2014 Houston Summer School on Dynamical Systems

May 14–22, 2014

Contents

Preface

This document began as a set of notes taken and typed by Colin Thomson, a participant in the summer school; we thank Colin for providing these notes and allowing us to use them. We are also grateful to Neetish Pradhan and Corinna Wendisch (2014 participants) and Rachel Meier and Dania Sheaib (2013 participants), who shared notes that we used in the editing process.

The lectures on 'Probability in dynamics' were given by Matt Nicol. The lectures on 'Hyperbolicity' were given by Vaughn Climenhaga. The lectures on 'Spectral Methods' were given by Andrew Török. The lectures on 'Cones' were given by Will Ott. The lectures on 'Statistical physics' were given by Renato Feres. The lectures on 'Symbolic dynamics and C^* algebras' were given by Mark Tomforde. Some of the chapters here (particularly 'Spectral methods') incorporate material covered in background lectures given during the school by Vaughn Climenhaga, who also compiled and edited these notes; some of this material also appeared at `vaughnclimenhaga.wordpress.com`.

The summer school was funded by NSF grant DMS-1363024. Further material from the summer school (slides, problem sets) may be found at:

`http://www.math.uh.edu/~climenha/2014-school.html`

Chapter 1

Probability in dynamics

1.1 Observations of dynamical systems as stochastic processes

1.1.1 Random variables and stochastic processes

Definition 1.1.1. A *random variable* is a function $X : \Omega \to \mathbb{R}^n$ on a probability space (Ω, P).

Definition 1.1.2. A *stochastic process* is a sequence X_n of random variables.

Definition 1.1.3. The *distribution function* of a random variable is $F(x) = P(X \leq x) = P(\omega \in \Omega : X(\omega) \leq x)$.

Definition 1.1.4. A stochastic process is *stationary* if for every $k \geq 0$ the sequence $(X_k, X_{k+1}, \dots, X_{k+n})$ has the same distribution as $(X_1, X_2, \dots, X_n)$.

Example 1.1.5. The normal (or Gaussian) distribution $N(\mu, \sigma)$ has distribution function

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma}} dt.$$

Example 1.1.6. The exponential distribution has distribution function $F(x) = \int_0^x \lambda e^{-\lambda t} dt$.

Suppose $T : X \to X$ is a map on a probability space (X, μ) that preserves μ: $\mu(T^{-1}A) = \mu(A)$ for all measurable $A \subset X$.

Example 1.1.7. $X = S^1$, $T(x) = 2x \bmod 1$, with $\mu =$ Lebesgue measure.

Definition 1.1.8. *Independent identically distributed random variables* are ones for which all X_i have the same distribution, and $P(X_i \in A_i, X_j \in A_j) = P(X_i \in A_i)P(X_j \in A_j)$ for all $i \neq j$.

Suppose $\phi : X \to \mathbb{R}$ is an observable on X. Then $X_n = \phi \circ T^n$ is a stationary stochastic process; in general it is not independent in general. Nevertheless, in systems such as the doubling map that exhibit 'enough' expansion, the process X_n often satisfies statistical laws such as the central limit theorem. The mechanism driving this is *decay of correlations.* In these lectures we examine how decay of correlations can be used to establish various statistical properties; the lectures on hyperbolicity, spectral gap, and cone techniques examine various ways to establish decay of correlations for examples of interest.

The first main result is the strong law of large numbers.

Theorem 1.1.9. *If $\{X_i\}$ are i.i.d. and $E[X_i] < \infty$ then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = E[X_1]$$

for P-almost every $x \in \Omega$.

In fact, the strong law of large numbers holds in a more general setting that does not require independence of the X_n. Recall that a measure-preserving transformation T is *ergodic* if every T-invariant set A (that is, $T^{-1}A = A$) has $\mu(A) = 0$ or 1. Birkhoff's ergodic theorem says that T is ergodic if and only if for all $\phi \in L^1(\mu)$ $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi \circ T^i(x) \to \int \phi d\mu$. In other words, the strong law of large numbers holds for $X_n = \phi \circ T^n$.

The next main result for i.i.d. variables is the central limit theorem.

Theorem 1.1.10. *If X_i are i.i.d. and $E[X_i^2] < \infty$ then*

$$\lim_{n\to\infty} \frac{S_n(x) - nE[X_1]}{\sqrt{nVar[X_1]}} \xrightarrow{\text{dist}} N(0,1),$$

where $S_n(x) = X_1(x) + \cdots + X_n(x)$ and $Var[X] = E[X - E[x]]^2$. In this sense, the normal distribution with mean zero and variance one is a universal attractor.

This result does not hold as broadly as the law of large numbers; one can easily produce examples of ergodic transformations that do not satisfy the CLT. Nevertheless, we will see that the CLT continues to hold for observations of a broad class of systems.

Example 1.1.11. The Manneville-Pomeau map is defined as

$$T(x) = \begin{cases} x(1+(2x)^{\alpha}) & \text{if } x \in \left[0, \frac{1}{2}\right] \\ 2x - 1 & \text{if } x \in \left(\frac{1}{2}, 1\right] \end{cases}$$

for parameter $\alpha \in (0,1)$. As $T'(0) = 1$, zero is an "indifferent fixed point" of T. If one chose $\alpha = 0$, T would degenerate to the doubling map.

Note that T is uniformly expanding away from the neutral fixed point at 0, so whether or not a trajectory appears 'chaotic' depends on whether or not we are currently close to this fixed poin. The Manneville-Pomeau map is used as a model of turbulence in fluid as it alternates between "laminar" and "bursting" behavior. Such alternation is call *intermittent behavior*.

Such intermittent-type maps have an absolutely continuous invariant measure with density $\frac{d\mu}{dx} = x^{-\alpha}$ near zero. If $\alpha = 1$, no such probability measure exists; only an infinite measure may be found. Moreover if $\alpha \in \left(0, \frac{1}{2}\right)$ then one can obtain the central limit theorem for Hölder continuous functions. If $\alpha \in \left(\frac{1}{2}, 1\right)$ then one obtains "stable laws".

1.1.2 Decay of correlation and mixing

Now we describe some conditions that are used to establish results like the CLT.

Definition 1.1.12. A dynamical system is *mixing* if

$$\lim_{n\to\infty} \left|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)\right| \to 0 \tag{1.1}$$

for all measurable $A, B \subset X$. Alternatively, we can say that the proportion of the image of A in B is the same as the measure of A:

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

This is a stronger requirement than ergodicity.

The idea is to quantify the rate of mixing; that is, the rate at which the quantity in (1.1) converges to 0. We must do this carefully, though. If we let A, B be any measurable sets, then the convergence can happen arbitrarily slowly. A better way is to reformulate (1.1) in terms of functions: mixing is equivalent to the condition that

$$\left| \int \phi\psi \circ T^n d\mu - \int \phi d\mu \int \psi d\mu \right| \to 0$$

for every $\phi, \psi \in L^2(\mu)$. The convergence can still happen arbitrarily slowly if we do not place further restrictions on ϕ, ψ. So the goal is the following: produce Banach spaces B_α, B_β such that when $\phi \in B_\alpha$ and $\psi \in B_\beta$, the quantity above decays quickly. This decay plays the role of 'asymptotic independence' of time series of observables.

The main techniques for establishing decay of correlations are spectral theory (via transfer operatores), convex cones, and coupling. We will discuss the first two of these in the other lectures.

Example 1.1.13. For intermittent maps, if ϕ is Lipschitz, $\psi \in L^\infty$, then

$$\left| \int \phi\psi \circ T_\alpha^n d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C n^{1-\frac{1}{\alpha}} ||\phi||_{Lip} ||\psi||_{L^1(\mu)}$$

where C is independent of ϕ, ψ. This is a sharp bound.

Example 1.1.14. For the doubling map, if $\phi \in BV[0,1]$, $\psi \in L^1(m)$ then

$$\left| \int \phi\psi \circ T^n d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C\theta^n ||\phi||_{BV} ||\psi||_{L^1(\mu)}$$

for $\theta \in (0,1)$.

1.1.3 Return-time statistics

Suppose (T, X, μ) is a measure-preserving transformation. Poincare recurrence implies that for measurable $A \subset X$, $T^n x \in A$ infinitely often for μ-almost every $x \in A$.

Theorem 1.1.15. *Kac's theorem states that if* $\tau_A(x) = \min\{n > 1 : T^n x \in A, x \in A\}$ *then* $E[\tau_A] = \frac{1}{\mu(A)}$.

Exponential return times laws: let A_n be a sequence of nested measurable sets based a point $p \in X$. Let $\tau_{A_n}(x)$ be the first return time of $x \in A_n$ to A_n.

Definition 1.1.16. T has *exponential return time* if

$$\lim_{n\to\infty} \frac{\mu\left(x \in A_n : \tau_{A_n}(x) > \frac{t}{\mu(A_n)}\right)}{\mu(A_n)} = e^{-t}.$$

For doubling maps, exponential return time holds for non-periodic points only. This partitions points in $[0,1]$ into periodic and non-periodic; in particular, there is no Cantor set where this property fails.

1.2 Martingale methods

1.2.1 Conditional expectation

See also "Martingale Limit Theory and its Applications," by P. Hall and C.C. Heyde. Let (Ω, P) be a probability space with Borel σ-algebra $\mathcal{B}$. Let $\mathcal{F} \subset \mathcal{B}$ be a sub-σ-algebra.

Example 1.2.1. For the doubling map, $\mathcal{F}_1 = T^{-1}\mathcal{B}$, and $\mathcal{F}_2 = T^{-2}\mathcal{B} \subsetneq T^{-1}\mathcal{B} \subsetneq \mathcal{B}$. Each is coarser than the next, as elements require more connected components. A function $\phi\colon \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if $\phi^{-1}(a,b) \in \mathcal{F}$ for all intervals (a,b). For example,

$$\phi(x) = \begin{cases} 1 & x \in [0,1/2] \\ -1 & x \in (1/2,1] \end{cases}$$

is not $T^{-1}\mathcal{B}$-measurable, since $\phi^{-1}((1-\epsilon, 1+\epsilon)) = [0,1/2] \notin T^{-1}\mathcal{B}$.

Assume $\int |\phi| dP < \infty$, so ϕ is integrable. The *conditional expectation* of ϕ given $\mathcal{F}$, written $E[\phi|\mathcal{F}]$, is any random variable Z (defined uniquely P-almost everywhere) that is $\mathcal{F}$-measurable and

$$\int_A E[\phi|\mathcal{F}] dP = \int_A \phi dP$$

for all $A \in \mathcal{F}$. In particular, $E[Z] \leq E[\phi]$. We also have linearity of expectation given $\mathcal{F}$.

Example 1.2.2. If ϕ is $\mathcal{F}$-measurable, then $E[\phi|\mathcal{F}] = \phi$.

Example 1.2.3. If ϕ is *independent* of $\mathcal{F}$ (so that $E[\phi I_A] = E[\phi]E[I_A] = E[\phi]P(A)$ for all $A \in \mathcal{F}$) then $E[\phi|\mathcal{F}] = E[\phi]$. We can check by examining integrals:

$$\int_A \phi dP = \int I_A \phi dP = E[\phi]P(A) = \int_A E[\phi]dP.$$

Example 1.2.4. Suppose $\{\Omega_i\}$ is a countable partition of Ω such that $P(\Omega_i) > 0$ for all i. Let $\mathcal{F}$ be the sub-σ-algebra generated by $\{\Omega_i\}$. Then $E[\phi|\mathcal{F}]$ must be constant on each Ω_i, with $E[\phi|\mathcal{F}] = \frac{1}{P(\Omega_i)} \int_{\Omega_i} \phi dP$. Indeed, by linearity of expectation it suffices to check that

$$\int_{\Omega_i} \left(\frac{1}{P(\Omega_i)} \int_{\Omega_i} \phi dP \right) = \frac{1}{P(\Omega_i)} \int_{\Omega_i} \int_{\Omega_i} \phi dP dP = \int_{\Omega_i} \phi dP.$$

An important application of conditional expectation to dynamics is the following: if P is the "transfer operator" and U is the Koopman operator $U\colon \phi \mapsto \phi \circ T$, then

$$UP\phi = (P\phi) \circ T = E[\phi|T^{-1}\mathcal{B}].$$

1.2.2 Martingales

Definition 1.2.5. Let $\mathcal{F}_n$ be an increasing sequence of σ-algebras (so that they become finer and finer). A sequence of random variables S_n is called a *martingale* with respect to the *filtration* $\{\mathcal{F}_n\}$ if $E[|S_n|] < \infty$, S_n is $\mathcal{F}_n$-measurable, and $E[S_{n+1}|\mathcal{F}_n] = S_n$.

Often, $\mathcal{F}_n$ is the σ-algebra generated by $S_1, \dots, S_n$. Then one can think of S_{n+1} as being a "fair game" given the first n outcomes.

Example 1.2.6. If X_n are i.i.d. tosses of a fair coin, and S_n the number of heads after n tosses, then $\mathcal{F}_n = \sigma(S_1, \dots, S_n)$ is generated by the n-cylinder sets of a shift on $\{0, 1\}$, and S_n is a martingale with respect to $\mathcal{F}_n$.

The above example holds whenever X_n are IID with $E[X_n] = 0$:

$$E[S_{n+1}|\mathcal{F}_n] = E[S_n + X_{n+1}|\mathcal{F}_n] = E[S_n|\mathcal{F}_n] + E[X_{n+1}|\mathcal{F}_n] = S_n + 0 = S_n.$$

Given a martingale S_n we call the terms $S_{n+1} - S_n$ the *martingale differences*. It is not always the case that martingale differences are IID, as they were in the previous example.

Example 1.2.7. *Polya's urn* is a stochastic process defined as follows. Consider an urn containing some number of red and blue balls. At each step, a single ball is drawn at random from the urn, and then returned to the urn, along with a new ball that matches the colour of the one drawn. Let Y_n be the fraction of the balls that are red after the nth iteration of this process.

Clearly the sequence of random variables Y_n is neither independent nor identically distributed. However, it is a martingale, as the following computation shows: suppose that at time n there are p red balls and q blue balls in the urn. (This knowledge represents knowing which element of $\mathcal{F}_n$ we are in.) Then at time $n+1$, there will be $p+1$ red balls with probability $\frac{p}{p+q}$, and p red balls with probability $\frac{q}{p+q}$. Either way, there will be $p+q+1$ total balls, and so the expected fraction of red balls is

$$\begin{aligned} E[Y_{n+1}|\mathcal{F}_n] &= \frac{p}{p+q} \cdot \frac{p+1}{p+q+1} + \frac{q}{p+q} \cdot \frac{p}{p+q+1} \\ &= \frac{p(p+q+1)}{(p+q)(p+q+1)} = \frac{p}{p+q} = Y_n. \end{aligned}$$

In fact for our purposes the following version of a martingale will be the most useful.

Definition 1.2.8. Let $\mathcal{F}_n$ be a decreasing sequence of σ-algebras. A sequence S_n of random variables is a *reverse martingale* if S_n is measurable with respect to $\mathcal{F}_n$, $E|S_n| < \infty$, and $E[S_n|\mathcal{F}_m] = S_m$ when $n < m$.

Although martingale differences need not be IID, they still satisfy the central limit theorem.

Theorem 1.2.9. *(Liverani, Neveu) Let $\{X_n\}_{n\geq 1}$ be a stationary ergodic sequence of martingale or reverse martingale differences with respect to the filtration $\{\mathcal{F}_n\}$. IF $X_1 \in L^2(P)$, $E[X_1] = 0$, then*

$$\lim_{n\to\infty} \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i \to^{dist} N(0,1)$$

where $\sigma = E[X_1^2]$.

1.2.3 Back to dynamics; martingale approximation

Now we discuss how time series of dynamical systems can be approximated by martingales provided correlations decay quickly enough. This allows us to deduce the CLT for a dynamical system as a consequence of Theorem 1.2.9. The ideas presented here were introduced by Gordin in 1969.

Suppose $T : X \to X$ is ergodic with respect to μ, $\phi : X \to \mathbb{R}$. The idea is to decompose $\phi \circ T^i$ by

$$\phi \circ T^i = \chi \circ T^i + g \circ T^i - g \circ T^{i+1}$$

and $S_n = \sum_{i=1}^n \chi \circ T^i$, where χ is chosen such that S_n is a (reverse) martingale. Then summing the above we get

$$\sum_{j=1}^n \phi \circ T^j = \left(\sum_{j=1}^n \chi \circ T^j \right) + g \circ T - g \circ T^{n+1}. \tag{1.2}$$

Divide by $\sqrt{n}\sigma$ and the g terms will vanish in the limit, so that the central limit theorem for $\phi \circ T^i$ will follow from the central limit theorem for martingales.

To find χ and g, we start by defining the *Koopman operator* $U_T\phi = \phi \circ T$ and its L^2-adjoint $P : L^2(\mu) \to L^2(\mu)$. This says that $\int (P\phi)\psi d\mu = \int \phi\psi \circ T d\mu$.

Lemma 1.2.10. *For every $\phi \in L^2(\mu)$ we have $PU\phi = \phi$ and $UP\phi = E(\phi|T^{-1}\mathcal{B}]$.*

Proof. For the first claim we observe that

$$\int (PU\phi) \cdot \psi \, d\mu = \int (U\phi) \cdot (\psi \circ T) \, d\mu = \int (\phi \circ T) \cdot (\psi \circ T) \, d\mu = \int \phi\psi \, d\mu$$

for every ψ. For the second, we start by observing that $UP\phi$ is $T^{-1}\mathcal{B}$-measurable, since for every $A \in \mathcal{B}$ we have

$$((UP)\phi)^{-1} A = T^{-1} \circ (P\phi)^{-1} A \in T^{-1}\mathcal{B}.$$

Thus the claim follows once we observe that for every $A \in T^{-1}\mathcal{B}$, say $A = T^{-1}C$, we have

$$\begin{aligned} \int_A UP\phi d\mu &= \int_{T^{-1}C} UP\phi d\mu = \int (I_{T^{-1}C}(x))((P\phi) \circ T)(x) d\mu(x) \\ &= \int (I_C \circ T)((P\phi) \circ T) d\mu = \int (I_C) \cdot (P\phi) d\mu = \int \phi \cdot (I_C \circ T) d\mu \\ &= \int \phi I_{T^{-1}C} d\mu = \int_{T^{-1}C} \phi d\mu. \end{aligned}$$

□

Now comes the key place where we make an assumption on decay of correlations. More precisely, we assume that ϕ is in some Banach space B_α (such as Lipschitz, Hölder, bounded variation, etc.) on which the transfer operator P has the property that

$$|P^n\phi|_{B_\alpha} \leq cp(n)|\phi|_{B_\alpha} \tag{1.3}$$

whenever $\int \phi d\mu = 0$, where $p(n)$ is such that $\sum_{n=1}^\infty p(n) < \infty$.

Our scheme is to let $\phi \in B_\alpha$ and define $g = \sum_{n=1}^\infty P^n\phi$. Then $g \in B_\alpha$ by convergence in the Banach space norm. Define $\chi = \phi - g \circ T + g$, so $\chi \in B_\alpha$. Then

$$\begin{aligned} P\chi &= P\phi - P(g \circ T) + P\left(\sum_{n=1}^\infty P^n\phi\right) \\ &= P\phi - PUg + \sum_{n=2}^\infty P^n\phi \\ &= P\phi - \sum_{n=1}^\infty P^n\phi + \sum_{n=2}^\infty P^n\phi = 0, \end{aligned}$$

where the third equality uses the first claim in Lemma 1.2.10. Using the second claim in that lemma we see that $E[\chi|T^{-1}\mathcal{B}] = (UP)\chi = 0$.

