that the spectral radius of $\mathcal{L}_{s,k}$ is strictly smaller than 1 uniformly. But now we need uniformity in two variables:

1. for $Re(s) > 1-\epsilon$ and $|Im(s)| > 1$, say; and
2. for $k \in \mathbb{Z}$

In order to deal with both variables at the same time we first restrict to the case that $\dim X \geq 2$. Moreover, we want to assume that there is a one-dimensional C^1 foliation $\{W_t\}_{t\in T}$ of local neighbourhoods of X given by level sets of the corresponding functions $\Theta_{ij}(x) := \Theta(T_i(x)) - \Theta(T_j(x))$. In particular, we require the following condition (using the notation from §5).

Hypothesis 6.15. *Along the leaves $\{W_t\}_{t\in T}$ we can assume that the derivative of $R_{ij}(x)$ is bounded away from zero.*

We can decompose the measure as $\mu_\sigma = \int_T \mu_{\sigma,t} d\overline{\mu_\sigma}$. In particular, we can replace the integral

$$\int_X |\widetilde{\mathcal{L}}_{s,k\Theta}^{n(t)} f|^2 d\mu_\sigma \text{ by } \int_T \left(\int_{W_t} |\widetilde{\mathcal{L}}_{s,k\Theta}^{n(t)} f|^2 d\mu_{\sigma,t}(u) \right) d\overline{\mu_\sigma}(t).$$

This was the approach used in the original article of Dolgopyat and in the article of Avila-Gouëzel-Yoccoz, corresponding to the choice $k = 0$. In the present context we have that the $L^2(\mu_t)$ bound on $|\widetilde{\mathcal{L}}_{s,k\Theta} f(x)|$ (or more generally on $|\widetilde{\mathcal{L}}_{s,k\Theta}^n f(x)|$) along the leaves W_t of the foliations contains contributions to the summation of the form

$$W_t \ni x \mapsto e^{-\sigma r(T_i x)} e^{-itr(T_i x)} e^{i\Theta(T_i x)} - e^{-\sigma r(T_i x)} e^{-itr(T_j x)} e^{j\Theta(T_j x)}.$$

However, on each leaf W_t the difference $\Theta(T_i x) - \Theta(T_j x)$ is constant and then the same argument as in the proof of *Step 3* applies, and thus the corresponding statement to *Claim 3* in §5 holds. Moreover, the bounds in *Claim 1* and *Claim 2* generalize to the present setting and thus we can deduce the analogous statement to Lemma 5.17 for the operators $\widetilde{\mathcal{L}}_{s,k\Theta}$. As in the case of semi-flows, one can now deduce exponential decay of correlations to the component functions in the expansions for F and G. Providing the individual coefficients a_k and a_l tend to zero sufficiently quickly (reflecting the regularity of the functions F and G) then this exponential mixing will hold for the original functions.

Remark 6.16. With a little extra work this method can be applied to frame flows. These are $SO(d-1)$-skew products over the geodesic flow $g_t : SV \to SV$ on the unit tangent bundle of a compact d-dimensional manifold. In particular, orthonormal d-frames are parallel transported along the geodesics defining the orbits of the geodesic flow. In particular, one can use this approach to show that the frame flow on a compact manifold with sectional curvatures close enough to -1 are exponentially mixing with respect to the natural invariant measure equivalent to volume. This again involves the reduction from the geodesic flow to a semi-flow. The condition for $\{W_t\}_{t\in T}$ that the derivative of R_{ij} is bounded away from zero can be checked explicitly in the case of metrics for which all of the sectional curvatures are equal to -1 and holds for nearby Riemannian metrics by continuity.

6.5 Euclidean Algorithm

Finally, we recall an application of the transfer operator results to algorithms. We begin with the following simple classical question on natural numbers.

Question 6.17. *Given two natural numbers $0 < u \leq v$ how do we find their greatest common divisor? (i.e. largest $\omega \in \mathbb{N}$ such that $\omega \mid u$ and $\omega \mid v$).*

This leads naturally to the oldest and best known algorithm in number theory, the Euclidean algorithm (due to Euclid circa 300 B.C.). According to Donald Knuth "The Euclidean algorithm is the granddaddy of all algorithms, because it is the oldest nontrivial algorithm that has survived to the present day."

Euclidean algorithm. Given a pair of natural numbers (u, v) with $0 < u \leq v$:

(i) Let $(u, v) \mapsto (r, u)$ where $v = qu + r$, where $r < u$; then

(ii) Repeat part (i) $N = N(u, v)$ times to get $(0, k)$, then k is the greatest common divisor of u and v.

