Omri Sarig

Lecture Notes on Thermodynamic Formalism for Topological Markov Shifts

Penn State, Spring 2009

May 5, 2009

(Prepared using the Springer `svmono` author package)

Special thanks to Vaughn Climenhaga, Van Cyr, Yair Daon, Ai Hua Fan, Andrey Gogolev, David Koslicki, Pavlo Tsytsura, Jisang Yoo for alerting me to mistakes and typos in the text. The many typos which no doubt remain are my responsibility, and mine alone. If you find any, and have the time, please let me know. My email is omsarig@gmail.com.

O.M.S.

Contents

List of Notation

α	The natural partition, $\alpha = \{[a] : a \text{ is a states}\}$
$\underline{a}$	A word
$[\underline{a}]$	A cylinder
$a \xrightarrow{n} b$	a connects to b in n steps ($\exists \xi_i$ s.t. $(a, \xi_1, \ldots, \xi_{n-1}, b)$ is admissible)
$\overline{a_0, \ldots, a_{n-1}}$	$(\underline{a}, \underline{a}, \underline{a}, \ldots)$, where $\underline{a} = (a_0, \ldots, a_{n-1})$
$\mathscr{B}$	Borel σ–algebra
δ_y	Dirac measure at y: $\delta_y(E) = 1$ if $E \ni y$, $\delta_y(E) = 0$ otherwise
log	$\ln(= \log_e)$
L_ϕ	Ruelle's operator
$M^{\pm 1}$	A quantity with values in $[M^{-1}, M]$ ($a = M^{\pm 1} b \Leftrightarrow M^{-1} b \leq a \leq Mb$).
$\pm M$	A quantity with values in $[-M, M]$ ($a = b \pm M \Leftrightarrow \lvert a - b \rvert \leq M$).
$\mathbb{N}$	$\{1, 2, 3, \ldots\}$
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$(\Omega, \mathscr{F}, \mu)$	a general measure space
$O(1)$	A bounded quantity
$o(1)$	A quantity which tends to zero
1_E	The indicator function of E (equal to one on E, and zero otherwise)
$\mathfrak{P}_T(X)$	The T–invariant Borel probability measures on X
$P_G(\phi)$	The Gurevich pressure of ϕ
s.t.	such that
S	The set of states
TMS	Topological Markov Shift (one–sided, unless states otherwise)
$\sigma(\gamma)$	The smallest σ–algebra which contains γ
T	The left shift map
$\widehat{T}$	The transfer operator
$\mathrm{var}_n \phi$	$\sup\{\lvert \phi(x) - \phi(y) \rvert : x_i = y_i \ (i = 0, \ldots, n-1)\}$ ("n–th variation")
w.l.o.g.	without loss of generality
w.r.t	with respect to
$\mathscr{W}_n$	The collection of admissible words of length n
X	(usually) a TMS
x_m^n	$(x_m, x_{m+1}, \ldots, x_n)$
x_m^∞	$(x_m, x_{m+1}, \ldots)$
$Z_n(\phi, a)$	$\sum_{T^n x = x} \exp \phi_n(x) 1_{[a]}(x)$
$Z_n^*(\phi, a)$	$\sum_{T^n x = x} \exp \phi_n(x) 1_{[\varphi_a = n]}(x)$, $\varphi_a(x) := 1_{[a]}(x) \inf\{n \geq 1 : T^n(x) \in [a]\}$
TMS	Topological Markov shift
$*$	A place holder for any possible symbol
$\ll$	$\mu \ll \nu \Leftrightarrow (\nu(E) = 0 \Rightarrow \mu(E) = 0)$
$\sim$	$a_n \sim b_n \Leftrightarrow a_n / b_n \to 1$, $\mu \sim \nu \Leftrightarrow (\mu \ll \nu$ and $\nu \ll \mu)$
$\asymp$	$a_n \asymp b_n \Leftrightarrow \exists M, N\ [\forall n > N\, (M^{-1} \leq a_n / b_n \leq M)]$

Chapter 1
Introduction and basic definitions

In this chapter we give an overview of the basic concepts of the thermodynamic formalism, and explain the heuristics which led to their creation. The discussion is non-rigorous, and meant as a motivation for the mathemtical work in the following chapters.

1.1 What is the thermodynamic formalism

The aim of ergodic theory is to understand the stochastic behavior of deterministic dynamical systems $T : X \to X$. It does so by studying the ergodic invariant probability measures of the system. Given such a measure μ, the "ergodic theorems" provide quantitative information on the behavior of $T^i(x)$ as $i \to \infty$ for μ–almost every $x \in X$.

But the results are sensitive to the choice of the measure μ. Different ergodic invariant probability measures of the same dynamical system are always mutually singular (Problem 1.1), so what holds almost surely with respect to one measure will often be an event of measure zero with respect to the other.

Since it is very common for a dynamical system to have many different ergodic invariant measures, we are faced with the following problem: *which invariant measure to choose to analyze the system*?

A similar problem exists in statistical physics. Statistical physics aims at explaining the large scale behavior of a many–particle system (such as a gas) from the microscopic behavior of its constituents (e.g. the gas molecules). An important idea in this theory is that macroscopic quantities (heat, pressure, etc.) can be viewed as weighted averages of quantities which are defined in terms of the constituents of the system (the sum of the energies of molecules, or the sum of the forces exerted by the molecules hitting a piston during a short interval of time per unit area). The averaging is done over the collection of all possible states of the system. Usually, it is a weighted average.

Again we run into a problem: the value of weighted average depends on the weights. What weighting scheme should we use?

Of course physics has mathematics at an advantage, because it has a source of experimental data to rely on. We know experimentally what macroscopic behavior to expect: that is what the collection of empirical laws of thermodynamics are about. Thus we can, at least in principle, know which choices of weights lead to the right answer, and which do not. The great physicist J.W. Gibbs came up with a list of recipes for the averaging schemes which work in various scenarios.[1]

We describe one such scenario. Imagine a many–particle system A which is contact with another much larger system B so that

1. A and B can exchange energy, but not particles;
2. B is at equilibrium and has temperature T;
3. B is much larger than A, so that it contact with A does not affect its equilibrium state.

Such systems B are called *heat baths*.

Let X denote the set of all possible states of system A: each element of X is a complete list of all the positions and velocities of all the particles in A. Assume for simplicity that the set of all possible states is a finite set $\{s_1,\dots,s_N\}$.[2] We think of the actual state of the system as constantly fluctuating, and random. Gibbs' rule for the probabilistic weights of the different states at equilibrium is

$$\Pr(s_i) := \frac{1}{Z(\beta)}e^{-\beta U(s_i)}, \text{ where } Z(\beta) := \sum_{i=1}^{N} e^{-\beta U(s_i)} \text{ and } \beta = \frac{1}{k_B T}. \qquad (1.1)$$

Here $U(s_i)$ is the total energy of system A when it is at state s_i, and k_B is a universal constant ("Boltzman's constant"). The constant β is sometimes called "inverse temperature".