We claim that $S_n = \sum_{j=1}^n \chi \circ T^j$ is a reverse martingale. Indeed, for every k and every $A \in T^{-(k+1)}\mathcal{B}$ we have $A = T^{-(k+1)}C$ for some $C \in \mathcal{B}$, so

$$\begin{aligned} \int_A \chi \circ T^k \, dP &= \int_{T^{-k}(T^{-1}C)} \chi \circ T^k \, dP = \int (\chi \circ T^k)(1_{T^{-1}C} \circ T^k) \, dP \\ &= \int \chi 1_{T^{-1}C} \, dP = \int_{T^{-1}C} E[\chi|T^{-1}\mathcal{B}] \, dP = 0. \end{aligned}$$

Moreover, as in (1.2) we have

$$\sum_{j=1}^n \phi \circ T^j = S_n + g \circ T - g \circ T^{n+1}. \tag{1.4}$$

Given $\phi \in B_\alpha \subset L^2(\mu)$ with $\int \phi \, d\mu = 0$, we can let χ, g be as given above and writing $X_n = \chi \circ T^n$. Then $S_n = \sum_{j=1}^n X_j$ is a martingale with respect to the decreasing sequence of σ-algebras $\mathcal{F}_n = T^{-n}\mathcal{B}$.

Moreover, $E[X_1] = \int \phi \, d\mu = 0$ and $\sigma := E[X_1^2] = \int \chi^2 \, d\mu$, so Theorem 1.2.9 and (1.4) imply that $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n \phi \circ T^n$ converges in distribution to the Gaussian $N(0,1)$ whenever $\sigma > 0$; this gives the central limit theorem for observables in B_α.

Note that if $\sigma = 0$, then $E[X_1^2] = 0$ implies that $X_1 = 0$, so in particular $\chi = 0$. This gives $\phi = g - g \circ T$; that is, ϕ is a *coboundary.* Notice that if ϕ is a coboundary then it sums to zero over every periodic orbit; in certain circumstances the reverse implication is always true, this is the *Livsic theorem.*

Example 1.2.11. Take T to be the doubling map and suppose $\int \phi dm = 0$. It can be shown using the spectral gap property (see the spectral theory lectures) that $P^n\phi|_{Lip} \leq C\theta^n|\phi|_{Lip}$ for some $\theta \in (0,1)$, and $P\phi(x) = \frac{1}{2}(\phi(y_1)+\phi(y_2))$. Hence $g = \sum_{n=1}^\infty P^n\phi$ is Lipschitz. We obtain the central limit theorem unless ϕ is a coboundary.

We conclude this section by deriving an explicit formula for σ in terms of ϕ (note that the formula above is in terms of χ).

Because $S_n = X_1 + \cdots + X_n$ is a (reverse) martingale, we have $E[X_iX_j] = 0$ for all $i \neq j$. Using (1.4), we have

$$\begin{aligned} \left(\sum_{j=1}^n \phi \circ T^j\right)^2 &= \left(S_n + g \circ T - g \circ T^{n+1}\right)^2 \\ &= S_n^2 + 2S_n(g \circ T - g \circ T^{n+1}) + (g \circ T - g \circ T^{n+1})^2, \\ E\left[\left(\sum_{j=1}^n \phi \circ T^j\right)^2\right] &= E[S_n^2] + O(\sqrt{n}), \end{aligned}$$

where we leave it as an exercise to show the $O(\sqrt{n})$ bound. Then we observe that

$$E[S_n^2] = E\left[\sum_{j=1}^n X_j^2\right] + 2\sum_{i<j} E[X_iX_j] = n\sigma,$$

where we use linearity of expectation and the fact that $E[X_iX_j] = 0$ for $i \neq j$. We conclude that

$$\lim_{n\to\infty} \frac{1}{n} E\left[\left(\sum_{j=1}^n \phi \circ T^j\right)^2\right] = \sigma.$$

Example 1.2.12. Given the Manneville-Pomeau map with exponent $\alpha \in (0,1)$, there is an absolutely continuous invariant probability measure μ with $\frac{d\mu}{dx} \sim x^{-\alpha}$ near zero. It has been shown that for every Lipschitz ϕ and $\psi \in L^\infty$, both with mean zero, we have

$$\left| \int \phi\psi \circ T^n d\mu \right| \leq c||\phi||_{Lip}|\psi|_\infty n^{-\beta}$$

where $\beta = \frac{1}{\alpha} - 1$. Plus $|P\phi|_\infty \leq |\phi|_\infty$. Let $\psi = sign P^n\phi$. Then $|P^n\phi|_1 \leq c||\phi||_{Lip}n^{-\beta}$. If $p = \beta - \delta$,

$$\int |P^n\phi|^p d\mu = \int |P^n\phi|^{p-1}|P^n\phi| d\mu \leq |P^n\phi|_\infty^{p-1}|P^n\phi|_1 \leq cn^{-\beta}.$$

This implies that $||P^n\phi||_p \leq c'n^{\frac{\beta}{p-\delta}}$, which is summable when $\beta > 2$, or when $\alpha < \frac{1}{2}$. Thus if $\alpha \in \left(0, \frac{1}{2}\right)$, we get the central limit theorem.

1.3 Axiom A systems

Suppose (X, T, μ) is Axiom A (μ a Gibbs measure). T invertible means that $P\phi = \phi \circ T^{-1}$, so there is no averaging under the transfer operator P. The idea is to reduce to one-sided (non-invertible) dynamics by quotienting out by the stable direction. This is the *Sinai trick*.

Use Markov partitions to code the Axiom A system by a two-sided shift (σ, Ω, m), with semi-conjugacy $\pi\colon \Omega \to X$. Lift $\phi : X \to \mathbb{R}$ to $\tilde{\phi}\colon \Omega \to \mathbb{R}$ by $\tilde{\phi}(\omega) = \phi(\pi\omega)$. Then $\tilde{\phi}$ is Hölder with the usual topology on (σ, Ω, m).

Our goal is to write $\tilde{\phi} = \psi + v - v \circ \sigma$ where ψ depends only on "future" coordinates $\omega_0\omega_1\omega_2\ldots$. Geometrically, consider the map $G : X \to X$ that maps points on a local stable manifold to the same point via holonomy: project (slide) points along stable manifolds onto a distinguished unstable manifold. Symbolically, this corresponds to the map $g\colon \Omega \to \Omega$ defined as follows: for each ω_0 in the alphabet of the shift, fix $\bar{\omega}^-(\omega_0) = \cdots \bar{\omega}_{-2}\bar{\omega}_{-1}$ such that $\bar{\omega}^-\omega_0$ is a legal sequence for the SFT. Then define g by

$$g(\omega) = \bar{\omega}^-(\omega_0)\omega_1\omega_2 \cdots .$$

The map g is Hölder continuous, and we write

$$v(\omega) = \sum_{k=0}^{\infty} \tilde{\phi}(\sigma^k\omega) - \tilde{\phi}(\sigma^k g\omega).$$

Note that $d(\sigma^k\omega, \sigma^k g\omega)$ decreases exponentially in k and so the sum converges by Hölder continuity of $\tilde{\phi}$. The function v is Hölder, and so $\psi := \tilde{\phi} + v\circ\sigma - v$ is Hölder as well. We have

$$\begin{aligned}\psi(\omega) &= \tilde{\phi}(\omega) + \sum_{k=0}^{\infty}\big(\tilde{\phi}(\sigma^{k+1}\omega) - \tilde{\phi}(\sigma^k g(\sigma\omega))\big) - \big(\tilde{\phi}(\sigma^k\omega) - \tilde{\phi}(\sigma^k g\omega)\big)\\ &= \sum_{k=0}^{\infty}\tilde{\phi}(\sigma^k g\omega) - \tilde{\phi}(\sigma^k g\sigma\omega)\end{aligned}$$

If ω, ω' have the same past ($\omega_n = \omega'_n$ for all $n < 0$), then we have $g\omega = g\omega'$, and so the above formula shows that $\psi(\omega) = \psi(\omega')$. Thus ψ can be viewed as a Hölder function on the one-sided shift corresponding to Ω, where the transfer operator has a spectral gap. This implies that $\psi\circ\sigma^n$, $\tilde{\phi}\circ\sigma^n$, and $\phi\circ T^n$ all satisfy the central limit theorem.

1.4 Non-stationary limit theorems

We stated the central limit theorem for reverse martingales and martingales. We will now use the "natural extension" to take a non-invertible system and make it invertible. In the process (which is sort of a reverse Sinai trick) we will lose the smooth structure but preserve the probabilistic properties.

Let (T, X, μ) be a non-invertible system. We define an invertible system related to it, the *natural extension* (σ, Ω, m). Let Ω be the set of one-sided sequences $x_0x_1x_2\cdots$ of elements in X with the property that $Tx_n = x_{n-1}$ for all $n \geq 1$. Define $\sigma\colon \Omega \to \Omega$ by $\sigma : x_0x_1x_2\ldots \mapsto (Tx_0)x_0x_1x_2\ldots$, so that σ^{-1} is the usual shift.

Functions and measures lift from X to Ω. Let $\phi : X \to \mathbb{R}$ be lifted to $\widetilde{\phi}$ where $\widetilde{\phi}(x_0x_1x_2\ldots) = \phi(x_0)$. The measures lift as well. (σ, Ω, m) is ergodic if and only if (T, X, μ) is ergodic.

Let $\mathcal{B}$ be the Borel sets on X and $\pi : \Omega \to X$ be the projection to the lead element. Lift $\mathcal{B}$ to $\mathcal{B}_0$ on Ω by $\mathcal{B}_0 = \pi^{-1}\mathcal{B}$. Then $\mathcal{F}_i = \sigma^i\mathcal{B}$ is an increasing sequence of σ-algebras and a filtration (resembling cylinder sets). $\widetilde{\phi}\circ\sigma^{-i}$ is $\mathcal{F}_i$-measurable.

By our earlier martingale approximation arguments we can write $\phi = \chi + g\circ T - g$ where $E[\chi|T^{-1}\mathcal{B}] = 0$. Lift χ to $\widetilde{\chi}$ on Ω: $E[\chi|T^{-1}\mathcal{B}] = 0$ implies that $E[\widetilde{\chi}|\sigma\mathcal{B}_0] = 0$. Now let $S_n = \sum_{i=1}^n \widetilde{\chi}\circ\sigma^{-i}$ so that

$$E[S_{n+1}|\mathcal{F}_n] = S_n + E[\tilde{\chi}\circ\sigma^{-(n+1)}|\mathcal{F}_n] = S_n + E[\chi|T^{-1}\mathcal{B}] = S_n.$$

This is enough to deduce distribution limits for reverse martingales from martingales: for example, for the CLT, we see that

$$\mu\left(x \in X : \frac{1}{\sigma\sqrt{n}}\left(\phi(x) + \phi(Tx) + \phi(T^2x) + \cdots + \phi(T^nx)\right) \in A\right)$$
$$= m\left(\omega \in \Omega : \frac{1}{\sigma\sqrt{n}}\left(\widetilde{\phi}(\omega) + \widetilde{\phi}(\sigma^{-1}\omega) + \widetilde{\phi}(\sigma^{-2}\omega) + \cdots + \widetilde{\phi}(\sigma^{-n}\omega)\right) \in A\right).$$

This method of reversing time cannot be used straightforwardly to obtain almost-sure results; for example, the law of the iterated logarithm (LIL) states that for P-a.e. ω we have

$$\limsup_{n\to\infty} \frac{X_1^{(\omega)} + \cdots + X_n^{(\omega)}}{\sqrt{n \log\log n}} = 1,$$

where $E[X_i] = 0$ and $0 < E[X_i^2] < \infty$. In order to establish the LIL for a given system, we need a stronger result on almost sure approximation by IID sums, such as the following recent result of C. Cuny and F. Merlevède (J. Theor. Probab. 2015).

Theorem 1.4.1. *Let $\{X_n\}$ be a sequence of reverse martingale differences, $\{\mathcal{F}_n\}$ an increasing sequence of σ-algebras, $E[X_n|\mathcal{F}_{n+1}] = 0$, $E[X_n] = 0$. In this case X_n is not necessarily $X \circ T^n$. Suppose*

$$\lim_{n\to\infty} \sigma_n^2 = \sum_{k=1}^{n} E[X_k^2] \to \infty$$

(in the stationary case, the sum on the right is $nE[X_n^2]$). Let $\{a_n\}$ be a non-decreasing sequence of positive numbers such that $\frac{a_n}{\sigma_n^2}$ is non-increasing and $\frac{a_n}{\sigma_n}$ is non-decreasing ($\sigma_n^2 \sim n$ in the stationary case). Assume

$$\sum_{k=1}^{n} E[X_k^2|\mathcal{F}_{k+1}] - E[X_k^2] = o(a_n)$$

and $\sum_{k=1}^{n} a_k^{-\nu} E[|X_n|^{2\nu}] < \infty$ for some $\nu \in [1, 2]$. Then there is a sequence of independent Gaussian random variables $\{Z_n\}$ with $E[Z_k^2] = E[X_k^2]$ such that

$$\sup_n \left|\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} Z_k\right| = o(a_n \log\log a_n).$$

Corollary 1.4.2. *In the stationary case*

$$\left|\sum_{k=1}^{n} X \circ T^k - \sum_{k=1}^{n} Z_k\right| = o(\sqrt{n \log\log n})$$

where $E[Z_k^2] = E[X \circ T^k]$ *and* $Z_i \sim N(0, \sigma^2)$*.*

We can often do much better than n; $a_n = n^{\frac{1}{4}+\epsilon}$ for expanding maps, for example.

This shows that the law of the iterated logarithm holds for $\{\phi \circ T^j\}$. Decomposing $\phi = \chi + g \circ T - g$ implies both the central limit theorem and the law of the iterated logarithm. Brownian motion is "lurking as a model" and corresponds with Birkhoff sums.

Next suppose that our observations change over time— we are looking at $\{\phi_n \circ T^n\}$. Suppose $\phi_n = I_{A_n}$ for $\{A_n\} \subset X$ (perhaps as nested balls about a point $p \in X$). Does $\sum_{k=1}^{n} I_{A_n} T^n(x)$ diverge for μ-almost every x? This would imply that $T^n x \in A_n$ infinitely often. More quantitatively, if $\sum_{n=1}^{\infty} \mu(A_n) = \infty$ is there a limit (as an i.i.d. process)

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} I_{A_j} \circ T^j x}{\sum_{j=1}^{n} \mu(A_j)} = 1$$

for μ-almost every x?

Suppose that the maps change over time: in particular, each map may not have the same invariant measure. If we choose a fixed sequence we have a "sequential dynamical system".

Example 1.4.3. (W. Philipp 1970) If T is the doubling map and $\{\phi_n\}$ is a positive sequence of functions bounded in the BV norm then

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} \phi_j \circ T^j x}{\sum_{j=1}^{n} \int \phi_j dm} = 1.$$

This is related to the Gal-Koksma lemma and strong Borel-Cantelli. The proof goes by non-stationary martingale approximation. Let $\widehat{\phi}_i = \phi_i - E[\phi_i]$, and consider $g_n = \sum_{j=1}^{n} P^j \widehat{\phi}_{n-j+1}$, where P is the transfer operator for the doubling map, so

$$\begin{aligned} g_1 &= P\hat{\phi}_1, \\ g_2 &= P\hat{\phi}_2 + P^2\hat{\phi}_1, \\ &\cdots \\ g_n &= P\hat{\phi}_n + \cdots + P^n\hat{\phi}_1. \end{aligned}$$

Let $\chi_{n+1} = \widehat{\phi}_{n+1} - g_{k+1} \circ T + g_k$. Observe that

$$\begin{aligned} P\chi_{n+1} &= P\widehat{\phi}_{n+1} - PUg_{k+1} + Pg_k \\ &= P\widehat{\phi}_{n+1} - \left(P\widehat{\phi}_{n+1} + \cdots + P^{n+1}\widehat{\phi}_1\right) + P(P\hat{\phi}_n + \cdots + P^n\widehat{\phi}_1) \\ &= 0 \end{aligned}$$

So $\{\chi_n \circ T^n\}$ is a sequence of reverse martingale differences and we have the telescoping sum

$$\frac{1}{\sigma_n}\sum_{k=1}^{n}\left(\phi_k \circ T^k - \int \phi_k dm\right) = \sum_{k=1}^{n}\left(\chi_k \circ T^k\right) + g_1 - g_{n+1} \circ T^n;$$

upon dividing by σ_n this converges to $N(0,1)$, as long as $\sigma_n \to \infty$.

If T varies, then instead of iterates of P one has compositions of different transfer operators. In this case, use cone techniques and the spectral gap to get limit theorems.

Chapter 2

Hyperbolicity

2.1 An abundance of measures

A *topological dynamical system* is a compact metric space X together with a continuous map $T\colon X \to X$. It is often the case that (X, T) has many invariant probability measures, and so it is not a priori clear which measure we ought to use. Thus we investigate the following questions:

1. Given a system (X, T), is there a distinguished invariant measure we ought to use?

2. If there is a distinguished measure μ, what are its statistical properties?

Let $\mathcal{M}(X)$ be the set of Borel probability measures on X. Within the class we consider the set $\mathcal{M}_T(X)$ of T-invariant probability measures, and the set $\mathcal{M}_T^e(X)$ of ergodic measures.

Definition 2.1.1. A measure μ is *ergodic* if for every measurable $E \subset X$ with $T^{-1}E = E$, we have $\mu(E) = 0$ or 1.

The ergodic measures are the extreme points of $\mathcal{M}_T(X)$, and every T-invariant measure is a convex combination of ergodic measures in a unique way. (Note that this convex combination may be infinite.) We say that $\mathcal{M}_T(X)$ is a *simplex*. A fact that is counterintuitive at first glance is that for many systems, the set of extreme points $\mathcal{M}_T^e(X)$ is dense in this simplex.

We list a number of examples for which $\mathcal{M}_T(X)$ has various types of behavior.

Example 2.1.2. The irrational rotation: $X = S^1 \subset \mathbb{C}$, $T(z) = e^{2\pi i\theta}z$, $\theta \in \mathbb{R} \setminus \mathbb{Q}$.

Example 2.1.3. (Example 1.1.7) The doubling map: $X = S^1$, $T(z) = z^2$.

Example 2.1.4. (Example 1.1.11) The Manneville-Pomeau map: exhibits intermittent behavior.

Example 2.1.5. The logistic map: $T : [0,1] \to [0,1]$, $T(x) = ax(1-x)$ for $a \in [0,4]$.

Example 2.1.6. $X = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, $T(x,y) = (x+\theta_1, y+\theta_2) \bmod \mathbb{Z}^2$ where $\theta_1, \theta_2 \in \mathbb{R} \setminus \mathbb{Q}$ are rationally independent.

Example 2.1.7. Shear on the torus: $X = \mathbb{T}^2$, $T(x,y) = (x+y, y)$. Preserves horizontal lines.

Example 2.1.8. The Arnold "cat" map: $X = \mathbb{T}^2$, $T(x,y) = (2x+y, x+y)$.

Example 2.1.9. The Hénon map: $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x,y) = (1 - ax^2 + y, bx)$ for $a, b \in \mathbb{R}$. 'Classical' values are $a = 1.4, b = .3$. One can view this as a logistic map with some "memory" coming from the y-coordinate, in the sense that x depends not only on its current position but also its previous position, which is saved as y.

We are looking for expansion (nearby trajectories diverge) and transitivity (every trajectory eventually comes close to every other one). When phase space is in $\mathbb{R}^p$ we can measure expansion by looking at $\|dT^n(x)(v)\|$ for $x \in \mathbb{R}^p$ and $v \in T_x\mathbb{R}^p$, and seeing how quickly this quantity grows. The idea then is to use the growth rate of this quantity to estimate the decay rate of correlations.