To relate this algorithm to transfer operators we can reformulate this in terms of the Gauss map $T : [0, 1] \to [0, 1]$, which we recall is defined by

$$T(x) = \begin{cases} \frac{1}{x} - \left[\frac{1}{x}\right] & \text{if } 0 < x \le 1 \\ 0 & \text{if } x = 1. \end{cases}$$

More precisely, $N = N(u, v) \ge 0$ is also the smallest natural number for which the rational number $\frac{u}{v}$ satisfies $T^N\left(\frac{u}{v}\right) = 0$. (In particular, this means that $\frac{u}{v}$ can be written as a finite continued fraction $\frac{u}{v} = [a_1, \dots, a_n]$, say).

One can ask about the distribution of the values $N(u, v)$ as we consider different pairs $1 \le u < v$. Given $n \ge 1$, we can associate the finite set of pairs

$$\Omega_n = \{(u, v) \colon 1 \le u < v \le n\}$$

and let $\mathbb{P}_n$ be the normalized counting probability on Ω_n, i.e.,

$$\mathbb{P}_n = \frac{1}{\#\Omega_n} \sum_{(u,v)\in\Omega_n} \delta_{(u,v)},$$

where $\delta_{(u,v)}$ is the Dirac measure supported on (u, v). The following result is a special case of the interesting work of Baladi-Vallée which describes the statistical properties of these measures [3]. In particular, this result takes the form of a Central Limit Theorem.

Theorem 6.18 (Baladi-Vallée)**.** *There exist constants $\mu = \frac{6\log 2}{\pi^2}$ and $\sigma^2 > 0$ such that for any $\alpha \in \mathbb{R}$:*

$$\lim_{n\to\infty} \mathbb{P}_n \left(\left\{ (u, v) \in \Omega_n \colon \frac{N(u, v) - \mu \log n}{\sqrt{\log n}} < \alpha \right\} \right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\alpha} e^{-\frac{t^2}{2\sigma}} dt + O\left(\frac{1}{\sqrt{\log n}}\right)$$

These, and the various generalizations, are nicely described in the books of Knuth [11] and Hensley [8].

The proof again uses transfer operators and generalizations of the the work of Dolgopyat, but in addition to Tauberian Theorems it also uses a Quasi-Powers theorem of Hwang.

References

[1] A.Avila, S.Goužel and J.-C. Yoccoz, Exponential mixing for the Teichmller flow. Publ. Math. Inst. Hautes tudes Sci. No. 104 (2006), 143-211.

[2] V. Baladi and B. Vallée, Exponential decay of correlations for surface semi-flows without finite Markov partitions. Proc. Amer. Math. Soc. 133 (2005), no. 3, 865874.

[3] V. Baladi and B. Vallée, Euclidean algorithms are Gaussian. J. Number Theory 110 (2005), no. 2, 331-386

[4] R. Bowen, Invariant measures for Markov maps of the interval (With an afterword by Roy L. Adler and additional comments by Caroline Series) Comm. Math. Phys. 69 (1979), no. 1, 117

[5] P. Collet, Pierre and J.-P. Eckmann, *Iterated maps on the interval as dynamical systems.* Progress in Physics, 1. Birkhuser, Boston, Mass., 1980.

[6] W. Doeblin and R. Fortet, Sur des chanes à liaisons complètes, Bull. Soc. Math. France 65 (1937), 132148.

[7] D. Dolgopyat, On decay of correlations in Anosov flows. Ann. of Math. (2) 147 (1998), no. 2, 357-390.

[8] D. Hensley, *Continued fractions*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006.

[9] H. Huber, Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen, Math. Ann. 138 1959 1-26.

[10] C.T.Ionescu Tulcea and G Marinescu, Théorie ergodique pour des classes d'opérations non complètement continues, Ann. of Math. (2) 52, (1950). 140-147.

[11] D. Knuth, Knuth, Donald E. *The art of computer programming, Vol. 2.* Seminumerical algorithms, Addison-Wesley, Reading, MA, 1998.

[12] A. Lasota and J. Yorke,On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc. 186 (1973) 481488

[13] C. Liverani, On contact Anosov flows. Ann. of Math. (2) 159 (2004), no. 3, 12751312.

[14] F. Naud, Dolgopyat's estimates for the modular surface, lecture notes from IHP June 2005, workshop "time at work".

[15] M. Pollicott and R. Sharp, Exponential error terms for growth functions on negatively curved surfaces. Amer. J. Math. 120 (1998) 10191042

[16] M. Reed and B. Simon, *Methods of modern mathematical physics. I*, Functional analysis. Academic Press, Inc. , New York, 1980.