The probability distribution (1.1) is called by physicists the *canonical ensemble*. Mathematicians call it the *Gibbs measure*.

Ya. Sinai and D. Ruelle had the idea of importing the ideas of Gibbsian statistical physics to the theory of smooth dynamical systems, so as to come up with principles for choosing "natural" invariant measures for study in this context. The theory of these measures and their properties is now known under the name *Thermodynamic Formalism* (following a monograph with this name as a title by D. Ruelle [3]).

At first sight this idea may seem a bit unnatural. Gibbs devised his averaging schemes with the experimental facts of thermodynamics in mind. But many of the important examples in the theory of dynamical systems have their origins in geometry, or number theory, or algebra, and have no relation to physics. Even the examples

[1] Apparently this was anticipated by L. Boltzmann, see [2], §1.1.9.

[2] This assumption is certainly not true. But it could be argued that due to resolution problems there are only finitely many *distinguishable* states, so there is no harm in working with a finite discretization of X rather than with X itself. The procedure of passing from a continuum of states to a finite discrete set is called *coarse graining*.

which arise out of classical mechanics or celestial mechanics physics usually deal with a small number of particles (three is already too difficult) and are therefore outside the scope of thermodynamics ($\sim 10^{24}$ particles).

The undisputable fact is, however, that thermodynamic formalism *is* relevant to the study of many of these systems! Thermodynamic formalism is one of the most ubiquitous parts of ergodic theory, with numerous applications to complex analysis, geodesic flows, geometric measure theory, and the theory of dependent stochastic processes.

The question is how come the theory is so useful? The reason is quite simple, although somewhat technical: It so happens that thermodynamic selection principles produce measures m whose Jacobian functions (see chapter 2) have strong regularity properties (the logarithm of the Jacobian is up to a negative constant an "energy function"). The tools of the theory apply to all measures with "regular" Jacobians. The reason this is so useful is that measures with regular (even smooth) Jacobians arise naturally in many mathematical contexts, with or without connections to physics. Form the technical point of view

Thermodynamic formalism is the part of ergodic theory which studies measures under assumptions on the regularity of their Jacobian functions.

The insight of Sinai and Ruelle was that the analogy with physics can be used to explore such measures, even if the original setting has absolutely nothing to do with statistical mechanics.

1.2 The basic notions of the thermodynamic formalism

The purpose of thermodynamic formalism is to define and analyze, for a given dynamical system $T : X \to X$, probability measures which satisfy (1.1). But in the dynamical context it is impossible to use (1.1) directly, for reasons which will be explained below. Therefore one needs to invent weak variants of (1.1).

We explain the problem and the approaches for its solution by analyzing a simple example, where the analogy with physics is easy to understand. Let $T : X \to X$ be the following dynamical system:

1. $X = \{(x_0, x_1, x_2, \ldots) : x_i \in \{0,1\}\}$, with the topology generated by the cylinder sets $[a_0, \ldots, a_{n-1}] := \{x \in X : x_0 = a_0, \ldots, x_{n-1} = a_{n-1}\}$ $(n \in \mathbb{N}, a_i \in \{0,1\})$;
2. $T : X \to X$ is defined by $T(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots)$.

It is useful to think of X as the collection of all possible configurations of a one–sided one dimensional lattice with sites $0, 1, 2, \ldots$, where each site can be in one of two states: 0 or 1. Thus $x \in X$ is the configuration where site zero is in state x_0, site one is in state x_1, and so on.

To define an analogue of the canonical ensemble, we need an energy function. Imagine that the different sites of the lattice "interact". Let

$$U(x) = U(x_0 | x_1, x_2, x_3, \ldots)$$

denote the total “potential energy” at site zero due its “interaction” with the other sites. It is useful to think about this quantity as minus the energy one needs to invest to break site zero from the lattice and move it to infinity. Using this interpretation it is easy to see that the total potential energy of the first n–sites is

$$U(x) + U(Tx) + \cdots + U(T^{n-1}x).$$

(To break off the first n–sites we first need to invest $-U(x)$ to take away site zero, then to invest $-U(Tx)$ to take away site one, and so one.)

If we apply (1.1) blindly, we get the following “formula” for the canonical ensemble on X:

$$\Pr(x) = \frac{\text{“}\exp\left(-\beta \sum_{n=0}^{\infty} U(T^n x)\right)\text{”}}{\text{“}\sum_{x\in X} \exp\left(-\beta \sum_{n=0}^{\infty} U(T^n x)\right)\text{”}}. \tag{1.2}$$

But this is meaningless:

1. the numerator has no reason to converge;
2. the denominator has very little hope of converging, because it is a sum over the uncountable collection of $x \in X$;
3. finally, the approach of trying to define the probability of each $x \in X$ is inadequate for the purposes of defining a measure on an uncountable set: if our measure is non-atomic, then $\Pr(x) = 0$ for all x!

We see that the naîve approach for defining the canonical ensemble leads nowhere.

Mathematical physicists have come up with several ingenious ways of making sense of (1.2). These lead to several reasonable ways of constructing, or defining, “Gibbs measures” in our setting. We describe them briefly.

A word on terminology: The various definitions of “Gibbs measures” which are about to follow are not all equivalent. This could cause some confusion in the literature, because what is known in the dynamics community as a “Gibbs measure” or a “Gibbs state” is not what the mathematical physicists call a “Gibbs state”. To prevent the confusion I have chosen to use other terminology. Although widely used, this terminology is not universally accepted.

1.2.1 Dobrushin–Lanford–Ruelle measures

The definition is based on a heuristic due to Dobrushin.[3] Instead of trying to guess the formula for the Gibbs distribution Pr as a whole, let us try to guess the formula for the following *conditional* measures:

$$\Pr(y_0, y_1, \ldots, y_{N-1} | x_N, x_{N+1}, \ldots).$$

[3] The same definition was discovered independently by Lanford and Ruelle a year later.

By this we mean the conditional distribution of the configuration of the first N sites $(y_0,\dots,y_{N-1})$ given that site N is in state x_N, site $N+1$ is in state x_{N+1} etc.

The point of this exercise is that the problems which prevented us from defining the Gibbs distribution on X do not arise when trying to define the Gibbs distribution on $\{y\in X : y_N^\infty = x_N^\infty\}$: (a) the energy content of the first N–sites is finite (equal to $U+U\circ T+\cdots+U\circ T^{N-1}$), and (b) the normalization factor $Z(\beta)$ is a finite sum (ranging over the 2^N possibilities for $(x_0,\dots,x_N)$). Thus formula (1.1) does make sense for the conditional measures.