For rotations (on the circle or torus), this quantity remains bounded, so there is no expansion and nearby trajectories remain nearby; these maps are called *elliptic*. For the shear map $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ on the torus, this quantity grows linearly in n, so we have sub-exponential expansion. The doubling map and cat map both have exponential expansions, with $\|dT^n(x)(v)\|$ growing like λ^n for some $\lambda > 1$. Broadly speaking, systems with this sort of exponential expansion are called *hyperbolic*.

There is a distinction to be made between *uniformly hyperbolic* systems, such as the doubling map and the cat map, and *non-uniformly hyperbolic* systems, such as the Manneville–Pomeau map, logistic map, and Hénon map.

Roughly speaking, in uniformly hyperbolic systems, the rate of growth and the time it takes to observe that growth do not depend on x, while in non-uniformly hyperbolic systems the rate of growth and the time it takes to observe it may depend on x, and the growth can become arbitrarily weak.

In all of these examples, the phase space X is a manifold and thus carries a natural notion of volume (length in one dimension, area in two), given by Lebesgue measure. We look for invariant measures that are absolutely continuous with respect to Lebesgue; if μ is an *absolutely continuous invariant measure* then something true for μ-a.e. x will also be true for Lebesgue-a.e. x. Some examples, such as the doubling map and cat map, preserve Lebesgue measure itself, while for others, such as the Manneville–Pomeau map and logistic map, Lebesgue measure is not preserved. For the time being, we restrict our attention to examples preserving Lebesgue measure.

2.2 Markov partitions

2.2.1 Expanding maps

In the case of the doubling map, assign to each $x \in [0,1]$ a sequence where the n^{th} number is 0 if $T^n(x) \in \left[0, \frac{1}{2}\right)$, and 1 otherwise. In general, given $T\colon X \to X$, let $\mathcal{A} = \{A_1, A_2, \ldots, A_d\}$ be a partition of X where only boundaries of distinct elements of $\mathcal{A}$ overlap. Given $x \in X$, define $\omega \in \Sigma_d^+ = \{1, 2, \ldots, d\}^{\mathbb{N}}$ by $A_{\omega_n} \ni T^n(x)$. Define Σ, a subset of the full shift Σ_d^+ to be the closure of a set of admissible sequences, where "admissible" is taken to mean "corresponding to a possible orbit of a point $x \in X$". In the case of the doubling map, ω is just the binary representation of x. Define $\sigma : \Sigma_d^+ \to \Sigma_d^+$ by $\sigma(\omega) = \sigma(\omega_0\omega_1\omega_2\ldots) = \omega_1\omega_2\omega_3\ldots$.

Note that there are typically 'forbidden' sequences. For example, if we partition $[0,1]$ into four intervals of equal length, labelled 0,1,2,3, and consider the doubling map, then the symbol 0 can only be followed by a 0 or 1, but not by a 2 or a 3, since no points in the partition element $A_0 = [0, 1/4]$ are mapped into $A_2 \cup A_3 = [1/2, 1]$.

Definition 2.2.1. A *subshift of finite type*, or a *topological Markov chain*, is defined by movements on a directed graph whose vertices are given distinct labels. Equivalently, one can characterize the shift $\Sigma \subset \Sigma_d^+$ in terms of an adjacency matrix A_{ij}, where $A_{ij} = 1$ iff there is an edge from vertex i to vertex j, and say that $\omega \in \Sigma$ iff $A_{\omega_n\omega_{n+1}} = 1$ for all n.

Here is an example. Given $\beta > 1$, let $T : [0,1] \to [0,1]$ be defined by $T : x \mapsto \beta x \bmod 1$. Set $\beta = \frac{1+\sqrt{5}}{2}$, and $A_0 = \left[0, \frac{1}{\beta}\right)$, $A_1 = \left[\frac{1}{\beta}, 1\right]$. Then $T(A_0) = A_0 \cup A_1$, while $T(A_1) = A_0$. In terms of $\Sigma = \{0,1\}^{\mathbb{N}}$, if $\omega \in \Sigma$ has $\omega_0 = 0$, then $0\omega \in \Sigma$ because $T(A_0) \supset A_0$. $1\omega \in \Sigma$ for the same reason. But if ω has $\omega_0 = 1$, then $0\omega \in \Sigma$ and $1\omega \notin \Sigma$. So $\mathcal{A} = \{A_0, A_1\}$ is a Markov partition of $[0,1]$ for T, and the corresponding Markov chain has adjacency matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

If $\beta > \frac{1+\sqrt{5}}{2}$, $T(A_1) \cap A_1 \neq \emptyset$, but $\beta < 2$ means $T(A_1) \not\supset A_1$. Now the block 11 may be legal, but 111 is not whenever $\beta < 2$. So it is not immediately clear whether this map can be coded by an SFT – this partition doesn't work, but maybe another one does.

Definition 2.2.2. If $\mathcal{A}$ is a partition for an expanding interval map, $\mathcal{A}$ is *Markov* if for all $A, A' \in \mathcal{A}$ we have $T(A) \supset A'$ or $T(A) \cap A' = \emptyset$ up to endpoints. Write $i \to j$ if $T(A_i) \supset A_j$.

2.2.2 Invertible maps

The Markov partition for the cat map requires $\mathcal{A}^{\mathbb{Z}}$ instead of just $\mathcal{A}^{\mathbb{N}}$. The latter codes forward trajectories of points on a common stable eigenline; these points have the same future. Write $\omega^+ = \omega_0\omega_1\cdots$ and let $h(\omega^+)$ be the set points such that $f^n(x)$ lies in A_{ω_n} for all $n \geq 0$. Then $h(\omega^+)$ is a segment of a stable eigenline. Similarly, given $\omega^- = \cdots\omega_{-1}\omega 0$ and writing $h(\omega^-)$ for the set of points such that $f^n(x) \in A_{\omega_n})$ for all $n \leq 0$, we get that $h(\omega^-)$ is a segment of an unstable eigenline (past is determined). Now to have a Markov chain, we need $h(\omega^+) \cap h(\omega^-) \neq \emptyset$ whenever $\omega_0^+ = \omega_0^-$. (Since every such pair $\omega^{\pm}$ corresponds to $\omega \in \mathcal{A}^{\mathbb{Z}}$ in which all pairs $\omega_n \to \omega_{n+1}$ are a legal transition.)

Here is another way of saying this. For every $A_i \in \mathcal{A}$, A_i must have the property that for all $x, y \in A_i$, $W^s_{\mathcal{A}}(x) \cap W^u_{\mathcal{A}}(y) \in A_i$ and in particular is not empty. Since W^s and W^u are eigenlines, partitions are going to be rectangles.

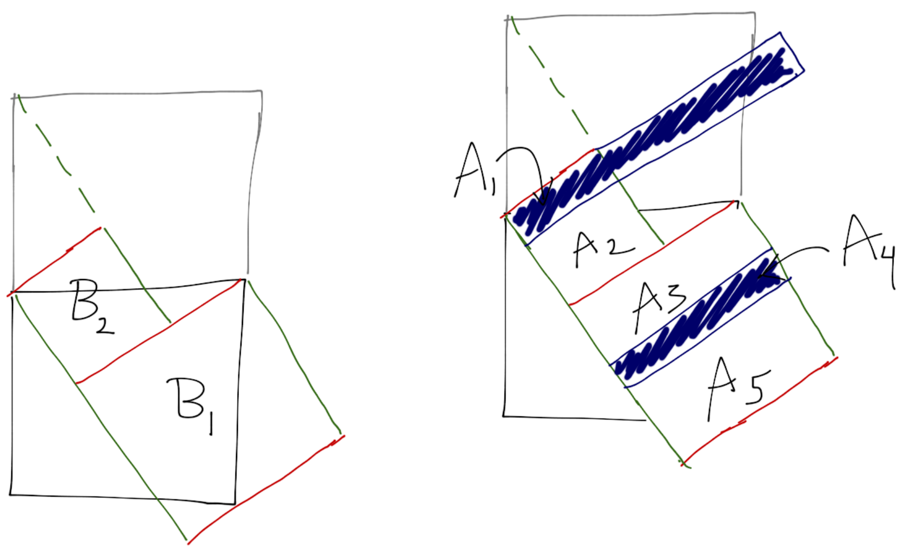
The Markov property for invertible maps states that if $x \in A_i$, $y \in A_j$ $(i \neq j)$ then

$$T(W^u_{\mathcal{A}}(x)) \supset W^u_{\mathcal{A}}(y) \text{ or } T(W^u_{\mathcal{A}}(x)) \cap W^u_{\mathcal{A}}(y) = \emptyset \tag{2.1}$$

and

$$T^{-1}(W^s_{\mathcal{A}}(y)) \supset W^s_{\mathcal{A}}(x) \text{ or } T^{-1}(W^s_{\mathcal{A}}(y)) \cap W^s_{\mathcal{A}}(x) = \emptyset. \tag{2.2}$$

To produce a Markov partition of the torus for the cat map, one can start with the squares B_1, B_2 shown in the picture. The sides of these squares are eigenlines of the map. Exercise: convince yourself that B_1 and B_2 satisfy (2.1) and (2.2). Note that the long skinny rectangle in the picture at the right is the image of B_2 under the map.



This is not quite a Markov partition yet because a point on the torus is not uniquely specified by its coding in terms of B_1, B_2. To remedy this we must partition further: the right-hand image shows $A_1 := f(B_2) \cap B_2$ and $A_4 := f(B_2) \cap B_1$, with A_2, A_3, A_5 the remaining partition elements as shown. One can check that this gives a generating Markov partition, so the cat map is (semi-)conjugate to an SFT on five letters.

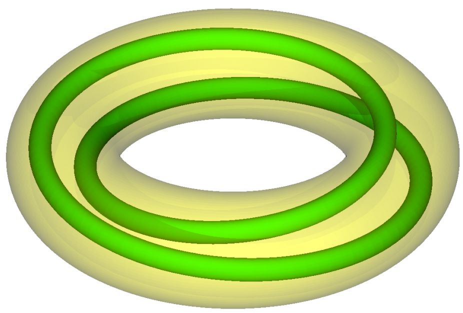
2.3 Nonlinear maps

Definition 2.3.1. Let $f : M \to M$ be a diffeomorphism. $\Lambda \subset M$ is a *hyperbolic set* if it is compact, f-invariant, and the tangent bundle decomposes as $T\Lambda = E^u \oplus E^s$ such that $df(E^u) = E^u$ and $df(E^s) = E^s$, where f contracts E^s uniformly and expands E^u uniformly.

Theorem 2.3.2. *A locally maximal hyperbolic set has a Markov partition.*

How do we find these Markov partitions for nonlinear hyperbolic maps? For example, one can consider a small perturbation of the cat map, or the following example.

Example 2.3.3. The *Smale–Williams solenoid* maps the solid torus $S^1 \times D^2$ into itself by stretching to twice the length and wrapping around. It expands in the S^1 "direction" and contracts in the D^2 "direction", as shown in the picture.[1]



Let M be a smooth manifold and $f : M \to M$ a diffeomorphism. (For now we just need C^1, although there are sometimes big differences between C^1 and $C^{1+\alpha}$.) If $f(p) = p$, and if $df_p : T_pM \to T_pM$ has no eigenvalues with modulus one, then

$$T_pM = E^u \oplus E^s, \quad E^u = \bigoplus_{|\lambda|>1} G_\lambda, \quad E^s = \bigoplus_{|\lambda|<1} G_\lambda,$$

where G_λ denotes the generalized eigenspace of λ. Then for all $v^u \in E^u$ we have $||df_p^n(v^u)|| \geq C\chi^{-n}||v^u||$, with a similar expression for E^s but in forward iteration ($\chi < 1$).

Theorem 2.3.4. *(Hadamard-Perron) There exists smooth submanifolds W^s, W^u which are tangent to E^s, E^u respectively at p such that $f(W^s) \subset W^s$ and $f^{-1}(W^u) \subset W^u$ where f, f^{-1} act as contractions on W^i by a factor of χ.*

In fact a neighborhood of the fixed point is foliated by stable/unstable manifolds: contracted along the stable direction and pushed away from fixed

[1]Picture from Wikipedia, created by user 'Ilya Voyager', dedicated to public domain.

point in unstable direction (to a different stable manifold). Consider the following local stable manifold through the fixed point:

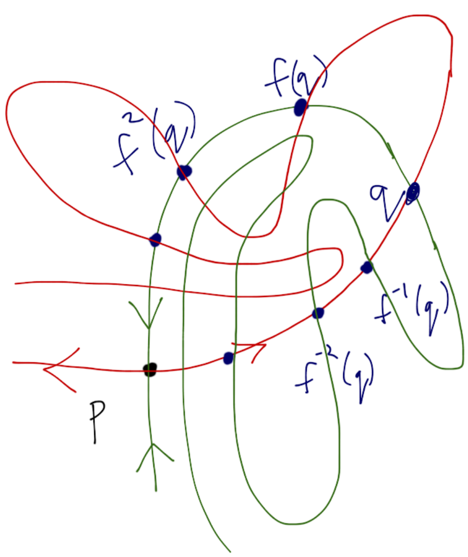
$$W_\delta^s = \{x \in M : d(f^n x, p) \le \delta \text{ for all } n \ge 0\}$$

and the same for W_δ^u for f^{-n}. Define, then, W^s as the *global stable manifold* as

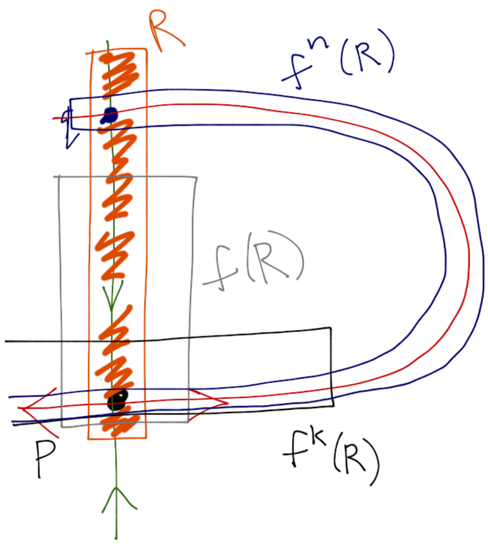
$$W^s = \{x \in M : f^n x \to p\} = \bigcup_{n \ge 0} f^{-n}(W_\delta^s)$$

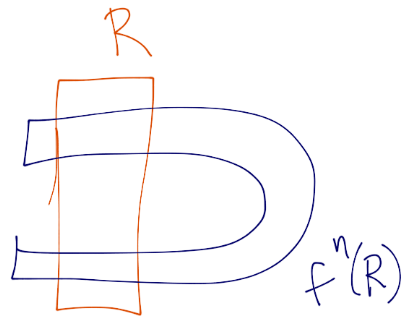
with W^u defined similarly.

If W^s, W^u intersect at a point $q \ne p$, then this point is a *homoclinic point*. It converges to p in both forward and backwards iterates, which induces an infinite number of intersections of W^u, W^s, the *homoclinic tangle*; the picture shows the beginning of this process.



In fact, one can show that there is some neighborhood R of p and some iterate $g = f^n$ such that $g \colon R \to g(R)$ is a 'horseshoe'; in particular, the maximal g-invariant set $\Lambda \subset R$ is a Cantor set on which g is topologically conjugate to the full shift on two symbols. The picture shows the rough idea behind this.





The Hadamard-Perron theorem also works if we replace p with "every $x \in \Lambda$," where Λ is a locally maximal hyperbolic set. In this case, the x are allowed to move and so we get a 'non-stationary' version of the theorem.

Definition 2.3.5. $R \subset \Lambda$ is a *rectangular set* if for all $x, y \in R$, $W^s_\delta \cap W^s_\delta \cap R$ is a single point, called $[x, y]$. So rectangles have "parallel" edges based on foliations by (un)stable manifolds. Elements of Markov partitions must be rectangles.

Definition 2.3.6. Let R be a rectangle. $A \subset R$ is a *u-set* if it is the union of unstable manifolds, and $B \subset R$ is an *s-set* if it the union of stable manifolds.

Example 2.3.7. The Smale horseshoe induced by looking at the images of a rectangle about a fixed and homoclinic point.

Definition 2.3.8. $\mathcal{A} = \{R_1, \ldots, R_m\}$ is a *Markov partition* for a hyperbolic set Λ if

1. each R_i is a rectangle and $R_i = \overline{\partial R_i}$, with $\partial R_i \cap \partial R_j = \emptyset$ for all $i \neq j$;

2. the partition is generating;

3. if $A \subset R_i$ is a u-set, then $f(A) \cap R_j$ is either empty or a u-set;

4. if $B \subset R_i$ is an s-set then $f^{-1} \cap R_j$ is empty or an s-set.

Definition 2.3.9. A hyperbolic set Λ is *locally maximal* if there exists an open set U containing it and $\Lambda = \{x \in M : f^n x \in U \text{ for all } n \in \mathbb{Z}\}$.

For the cat map, $\Lambda = \mathbb{T}^2$ is a hyperbolic set. In fact, it continues to be a hyperbolic set for small perturbations of the cat map; small open cones around the stable and unstable eigenvectors are invariant for f, and hence for small perturbations of f (since they are open), which can be used to deduce uniform hyperbolicity.

The key tools in the proof of existence of Markov partitions are the following.

Definition 2.3.10. Given $\epsilon > 0$, an *ϵ-pseudo-orbit* is $\{x_n\}_{n\in\mathbb{Z}}$ such that $d(x_{n+1}, f(x_n)) < \epsilon$ for all $n \in \mathbb{Z}$.

Lemma 2.3.11. *(shadowing lemma) For all $\delta > 0$ there exists an $\epsilon > 0$ such that if $\{x_n\}$ is an ϵ-pseudo-orbit, then there exists a unique $y \in \Lambda$ such that $d(f^n y, x_n) < \delta$ for all $n \in \mathbb{Z}$.*

The shadowing lemma may be proved by using the Hadamard–Perron theorem to produce "s-sets and u-sets" through the points x_n and getting good intersection properties.

To produce a Markov partition, since Λ is compact, we can fix an ϵ and let $a_1, a_2, \ldots, a_d$ be an ϵ-dense set. Use $\{a_1, \ldots, a_d\}$ as the alphabet for a coding: $i \to j$ if $d(f(a_i), a_j) < \epsilon$. Then Σ is a topological Markov chain on $\{1, \ldots, d\}$ with this relation. $\omega \in \Sigma$ if and only if $\{a_{\omega_n}\}_{n\in\mathbb{Z}}$ is a valid ϵ-pseudo-orbit. By the uniqueness of this coding, we can find $y \in \Lambda$ coded by a_{ω_n}.

In fact this does not quite give a Markov partition: writing A_i for the set of points in Λ coded by some sequence with $\omega_0 = i$, one actually obtains an open cover, but the intersections between A_i and A_j may be too large. The way around this is to use *Sinai's trick* to reduce the cover to a partition while retaining the Markov property.

2.4 Absolute continuous invariant measures

Markov partitions can be used to produce SRB measures, which are the appropriate 'physical' measures for ergodic theory.

Consider a map $f : [0,1] \to [0,1]$, we want to find an absolutely continuous invariant measure (acim) $d\mu = \psi dx$.

Example 2.4.1. Take f to be the doubling map, $p \in (0,1)$ and $\mu_p = (p, 1-p)$-Bernoulli measure. $\mu_p[\omega_1\omega_2 \ldots \omega_n] = p^{\#\text{ of 0s}}(1-p)^{\#\text{ of 1s}}$. If $p = \frac{1}{2}$, we get Lebesgue measure. Given μ, $x \in I$,

$$d_\mu(x) = \lim_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}.$$

Then $\mu(B(x,r)) \approx rd_\mu(x)$. (Note that the limit may not exist for all x, but ignore this for the moment.)

Definition 2.4.2. Let $B_n(x,\delta) = \{y \in X : d(f^k y, f^k x) < \delta, 1 \le k \le n\}$. This is the *Bowen ball* or *dynamical ball* of order n; as $n \to \infty$ it is equal to $\{x\}$.

For an interval map, $f^n(B(x,r)) \approx B\left(f^n x, re^{S_n\phi(x)}\right)$ where $\phi(x) = \log|f'(x)| > 0$ and S_n are the ergodic sums. So roughly $B_n(x,\delta) \approx B\left(x, \delta e^{-S_n\phi(x)}\right)$. When $r_n = \delta e^{-S_n\phi(x)}$ we get

$$d_\mu(x) = \lim_{n\to\infty} \frac{\log \mu(B_n(x,\delta))}{\log(\delta) - S_n\phi(x)} = \lim_{n\to\infty} \frac{-\log \mu(B_n(x,\delta))}{n} \div \frac{S_n\phi(x)}{n}. \quad (2.3)$$

The Brin-Katok entropy formula states that if μ is ergodic, then

$$\lim_{n\to\infty} -\frac{1}{n} \log \mu(B_n(x,\delta))$$

exists and is constant μ-almost everywhere. (In general one must also take $\delta \to 0$, but for expanding interval maps we can omit this step.)

Call the common value of the limit $h(\mu)$; this is equal to the Kolmogorov-Sinai entropy (which is usually defined differently). Write

$$\lambda(\mu) = \int \log|f'(x)| d\mu$$

for the Lyapunov exponent of μ. Then we see from (2.3) and the Birkhoff ergodic theorem that for μ-a.e. x, we have

$$d_\mu(x) = \frac{h(\mu)}{\lambda(\mu)}.$$

Exercise: show that the measure-theoretic entropy of the doubling map with the Bernoulli measure μ_p is

$$h(\mu_p) = -p\log p - (1-p)\log(1-p).$$

Now we return to the definition of $d_\mu(x)$ and look at how quickly $\mu(B(x,r))$ decays with r. Given an ergodic measure μ on $[0,1]$, we must have $d_\mu(x) \leq 1$ for μ-a.e. x, and the idea is that we have equality iff μ is an acim. In other words, $h(\mu) \leq \lambda(\mu)$, with equality iff μ is an acim, and so with $\phi(x) = -\log|f'(x)|$, we have

$$\sum_{\mathcal{M}_f^e}\{h(\mu) - \lambda(\mu)\} = \sup_{\mathcal{M}_f^e}\left\{h(\mu) + \int \phi d\mu\right\} = 0.$$

Definition 2.4.3. If $f : X \to X$, $\phi : X \to \mathbb{R}$, define the *topological pressure* P as

$$P(\phi) = \sup_{\mathcal{M}_f^e}\left\{h(\mu) + \int \phi d\mu\right\}.$$

A measure achieving this supremum is an *equilibrium state*.

Statistical properties of equilibrium states are related to the analytic properties of $\phi \mapsto P(\phi)$. For example, the first and second derivatives of the pressure function $P : C(X) \to \mathbb{R}$ then are related to the mean and variance of the time averages for ϕ.

Let $f : M \to M, U \subset M$ be open such that $\overline{f(U)} \subset U$, and $\Lambda = \bigcap_{n\geq 0} f^n U$ be a (hyperbolic) attracting set. Axiom A systems are an example.

Example 2.4.4. The cat map (Example 2.1.8) with a small perturbation.

Example 2.4.5. The solenoid (Example 2.3.3) gives an attractor that is a Cantor set; in particular, the Lebesgue measure of Λ is zero, so there is not a.c.i.m., but there are physical measures.

The 'physical measures' mentioned above are obtained as 'SRB measures' (for Sinai–Ruelle–Bowen). The idea is that a measure μ can be 'decomposed' into conditional measures on the unstable manifolds (this is a version of Fubini's theorem); if these conditional measures are absolutely continuous, μ is an *SRB measure*. If $\phi(x) = -\log\left|df|_{E^u(x)}\right|$, $d_\mu = \frac{h(\mu)}{\lambda(\mu)}$ along the unstable directions.

Chapter 3

Spectral Methods

3.1 Transfer operator and spectrum

3.1.1 The doubling map

Our goal in these lectures is to prove existence of absolutely continuous invariant measures, and exponential decay of correlations, for certain classes of systems by establishing the existence of a spectral gap.

Some systems for which we can do this are piecewise expanding maps, including the doubling map, some non-uniformly expanding maps, and some (non-uniformly) hyperbolic maps. For example, we will prove the following.

Theorem 3.1.1. *Given the doubling map, μ Lebesgue measure, $\phi \in L^\infty$, ψ Lipschitz,*

$$\left| \int \phi \circ T^n \psi d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C \frac{1}{2^n} ||\phi||_{L^\infty} ||\psi||_{Lip}.$$

A similar result will be proven for piecewise-expanding maps on $[0,1]$: we will show that there is an invariant measure μ with $d\mu = h\,dx$ (this is the acip). It is unique if the map is topologically mixing. Moreover, there is $\lambda \in (0,1)$ such that

$$\left| \int \phi \circ T^n \psi d\mu - \int \phi d\mu \int \psi d\mu \right| \leq C \lambda^n ||\phi||_{L^\infty} ||\psi||_{Lip}.$$

Both of the above results will be proved using the (Ruelle) transfer operator. Assume $T : [0,1] \to [0,1]$ that is nonsingular, i.e. $Leb(A) \neq$

$0 \Rightarrow Leb(TA) \neq 0$. (Equivalently, assume that the Koopman operator $U_T : L^\infty(Leb) \to L^\infty(Leb)$ given by $U_T(\phi) = \phi \circ T$ is well-defined.)

Then there exists a $P_T : L^1(Leb) \to L^1(Leb)$ given by

$$\int \phi \circ T \psi dx = \int \phi P_T(\psi) dx \text{ for all } \phi \in L^\infty(Leb), \psi \in L^1(Leb). \tag{3.1}$$

It is easy to show that $||P_T||_{L^1} \leq 1$, i.e. $||P_T\phi||_{L^1} \leq ||\phi||_{L^1}$.

Given an absolutely continuous measure $d\mu = h\,dx$, where $h \in L^1(Leb)$, the transfer operator P_T produces the new absolutely continuous measure $d(T^*\mu) = P_T(h)dx$ where $T^*\mu(A) = \mu(T^{-1}A)$ is the adjoint. So $d\mu = hdx$ is T-invariant iff $P_T h = h$.

Thus finding an absolutely continuous invariant measure boils down to finding a fixed point of P_T. Once we have found the fixed point $h = P_T h$ and set $d\mu = h\,dx$, to study decay of correlations we can consider $\psi \in L^1_0(\mu) = \{\psi \in L^1(\mu) : \int \psi d\mu = 0\}$, for which

$$\left| \int \phi \circ T^k \psi d\mu \right| = \left| \int \phi \circ T^k \psi h\, dx \right| = \left| \int \phi P_T^k(\psi h) d\mu \right| \leq ||\phi||_{L^\infty} ||P_T^k(\psi h)||_{L^1}$$

In other words, one wants to study the rate of decay of $\|P_T^k|_{L^1_0(dx)}\|$. So overall, the idea is to consider P_T acting on $L^1(dx)$, find an eigenfunction corresponding to the eigenvalue 1, and then show that P_T restricted to a subspace transverse to thie eigenspace has norm strictly smaller than 1.

As mentioned before, this does not work if we just consider L^1 functions, but if we introduce a little more regularity then we can hope for success. To this end, suppose B is a Banach space such that $B \hookrightarrow L^1$; this means that $\exists c > 0$ such that $\phi \in B \Rightarrow ||\phi||_{L^1} \leq c||\phi||_B$. For example $B = Lip, BV$.

Now if there is $\lambda \in (0,1)$ such that $||P_T|_{B_0}|| < \lambda$, then we can carry out the plan described above: for all $\psi \in B_0$, $||P_T^k\psi||_{L^1} \leq c||P_T k\psi||_B \leq c'\lambda^k||\psi||_B$. In conclusion, $\phi \in L^\infty, \psi \in B_0$ implies that

$$\left| \int \phi \circ T^k \psi d\mu \right| \leq ||\phi||_{L^\infty} ||P_T^k \psi||_{L^1} \leq C\lambda^k ||\phi||_{L^\infty} ||\psi||_B.$$

Definition 3.1.2. P_T has a *spectral gap* on B if $\|P_T|_{B_0}\| < 1$. We also say that P_T is *quasicompact*.

Given a piecewise C^1 expanding interval map, one can show that $P_T\phi|_x = \sum_{Ty=x} \frac{\phi(y)}{|T'(y)|}$. For the doubling map,

$$(P_T\phi)(x) = \frac{1}{2}\phi\left(\frac{x}{2}\right) + \frac{1}{2}\phi\left(\frac{x+1}{2}\right).$$

Here we can skip the first step of producing an invariant measure, because we already know Lebesgue is invariant (in other words, the constant function 1 is a fixed point of P_T).

Consider the semi-norm

$$|\phi|_{Lip} = \sup_{x \neq y} \frac{|\phi(x) - \phi(y)|}{|x - y|}$$

For the doubling map we have

$$\begin{aligned} |P_T\phi|_x - P_T\phi|_y| &= \left| \frac{1}{2}\left(\phi\left(\frac{x}{2}\right) - \phi\left(\frac{y}{2}\right)\right) + \frac{1}{2}\left(\phi\left(\frac{x+1}{2}\right) - \phi\left(\frac{y+1}{2}\right)\right)\right| \\ &\leq |\phi|_{Lip}\frac{|x-y|}{2} \end{aligned}$$

where we make use of the fact that $|\phi(x) - \phi(y)| \leq |\phi|_{Lip}|x - y|$ in the last line. Thus $|P_T\phi|_{Lip} \leq \frac{1}{2}|\phi|_{Lip}$.

We now define

$$Lip = Lip_{\mathbb{C}}[0,1] = \{\phi : [0,1] \to \mathbb{C} : |\phi|_{Lip} < \infty\} \subset C^0[0,1]$$

with the norm $||\cdot||_{Lip} = ||\cdot||_{C^0} + |\phi|_{Lip}$ so that Lip is a Banach space. We claim that P_T has a spectral gap on $Lip_{\mathbb{C}}[0,1]$. Write

$$Lip = \mathbb{C}1 + \left\{\phi \in Lip \,\middle|\, \int_0^1 \phi dx = 0\right\}.$$

There exists a C such that $\phi \in Lip_0 \Rightarrow ||\phi||_{C^0} \leq C||\phi||_{Lip}$. That $Re(\phi)$, $Im(\phi)$ have zero integral means that ϕ vanishes somewhere, so for $\phi \in Lip_0$, we have $||P_T^k\phi||_{Lip} \leq C\frac{1}{2^k}||\phi||_{Lip}$.

P_T on Lip is quasicompact: $P_T(1) = 1$. $P_T(Lip) \subset Lip_0$:

$$\int P_T\phi dx = \int P_T\phi \circ 1dx = \int \phi 1 \circ T dx = \int \phi dx$$

using Lebesgue as the invariant measure. The spectral radius of P_T on Lip_0 is less than one.

3.1.2 Spectral properties

Let us be a little more explicit about the role of the spectrum of P_T. Recall that the spectrum of the operator $P_T\colon Lip \to Lip$ is the set

$$\sigma(P_T) = \{\lambda \in \mathbb{C} \mid P_T - \lambda I \text{ is not an invertible operator on } Lip\},$$

which contains (but is not necessarily equal to) the set of eigenvalues of P_T (the *point spectrum*). We emphasise that this is a very general definition, valid for any bounded linear operator on any Banach space, not just P_T acting on *Lip*. A basic fact in functional analysis is that the spectrum is always compact and non-empty.

In the example above, the constant function 1 is an eigenfunction with eigenvalue 1, and using this invariant decomposition $Lip = \mathbb{C}1 \oplus H$ from before (where H is the space of Lipschitz functions with zero mean), we have $\sigma(P_T) = \{1\} \cup \sigma(P_T|_H)$. That is, apart from the eigenvalue at 1, the spectrum of P_T is determined by its action on the subspace H.

Recall from functional analysis that if we write $\rho(P_T) = \sup\{|\lambda| \mid \lambda \in \sigma(P_T)\}$ for the *spectral radius* of P_T, we have

$$\rho(P_T) = \lim_{n\to\infty} \|P_T^n\|^{1/n} \le \|P_T\|. \tag{3.2}$$

To determine the spectrum of $P_T|_H$ we can use either the Lipschitz norm $\|\cdot\|_{Lip}$ or the semi-norm $|\cdot|_{Lip}$, because on the subspace H the semi-norm becomes a norm and the two are equivalent:

$$|\hat{\psi}|_{Lip} \le \|\hat{\psi}\|_{Lip} = \|\hat{\psi}\|_\infty + |\hat{\psi}|_{Lip} \le 2|\hat{\psi}|_{Lip}.$$

(This fails outside of H, where to apply (3.2) we would need to use $\|\cdot\|_{Lip}$.) From the previous section and (3.2) we see that $\rho(P_T|_H) \le \frac{1}{2}$. Thus the spectrum of P_T has a single eigenvalue at 1, while the rest of the spectrum is contained in the disc with centre 0 and radius 1/2.

Going beyond the doubling map to such examples as the piecewise expanding interval maps discussed above, the goal is to carry out a similar procedure by finding a suitable Banach space $\mathcal{B}$ of functions on which the transfer operator acts with a *spectral gap*: that is, where there is a single eigenvalue (or at most finitely many) lying on the unit circle, and the rest of $\sigma(P_T)$ is contained in a disc of radius $\rho < 1$. Then one is able to draw the following conclusions.

1. The eigenfunction(s) corresponding to the eigenvalue 1 are the densities for the absolutely continuous invariant measures.

2. Given any $r \in (\rho, 1)$, there is a constant C_r such that $\|P_T^k\|_{\mathcal{B}} \leq C_r r^k$, and so the correlations $C_k(\varphi, \psi)$ decay like r^k when the observables φ and ψ are chosen from suitable function spaces.

Eventually it is also interesting to consider a more general class of transfer operators associated to *potential functions* for which the largest eigenvalue may not be 1, but for now we stick to the setting described so far.

3.2 Function spaces and compactness

Before moving on to more general piecewise expanding interval maps, we recall some background material on functional analysis, and in particular on compactness properties that will be important.

3.2.1 Function spaces and extra structure

It is useful to treat real-valued functions (or complex-valued functions, or vector space-valued functions) as elements of a vector space, so that the tools from linear algebra can be applied. Given a set X one may consider the vector space $\mathbb{R}^X$ of all real-valued functions with domain X. If X is finite, say with n elements, then this is just the familiar vector space $\mathbb{R}^n$. The more interesting examples are when X is infinite, and so $\mathbb{R}^X$ is infinite-dimensional. We will focus on the case $X = [0, 1]$, which is reasonably representative.

Generally speaking, the functions $[0, 1] \to \mathbb{R}$ that arise from some application are not entirely arbitrary, but have some degree of regularity – maybe they are continuous, or piecewise continuous, or measurable, or integrable, etc. It turns out that the vector space $\mathbb{R}^{[0,1]}$ is "too large" for many applications, and that it is more suitable to consider a smaller space, whose elements are functions with some extra properties. We will consider some of the ways to do this, paying particular attention to how those choices let us recover certain properties of $\mathbb{R}^n$ that involve extra structure beyond that of the vector space itself:

- *Topology:* We know what it means for a sequence $\vec{x}_k \in \mathbb{R}^n$ to converge to some $\vec{x} \in \mathbb{R}^n$, and we want a similar notion of convergence in a vector space $V \subset \mathbb{R}^{[0,1]}$.

- *Metric and norm:* We want the notion of convergence to come from a metric (distance function) that is compatible with the vector space structure of V – that is, a norm, with respect to which the vector space V becomes a Banach space.

- *Compactness:* A subset of $\mathbb{R}^n$ is compact if every sequence in that subset has a convergent subsequence, and this property is important in many applications and proofs. By the Heine–Borel theorem compactness in $\mathbb{R}^n$ is equivalent to being closed and bounded. How can we determine when a set of functions in V is compact?

3.2.2 Continuous functions and Arzelà–Ascoli

The extra structure we seek to place on $V \subset \mathbb{R}^{[0,1]}$ should leverage some of the extra structure that $[0, 1]$ has, beyond simply being an uncountable set. In particular, we may use either the topology of $[0, 1]$ or Lebesgue measure on $[0, 1]$ to define properties of functions $f : [0, 1] \to \mathbb{R}$. First we discuss the topological option – later we see what happens when we use the measure-theoretic structure to define the L^p spaces (and others).

The natural space to use is $C(X)$, the space of continuous real-valued functions on $X = [0, 1]$, with the norm $\|f\|_{C^0} = \sup_{x\in[0,1]} |f(x)|$. The space of continuous functions is complete with respect to this norm, and so we have a Banach space. What about compactness? How do we tell if a set $A \subset C(X)$ is compact? Of course A should be closed, but what else do we need? Boundedness is no longer enough: the unit ball in $C(X)$ is not compact, as can be seen by considering the sequence of functions shown in Figure 3.1.

The solution here is given by the Arzelà–Ascoli theorem: a set $A \subset C(X)$ is *pre-compact* (has compact closure) if and only if the following conditions are satisfied.

- A is uniformly bounded: $\sup_{f\in A} \sup_{x\in X} |f(x)| < \infty$.

- A is equicontinuous: for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for every $f \in A$ and $|x - y| < \delta$.

Remark 3.2.1. The proof that these conditions guarantee compactness uses the following strategy, which it is a useful exercise to complete:

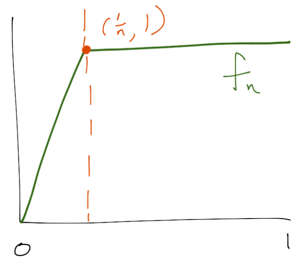


Figure 3.1: Uniformly bounded but no convergent subsequence.

1. Given any sequence $f_n \in A$, use uniform boundedness and a diagonalisation argument to find a subsequence that converges at every rational number. (Or on some other countable dense set.)

2. Use equicontinuity to guarantee that $\{f_{n_k}(x)\}_{k\geq 1}$ is Cauchy for every $x \in [0,1]$, and hence converges.

In particular, one can consider the subspace $C^\alpha(X) \subset C(X)$ of Hölder continuous functions with exponent $\alpha \in (0,1)$ – this is a Banach space with norm

$$\|f\|_{C^\alpha} = \|f\|_{C^0} + |f|_\alpha, \qquad |f|_\alpha = \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha}.$$

When $\alpha = 1$ this is the space of Lipschitz functions. If $A \subset C^\alpha(X)$ is uniformly bounded *in the C^α norm*, then it is uniformly bounded in the C^0 norm and equicontinuous, and hence it is pre-compact *in the C^0 norm*.

It is important to note here the structure of the last statement – we have two norms, $\|\cdot\|_{C^\alpha}$ and $\|\cdot\|_{C^0}$, such that **uniform boundedness in one norm implies pre-compactness in the other**. This is the closest that we can come to an infinite dimensional analogue of Heine–Borel: as a consequence of Riesz's lemma, every infinite-dimensional Banach space has a uniformly bounded sequence with no convergent subsequence.

In our study of spectral methods in dynamics, an important step is always to find two norms with this relationship: uniform boundedness in one implies pre-compactness in the other. We remark that the Arzelà–Ascoli theorem actually gives just a little bit more than this: given a sequence $f_n \in C(X)$ that is uniformly bounded in the C^α norm, pre-compactness only guarantees the existence of a limit point $f_{n_k} \xrightarrow{C^0} f \in C^0$, but in fact the limit point f is in C^α as well, because any modulus of continuity for the sequence f_n is also a modulus of continuity for any limit point.