This suggests the following alternative definition of a "Gibbs measure":

Definition 1.1. Suppose $U : X\to\mathbb{R}$ is a function, and $\beta>0$. A probability measure m on X is called a *Dobrushin–Lanford–Ruelle (DLR) measure* with potential U and inverse temperature β if for every N,

$$m(x_0,\dots,x_{N-1}|x_N,x_{N+1},\dots) = \frac{\exp\left(-\beta\sum_{n=0}^{N-1}U(T^n x)\right)}{Z_N(\beta,x_N^\infty)} \text{ for a.e. } x, \tag{1.3}$$

where $Z_N(\beta,x_N^\infty)=\sum_{(x_0,\dots,x_{N-1})}\exp\left(-\beta\sum_{j=0}^{N-1}U(T^j x)\right)$ a normalization constant.

Equations (1.3) are called the *DLR equations*. DLR measures also appear in the literature under the name *Gibbs states*. The real power of Dobrushin's idea reveals itself when the definition is made for higher dimensional lattices (i.e. with $\mathbb{Z}^2$ or $\mathbb{Z}^3$ replacing $\mathbb{N}$). In this context DLR measures are often referred to as *Gibbsian Random Fields*.

Remark: It does *not* follow from the Carathéodory or Kolmogorov extensions theorems that the DLR equations define a measure. The existence of a measure satisfying (1.3) requires proof. It is also not clear (and in some cases not true) that equations (1.3) characterize the measure: there are U and β with more than one DLR measure.

1.2.2 Thermodynamic Limits

Here the idea is to approximate the (uncountable) configuration space X by finite (or at least countable) discretizations X_N, define the Gibbs distribution on X_N, and pass to the limit.

The most popular way of achieving this is to use *boundary conditions*. Fix $x\in X$ and consider the following (finite) set

$$X_N := T^{-N}\{x\} = \{(y_0,\dots,y_{N-1};x_0,x_1,x_3,\dots) : y_0,\dots,y_{N-1}\in\{0,1\}\},$$

which represents all possible configurations subject to the boundary condition that the site N is in state x_0, site $N+1$ is in state x_1 etc.

We determine the Gibbs distribution on this set using the following heuristic (what follows is not valid mathematically): For each $y\in T^{-N}\{x\}$,

$$
\begin{aligned}
\Pr(y) &= \frac{\text{“}\exp\left(-\beta\sum_{n=0}^{\infty} U(T^n y)\right)\text{”}}{\text{“}\sum_{T^N y = x}\exp\left(-\beta\sum_{n=0}^{\infty} U(T^n y)\right)\text{”}} \\
&= \frac{\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)}{\sum_{T^N y = x}\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)} \cdot \frac{\text{“}\exp\left(-\beta\sum_{n=N}^{\infty} U(T^n x)\right)\text{”}}{\text{“}\exp\left(-\beta\sum_{n=N}^{\infty} U(T^n x)\right)\text{”}} \\
&= \frac{\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)}{\sum_{T^N y = x}\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)}.
\end{aligned}
$$

This leads to the following definition. Let δ_y denote the point mass measure at y, defined by $\delta_y(E) = 1$ when $E \ni y$ and $\delta_y(E) = 0$ otherwise.

Definition 1.2. Suppose $x \in X$ and define for $N \in \mathbb{N}$,

$$
\nu_N^x := \frac{\sum_{T^N y = x}\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)\delta_y}{\sum_{T^N y = x}\exp\left(-\beta\sum_{n=0}^{N-1} U(T^n y)\right)}.
$$

Any weak star limit point of the sequence $\{\nu_N^x\}_{N\geq 1}$ is called a *thermodynamic limit*.

Thermodynamic limits are also called by some people “Gibbs states”.

In the example we are studying X is compact, so thermodynamic limits exist. But it is not clear whether they are unique, and whether they depend on x.

1.2.3 Equilibrium measures

Suppose $(U_1, \ldots, U_N)$ is a vector of real numbers, then the probability vector $(p_1, \ldots, p_N)$ given by $p_i := e^{-\beta U_i}/\sum_{i=1}^N e^{-\beta U_i}$ $(i = 1, \ldots, N)$ is the (unique) probability vector which brings to a minimum the quantity

$$
F := \sum_{i=1}^N p_i U_i - \frac{1}{\beta}\left(-\sum_{i=1}^N p_i \log p_i\right).
$$

This can be proved using Lagrange multipliers.

1. $U := \sum_{i=1}^N p_i U_i$ is called the *energy* of $\underline{p} = (p_1, \ldots, p_N)$;
2. $S := -\sum_{i=1}^N p_i \log p_i$ is the *entropy* of $\underline{p}$;[4]
3. and $F = U - \frac{1}{\beta} S$ is called the *Helmholtz free enegy* of $\underline{p}$ (or just “free energy”).

This variational characterization of the Gibbs distribution is often interpreted as saying that “nature minimizes the free energy at equilibrium”.

[4] Up to a constant which depends on the context. In information theory, the entropy is measured logarithms to base 2, not natural logarithms. It is dimensionless. In statistical physics, the average Boltzmann entropy of $\underline{p}$ is $k_B \sum p_i \log \frac{1}{p_i}$ where k_B (Boltzmann's constant) is a constant which gives S the dimensions of energy/temperature.

We now use these ideas to choose a measure on X, the configuration of the infinite lattice. It is natural to ask for a measure μ which minimizes the *free energy per site*:

$$\lim_{n\to\infty}\frac{1}{n}\left(\int\sum_{k=0}^{n-1}U(T^k x)d\mu-\frac{1}{\beta}\sum_{x_0,\dots,x_{n-1}}\mu[x_0,\dots,x_{n-1}]\log\frac{1}{\mu[x_0,\dots,x_{n-1}]}\right)$$

If μ is T–invariant, then the value of this limit is $\int Ud\mu-\frac{\log 2}{\beta}h_\mu(T)$, where $h_\mu(T)$ is the metric entropy of μ (we have used here Sinai's Generator Theorem). The $\log 2$ factor is because the metric entropy is defined using logarithms to the base 2.

We are led to the following definition: A T–invariant probability measure μ is called an *equilibrium measure* for the potential U and inverse temperature β, if it minimizes $\int Ud\mu-\frac{\log 2}{\beta}h_\mu(T)$.

In the dynamical context it is customary to set $\phi:=-\frac{\beta}{\log 2}U$ and write $\int Ud\mu-\frac{\log 2}{\beta}h_\mu(T)=-\frac{\log 2}{\beta}\left(h_\mu(T)+\int\phi d\mu\right)$. Minimizing this expression amounts to maximizing the term in the brackets (because $\beta>0$). The definition then becomes

Definition 1.3. An invariant probability measure μ is called an *equilibrium measure* for a function $\phi:X\to\mathbb{R}$ if it maximizes $h_\mu(T)+\int\phi d\mu$.

Remark: An equilibrium measure is invariant by definition. But DLR measures and thermodynamic limits do not need to be T–invariant.