Another important family of function spaces, which leverages not only the topological but also the differentiable structure of the unit interval, are the spaces C^r, defined inductively as

$$C^{r+1} = \{f \colon [0,1] \to \mathbb{R} \mid f \text{ is differentiable and } f' \in C^r\}.$$

Here r need not be an integer (the base case for the induction is $0 \le r < 1$), so for example, for $0 < \alpha < 1$, $C^{1+\alpha}$ is the space of differentiable functions whose derivatives are Hölder continuous with exponent α. The space C^r becomes a Banach space when endowed with the norm inductively given by

$$\|f\|_{C^{r+1}} = \|f\|_{C^0} + \|f'\|_{C^r}.$$

For example, on C^1 the appropriate norm is

$$\|f\|_{C^1} = \|f\|_{C^0} + \|f'\|_{C^0}. \tag{3.3}$$

The relationship discussed above between uniform boundedness in one norm and pre-compactness in another can be stated quite generally for this family of norms: **uniform boundedness in the C^r norm implies pre-compactness in the C^s norm for any $0 \le s < r$.** This relationship is often expressed by saying that "C^r is compactly embedded in C^s for $r > s$".

3.2.3 Lp spaces

In terms of the measure-theoretic structure of the unit interval, the most important function spaces are the L^p spaces

$$\begin{aligned} L^p &= L^p([0,1], dx) \\ &= \Bigg\{ f \colon [0,1] \to \mathbb{R} \,\Big|\, f \text{ is measurable and} \\ &\qquad\qquad \|f\|_p := \left(\int_{[0,1]} |f(x)|^p \, dx \right)^{\frac{1}{p}} < \infty \Bigg\}, \end{aligned}$$

where $1 \le p < \infty$, and

$$L^\infty = \{f \colon [0,1] \to \mathbb{R} \mid f \text{ is measurable and } \|f\|_\infty < \infty\},$$

where $\|f\|_\infty = \sup\{L \geq 0 \mid \{x \in [0,1] \mid |f(x)| > L\}$ has positive Lebesgue measure$\}$ is the essential supremum of f.

In fact, this definition cheats a little bit, because elements of an L^p space are actually *equivalence classes* of functions, where two functions are equivalent if they agree on a set of full Lebesgue measure. This throws a small technical monkey wrench into many arguments involving L^p spaces, since strictly speaking an expression like $f(x)$ for $f \in L^p$ has no meaning unless it is inside an integral sign. One way to avoid these technicalities is to emphasise the role of elements of L^p not necessary as functions, but rather as linear functionals.

Recall that if $\mathcal{B}$ is a Banach space, then $\mathcal{B}^*$ is the dual space of continuous linear functionals $\mathcal{B} \to \mathbb{R}$. The L^p spaces have the property that

$$(L^p)^* = L^q \text{ for } 1 < p, q < \infty \text{ such that } \frac{1}{p} + \frac{1}{q} = 1,$$

where $f \in L^p$ defines a linear functional on L^q by

$$g \mapsto \int f \cdot g \, dx \text{ for } g \in L^q. \tag{3.4}$$

Thus instead of thinking of a function $f \in L^p$, we may think of the associated functional in (3.4), which is obtained by integrating the function f against *test functions* from a suitable space. In this case the space of test functions is taken to be L^q, but there are many other examples we could consider – eventually this leads to the idea of considering *distributions* in place of functions, but we will not go this far here.

Remark 3.2.2. Before moving on, we note that $(L^1)^* = L^\infty$, but $(L^\infty)^*$ is a larger space than L^1.

3.2.4 Weak derivatives

An important use of this alternate viewpoint – functions as continuous linear functionals – is to define the *weak derivative* of a function. If $f\colon [0,1] \to \mathbb{R}$ is differentiable, then for any differentiable $g\colon [0,1] \to \mathbb{R}$ with $g(0) = g(1) = 0$, integration by parts gives

$$\int f' \cdot g \, dx = -\int f \cdot g' \, dx. \tag{3.5}$$

Equation (3.5) characterises the derivative f', which motivates the following definition: $h \in L^1$ is the weak derivative of $f \in L^1$ if

$$\int h \cdot \varphi \, dx = -\int f \cdot \varphi' \, dx \text{ for all } \varphi \in \mathcal{G}, \tag{3.6}$$

where the space of test functions is $\mathcal{G} = \{\varphi \in C^1([0,1], \mathbb{R}) \mid \varphi(0) = \varphi(1) = 0\}$. Write $h = Df$ in this case.

Example 3.2.3. The absolute value function $f(x) = |x|$ has as its derivative the step function $Df(x) = -1(x < 0), 1(x > 0)$. Note that the value of $Df(0)$ is not uniquely defined because Df is considered as an element of L^1.

Writing $g(x) = Df(x)$ for the step function just described, we see that g does not have a weak derivative in L^1. Indeed, this is true for any function with a jump discontinuity.

Using *mollifiers* one can show that any L^1 function f can be L^1 approximated by (infinitely) differentiable functions f_ϵ such that f'_ϵ approximates Df in L^1. This can be used to show that the usual product rule for derivatives holds for weak derivatives as well: $D(fg) = (Df) \cdot g + f \cdot (Dg)$, as long as f and g both have weak derivatives. The space of L^1 functions with a weak derivative in L^1 is denoted $W^{1,1}$ and is an important example of a *Sobolev space*. Here the norm is

$$\|f\|_{W^{1,1}} = \|f\|_{L^1} + \|Df\|_{L^1},$$

which can be viewed as an analogue of the definition of the C^1 norm in (3.3). Moreover, just as the C^1 unit ball is C^0 compact, so also the $W^{1,1}$ unit ball is L^1 compact, as we will see.

3.2.5 Kolmogorov–Riesz compactness theorem

In understanding compactness for subsets of function spaces, it is useful to recall that the Heine–Borel theorem can be generalised to arbitrary complete metric spaces as follows: a set is compact if and only if it is closed and *totally bounded*. In particular, for Banach spaces, pre-compactness is equivalent to being totally bounded.

The Arzelà–Ascoli theorem gives a necessary and sufficient condition for a set in C^0 to be totally bounded (and hence pre-compact). A similar result in the L^p spaces is the Kolmogorov–Riesz compactness theorem – an expository account of this theorem and its relationship to the Arzelà–Ascoli theorem is given in a recent paper by H. Hanche–Olsen and H. Holden, 'The Kolmogorov–Riesz compactness theorem' (*Expo. Math.* **28** (2010), 385–394).

In our setting (where we consider L^p spaces with respect to a *finite* measure), the Kolmogorov–Riesz theorem can be stated as follows: a set $\mathcal{F} \subset L^p$ is totally bounded (in the L^p norm) if and only if

1. $\mathcal{F}$ is bounded, and
2. for every $\varepsilon > 0$ there is $\delta > 0$ such that $\|f \circ T_\gamma - f\|_p < \varepsilon$ for every $f \in \mathcal{F}$ and $|\gamma| < \delta$, where $T_\gamma \colon x \mapsto x + \gamma$.

In other words, to go from bounded to totally bounded one needs the added condition that small changes to the argument result in (uniformly) small changes in the function, with respect to the L^p norm.

Roughly speaking the idea is that if a set can be "approximately embedded" into a totally bounded set, then it must itself be totally bounded – this is Lemma 1 in the paper referred to above. Then the condition on $f \circ T_\rho - f$ for $f \in \mathcal{F}$ allows the set $\mathcal{F}$ to be "approximately embedded" into a bounded set in $\mathbb{R}^n$ by averaging f over small neighbourhoods in its domain. This is of course a very rough description and one should read the paper for the complete proof and precise formulation of what it means to be "approximately embedded".

3.2.6 Bounded variation and Helly's theorem

One can use the Kolmogorov–Riesz theorem to show that $W^{1,1}$ is compactly embedded in L^1. (This is a special case of the Rellich–Kondrachov theorem.) However, since functions with jump discontinuities are not in $W^{1,1}$, we want to use a bigger function space in order to study spectral properties of the transfer operator.

The definition of weak derivative can be generalised if one is willing to allow Df to live somewhere besides L^1. Recall that we want Df to satisfy

$$\int (Df) \cdot \varphi \, dx = -\int f \cdot \varphi' \, dx$$

for every test function $\varphi \in \mathcal{G}$, the space of C^1 functions on the interval that vanish at the endpoints. The left-hand side defines a linear functional

$\mathcal{G} \to \mathbb{R}$, and given any $f \in L^1$ we may define Df as such a linear functional by setting

$$(Df)(\varphi) = -\int f \cdot \varphi' \, dx.$$

If $f \notin W^{1,1}$, this functional is not given by integration against an L^1 function, but now the definition makes sense for any $f \in L^1$. Moreover, the space of linear functionals on $\mathcal{G}$ carries a natural norm: the norm of $\ell \colon \mathcal{G} \to \mathbb{R}$ is

$$\|\ell\|_{\mathcal{G}^*} = \sup\{|\ell(\varphi)| \mid \varphi \in \mathcal{G}, \|\varphi\|_{C^0} \leq 1\}.$$

A functional ℓ is continuous if and only if $\|\ell\| < \infty$. One can show that $\|Df\|_{\mathcal{G}^*} = |f|_{BV}$, and so

$$BV = \{f \in L^1 \mid \|Df\|_{\mathcal{G}^*} < \infty\}.$$

The BV norm can be written as $\|f\|_{BV} = \|f\|_{L^1} + \|Df\|_{\mathcal{G}^*}$. Note that BV is exactly the set of functions $f \in L^1$ for which Df is a *continuous* linear functional on $\mathcal{G}$.

Helly's selection theorem states that BV is compactly embedded in L^1. (This is not to be confused with Helly's theorem in geometry.) This is a consequence of the Kolmogorov–Riesz compactness theorem, because a relatively straightforward computation shows that

$$\|f \circ T_\gamma - f\|_{L^1} \leq |f|_{BV} |\gamma|.$$

(See Lemma 11 and Theorem 12 in the paper of Hanche–Olsen and Holden referenced above.) We remark that one can also give a direct proof following the hint given in Footnote 8 of Keller and Liverani's 'A spectral gap for a one-dimensional lattice of coupled piecewise expanding interval maps': given $f \in BV$, let f_n be the step function that is constant on each dyadic interval $[k, k+1]/2^n$, with value equal to the average of f on that interval. Then the functions f_n approach f in L^1, and the problem reduces to finding a suitable subsequence of step functions.

3.3 Expanding interval maps

3.3.1 General strategy

Now we consider general piecewise expanding interval maps T. The map T is assumed to be C^2 on each of finitely many intervals whose union is $X = [0, 1]$

– these are called the *basic intervals* for T. Moreover, we assume that $\lambda > 1$ is such that $|T'(x)| \geq \lambda$ for every $x \in X$.

Our goal is to show that the transfer operator for such maps has a spectral gap when it acts on suitable Banach spaces. Existence of a spectral gap can be interpreted as the statement that apart from functions which are densities of absolutely continuous invariant measures (and hence are fixed by P_T), the transfer operator acts as a contraction on a certain space of functions; the mechanism driving this contractive property is the fact that T expands distances on the phase space $[0, 1]$. We note that the action of P_T on L^1 satisfies

$$\begin{aligned} \|P_T\varphi\|_1 &= \sup\left\{\int (P_T\varphi)\cdot\psi\,dx \,\Big|\, \psi \in L^\infty, \|\psi\|_\infty \leq 1\right\} \\ &= \sup\left\{\int \varphi\cdot(\psi\circ T)\,dx \,\Big|\, \psi \in L^\infty, \|\psi\|_\infty \leq 1\right\} \\ &\leq \|\varphi\|_1. \end{aligned} \tag{3.7}$$

In fact, (3.7) holds for any measurable transformation T that is *non-singular* – that is, T does not map a set of positive Lebesgue measure into a set of zero measure. Non-singular maps are precisely those maps for which every $\psi \in L^\infty$ has $\|\psi \circ T\|_\infty \leq \|\psi\|_\infty$. In other words, non-singularity of T implies that the Koopman operator does not expand distances in L^∞, which in turn implies that the transfer operator does not expand distances in L^1. However, (3.7) is not enough to deduce any information on decay of correlations for T, because the contraction is not strict.

In fact, (3.7) does not even let us deduce the existence of an absolutely continuous invariant measure. How might we hope to find such a measure? Recall the proof of the Krylov–Bogolyubov theorem, which establishes the existence of an invariant measure for a continuous map on a compact metric space (though there is no mention of absolute continuity): one starts with a measure μ that is not necessarily invariant, and then considers the sequence of Cesàro averages $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \mu \circ T^{-k}$. Any limit point of this sequence is an invariant measure, and compactness of the space of measures shows that such limit points exist.

In our setting we want an *absolutely continuous* invariant measure, which means we should play the same game on the set of density functions: starting with the constant function 1, representing Lebesgue measure, we may consider the sequence

$$\varphi_n = \frac{1}{n} \sum_{k=0}^{n-1} P_T^k 1. \tag{3.8}$$

If $\varphi_{n_j} \to \varphi \in L^1$, then $d\mu = \varphi\, dx$ defines an invariant measure μ, which is an acip. (Note that $\int \varphi_n\, dx = 1$ and $\varphi_n \geq 0$ for all n.) But how do we obtain a convergent subsequence? Thanks to (3.7) we know that every φ_n is contained in the unit ball in L^1 – but this ball is not compact.

The solution is to consider an auxiliary Banach space $\mathcal{B} \subset L^1$ such that the unit ball of $\mathcal{B}$ is relatively compact in L^1. If $\mathcal{B}$ can be chosen such that the sequence φ_n is uniformly bounded in the $\mathcal{B}$-norm, then relative compactness implies the existence of a subsequence that converges (in L^1) to some $\varphi \in L^1$, which is the desired density. (Indeed, it is often the case that $\varphi \in \mathcal{B}$.)

For the doubling map, which we studied earlier, the appropriate Banach space to use was the space of Lipschitz functions, whose unit ball embeds compactly into L^1 by the Arzelà–Ascoli theorem. However, this choice does not fare so well for general piecewise expanding interval maps.

Say that the map T is *full-branched* if $T(J_i) = [0,1]$ for each basic interval J_i. If T is not full-branched, then one can choose points x_1, x_2 that are arbitrarily close together but have different numbers of pre-images, and so in particular the quantities $\sum_{y \in T^{-1}(x_j)} |T'(y)|^{-1}$ for $j = 1, 2$ do not approach each other as $x_1 \to x_2$. This means that $P_T\mathbf{1}$ has a discontinuity at the endpoints of a non-full branch of T, and so the space of continuous functions is not P_T-invariant.

We deal with the situation by replacing the space of Lipschitz functions with a different space, which *is* invariant under the action of P_T.

3.3.2 Functions of bounded variation

We recall some more facts about functions of bounded variation, which we discussed earlier. Recall that the total variation of a function $\varphi\colon [0,1] \to \mathbb{C}$ is

$$|\varphi|_{BV} = \sup \left\{ \sum_{k=1}^{n} |\varphi(x_k) - \varphi(x_{k-1})| \,\middle|\, 0 = x_0 < x_1 < \cdots < x_n = 1 \right\}, \tag{3.9}$$

and that φ has bounded variation if $|\varphi|_{BV} < \infty$. We denote by BV the vector space of such functions. A useful example to keep in mind is the following:

Given any $\alpha \geq 0$, the function $\varphi_\alpha(x) = x^\alpha \sin(1/x)$ is defined on $(0, 1]$ and can be extended to $[0, 1]$ by $\varphi_\alpha(0) = 0$. It has bounded variation if and only if $\alpha > 1$.

Remark 3.3.1. A bounded variation function is continuous except perhaps on a countable set of jump discontinuities, and differentiable Lebesgue-a.e. (Think of the examples just mentioned – the function φ_α is continuous at 0 as long as $\alpha > 0$, and is differentiable at 0 precisely when $\alpha > 1$, that is, when it is of bounded variation.)

The total variation as defined in (3.9) is a semi-norm on BV. We want to think of BV as a subspace of L^1, but we must be careful to remember that elements of L^1 are *equivalence classes* of functions (mod zero w.r.t. Lebesgue measure), and note that the quantity in (3.9) depends on which representative of the equivalence class we choose. Thus to define $|\cdot|_{BV}$ on L^1 we put (abusing notation slightly)

$$|\varphi|_{BV} = \inf\{|\hat{\varphi}|_{BV} \mid \varphi = \hat{\varphi} \text{ Lebesgue-a.e.}\}. \tag{3.10}$$

An alternate approach that allows us to avoid this step is to define the BV-semi-norm through integration: it can be shown that (3.9) is equivalent to

$$|\varphi|_{BV} = \sup\left\{\left|\int_{[0,1]} \varphi \cdot g' \, dx\right| \mid g \in \mathcal{G}\right\}, \tag{3.11}$$

where $\mathcal{G} = \{g \in C^1([0, 1], \mathbb{C}) \mid \|g\|_\infty \leq 1, g(0) = g(1) = 0\}$. The idea behind this equivalence is the following.

- When φ is differentiable, (3.9) is equivalent to $|\varphi|_{BV} = \int_{[0,1]} |\varphi'| \, dx$.
- Choosing $g \in \mathcal{G}$ such that $\varphi' \cdot g \approx |\varphi'|$, one gets $\int |\varphi'| \, dx \approx \varphi' \cdot g \, dx$.
- Integrating by parts yields the expression in (3.11).

Although the expression (3.11) does not make the heuristic interpretation of "total variation" as obvious as (3.9) does, it nevertheless has two important advantages over that definition:

1. it does not depend on the choice of representative function in an equivalence class of L^1, and so allows us to define $|\cdot|_{BV}$ on L^1 without an extra step along the lines of (3.10);

2. it generalises more readily to functions on higher-dimensional domains.

As with the Lipschitz semi-norm that we used last time for the doubling map, we can define a BV-norm by adding the L^1-norm to the BV-semi-norm:

$$\|\varphi\|_{BV} = \|\varphi\|_1 + |\varphi|_{BV}.$$

The space of BV functions is appropriate for us to study because its unit ball is relatively compact in L^1 – this is Helly's selection theorem, which states that if $\varphi_n \in BV$ is such that $\|\varphi_n\|_{BV}$ is uniformly bounded, then there is $\varphi \in BV$ such that $\varphi_{n_j} \xrightarrow{L^1} \varphi$ for some subsequence n_j.

In particular, if we can show that the sequence φ_n defined in (3.8) is uniformly bounded in the BV norm, then Helly's theorem will yield a BV limit point φ, and the measure μ defined by $d\mu = \varphi\, dx$ will be an acip for T.

3.3.3 A Lasota–Yorke inequality

In order to proceed further, we must investigate the properties of the transfer operator P_T with respect to the BV norm. Along the way we will see that BV is invariant under P_T. We give an argument using the definition (3.11) to derive a bound that was first given by A. Lasota and J. Yorke in a 1974 paper – the argument there is equivalent to the one here, but uses the definition (3.9).

Given a function $g \in \mathcal{G}$, we need to estimate $\int (P_T\varphi) \cdot g'\, dx$. To this end we recall that by the definition of the transfer operator, we have

$$\int (P_T\varphi) \cdot g'\, dx = \int \varphi \cdot (g' \circ T)\, dx = \int \varphi \cdot (g \circ T)' \cdot (T')^{-1}\, dx,$$

where the second equality is valid because T is differentiable at all but finitely many points. Recalling the definition (3.11), this gives

$$|P_T\varphi|_{BV} \leq \sup \left\{ \left| \int \varphi \cdot (g \circ T)' \cdot (T')^{-1}\, dx \right| \;\middle|\; g \in \mathcal{G} \right\}. \tag{3.12}$$

It is tempting to try and use the bound $|T'(x)| \geq \lambda$ to conclude that this quantity is $\leq \lambda^{-1} \sup \left\{ \left| \int \varphi \cdot (g \circ T)'\, dx \right| \;\middle|\; g \in \mathcal{G} \right\}$, but we must take care – the argument of the integrand may vary, and so we cannot proceed quite so directly. Rather, we use the identity

$$\frac{d}{dx}\left(\frac{g \circ T}{T'}\right) = (g \circ T)'(T')^{-1} - (g \circ T)\frac{T''}{(T')^2}$$

to obtain

$$\left|\int \varphi \cdot (g \circ T)' \cdot (T')^{-1}\, dx\right| \le \left|\int \varphi \left(\frac{g \circ T}{T'}\right)' dx\right| + \int |\varphi| \cdot |g \circ T| \cdot \frac{|T''|}{|T'|^2}\, dx$$
$$\le \lambda^{-1} \left|\int \varphi \tilde{g}'\, dx\right| + K\|\varphi\|_1,$$

where $\tilde{g} = \lambda \frac{g \circ T}{T'}$ has $\|g\|_\infty \le 1$ and $K = \max(|T''|/|T'|^2)$. (Note that it is at this point that we use the hypothesis that T is C^2 – elsewhere only C^1 is used.)