1.2.4 Gibbs measures in the sense of Bowen

This is the definition most frequently used by dynamicists. It is taken from Bowen's influential monograph [1]. We continue to use the notation $\phi:=-\frac{\beta}{\log 2}\beta U$.

Definition 1.4. A *Gibbs measure (in the sense of Bowen)* for a function $\phi:X\to\mathbb{R}$ is an invariant probability measure μ for which there are constants $M>1$ and $P\in\mathbb{R}$ s.t. the following holds for every cylinder $[a_0,\dots,a_{n-1}]$ and $n\in\mathbb{N}$:

$$M^{-1}\le\frac{\mu[a_0,\dots,a_{n-1}]}{\exp\left(\sum_{k=0}^{n-1}\phi(T^k x)-nP\right)}\le M\quad(x\in[a_0,\dots,a_{n-1}]).$$

This should be thought of as the approximate identity

$$m[a_0,\dots,a_{n-1}]\asymp\frac{1}{Z_n}\exp\sum_{k=0}^{n-1}\phi(T^k x),\;\; Z_n:=e^{nP}$$

which is certainly in the spirit of (1.1). The point of this definition is that it is indeed sometimes possible to find measure like that.

Remark: The term "Gibbs measure (in the sense of Bowen)" is not standard at all. In the dynamical literature such measures are simply called "Gibbs measures".

1.3 Topological Markov shifts

We will develop the thermodynamic formalism in the context of *topological Markov shifts*, as this is the context for which the theory is most developed.

1.3.1 Definition and examples

Let S be a countable set and $\mathbb{A} = (t_{ij})_{S\times S}$ be a matrix of zeroes and ones with no columns or rows which are all zeroes. Out of this data one can construct a directed graph with set of vertices S, and set of edges $\{a \to b : t_{ab} = 1\}$. The set of all one–sided infinite allowed paths on the graph is called a topological Markov shift. Here is the formal definition:

Definition 1.5 (Topological Markov shift). The *topological Markov shift (TMS)* with set of *states* S and *transition matrix* $\mathbb{A} = (t_{ab})_{S\times S}$ is the set

$$X := \left\{x \in S^{\mathbb{N}_0} : t_{x_i x_{i+1}} = 1\,,\ \forall i \geq 0\right\} \quad (\mathbb{N}_0 = \{0,1,2,\ldots\}),$$

equipped with the topology generated by the collection of *cylinders*

$$[a_0,\ldots,a_{n-1}] := \{x \in X : x_i = a_i\,,\ 0 \leq i \leq n-1\} \quad (n \in \mathbb{N}, a_0,\ldots,a_{n-1} \in S),$$

and endowed with the action of the *left shift* map $T : (x_0,x_1,\ldots) \mapsto (x_1,x_2,\ldots)$.

The topology of a TMS is metrizable, for example it is given by the metric $d(x,y) := 2^{-\min\{n:x_n\neq y_n\}}$ when $x \neq y$, and $d(x,y) := 0$ when $x = y$. With respect to this metric, X is a complete and separable metric space. But it need not be compact, or even locally compact (Problem 1.2).

Definition 1.6 (Subshift of Finite Type (SFT)). A TMS with a finite set of states is called a *subshift of finite type (SFT)*.

Equivalently, a subshift of finite type is a compact topological Markov shift.

Definition 1.7 (Words). A *word* on an alphabet S is an element $(a_0,\ldots,a_{n-1}) \in S^n$ $(n \in \mathbb{N})$. The *length* of the word is n. A word is called *admissible* (w.r.t. to a transition matrix $\mathbb{A}$) if the cylinder it defines is non-empty.

Equivalently, $(a_0,\ldots,a_{n-1})$ is admissible iff $t_{a_0,a_1}t_{a_1,a_2}\cdots t_{a_{n-2},a_{n-1}} = 1$.

Any $x \in X$ gives rise to the following collection of admissible words:

$$x_m^n = (x_m,x_{m+1},\ldots,x_n) \quad (m < n).$$

The word x_0^{n-1} is called the *n–prefix* of x.

Recall that a continuous map T on a topological space X is called *topologically transitive* if it has a point x with a dense (forward) orbit,[5] and *topologically mixing* if for every pair of open sets $U,V \subseteq X$ there is a number $N(U,V) \in \mathbb{N}$ s.t.

$$n \geq N(U,V) \Longrightarrow U \cap T^{-n}V \neq \varnothing.$$

We shall often abuse terminology and say that a TMS X is topologically transitive (resp. topologically mixing) when the left shift $T : X \to X$ is topologically transitive (resp. topologically mixing).

Write $a \xrightarrow{n} b$ if there is an admissible word of length $n+1$ which starts at a and ends at b.

Proposition 1.1. *Let X be a topological Markov shift with set of states S and transition matrix $\mathbb{A} = (t_{ab})_{S\times S}$.*

1. *X is topologically transitive iff for all $a,b \in S$ there is an n s.t. $a \xrightarrow{n} b$.*
2. *X is topologically mixing iff for all $a,b \in S$ there is a number N_{ab} s.t. for all $n \geq N_{ab}$, $a \xrightarrow{n} b$.*
3. *A topologically transitive TMS is topologically mixing iff it contains two points x,y s.t. $T^p(x) = x, T^q(y) = y$, and $(p,q) = 1$.*

We leave the proof as an exercise.

Example 1 (Ising Model): This is a crude model for a magnet. Imagine a one dimensional array of sites $0,1,2,\ldots$. Each of the sites is magnetized in one of two possible ways: spin "up" (+) or spin "down" (-). The configuration space of the entire array is $X = \{+,-\}^{\mathbb{N}_0}$. This is a subshift of finite type with set of states $S = \{+,-\}$ and transition matrix all of whose entries are made of ones.

A TMS whose transition matrix has no zero entries is called a *full shift*. The full shift is obviously topologically mixing.

Example 2 (Hard Core Lattice Gas): Imagine an array of sites $0,1,2,\ldots$. Each of the sites can be in one of two states: empty (0) or occupied by one particle (1). We do not allow particles to occupy adjacent sites ("hard core assumption"). Here the set of all possible configurations is $X = \{x \in \{0,1\}^{\mathbb{N}_0} : \text{no adjacent ones}\}$.

This is the TMS with set of states $\{0,1\}$ and transition matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. This TMS is topologically mixing.

Example 3 (Random Walk on $\mathbb{Z}$): Here the set of states is $S = \mathbb{Z}$ and the transition matrix is $\mathbb{A} = (t_{ij})_{\mathbb{Z}\times\mathbb{Z}}$, where

$$t_{ij} = 1 \Leftrightarrow |i - j| = 1$$

[5] The standard definition for invertible continuous maps $T : X \to X$ is weaker: the existence of a point x such that the *full* orbit $\{T^k(x) : k \in \mathbb{Z}\}$ is dense. The two definitions are equivalent whenever X is a complete separable metric space without isolated points (a point x is called isolated if there is some $\varepsilon > 0$ s.t. $d(x,y) < \varepsilon \Rightarrow y = x$).