If the map T were differentiable on the entire interval $[0,1]$ and fixed the endpoints, then we would have $\tilde{g} \in \mathcal{G}$ and so (3.12) would immediately imply $|P_T\varphi|_{BV} \le \lambda^{-1}|\varphi|_{BV} + K\|\varphi\|_1$. Unfortunately, as shown in Figure 3.2, $\tilde{g}$ is discontinuous at each of the discontinuity points of T, and moreover does not vanish at the endpoints of $[0,1]$ if those endpoints are not fixed by T. Thus we must be more careful.

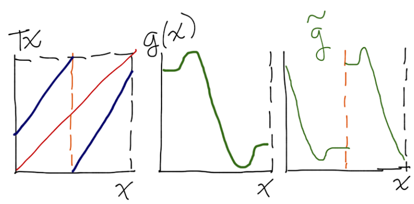


Figure 3.2: $\tilde{g}$ may not be in $\mathcal{G}$.

The idea is to approximate $\tilde{g}$ with functions from $\mathcal{G}$, as shown in Figure 3.3. Let $0 = b_0 < b_1 < \cdots < b_n = 1$ be the endpoints of the intervals on which the map T is C^2. Given $\varepsilon > 0$, let $h\colon [0,1] \to \mathbb{C}$ be a continuous function such that $h(0) = h(1) = 0$, $h(x) = \tilde{g}(x)$ when $|x - b_k| \ge \varepsilon$ for each k, and h is linear on $B(b_k, \varepsilon)$.

Finally, let $\tilde{h} \in \mathcal{G}$ be close to h in the uniform metric and agree with h except on an ε^2-neighbourhood of each point where h is non-differentiable.

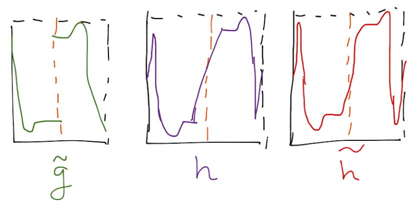


Figure 3.3: Approximating $\tilde{g}$ with elements of $\mathcal{G}$.

We get

$$\begin{aligned}\int \varphi \cdot \tilde{g}' \, dx &\leq \int \varphi \cdot \tilde{h}' \, dx + \int \varphi \cdot |\tilde{h}' - \tilde{g}'| \, dx \\ &\leq |\varphi|_{BV} + \sum_{k=0}^{n} \Bigg(\int_{B(b_k,\varepsilon)} \varphi \cdot |\tilde{h}'| \, dx \\ &\qquad\qquad + \int_{B(b_k,\varepsilon)} \varphi \cdot |\tilde{g}'| \, dx \Bigg).\end{aligned} \tag{3.13}$$

The second integral in the sum goes to 0 as $\varepsilon \to 0$. (This uses the assumption that $T' \in L^1$.) For the first integral, we use the fact that $h' = \frac{1}{2\varepsilon}(\tilde{g}(b_k + \varepsilon) - \tilde{g}(b_k - \varepsilon))$ to conclude that as $\varepsilon \to 0$, the integral goes to $\varphi(b_k)|\tilde{g}(b_k^+) - \tilde{g}(b_k^-)|$, where $\varphi(b_k)$ is understood as $\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{B(b_k,\varepsilon)} \varphi \, dx$, so that in particular we choose the representative of the L^1-equivalence class that minimises the total variation, as in (3.10).

Since $\|g\|_\infty \leq 1$ and $g(0) = g(1) = 0$, we conclude that

$$\sum_{k=0}^{n} \int_{B(b_k,\varepsilon)} \varphi \cdot |\tilde{h}'| \, dx \leq \sum_{k=1}^{n} |\varphi(b_{k-1})| + |\varphi(b_k)|. \tag{3.14}$$

We can bound this sum in terms of $|\varphi|_{BV}$ and $\|\varphi\|_1$. Let $m_k = \inf_{x \in [b_{k-1}, b_k]} |\varphi(x)|$, then

$$|\varphi(b_{k-1})| + |\varphi(b_k)| \leq 2m_k + \big|\varphi|_{[b_{k-1},b_k]}\big|_{BV},$$

as suggested by Figure 3.4.

Moreover, $\int_{[b_{k-1},b_k]} |\varphi| \, dx \geq m_k(b_k - b_{k-1}) \geq m_k \Delta$, where $\Delta = \min_k(b_k -$

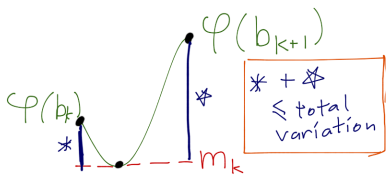


Figure 3.4: Bounding $|\varphi(b_{k-1})| + |\varphi(b_k)|$.

b_{k-1}), and so we can sum over k to get

$$\sum_{k=1}^{n} |\varphi(b_{k-1})| + |\varphi(b_k)| \leq 2\Delta^{-1}\|\varphi\|_1 + |\varphi|_{BV}.$$

Together with (3.13) and (3.14), this gives

$$\int \varphi \cdot \tilde{g}'\,dx \leq 2|\varphi|_{BV} + 2\Delta^{-1}\|\varphi\|_1,$$

so that (3.12) and the discussion following it gives us

$$|P_T\varphi|_{BV} \leq 2\lambda^{-1}|\varphi|_{BV} + (2\Delta^{-1} + K)\|\varphi\|_1.$$

In terms of the BV norm we have

$$\|P_T\varphi\|_{BV} \leq 2\lambda^{-1}\|\varphi\|_{BV} + (2\Delta^{-1} + K + 1)\|\varphi\|_1;$$

using the assumption that $\lambda > 2$, we can write this in the form

$$\|P_T\varphi\|_{BV} \leq r\|\varphi\|_{BV} + R\|\varphi\|_1 \tag{3.15}$$

for $r \in (0,1)$ and $R > 0$. This is a *Lasota–Yorke inequality*, and turns out to have important implications for the statistical properties of the map T.

3.3.4 Existence of an acip

Now we can return to the sequence φ_n defined in (3.8) as $\frac{1}{n}\sum_{k=0}^{n-1} P_T^k 1$, and show that it is uniformly bounded in BV. Indeed, iterating the Lasota–Yorke inequality (3.15) gives

$$\begin{aligned}\|P_T^2\varphi\|_{BV} &\le r\|P_T\varphi\|_{BV} + R\|P_T\varphi\|_1 \\ &\le r^2\|\varphi\|_{BV} + (1+r)R\|\varphi\|_1,\end{aligned}$$

where we use the inequality $\|P_T\varphi\|_1 \le \|\varphi\|_1$ from (3.7). Writing $\bar{R} = R(1 + r + r^2 + \cdots) = R(1-r)^{-1}$, we have by induction

$$\|P_T^k\varphi\|_{BV} \le r^k\|\varphi\|_{BV} + \bar{R}\|\varphi\|_1. \tag{3.16}$$

In particular, we conclude that the sequence φ_n is uniformly bounded in BV, since

$$\|\varphi_n\|_{BV} \le r^n + \bar{R} \le 1 + \bar{R}.$$

As discussed above, Helly's theorem shows that there is $\varphi \in BV$ such that $\varphi_{n_j} \xrightarrow{L^1} \varphi$ for some subsequence n_j, and the measure μ defined by $d\mu = \varphi\,dx$ is an acip for T.

Note that this proves the *existence* of an acip for T, but it does not prove *uniqueness*. For the doubling map there is only one acip, Lebesgue measure, but for other piecewise expanding interval maps there may be more than one. For example, the map shown in Figure 3.5 has two ergodic acips, one supported on $[0, 1/2]$ and the other supported on $[1/2, 1]$.

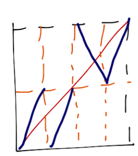


Figure 3.5: Non-uniqueness of an acip.

3.3.5 The spectrum of the transfer operator

The Lasota–Yorke inequality (3.15) also lets us deduce spectral information about P_T. First we observe that by the spectral radius formula and the iterated inequality (3.16), the spectral radius of $P_T \colon BV \to BV$ is bounded above by the inequality

$$\rho(P_T) = \lim_{n\to\infty} \|P_T^n\|^{1/n} \leq \lim_{n\to\infty} (r^n + \bar{R})^{1/n} = 1,$$

where we use the fact that $\|\varphi\|_1 \leq \|\varphi\|_{BV}$. The previous section shows that $1 \in \sigma(P_T)$, and we conclude that $\rho(P_T) = 1$.

In fact, one can also use the Lasota–Yorke inequality to show that the essential spectral radius of P_T is $\leq r < 1$, so that $\sigma(P_T)$ only has finitely many elements outside of $B(0, r)$. Once it is shown that the only element of the spectrum lying on the unit circle is 1, and that 1 is a simple eigenvalue, it follows that we have exponential decay of correlations.

Chapter 4

Cones

4.1 Non-equilibrium dynamics

Let $T : X \to X$ be a map. It induces a map T_* on probability measures defined by $(T_*\mu)(A) = \mu(T^{-1}A)$ for measurable sets $A \subset X$.

Example 4.1.1. $\mu = \delta_p$, $T_*\mu = \delta_{T(p)}$.

Example 4.1.2. $\mu = \frac{1}{k}\sum_{i=1}^{k}\delta_{p_i}$ then $T_*\mu = \frac{1}{k}\sum_{i=1}^{k}\delta_{T(p_i)}$ provided that T is one-to-one.

An invariant measure is some kind of equilibrium state. Does T admit any invariant probability measures? Yes— if X is a compact metric space and if T is continuous. (This is a consequence on the Krylov-Bogoliubov theorem.) If T is continuous then it is possible that there is no invariant Borel probability measures.

Within the class of invariant measures, one can ask whether T admits any "important" measures? By "important," we mean measures with some physicality property, which leads to the Sinai-Ruelle-Bowen (SRB) measures considered in the hyperbolicity lectures.

Rather than considering iterates of a single map, we could also consider a situation where the dynamics change with time. That is, we let $\{f_i\}_{i=1}^{\infty}$ be a sequence of maps on X and study the compositions $f_n \circ \cdots \circ f_1$.

In this case it is not clear what "invariant measure" should mean, and so rather than search for invariant measures we try to characterise the correlation decay properties discussed earlier in an alternate way. The object of study is the memory loss from the "time-dependent" dynamics. Trajectories coalesce (in the case of contractive systems) or we may use statistical properties, in the case of systems with hyperbolic or expanding behavior.

Recall that when we consider iterates of a single map, density functions ρ are transformed according to the Perron–Frobenius operator (the transfer operator) P_T. Thus given a density function ρ_0 at time 0, the density function at time n is $\rho_n = P_T^n \rho_0$. In the time-dependent (non-equilibrium) case one defines ρ_n by

$$\rho_n = P_{f_n} \circ \cdots \circ P_{f_1} \rho_0$$

where P_{f_i} is the Perron-Frobenius operator associated with the map f_i.

Definition 4.1.3. We say that the time-dependent dynamical system $\{f_i\}_{i=1}^{\infty}$ exhibits *exponential loss of memory* in the statistical sense if there exists an $\alpha > 0$ such that

$$\int |\rho_n - \hat{\rho}_n| dm \leq Ce^{-\alpha n}$$

for all probability density functions $\rho_0, \hat{\rho}_0$ in some suitable class and absolutely continuous with respect to the reference measure m.

4.2 Convex cones

A useful tool for establishing memory loss are the notions of 'convex cones' and 'Hilbert metric', which we now introduce. Let V be a vector space over the reals. Ultimately we will be most interested in the case when V is a function space, such as L^1 or BV, but for now we make the definitions in the general context.

Definition 4.2.1. A subset $C \subset V$ is a *convex cone* (or *positive cone*) if

1. $C \cap (-C) = \emptyset$;

2. $\lambda C = C$ for each $\lambda > 0$;

3. C is convex; and

4. for all $f, g \in C$ and $\alpha \in \mathbb{R}$, we have the following property: if $\alpha_n \to \alpha$ and $g - \alpha_n f \in C$ for every n, then $g - \alpha f \in C \cup \{0\}$.

The first three conditions are very geometric and in some sense guarantee that C "looks like a cone should look". The last condition is more topological; if V is a topological vector space and $C \cup \{0\}$ is a closed subset of V, then this condition holds, but we stress that the condition itself is actually weaker than this and is phrased without reference to any topology on V.

Remark 4.2.2. The relation $\leq$ on V defined by $\phi \leq \psi$ if and only if $\psi - \phi \in C \cup \{0\}$ is a partial order that is compatible with the algebraic structure on V.

Example 4.2.3. Let $V = BV([0,1], \mathbb{R})$ be the space of all real-valued functions on the unit interval with bounded variation, and let $C = \{\varphi \in V \mid \varphi \geq 0, \varphi \not\equiv 0\}$. Then C is a convex cone.

We see immediately from this example that the notion of convex cone is relevant to the sorts of questions we want to ask about invariant measures of a dynamical system, because this set C is exactly the set of density functions that arises when we are searching for an absolutely continuous invariant measure.

This suggests that we will ultimately want to consider the action of some operator $L\colon C \to C$, and in particular may want to find a fixed point of this action (for a suitable operator L). One of the most powerful methods for finding a fixed point is to find a metric in which L acts as a contraction, and this is accomplished by the *Hilbert metric*, which we now introduce.

Definition 4.2.4. Fix a convex cone $C \subset V$. Given $\varphi, \psi \in C$, let

$$\begin{aligned} \beta(\varphi, \psi) &= \inf\{\mu > 0 \mid \mu\varphi - \psi \in C\}, \\ \alpha(\varphi, \psi) &= \sup\{\lambda > 0 \mid \psi - \lambda\varphi \in C\}, \end{aligned} \tag{4.1}$$

with $\alpha = 0$ and/or $\beta = \infty$ if the corresponding set is empty. The *cone distance* between φ and ψ is

$$d_C(\varphi, \psi) = \log\left(\frac{\beta(\varphi, \psi)}{\alpha(\varphi, \psi)}\right). \tag{4.2}$$

The distance d_C is also called the *Hilbert (projective) metric.*

Several remarks are now in order. First we observe that although V may be infinite-dimensional, the distance $d_C(\varphi, \psi)$ is completely determined in terms of the two-dimensional subspace spanned by φ and ψ, and in particular by the points shown in Figure 4.1 – in the figure, the lines $0A$ and $0B$ are the boundary of this two-dimensional cross-section of C. The lines $0X$ and $Y\psi$ are parallel, as are the lines $0A$ and ψX; then we have

$$\alpha = \frac{|\psi Y|}{|0\varphi|} \text{ and } \beta = \frac{|0X|}{|0\varphi|}.$$

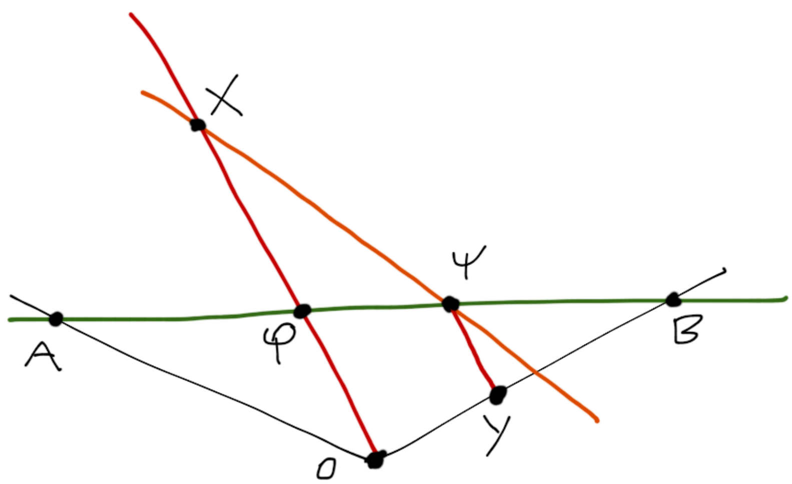


Figure 4.1: Determining the cone distance between φ and ψ.

An alternate description of d_C is available in terms of this more geometric description. Let ℓ be the line through φ and ψ, and let A, B be the points where this line intersects the boundary of C. We see from Figure 4.1 that the triangles $BY\psi$ and $B0\varphi$ are similar, so

$$\alpha = \frac{|\psi Y|}{|0\varphi|} = \frac{|B\psi|}{|B\varphi|}.$$

Furthermore, $\varphi 0A$ and $\varphi X\psi$ are similar, so

$$\beta = \frac{|0X|}{|0\varphi|} = 1 + \frac{|\varphi X|}{|0\varphi|} = 1 + \frac{|\psi\varphi|}{|A\varphi|} = \frac{|A\psi|}{|A\varphi|}.$$

Thus d_C can be given in terms of the cross-ratio of the points φ, ψ, A, B:

$$\frac{\beta}{\alpha} = \frac{|A\psi|}{|A\varphi|}\frac{|B\varphi|}{|B\psi|} = (\varphi, \psi; A, B).$$

We have

$$d_C(\varphi, \psi) = \log(\varphi, \psi; A, B). \tag{4.3}$$

Note that it is possible that the line ℓ does not intersect the boundary of C twice; this corresponds to the case when either $\alpha = 0$ or $\beta = \infty$ (or both) in (4.1), and in this case $d_C(\varphi, \psi) = \infty$.

This situation occurs, for example, when we take $V = BV([0,1], \mathbb{R})$ and C as in the example above, and consider $\varphi, \psi \in C$ with disjoint supports – that is, $\varphi(x)\psi(x) = 0$ for all x. In this case $\alpha = 0$ and $\beta = \infty$ so the cone distance between φ and ψ is infinite.

Because of this phenomenon, d_C is not a true metric. Moreover, we observe that d_C is projective: $d_C(\varphi, \lambda\varphi) = 0$ for every $\lambda > 0$.

An important property of the Hilbert metric is the following theorem, due to Birkhoff, which states that a linear map from one convex cone to another is a contraction whenever its image has finite diameter.

Theorem 4.2.5. *Let $C_1 \subset V_1$ and $C_2 \subset V_2$ be convex cones, and let $L\colon V_1 \to V_2$ be a linear map such that $L(C_1) \subset C_2$. (This is a sort of 'positivity' condition.) Let*

$$\Delta = \sup_{\hat{\varphi}, \hat{\psi} \in L(C_1)} d_{C_2}(\hat{\varphi}, \hat{\psi}).$$

Then for all $\varphi, \psi \in C_1$, we have

$$d_{C_2}(L\varphi, L\psi) \leq \tanh\left(\frac{\Delta}{4}\right) d_{C_1}(\varphi, \psi), \tag{4.4}$$

where we use the convention that $\tanh \infty = 1$.

We also want to relate d_C to a more familiar norm. Say that a norm $\|\cdot\|$ on V is *adapted* if the following is true: whenever $\varphi, \psi \in V$ are such that $\varphi - \psi \in C$ and $\varphi + \psi \in C$, we have $\|\psi\| \leq \|\varphi\|$.

Example 4.2.6. On BV, the L^1 norm is adapted, but the BV norm is not.