(encoding the rule that if we are at state i we can only move to $i \pm 1$). This TMS is clearly topologically transitive. But it is not topologically mixing, because there is no way to move from the origin to itself with an odd number of states.

Example 4 (Cayley graphs): Suppose G is a finitely generated group with a finite set of generators R.[6] The *Cayley graph TMS* associated to R is the TMS with set of states G and transition matrix

$$t_{ab} = 1 \Leftrightarrow b = ar \text{ for some } r \in R.$$

For example the Cayley graph TMS of $\mathbb{Z}$ with $R = \{\pm 1\}$ is the TMS described in example 3.

Cayley graph TMS are always topologically transitive, because for any pair of states $a, b \in G$ one can find an admissible word $\underline{w}$ which starts at a and ends at b by expanding $a^{-1}b = r_1 \cdots r_n$ and setting $\underline{w} = (a, ar_1, ar_1r_2, \ldots, ar_1r_2 \cdots r_{n-1}, b)$.

Example 5 (Angle Doubling Map): $z \mapsto z^2$ on $\{z \in \mathbb{C} : |z| = 1\}$. Write $z = \exp(2\pi i\theta)$ and express $0 \leq \theta \leq 1$ in binary coordinates

$$\theta = \sum_{j=1}^{\infty} \frac{x_j}{2^j}, \ \ (x_j \in \{0, 1\}).$$

These coordinates are uniquely defined for all except a countable set of z's which have more than one expansion (the dyadic points). It is easy to verify that in these coordinates the system becomes the left shift on $X = \{0, 1\}^{\mathbb{N}_0}$. We got the same SFT as in the Ising model case: the full shift on two symbols.

Example 6 (Gauss Map): $x \mapsto \frac{1}{x} \mod 1$ on $[0, 1]$. Here we can use the continued fraction expansion

$$x = \cfrac{1}{x_0 + \cfrac{1}{x_1 + \cdots}}, \ \ (x_i \in \mathbb{N})$$

which is uniquely defined for all irrational $x \in [0, 1]$. Ignoring the countable (whence negligible) set of rational points, we can represent the system as a full shift, except that this time the set of states is infinite.

Example 7 (Expanding Markov Interval Maps) Here is a generalization of the last two examples. Let $f : I \to I$, $I = [0, 1]$, be a map for which there exists a finite or countable collection of pairwise disjoint open intervals $\{I_a\}_{a \in S}$ s.t.:

1. $I = \bigcup_{a \in S} \overline{I}_a \cup \{0, 1\}$.
2. $f|_{I_a}$ extends to a C^1 monotonic map on an open neighborhood of $\overline{I}_a$ $(a \in S)$.
3. *Uniform expansion*: There are constants $N \in \mathbb{N}$ and $\lambda > 1$ s.t. $|(f^N)'| > \lambda > 1$ on $\bigcup_{a \in S} \overline{I}_a$.
4. *Markov Partition*: For every $a, b \in S$, if $f(I_a) \cap I_b \neq \varnothing$, then $f(I_a) \supseteq I_b$.
5. For every $a \in S$ there are $b, c \in S$ s.t. $f(I_a) \supseteq I_b$ and $f(I_c) \supseteq I_a$.

[6] This means that every element of G can be written in the form $r_1 r_2 \cdots r_n$ with $r_i \in R$ and $n \in \mathbb{N}$.

Let X denote the TMS with set of states S and transition matrix $(t_{ab})_{S\times S}$ given by $t_{ab} = 1 \Leftrightarrow f(I_a) \supseteq I_b$. This matrix has no rows or columns made only of zeroes, because of property 5.

Proposition 1.2. *Set* $\mathscr{N} := \bigcup_{n\geq 0} f^{-n}[\partial I \cup \bigcup_{a\in S} \partial I_a]$ *(a countable set). There is a Hölder continuous map* $\pi : X \to I$ *with the following properties:*

1. *The image of* π *contains* $(0,1)\setminus\mathscr{N}$*;*
2. *Every* $t \in (0,1)\setminus\mathscr{N}$ *has a unique pre-image* $(x_0,x_1,\dots)\in X$ *and this pre-image is determined by* $f^n(t) \in I_{x_n}$ $(n\in\mathbb{N}_0)$*;*
3. *If* $x \in X$ *and* $\pi(x) \in (0,1)\setminus\mathscr{N}$*, then* $(\pi\circ T)(x) = (f\circ\pi)(x)$.

Proof. We show that for every $(x_0,x_1,\dots)\in X$, $\bigcap_{n\geq 0}\overline{\bigcap_{k=0}^{n} f^{-k}(I_{x_k})}$ contains exactly one point, and then define $\pi(x_0,x_1,\dots)$ to be this point.

Step 1. If $(x_0,x_1,\dots)\in X$, then $\bigcap_{n\geq 0}\overline{\bigcap_{k=0}^{n} f^{-k}(I_{x_0})}$ contains exactly one point.

The map $f|_{I_{x_0}}$ is monotonic, one-to-one, and differentiable. Since $t_{x_0,x_1} = 1$, $f(I_{x_0}) \supseteq I_{x_1}$, whence $f(I_{x_0}\cap f^{-1}(I_{x_1})) = I_{x_1}$. Thus $I_{x_0}\cap f^{-1}(I_{x_1})$ is an interval of positive length which is mapped by f onto I_{x_1}.

Since $t_{x_1,x_2} = 1$, $f(I_{x_1}) \supseteq I_{x_2}$, so $f^2[I_{x_0}\cap f^{-1}(I_{x_1})] = f(I_{x_1}) \supseteq I_{x_2}$. The restriction of f^2 to $I_{x_0}\cap f^{-1}(I_{x_1})$ is equal to the the restriction of $f|_{I_{x_1}}\circ f|_{I_{x_0}}$ to that interval, which is clearly monotonic and C^1. It follows that $I_{x_0}\cap f^{-1}(I_{x_1})\cap f^{-2}(I_{x_2})$ is a non-empty interval which is mapped by f^2 to I_{x_2}.

Continuing in this way we see that $J_n := \bigcap_{k=0}^{n} f^{-k}(I_{x_k})$ is a non-empty interval which is mapped by f^n in a C^1 way onto I_{x_n}.

If $n = kN$, then $|(f^n)'| \geq \lambda^{n/N}$, and we get that the length of J_n is at most $\lambda^{-n/N}$. Since J_n are decreasing, we can extrapolate to all n and obtain that the length of $\overline{J_n}$ is at most $\lambda^{-\lfloor n/N\rfloor} \leq \lambda\cdot\lambda^{-n/N}$. Passing to closures, we see that $\bigcap_{n\geq 0}\overline{J_n}$ is a decreasing intersection of compact intervals with lengths no more that $\lambda\cdot\lambda^{-n/N}$. The intersection of such a sequence must equal a single point.

Step 2. π is Hölder continuous, $\pi(X)\supset(0,1)\setminus\mathscr{N}$, and each $t\in(0,1)\setminus\mathscr{N}$ has exactly one pre–image.

Suppose $x,y\in X$ and $x\neq y$, then $d(x,y) = e^{-n}$ where $x_0^{n-1} = y_0^{n-1}$ and $x_n \neq y_n$. It follows that $\pi(x),\pi(y) \in \overline{I_{x_0}\cap f^{-1}(I_{x_1})\cap\cdots\cap f^{-(n-1)}(I_{x_{n-1}})}$, whence by the calculations done in the previous step, $|\pi(x)-\pi(y)| \leq \lambda\cdot\lambda^{-n/N} = \lambda\cdot d(x,y)^{(\log\lambda)/N}$.

We show that every $t\in(0,1)\setminus\mathscr{N}$ has at least one pre-image. Such a t satisfies $f^n(t)\in\bigcup_{a\in S} I_a$ for all n, so there are $x_n\in S$ s.t. $f^n(t)\in I_{x_n}$. This means that

$$t \in \bigcap_{k=0}^{\infty} f^{-k}(I_{x_k}).$$

The sequence $(x_0,x_1,\dots)$ belongs to X, because for every i, $f(I_{x_i})\cap I_{x_{i+1}} \ni f^{i+1}(x)$, so $f(I_{x_i})\cap I_{x_{i+1}} \neq \varnothing$, whence by the Markov property $f(I_{x_i}) \supseteq I_{x_{i+1}}$ for all i.

Next suppose the t above is also the image of $\pi(y_0,y_1,\dots)$. By the definition of π, $t\in\overline{f^{-n}(I_{y_n})}$ for all n.

Since $t \notin \mathscr{N}$, $t \in \bigcap_{n=0}^{\infty} f^{-n}(I_{y_n})$, whence

$$t \in \bigcap_{n=0}^{\infty} f^{-n}(I_{x_n}) \cap \bigcap_{n=0}^{\infty} f^{-n}(I_{y_n}).$$

It follows that for all n, $f^n(t) \in I_{x_n} \cap I_{y_n}$. Since the intervals $\{I_a\}_{a\in S}$ are pairwise disjoint, $x_n = y_n$ for all n.

Step 3. If $\pi(x) \in (0,1)\setminus\mathscr{N}$, then $(f\circ\pi)(x) = (\pi\circ f)(x)$.

The proof of step 2 shows that if $t = \pi(x_0, x_1, \ldots) \notin \mathscr{N}$, then $f^n(t) \in I_{x_n}$ for all $n \geq 0$. Thus $f^n(f(t)) \in I_{x_{n+1}}$ for all $n \geq 0$, proving that $f(t) = \pi(x_1, x_2, \ldots)$. □

1.3.2 If we induce a TMS on a partition set, we get a full shift

Suppose T is a measurable map on a measurable space $(X, \mathscr{B})$. Let $A \in \mathscr{B}$, and set $A' := \{x \in A : T^n(x) \in A$ infinitely often$\}$. The *induced map* on A is the map $T_A : A' \to A'$, $T_A(x) = T^{\varphi_A(x)}(x)$, where $\varphi_A(x) := \inf\{n \geq 1 : T^n(x) \in A\}$.

If we apply this construction to a general TMS, with $A = [a]$ for some state $a \in S$, then the resulting dynamical system is topologically conjugate to a full shift on a countable alphabet. This construction is very useful, because the combinatorics of the full shift are much simpler handle than those of a general TMS.

Here is the topological conjugacy. Fix $a \in S$ (the state on which we induce). The domain of the induced map is $A := \{x \in [a] : x_i = a$ for infinitely many $i \in \mathbb{N}\}$. Define

1. $\overline{S} := \{[a, \xi_1, \ldots, \xi_t] : t \geq 0, \xi_i \in S \setminus \{a\}, [a, \xi_1, \ldots, \xi_t, a] \neq \varnothing\}$;
2. $\overline{X} := \overline{S}^{\mathbb{N}_0}$, $\overline{T} : \overline{X} \to \overline{X}$ is the left shift;
3. $\pi : \overline{X} \to A$, $\pi([a, \underline{\xi}^1], [a, \underline{\xi}^2], \ldots) = (a, \underline{\xi}^1, a, \underline{\xi}^2, \ldots)$.

Then $\pi \circ \overline{T} = T_A \circ \pi$, and π is a conjugacy between $T_A : A \to A$ and $\overline{T} : \overline{X} \to \overline{X}$.

1.4 Functions on topological Markov shifts

1.4.1 Regularity properties

Throughout this section, X is a fixed TMS. The *variations* of $\phi : X \to \mathbb{R}$ are

$$\mathrm{var}_n(\phi) := \sup\{|\phi(x) - \phi(y)| : x, y \in X\,,\, x_i = y_i\,,\, 0 \leq i \leq n-1\}$$

(this could be infinite). The n–th *ergodic sum* of $\phi : X \to \mathbb{C}$ is

$$\phi_n(x) := \phi(x) + \phi(Tx) + \cdots + \phi(T^{n-1}x).$$

We will be mainly interested in the following two regularity conditions:

Definition 1.8 (Weak Hölder continuity). We say that ϕ is *weakly Hölder continuous* (with parameter θ) if there exist $A > 0$ and $\theta \in (0,1)$ such that for all $n \geq 2$, $\mathrm{var}_n(\phi) \leq A\theta^n$.

Definition 1.9 (Summable variations). We say that ϕ has *summable variations* if $\sum_{n=2}^{\infty} \mathrm{var}_n(\phi) < \infty$.

Weak Hölder continuity is obviously stronger than summable variations. Each of these conditions implies that ϕ is uniformly continuous. But if $|S| = \infty$, then these conditions do not imply that ϕ is bounded. Moreover, they allow ϕ to have infinite variation on cylinders of length one.

Occasionally, we will consider the following weaker regularity condition introduced by P. Walters:[7]

Definition 1.10 (Walters condition). A function ϕ on a TMS X is said to satisfy the *Walters condition*, if for every $k \geq 1$, $\sup_{n\geq 1}[\mathrm{var}_{n+k}\phi_n] < \infty$ and $\sup_{n\geq 1}[\mathrm{var}_{n+k}\phi_n] \xrightarrow[k\to\infty]{} 0$.