The following lemma, due to Liverani, Saussol, and Vaienti, relates the cone metric to an adapted norm.

Lemma 4.2.7. *Let $\|\cdot\|$ be an adapted norm on V and $C \subset V$ a convex cone. Then for all $\varphi, \psi \in C$ with $\|\varphi\| = \|\psi\| > 0$, we have*

$$\|\varphi - \psi\| \le \left(e^{d_C(\varphi,\psi)} - 1\right) \|\varphi\|. \tag{4.5}$$

Proof. If $d_C(\phi, \psi) = \infty$, we are done. Otherwise, $d_C(\phi, \psi)$ is finite, and in that case, $d_C(\phi, \psi) = \log \frac{\beta}{\alpha}$, where $\alpha\phi \le \psi \le \beta\phi$. This implies that

$$\alpha||\phi|| \le ||\psi|| \le \beta||\phi||$$

since $||\cdot||$ is adapted. This gives us that $\alpha \le 1 \le \beta$. We therefore have

$$(\alpha - \beta)\phi \le (\alpha - 1)\phi \le \psi - \phi \le (\beta - 1)\phi \le (\beta - \alpha)\phi.$$

Since the norm is adapted, $||\psi - \phi|| \le ||(\beta - \alpha)\phi||$. Just pull out the $\beta - \alpha$ and exponentiate in order to obtain the inequality. □

4.3 Perron–Frobenius theorem

Before returning to our discussion of dynamical systems and density functions, we see how convex cones and the Hilbert metric can be used to obtain an explicit estimate on the rate of convergence in the Perron–Frobenius theorem.

4.3.1 The theorem

We start by recalling the statement of the Perron–Frobenius theorem. Let A be a $d \times d$ stochastic matrix, where here we use this to mean that the entries of A are non-negative, and every column sums to 1: $A_{ij} \in [0, 1]$ for all i, j, and $\sum_{i=1}^{d} A_{ij} = 1$ for all j. Thus the columns of A are probability vectors.

Such a matrix A describes a weighted random walk on d sites: if the walker is presently at site j, then A_{ij} gives the probability that he will move to site i at the next step. Thus if we interpret a probability vector v as giving the probability of the walker being at site j with probability v_j, then $v \mapsto Av$ gives the evolution of this probability under one step of the random walk.

Now one version of the Perron–Frobenius theorem is as follows: If A is a stochastic matrix with $A > 0$ (that is, $A_{ij} > 0$ for all i, j), then there is exactly one probability vector π that is an eigenvector for A. Moreover, the eigenvalue associated to this eigenvector is 1, the eigenvalue 1 is simple, and all other eigenvalues have modulus < 1. In particular, given any $v \in [0, \infty)^2$ we have $A^n v \to \pi$ exponentially quickly.

The eigenvector π is the stationary distribution for the random walk (Markov chain) given by A, and the convergence result states that any initial distribution converges to the stationary distribution under iteration of the process.

The assumption that $A > 0$ is quite strong: for the random walk, this says that the walker can get from any site to any other site in a single step. A more general condition is that A is *primitive*: that is, there exists $N \in \mathbb{N}$ such that $A^N > 0$. This says that there is a time N such that by taking N steps, the walker can get from any site to any other site. The same result as above holds in this case too.

In fact, the result holds in the even more general case when A is *irreducible*: for every i, j there exists N such that $(A^N)_{ij} > 0$. This says that the walker can get from every site to every other site, but removes the assumption that there is a single time N that works for all site. For example, consider a random walk on a chessboard, where the walker is allowed to move one square horizontally or vertically at each step. Then for a sufficiently large even value of N, the walker can get from any white square to any other white square, but to get to a black square requires an odd value of N.

4.3.2 A cone and a metric

As stated above, the Perron–Frobenius theorem does not give any result on the rate with which $A^n v$ converges to π. One way to give an estimate on this rate is to use convex cones and the Hilbert metric (this also gives a proof of the theorem).

Let $\mathcal{C}$ be the convex cone $[0, \infty)^d \subset \mathbb{R}^d$. We want an estimate on the diameter of $A(\mathcal{C})$ in the Hilbert metric $d_{\mathcal{C}}$. Recall that this metric is given by $d_{\mathcal{C}}(v, w) = \log(\beta/\alpha)$, where

$$\beta = \inf\{\mu > 0 \mid \mu v - w \in \mathcal{C}\},$$
$$\alpha = \sup\{\lambda > 0 \mid w - \lambda v \in \mathcal{C}\}.$$

Another way of interpreting the cone $\mathcal{C}$ is in terms of the partial order it places on V, which is given by $v \preceq w \Leftrightarrow w - v \in \mathcal{C} \cup \{0\}$. We see that β and

α can be characterised as

$$\alpha = \sup\{\lambda \mid \lambda w \preceq v\}, \qquad \beta = \inf\{\mu \mid v \preceq \mu w\}.$$

In our present example, we see that the cone $\mathcal{C} = [0, \infty)^d$ induces the partial order $v \preceq w \Leftrightarrow v_i \leq w_i \ \forall i$. Thus

$$\alpha = \sup\{\lambda \mid \lambda w_i \leq v_i \ \forall i\} = \min_{1 \leq i \leq d} \frac{v_i}{w_i}, \tag{4.6}$$

and similarly $\beta = \max_{1 \leq i \leq d} \frac{v_i}{w_i}$.

4.3.3 Diameter of $A(\mathcal{C})$

Now we need to determine the diameter Δ of $A(\mathcal{C})$ in the Hilbert metric $d_{\mathcal{C}}$. If $\Delta < \infty$, then the theorem of Birkhoff from the previous post will imply that $d_{\mathcal{C}}$ contracts distances by a factor of $\tanh(\Delta/4) < 1$.

Let e_i be the standard basis vectors in $\mathbb{R}^d$. Because $d_{\mathcal{C}}$ is projective we can compute Δ by considering $d_{\mathcal{C}}(Av, Aw)$ where $\sum v_i = \sum w_j = 1$. Using the triangle inequality, we have

$$\begin{aligned} d_{\mathcal{C}}(Av, Aw) &= d_{\mathcal{C}}\left(A \sum v_i e_i, A \sum w_j e_j\right) = d_{\mathcal{C}}\left(\sum v_i (Ae_i), \sum w_j (Ae_j)\right) \\ &\leq \sum_{i,j} v_i w_j d_{\mathcal{C}}(Ae_i, Ae_j) \leq \max_{i,j} d_{\mathcal{C}}(Ae_i, Ae_j), \end{aligned}$$

so it suffices to consider $d_{\mathcal{C}}(Ae_i, Ae_j)$ for $1 \leq i, j \leq d$. But Ae_i is just the ith column of the matrix A, so writing $A = [v^1 \cdots v^n]$, where v^i is the ith column vector, we see that

$$\Delta \leq \max_{i,j} d_{\mathcal{C}}(v^i, v^j). \tag{4.7}$$

4.3.4 Contraction under multiplication by A

Now we have a very concrete procedure for estimating the amount of contraction in the $d_{\mathcal{C}}$ metric under multiplication by A:

1. estimate Δ using (4.7) and the expression for $d_{\mathcal{C}}$ in (4.6) and the discussion preceding it;

2. get a contraction rate of $\tanh(\Delta/4) < 1$.

From (4.6) and the discussion preceding it, the distance $d_{\mathcal{C}}(v^i, v^j)$ is given as

$$d_{\mathcal{C}}(v^i, v^j) = \log \beta - \log \alpha = \log \left(\max_{1 \le k \le d} \frac{v_k^i}{v_k^j} \cdot \max_{1 \le k \le d} \frac{v_k^j}{v_k^i} \right). \tag{4.8}$$

Let $\Lambda = \tanh(\Delta/4)$. To write an explicit estimate for Λ, we use

$$\Lambda = \frac{e^{\Delta/4} - e^{-\Delta/4}}{e^{\Delta/4} + e^{-\Delta/4}} = \frac{1 - e^{-\Delta/2}}{1 + e^{-\Delta/2}} \le \frac{1-s}{1+s}, \tag{4.9}$$

where $s < 1$ is any estimate we can obtain satisfying $e^{-\Delta/2} \ge s$. From (4.8) and (4.7), we have

$$e^{-\Delta/2} \ge \max_{i,j} \sqrt{ \min_k \left(\frac{v_k^i}{v_k^j} \right) \min_k \left(\frac{v_k^j}{v_k^i} \right) } =: s. \tag{4.10}$$

This allows us to obtain estimates on $d_{\mathcal{C}}(A^n v, A^n w)$. However, we want to estimate $d(A^n v, A^n w)$ in a more familiar metric, such as one coming from a norm. We can relate the two by observing that if $v, w \in (0,1]^d$, then

$$\begin{aligned} d_{\mathcal{C}}(v, w) &= \log \max_k \left(\frac{v_k}{w_k} \right) + \log \max_k \left(\frac{w_k}{v_k} \right) \\ &\ge \max_k |\log v_k - \log w_k| \ge \max_k |v_k - w_k| = \|v - w\|_{L^\infty}, \end{aligned}$$

where the last inequality uses the fact that log has derivative ≥ 1 on $(0,1]$. Since A maps the unit simplex to itself (because A is stochastic), we see that

$$\|A^n v - A^n w\|_{L^\infty} \le d_{\mathcal{C}}(A^n v, A^n w) \le C \Lambda^n, \tag{4.11}$$

where Λ is given by (4.9) and (4.10), and where we can take either $C = d_{\mathcal{C}}(v, w)$ or $C = \Delta/\Lambda$ (since $d_{\mathcal{C}}(Av, Aw) \le \Delta$), whichever gives the better bound. Since all norms on $\mathbb{R}^d$ are equivalent, we have a similar bound in any norm.

4.3.5 Nonnegative matrices

The analysis in the previous section required A to be positive ($A_{ij} > 0$ for all i, j). A more general condition is that A is nonnegative and primitive: that is, $A_{ij} \ge 0$ for all i, j, and moreover there exists N such that $A^N > 0$.

If $A_{ij} = 0$ for some i, j, then it is easy to see from the calculations in the previous section that $A(\mathcal{C})$ has infinite diameter in the Hilbert metric, so the above arguments do not apply directly. However, they do apply to A^N when $A^N > 0$, and so we fix N for which this is true, and we obtain $\Lambda < 1$ such that $d_{\mathcal{C}}(A^N v, A^N w) \leq \Lambda d_{\mathcal{C}}(v, w)$ for all $v, w \in \mathcal{C}$.

Moreover, let $L \in \mathbb{R}$ be such that $\|A^r\| \leq L$ for all $0 \leq r < N$. Then for any $n \in \mathbb{N}$ we can write $A^n = A^{kN+r}$ for some $0 \leq r < N$, so that

$$\|A^n v - A^n w\| = \|A^r (A^{kN} v - A^{kN} w)\| \leq LC\Lambda^k,$$

where C is as in (4.11). Thus we conclude that asymptotically, $A^n v$ approaches the eigenvector with contraction rate $\Lambda^{1/N}$.

To see this in action, consider a Markov chain with transition matrix

$$A = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}.$$

That is, from the first state the walker transitions to either state with probability 1/2, while from the second state the walker always returns to the first state. Since the transition from the second state to itself is forbidden, $A(\mathcal{C})$ has infinite diameter. However, the two-step transition matrix is

$$A^2 = \begin{pmatrix} \frac{3}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

for which we can compute

$$s = \sqrt{\frac{1/4}{1/2} \cdot \frac{1/2}{3/4}} = \frac{1}{\sqrt{6}} \quad \Rightarrow \quad \Lambda \leq \frac{\sqrt{6}-1}{\sqrt{6}+1}.$$

Thus the estimate on A^2 gives us a definite rate of contraction, which the estimate from A does not.

It can be useful to use the estimate on A^N even when $A > 0$. For example, if we consider the Markov chain with transition matrix

$$A = \begin{pmatrix} \frac{1}{5} & \frac{9}{10} \\ \frac{4}{5} & \frac{1}{10} \end{pmatrix},$$

then we have

$$s = \sqrt{\frac{1/5}{9/10} \cdot \frac{1/10}{4/5}} = \sqrt{\frac{2}{9} \cdot \frac{1}{8}} = \frac{1}{6} \quad \Rightarrow \quad \Lambda \leq \frac{5}{7} \approx .714$$

as the rate of contraction, while considering

$$A^2 = \begin{pmatrix} \frac{19}{25} & \frac{27}{100} \\ \frac{6}{25} & \frac{73}{100} \end{pmatrix}$$

gives

$$s = \sqrt{\frac{27/100}{19/25} \cdot \frac{6/25}{73/100}} \approx .3418 \quad \Rightarrow \quad \Lambda \leq \frac{.6582}{1.3418} \approx .4906 \approx (.7)^2,$$

a better estimate than we obtained from considering A itself.

4.4 Non-equilibrium open systems

Convex cones and the Hilbert metric are well suited to studying nonequilibrium open systems. Consider the following setting. Let X be a Riemannian manifold, λ volume on X, and $\hat{f}_i \colon X \to X$ a diffeomorphism. For $m \in \mathbb{N}$, let $\hat{F}_m = \hat{f}_m \circ \cdots \circ \hat{f}_1$. This is a nonequilibrium closed system. (Nonequilibrium because the map changes at each time step, closed because every point can be iterated arbitrarily many times.)

Now consider sets $H_j \subset X$, which we interpret as a "hole" at time j. The *time-m survivor set* is

$$S_m = X \setminus \bigcup_{i=1}^{m} \hat{F}_i^{-1}(H_i),$$

the set of points that do not fall into a hole before time m. Let $F_m = \hat{F}_m|_{S_m}$. We refer to the pair (F_m, H_m) as a *nonequilibrium open dynamical system.*

We would like an analogue of decay of correlations for such systems. Let φ_0, ψ_0 be two probability density functions on X, and evolve these under (F_m, H_m). We expect that $\|\varphi_t\|_{L^1(\lambda)} < 1$ because there is a positive probability of falling into a hole.

Let $\hat{\mathcal{P}}_j$ be the Perron–Frobenius operator for the closed system $\hat{f}_j$ (with respect to λ). Then to the open system f_j we can associate the operator

$$\mathcal{P}_j(\varphi) = \hat{\mathcal{P}}_j(\varphi) 1_{X \setminus H_j}.$$

Definition 4.4.1. We say that (F_m, H_m) exhibits *conditional memory loss in the statistical sense* if for all suitably chosen φ_0, ψ_0, we have

$$\lim_{t\to\infty} \left\| \frac{\varphi_t}{\|\varphi_t\|_{L^1(\lambda)}} - \frac{\psi_t}{\|\psi_t\|_{L^1(\lambda)}} \right\|_{L^1(\lambda)} = 0.$$

The idea of this definition is that before comparing the probabilities, we need to first condition on the event that the trajectory survives.

In the one-dimensional case, our space is $[0,1]$ and λ Lebesgue is our reference measure.

Definition 4.4.2. An *underlying closed system* M consists of maps $\hat{g}$ on $[0,1]$ such that there exists a finite partition $\mathcal{A}(\hat{g})$ of $[0,1]$ into subintervals such that for each interval $J \in \mathcal{A}(\hat{g})$, $\hat{g}$ is C^2 on J and extends to a C^2 map on $\overline{J}$, and

$$\max_{J\in\mathcal{A}(\hat{g})} \sup_{x\in J} |(\hat{g}')^{-1}| \leq s < 1.$$

There are no Markov assumptions on $\mathcal{A}$. Note that expansion alone is not enough for memory loss— two subsystems that never "communicate" are an example. Such a system is not ergodic for Lebesgue measure.

Definition 4.4.3. (a type of mixing) Let $z_1 \in (0,1)$ and $z_2 \in (1,\infty)$. We say a map $\hat{g} : [0,1] \to [0,1]$ belongs to $E(z_1, z_2)$ if for every partition Q of $[0,1]$ into equal subintervals there exists a time $T_{mix}(Q, z_1, z_2)$ such that for $J_i, J_j \in Q$

$$z_1 \leq \frac{\lambda(J_i \cap \hat{g}^{-k} J_i)}{\lambda(J_i)\lambda(J_j)} \leq z_2$$

for all $k \geq T_{mix}(Q, z_1, z_2)$.

Definition 4.4.4. We say $\hat{f} \in M$ is a δ-perturbation of $\hat{g} \in M$ and we write $\hat{f} \in N(\hat{g}, \delta)$ if $\delta < \frac{1}{4}\left(\min_{1\leq i\leq j-1} x_{i+1} - x_i\right)$ where x_i are partition points associated with the base map $\hat{g}$; if $\{0 = y_1, \ldots, y_k = 1\}$ is the set of partition points associated with $\hat{f}$, then $|y_i - x_i| < \delta$ for all $i = 1, 2, \ldots, k$; and if $\xi_{\hat{f}_{\hat{g}}}$ maps each $[x_i, x_{i+1}]$ onto $[y_i, y_{i+1}]$ in an affine way; then we have

$$\left|\left|\hat{f} \circ \xi_{\hat{f}_{\hat{g}}} - \hat{g}\right|\right|_{C^2(J)} < \delta$$

for all $J \in \mathcal{A}(\hat{g})$.

Why the restriction on δ? It defines the basis for a topology.

Let our space of densities D be non-negative functions that integrate to one. What should holes look like? We need to constrain the complexity somehow. Say each H_j is a finite union of open intervals and that the number is uniformly bounded in j.

Theorem 4.4.5. *(Mohapatra, Ott 2014) Let* $\widehat{g} \in M \cap E(z_1, z_2)$, $L \in \mathbb{N}$. *There exists a* $\delta_0 > 0$, $\epsilon > 0$ *and* $\Lambda < 1$ *such that for any sequence* $\{\widehat{f_i}\}$ *in* $N(\widehat{g}, \delta_0)$ *and sequence of holes* $\{H_j\}$ *where each* H_j *consists of at most* L *open subintervals, and* $\lambda(H_j) \leq \epsilon_0$. *Then there exists a convex cone* C_a *in* $BV([0,1], \mathbb{R})$ *and a constant* $C_1 > 0$ *such that for all* $\phi, \psi \in C_a \cap D$ *we have*

$$||R_{F_m}(\phi) - R_{F_m}(\psi)||_{L^1(\lambda)} \leq C_1 \Lambda^m.$$

The theorem uses the notation that $\widehat{F}_m = \widehat{f}_m \circ \cdots \circ \widehat{f}_1$. Taking the hats off takes into account the loss of trajectories due to holes. We must also define the operators L and R. L gives the evolution of densities under the open dynamics F_m:

$$L_{F_m}(\phi)x = \sum_{z : F_m(z) = x} \frac{\phi(z)}{|F'_m(z)|}.$$

This is an "open transfer operator" analogous to the Perron-Frobenius operator. R_{F_m} is given by renormalizing:

$$R_{F_m}(\phi) = \frac{L_{F_m}}{||L_{F_m}(\phi)||_{L^1(\lambda)}}.$$

R_{F_m} is not linear, a fact that we must juggle in the results to come.

Note that the theorem does not hold for all BV densities one ϵ_0 is fixed (since BV functions can be supported on arbitrarily small sets).

Proof. Define a good cone and show it contracts.

$$C_a = \{\phi \in BV : \phi \geq 0, \phi \neq 0, Var(\phi) \leq aE[\phi|Q]\}$$

where

$$E[\phi|Q](x) = \frac{1}{\lambda(J)} \int_J \phi d\lambda$$

for $x \in J$. We want to show that for some time T, L_{F_T} takes C_a strictly into itself: for some $\sigma < 1$, $L_{F_T} C_a \subset C_{\sigma a}$. We control variation using a Lasota-Yorke-type inequality: $Var(L_{F_T}(\phi)) \leq \theta^T Var(\phi) + K_{LY}||\phi||_{L^1(\lambda)}$ for some $\theta < 1$ and all $\phi \geq 0$, $\phi \in BV$.