This is weaker than summable variations or weak Hölder continuity, because of the following lemma:

Lemma 1.1. *Suppose $\phi : X \to \mathbb{C}$ has summable variations. For any $n \geq 1$ and any admissible word $\underline{a} = (a_0, \ldots, a_{n-1})$ of length n, if $x, y \in T[a_{n-1}]$ and $x_0^{m-1} = y_0^{m-1}$, then $|\phi_n(\underline{a}x) - \phi_n(\underline{a}y)| \leq \sum_{k\geq m+1} \mathrm{var}_k \phi$.*

Proof. Under these assumptions, for all $0 \leq j \leq n-1$ $T^j(\underline{a}x) = (a_j, \ldots, a_{n-1}x)$ and $T^j(\underline{a}y) = (a_j, \ldots, a_{n-1}y)$ agree on (at least) the first $n - j + m$ coordinates, so $|\phi(T^j\underline{a}x) - \phi(T^j\underline{a}y)| \leq \mathrm{var}_{n+m-j}\phi$. Thus $|\phi_n(\underline{a}x) - \phi_n(\underline{a}y)| \leq \sum_{j=0}^{n-1} \mathrm{var}_{n+m-j}\phi \leq \sum_{j=m+1}^{\infty} \mathrm{var}_j \phi$. $\square$

1.4.2 Cohomology

Much of the thermodynamic formalism deals with the following problem: Given a function ϕ (thought of as $-\beta U$ with U a "potential"), construct and analyze the DLR/thermodynamic limit/equilibirum/Gibbs measures of ϕ. For reasons which shall be explained in later chapters, if we are able to answer these questions for ϕ, then we are able to answer them for functions of the form $\phi + h - h \circ T$ with h measurable, at least if h is bounded. The following definition is therefore important to us:

[7] The original definition was made for compact TMS, and did not include the condition that $\sup_n[\mathrm{var}_{n+k}\phi_n] < \infty$ for all $k \geq 1$. In the compact case the finiteness of $\sup_n[\mathrm{var}_{n+k}\phi_n]$ for all $k \geq 1$ follows from $\sup_n[\mathrm{var}_{n+k}\phi_n] \xrightarrow[k\to\infty]{} 0$, because that condition implies that ϕ is continuous, whence bounded.

Definition 1.11 (Cohomology). Two functions $\phi, \psi : X \to \mathbb{C}$ are said to be *cohomologous* via a *transfer function* h, if $\phi = \psi + h - h \circ T$. A function which is cohomologous to zero is called a *coboundary*.

Theorem 1.1 (Livsic Theorem). *Suppose X is a topologically transitive TMS, and $\phi, \psi : X \to \mathbb{C}$ have summable variations. Then ϕ and ψ are cohomologous iff for every $x \in X$ and $p \in \mathbb{N}$ s.t. $T^p(x) = x$, $\phi_p(x) = \psi_p(x)$.*

Proof. It is enough to treat the case $\psi = 0$ and show that ϕ is a coboundary iff $\phi_p(x) = 0$ for all x and p s.t. $T^p(x) = x$. (If $\psi \neq 0$, work with $\phi - \psi$, $\psi - \psi$.)

If $\phi = h - h \circ T$, then $\phi_p = h - h \circ T + h \circ T - h \circ T^2 + \cdots h \circ T^{p-1} - h \circ T^p = h - h \circ T^p$, so its clear that $\phi_p(x) = 0$ for all p–periodic points x.

Now suppose for all p and x s.t. $T^p(x) = x$, $\phi_p(x) = 0$. We are assuming that X is topologically transitive, so there exists $x \in X$ with a dense forward orbit.

Suppose there were a function h s.t. $\phi = h - h \circ T$ and s.t. (w.l.o.g.) $h(x) = 0$, then the following equation must hold along the orbit of x:

$$h(x) = 0, \; h(T^j x) = -\phi_j(x). \tag{1.4}$$

The idea of the proof is to use (1.4) to define $h(\cdot)$ on the orbit of x, show that this function has a continuous extension $h^* : X \to X$, and then check that $h^* - h^* \circ T = \phi$.

Let $\mathscr{O}(x) := \{x, Tx, T^2x, \ldots\}$ and define $h : \mathscr{O}(x) \to \mathbb{C}$ by (1.4). We claim that h is uniformly continuous on $\mathscr{O}(x)$. To see this suppose $T^m(x), T^n(x)$ agree on their first k–coordinates and write $(x_m, \ldots, x_{m+k-1}) = (x_n, \ldots, x_{n+k-1}) = \underline{a}$. Without loss of generality $n > m$.

Case 1. $k < n - m$.

In this case $x_m^\infty = (\underline{a}, * \cdots, *, x_n^\infty)$ and $x_n^{n+k-1} = \underline{a}$, so

$$\begin{aligned}
|h(T^n x) - h(T^m x)| &= |\phi_n(x) - \phi_m(x)| = |\phi_{n-m}(T^m x)| \\
&= |\phi_{n-m}(x_m, \ldots, x_{n-1}; x_n, x_{n+1}, \ldots)| \\
&= |\phi_{n-m}(\underline{a}, x_{m+k}, \ldots, x_{m+(n-m)-1}; \underline{a}, x_{n+k}, x_{n+k+1}, \ldots)| \\
&\leq |\phi_{n-m}(z)| + \sum_{j>k} \mathrm{var}_j \phi, \quad \text{by lemma 1.1,}
\end{aligned}$$

where z is the periodic point with period $(\underline{a}, x_{m+k}, \ldots, x_{m+(n-m)-1})$. Since $T^{n-m}(z) = z$, we get that $|h(T^n x) - h(T^m x)| \leq \sum_{j>k+1} \mathrm{var}_j \phi$. $\diamondsuit$

Case 2. $k \geq n - m$.

Write $k = \ell(n-m) + r$ where $0 \leq r < n - m$, and let $\underline{p} = x_m^{n+m-1}$, then

$$x_m^\infty = (\underbrace{\underline{p}, \ldots, \underline{p}}_{\ell+1}, p_0^{r-1}, x_{n+k}^\infty)$$

(check!). The same argument as before, but with z equal to the periodic point with period $\underline{p}$, shows that again $|h(T^n x) - h(T^m x)| \leq \sum_{j>k} \mathrm{var}_j \phi$. $\diamondsuit$

Since ϕ has summable variations, this establishes a uniform modulus of continuity for h on $\mathscr{O}(x)$. Since h is uniformly continuous on $\mathscr{O}(x)$, it has a continuous extension to the closure $\overline{\mathscr{O}(x)}$. Since x has a dense orbit this closure is equal to X. Thus h extends continuously to some continuous function $h^* : X \to \mathbb{C}$.

We claim that $\phi = h^* - h^* \circ T$. This is certainly the case on $\mathscr{O}(x)$, because for every $T^n x \in \mathscr{O}(x)$,

$$h^*(T^n x) - h^*(T^{n+1}x) = h(T^n x) - h(T^{n+1}x) = \phi_{n+1}(x) - \phi_n(x) = \phi(T^n x).$$

Since the equation $\phi = h^* - h^* \circ T$ holds on $\mathscr{O}(x)$ and h^*, ϕ are continuous, this equation holds on $\overline{\mathscr{O}(x)} = X$. $\square$

Remark: The proof that the cohomology can be achieved by a transfer function h s.t. $\mathrm{var}_k h \leq \sum_{j>k} \mathrm{var}_j \phi$. In particular, if ϕ is weakly Hölder, then one can take h weakly Hölder with finite first variation.