Bound $E[L_{F_T}|Q]$ from below for $\phi \in C_a$ by using the mixing assumption on $\widehat{g}$. Bound the diameter of $L_{F_T}(C_a)$. By the Birkhoff theorem, this implies

L_{F_T} contracts C_a. $||\cdot||_{L^1(\lambda)}$ is adapted to C_a so contractions may be carried over. In addition,

$$d_{C_a}(L_{F_T}(\phi), L_{F_T}(\psi)) = d_{C_a}(R_{F_T}(\phi), R_{F_T}(\psi))$$

by the projectivity of the cone. □

Chapter 5

Statistical Physics

Some supplemental material to these notes is found in the slides for Renato Feres's talks, which are on the summer school website at

`http://www.math.uh.edu/~climenha/2014-school.html`

5.1 Deterministic mechanical systems

Definition 5.1.1. A *mechanical system* consists of

1. a body B together with a distribution of mass given by a measure μ on B;

2. a configuration manifold M such that every configuration of the body B corresponds to some point $q \in M$;

3. a position map $\phi\colon M \times B \to \mathbb{R}^3$ such that $\phi(q, b)$ gives the position in $\mathbb{R}^3$ of the point $b \in B$ when the system is in configuration q.

Given a tangent vector $v \in T_qM$, say that a curve γ in M represents v if $\gamma(0) = q$ and $\gamma'(0) = v$. Then given $b \in B$, we write

$$v(b) = \frac{d}{ds}\bigg|_{s=0} \Phi(\gamma(s), b).$$

Define the inner product $\langle v, w \rangle_q = \int_B v(b)w(b)d\mu(b)$, and interpret $\frac{1}{2}||v||_q^2$ as the kinetic energy of the system in state q. Non-collision movement is just geodesic motion on M. Assume in the gas case that particles only interact with the boundary, and not with one another.

Example 5.1.2. A one-dimensional billiard system with two masses has $B = \{1,2\}$, $\mu(1) = m_1, \mu(2) = m_2$. $M = \{(x_1, x_2) \in [0, L]^2 : x_1 \leq x_2\}$ is a triangle (a "manifold with corners") where each boundary represents a different type of collision. Define the metric using $\langle u, v\rangle = m_1 u_1 v_1 + m_2 u_2 v_2$. The energy of the system in state q is $E(q, v) = \frac{1}{2}||v||_q^2 = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2$.

Assume energy and momentum are conserved in collisions. The collision map $C_q : T_q M \to T_q M$ preserves the norm (the energy) and the momentum (the inner product); thus if we write $\hat{v} = C_q(v)$, we get

$$m_1 v_1^2 + m_2 v_2^2 = m_1 \hat{v}_1^2 + m_2 \hat{v}_2^2,$$
$$m_1 v_1 + m_2 v_2 = m_1 \hat{v}_1 + m_2 \hat{v}_2.$$

By rescaling the triangle one can assume that $m_1 = m_2 = 1$ and thus the collision map gives a new velocity reflected across the normal line of the boundary.

Example 5.1.3. Consider two particles moving on a semi-infinite line $[0, \infty)$. Let $m_1 < m_2$ be the masses, with mass m_1 closer to 0 and both masses moving left, towards the wall at 0. Question: How many collisions occur? One can show that the total number of collisions is bounded above by $\lceil (\arctan \sqrt{m_1/m_2})^{-1}\pi \rceil$.

5.2 Billiard systems

We will be mostly interested in Euclidean billiard systems, where we fix a region $M \subset \mathbb{R}^n$ in which a particle moves freely, with its velocity reflecting around $T_q M$ when it hits the boundary ∂M. We will be particularly interested in open billiards: An *open system* is one for which part of the boundary $\Gamma \subset \partial M$ is "open to the world," (as opposed to topologically open). We are looking for the return map for Γ: how long does a particle spend inside the system?

To describe the reflection at the boundary more precisely, let n_q be the inward-pointing normal at $q \in \partial M$, and then consider the set $N^+ = \{(q, v) : q \in \partial M, |v| = 1, v \cdot n_q > 0\}$ of inward-pointing vectors, and the set $N^- = \{(q, v) : q \in \partial M, |v| = 1, v \cdot n_q < 0\}$ of outward-pointing vectors. Identify N^+ and N^- via reflection over the $T_q M$; that is, identify N^- with N^+ for the continuation of a particle path. Let N_Γ be the tangent vectors corresponding to the escaping region $\Gamma \subset \partial M$.

Example 5.2.1. *Cook's billiard* In two dimensions, a particle with mass m_1 and a barrier with mass $m_2 \gg m_1$; the barrier moves up and down in a limited vertical range (held by a perfectly flexible string). The top is open to particles moving in and out.

Let x, y be the coordinates of m_1 and z the height of m_2. Under proper rescaling we obtain a configuration manifold where the dynamics are regular billiard dynamics as a subset of $\mathbb{R}^3$: this is a cube with open top and a diagonal running from one top edge to the opposite bottom edge, representing the different positions of the bottom mass.

One can consider a particle moving in a channel ($\mathbb{R} \times [0,1]$) with random scattering off the boundaries by putting copies of the Cook billiard at microscopic scale along the boundary.

Example 5.2.2. Equilateral triangle with circular scatterers (centered at vertices of triangle, taking 'bites' out of the corners), and one edge open. "Trapped" paths form a Cantor set: one way of seeing this is to consider a light ray entering through the open edge, and consider the region illuminated it after 1 reflection, 2 reflections, etc.

Definition 5.2.3. The *Knudsen measure* on ∂M is invariant under the billiard map, and is defined

$$d\mu(q, v) = cv \cdot n_q dVol^{n-1}(q) dVol^{2n-2}(v),$$

where Vol^k denotes k-dimensional volume.

Example 5.2.4. If L is the circumference of a circle and θ the angle of impact, then $S_+^{n-1} \times \partial M = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, L]$ is a rectangle and $d\mu(r, \theta) = \frac{1}{2L} \cos\theta dr d\theta$.

For flat boundaries, trajectories do not expand, but inside a sphere parallel trajectories are scattered.

We now apply the Poincaré recurrence $B : N_\Gamma \to N_\Gamma$ for the first return to Γ. The restriction of μ to N_Γ is B-invariant, and B is well-defined for μ-almost every element. We can, in fact, define the expected number of collisions before a return and expected time of return for ergodic particles by $\frac{L}{e}$ and $\frac{A\pi}{e|v|}$, respectively, where e is the length of Γ.

5.3 Statistical properties of billiards

We will not introduce temperature quite yet. We think of the microscopic scale as completely separate from the macroscopic: once inside a microscopic cell, the macroscopic knows nothing until it comes out, and the microscopic cell knows only the particle's entry position and velocity. This allows the (macroscopically) smooth boundary to acquire scattering properties using, for example, Cook's billiard as a microscopic cell. Thus we will model the macroscopic process as a Markov chain, where the transition probabilities at collision times are determined by the microscopic structure of the boundary.

Definition 5.3.1. Let $V(r, v)$ be the exit velocity (for the microsystem) associated to entering at position $r \in [0, 1]$ with direction v. Then the *transition probability operator* P is

$$(Pf)v = \int_0^1 f(V(r, v))dr = E[f(V)|v].$$

where $f : S_+^{n-1} \to \mathbb{R}$ is bounded.

One can check that $(P\mathbf{1}_U)(v)$ is the probability that the reflected velocity V is in U given an initial velocity v.

The velocity V of an particle leaving a cell is a random variable $V(q, v)$ where the entry point q has a uniform distribution. We will use the shorthand $\mu(f) = \int f d\mu$ and $(\mu P)f = \mu(Pf) = \int Pf d\mu$. So $\delta_v P$ gives the distribution of the scattered trajectories with initial velocity v. General properties for P are that μ is *stationary* for P if $\mu P = \mu$. The measure $d\mu(v) = c \cos\theta dVol^{S_+^{n-1}}(v)$ is stationary for P.

Define P on $\mathcal{H} = L^2(S_+^{n-1}, \mu)$. P is a self-adjoint operator norm on $\mathcal{H}$: $\langle f, g\rangle = \int_{S_+^{n-1}} f\overline{g}d\mu$, and $\langle Pf, g\rangle = \langle f, Pg\rangle$. In particular P has real spectrum in $[-1, 1]$.

Our model system will be a disk of radius r times an interval, either $\mathbb{R}$ if infinite in both directions of $[0, 2L]$ if finite. Let s be the speed of the particle. The velocities $V_0, V_1, \ldots$ have the Markov property and the collisions form a random walk along the z direction, modeled by Brownian motion. Let Z_j be the displacement along the horizontal after collision j. Define $X_t = Z_0 + Z_1 + \cdots + Z_{N_t} +$ (small error). Our claim is that, asymptotically (for large L) the mean exit time τ is

$$\tau(L,r,s) = \begin{cases} \frac{L^2}{D} & \textit{if } n \geq 3 \\ \frac{1}{D}\frac{L^2}{\ln\left(\frac{L}{r}\right)} & \text{if } n = 2 \end{cases}$$

where D is the diffusion constant: the position of a random walker at time t follows a distribution $N(0, tD)$.

Let $a \gg 0$. Then $\tau(aL, r, s) = a^2\tau\left(L, \frac{r}{a}, as\right)$. This is because $\tau(L, r, s) = \tau(aL, ar, as)$ and $\tau(L, r, as) = \frac{1}{a}\tau(L, r, s)$. $\frac{r}{a}$ means a smaller channel, and as means faster particles. Thus as a scaled random walk, $Z_j^a = \frac{1}{a}Z_j$ and $X_{a,t} = Z_0^a + \cdots + Z_{N_{a,t}}^a$. Note that $N_{a,t}$ is the number of collisions, and it different from N_t. In fact, $N_{a,t} = N_{a^2t}$.

Theorem 5.3.2. *Central Limit Theorem. As $a \to \infty$, $X_{a,t} \to^{dist} N(0, tD)$. In addition, $X_{a,t} \to^{weak}$ Brownian motion with diffusion constant D.*

Here is how to compute D: tD is the variance of $X_{a,t}$, so $tD = \lim_{a\to\infty} E[X_{a,t}^2]$. Say D_0 is the diffusion constant for the i.i.d. case.

$$\frac{D}{D_0} = \int_{-1}^{1} \frac{1+\lambda}{1-\lambda} d\pi_z(\lambda),$$

the integral over the spectrum with respect to the projection-valued measure for self-adjoint operators.

$$\pi_z(U) = \frac{1}{||z||^2}\langle z, \pi(U)z\rangle$$

where π is the orthogonal projection onto the Hilbert space.

Example 5.3.3. For half-circles (concave relative to the particle),

$$\frac{D}{D_0} = \frac{1 - \frac{1}{4}\ln 3}{1 + \frac{1}{4}\ln 3} < 1.$$

5.4 Non-equilibrium systems

Under our current model, energy doesn't change in cells— we should take this into account in a non-equilibrium system, which is more thermodynamically interesting. We will alter Cook's billiard so that the momentum of a particle may increase or decrease: v and V ($\dot{x}_1$, $\dot{x}_2$) are observable variables in the upper-half plane. Then x_1, x_2, x_3 and $\dot{x}_3$ are hidden variables. Construct a Markov chain that assigns to each hidden variable a fixed probability distribution η. Now $N_H = \Gamma \times \mathbb{R}^{m-n}$ where there are n observable and m total dimensions, and η_H is the probability measure on N_H. The observable variables form $N_O = \mathbb{H}^n$, the upper-half plane of $\mathbb{R}^n$. The billiard map B takes $v \in N_O$, $z \in N_H$ based on η, and produces a new point in $N = N_H \times N_O$. V is the projection of this new point onto N_O.

What is the probability distribution of V?

$$\delta_V P = (\pi \circ B)_* \delta_v \otimes \eta_H.$$

This formula relies on the assumption that η_H is a *Gibbs state*, which means $E = E_O + E_H$ and

$$d\eta_H(q, w) = \rho_H(\epsilon_H) d\mu_H^{\epsilon_H}(q, w) d\epsilon_H$$

where ρ_H maximizes the Boltzmann entropy for constant ϵ_H, (q, w) are the velocities of the hidden variables, ϵ_H is the energy level, and $\mu_H^{\epsilon_H}$ is a reference measure. So $d\eta_H$ is constant along observable energy levels. The *Boltzmann entropy* is

$$H(\rho_H) = -\iint \rho_H(\epsilon) \ln \rho_H(\epsilon) d\mu_H^{\epsilon}(q, w) d\epsilon.$$

By Jensen's inequality, $\rho_0, \rho_1, \rho_2, \ldots$ must increase to ρ_H and, by Lagrange multipliers, imply that

$$\rho_H(\epsilon) = Ce^{-\beta_H \epsilon}.$$

Let P be the scattering operator defined using the (hidden) Gibbs state η_H with inverse temperature parameter $\beta_H = \frac{1}{kT}$. Let η_O be the Gibbs state on N_O with parameter $\beta_O = \beta_H$. (This equality characterizes thermal equilibrium.) Then η_O is stationary for P; that is, $\eta_O = \eta_O P$.

P is an operator on $L^2(N_O, \eta_O)$. Suppose we can prove that P has spectral gap: there exists a $\chi < 1$ such that for all $f \in L^2(N_O, \eta_O)$ with $\int f d\eta_O = 0$, $||Pf||_{L^2} < \chi ||f||_{L^2}$. Then for arbitrary distributions μ,

$$||\mu P^n - \eta_O||_{TV} \leq C_\mu \chi^n$$

where the *total variation norm* $||\mu - \nu||_{TV} = \sup\{|\mu(A) - \nu(A)| : A \in \mathcal{F}\}$.

Example 5.4.1. two-particle, one on string.

$$d\eta_H(x_1, \dot{x}_1) = \frac{I_{[0,l]}(x_1)}{l}\sqrt{\frac{m_1}{(m_1+m_2)\sigma^2 2\pi}}e^{-\frac{1}{2}\frac{m_1\dot{x}_1^2}{(m_1+m_2)\sigma^2}}dx_1 d\dot{x}_1$$

$$d\eta_O(\dot{x}_2) = \frac{m_2}{(m_1+m_2)\sigma}\dot{x}_2 e^{-\frac{1}{2}\frac{m_2\dot{x}_2^2}{(m_1+m_2)\sigma^2}}d\dot{x}_2$$

in this case, $\beta = \frac{1}{(m_1+m_2)\sigma^2}$.

P acting on $L^\infty((0,\infty), \eta_O)$ is a self-adjoint, norm one compact (Hibert-Schmidt) operator.

Conj: the spectral gap of the above example is about $4\frac{m_2}{m_1}$ assuming that $m_2 \ll m_1$. In addition, $P_{\frac{m_2}{m_1}} - id$ can be estimated by the Sturm-Liousville operator.

Thermodynamics are stationary but irreversible.

Example 5.4.2. Particle is accelerated by T_H bumper and slowed by T_C bumper: the particle transfers energy from one side to the other. One can construct a "motor" that takes advantage of the velocity disparity for each direction. This motor system is an example of Carnot thermodynamics and may be modeled by Brownian motion with drift.

Chapter 6

Symbolic dynamics and C^* algebras

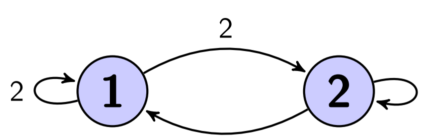
For slides from the lectures, please see the summer school homepage. Here we give a list of some of the examples discussed.

Example 6.0.3. Let

$$M = \begin{pmatrix} 3 & -2 & 3 \\ 0 & 1 & 0 \\ 3 & -3 & 3 \end{pmatrix} \sim \begin{pmatrix} 3 & & \\ & 1 & \\ & & 0 \end{pmatrix}$$

which implies that the cokernel of M is $\mathbb{Z}_3 \oplus \mathbb{Z}$.

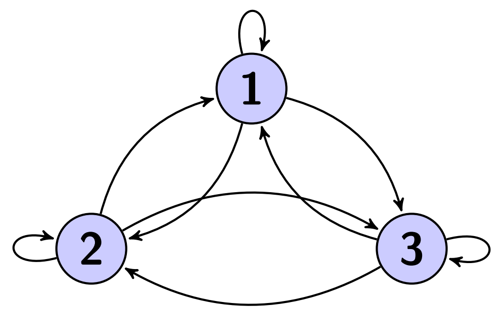
Example 6.0.4. E_1



$$A = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}, I - A^T = \begin{pmatrix} -1 & -1 \\ -2 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & \\ & 2 \end{pmatrix}$$

so the cokernel is $\mathbb{Z}_2$, $det(I - A^T) = -2 < 0$, and $(1, 1)^T \notin im(I - A^T)$.

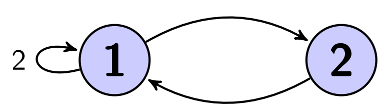
Example 6.0.5. E_2



$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, I - A^T \sim \begin{pmatrix} 1 & & \\ & 1 & \\ & & 2 \end{pmatrix}$$

so the cokernel is $\mathbb{Z}_2$ $det(I - A^T) = -2 < 0$, and $(1,1)^T \notin im(I - A^T)$, just as in E_1.

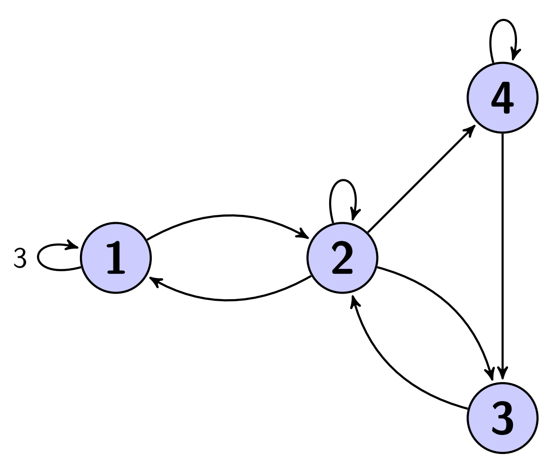
Example 6.0.6. E_3



$$A = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}, I - A^T \sim \begin{pmatrix} 1 & \\ & 2 \end{pmatrix}$$

Again, the cokernel is $\mathbb{Z}_2$, and the determinant of $I - A^T$ is negative, but $(1,1)^T \in im(I - A^T)$, which is not the case in the previous two examples.

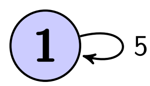
Example 6.0.7. E_4



$$A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, I - A^T \sim \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 2 \end{pmatrix}$$

E_4 has cokernel $\mathbb{Z}_2$ but the determinant of $I - A^T = 2 > 0$ is positive.

Example 6.0.8. E_5



$$A = (5), I - A^T \sim (4)$$

The cokernel is $\mathbb{Z}_4$ and $I - A^T = (-4)$ has negative determinant.

What conclusions can we draw from these examples? X_{E_1} and X_{E_2} are continuously orbit equivalent, $\left(C^*(E_1), \overline{D(E_1)}\right) \cong_* \left(C^*(E_2), \overline{D(E_2)}\right)$, and $(L_k(E_1), D(E_1)) \cong (L_k(E_2), D(E_2))$. On the other hand, even though $\left(C^*(E_3), \overline{D(E_3)}\right) \cong_* \left(C^*(E_4), \overline{D(E_4)}\right)$, the signs of the determinants are different so they are not continuously orbit equivalent. (We know nothing about $(L_k(E_3), D(E_3))$ and $(L_k(E_4), D(E_4))$.)

X_{E_1} and X_{E_3} are not continuously orbit equivalent, but they are flow equivalent, which we can see by transforming E_3 by reduction and outsplitting. We know, based on the signs of determinants, that transforming E_1 into E_4 will require a Cuntz splice— in fact, the sequence of operations is outamalgamation, Cuntz splice, and outsplitting.