1.5 Two sided topological Markov shifts

TMS arise as models for certain non-invertible transformations such as expanding Markov maps of the interval. The invertible case calls for an invertible analogue of a TMS, which we proceed to define.

Suppose S is a finite or countable set of states and $\mathbb{A} = (t_{ab})_{S\times S}$ is a matrix of zeroes and ones without rows which are made solely of zeroes.

Definition 1.12 (Two sided TMS). The *two–sided topological Markov shift* with set of *states* S and transition matrix $\mathbb{A} = (t_{ab})_{S\times S}$ is the set

$$\widetilde{X} := \{(\dots, x_{-1}, x_0, x_1, \dots) \in S^{\mathbb{Z}} : \forall i \in \mathbb{Z},\ t_{x_i x_{i+1}} = 1\},$$

equipped with the topology generated by the two sided cylinders

$${}_m[a_0, \dots, a_{n-1}] := \{x \in \widetilde{X} : (x_m, \dots, x_{m+n-1}) = (a_0, \dots, a_{n-1})\} \quad (m \in \mathbb{Z}, n \in \mathbb{N}).$$

and the action of the *left shift* $\widetilde{T} : \widetilde{X} \to \widetilde{X}$, $(Tx)_i = x_{i+1}$.

This topology is induced by the metric defined for $x \neq y$ by

$$d(x,y) = \exp(-\min\{n \geq 0 : x_n \neq y_n \text{ or } x_{-n} \neq y_{-n}\}).$$

The *variations* of a function $\phi : \widetilde{X} \to \mathbb{R}$ are defined by

$$\mathrm{var}_n \phi := \sup\{|\phi(x) - \phi(y)| : x_i = y_i \ \ (|i| \leq n-1)\}.$$

A function is called *weakly Hölder* continuous, if there are constants $A > 0$ and $0 < \theta < 1$ s.t. $\mathrm{var}_n \phi \leq A\theta^n$ for all $n \geq 2$.

Many results in the thermodynamic formalism for TMS can be extended to two–sided TMS using the following fact: Let X and $\widetilde{X}$ be the one sided and two sided TMS with set of states S and transition matrix $\mathbb{A}$.

Theorem 1.2 (Sinai). *If $\phi : \widetilde{X} \to \mathbb{C}$ is weakly Hölder and $\mathrm{var}_1\phi < \infty$, then there exists a bounded weakly Hölder $h : \widetilde{X} \to \mathbb{C}$ and a weakly Hölder $\phi^* : X \to \mathbb{C}$ s.t. $\mathrm{var}_1\phi^* < \infty$ and $\phi(x) + h(x) - h(\widetilde{T}x) = \phi^*(x_0, x_1, \dots)$.*

Proof. For every $a \in S$, let $\underline{z}^a = \left(\dots, z^a_{-2}, z^a_{-1}\right)$ be a one–sided infinite admissible sequence s.t. $\left(z^a_{-1}, a\right)$ is admissible. Given $x \in \widetilde{X}$, let x^* denote the point s.t. $(x^*)^{-1}_{-\infty} := \underline{z}^{x_0}$ and $(x^*)^{\infty}_0 := x^{\infty}_0$. We show that the following transfer function works:

$$h(x) := \sum_{k=0}^{\infty} \left(\phi(\widetilde{T}^k x^*) - \phi(\widetilde{T}^k x) \right).$$

Step 1. h is well defined, bounded, and continuous.

The map $x \mapsto x^*$ is continuous, therefore the summands of the series are continuous. The k–th summand in the infinite sum is bounded by $\mathrm{var}_{k+1}\phi \le A\theta^{k+1}$, because $\widetilde{T}^k(x)$ and $\widetilde{T}^k(x^*)$ have equal j coordinates for all $j \ge -k$. Thus the series converges uniformly to a bounded continuous function. ♢

Step 2. Calculation of $\mathrm{var}_n h$ and proof that h is (weakly) Hölder.

Since h is bounded, all its variations are finite, and it is enough to bound $\mathrm{var}_n\phi$ just for $n \ge 2$. Fix $n \ge 2$ and suppose $x^{n-1}_{-(n-1)} = y^{n-1}_{-(n-1)}$, then

$$\begin{aligned} |h(x) - h(y)| \le & \sum_{k=0}^{\lfloor n/2 \rfloor - 1} \left| [\phi(\widetilde{T}^k x^*) - \phi(\widetilde{T}^k x)] - [\phi(\widetilde{T}^k y^*) - \phi(\widetilde{T}^k y)] \right| \\ & + \sum_{k=\lfloor n/2 \rfloor}^{\infty} \left| \phi(\widetilde{T}^k x^*) - \phi(\widetilde{T}^k x) \right| + \left| \phi(\widetilde{T}^k y^*) - \phi(\widetilde{T}^k y) \right|. \end{aligned}$$

The first sum can be estimated by noting that if $k < n/2$, then

$$|\phi(\widetilde{T}^k x) - \phi(\widetilde{T}^k y)|, |\phi(\widetilde{T}^k x^*) - \phi(\widetilde{T}^k y^*)| \le \mathrm{var}_{\lfloor n/2 \rfloor}\phi.$$

The second sum can be estimated by noting that if $k \ge \lfloor n/2 \rfloor$, then

$$\left| \phi(\widetilde{T}^k x^*) - \phi(\widetilde{T}^k x) \right|, \left| \phi(\widetilde{T}^k y^*) - \phi(\widetilde{T}^k y) \right| \le \mathrm{var}_k\phi.$$

Thus $\mathrm{var}_n h \le n\,\mathrm{var}_{\lfloor n/2 \rfloor}\phi + 2\sum_{k \ge \lfloor n/2 \rfloor} \mathrm{var}_k\phi$ for all $n \ge 2$ (and $\mathrm{var}_1 h \le 2\sup|h| < \infty$). It follows that h is weakly Hölder continuous. ♢

Step 3. $\phi(x) + h(x) - h(Tx)$ only depends on $(x_0, x_1, \dots)$.

$$\phi(x)+h(x)-h(Tx)=$$

$$\begin{aligned}
&=\phi(x)+\sum_{k=0}^{\infty}\Big(\phi(\widetilde{T}^k x^*)-\phi(\widetilde{T}^k x)\Big)-\sum_{k=0}^{\infty}\Big(\phi(\widetilde{T}^k (Tx)^*)-\phi(\widetilde{T}^{k+1}x)\Big)\\
&=\phi(x)+\left[\phi(x^*)-\phi(x)+\sum_{k=1}^{\infty}\Big(\phi(\widetilde{T}^k x^*)-\phi(\widetilde{T}^k x)\Big)\right]-\sum_{k=0}^{\infty}\Big(\phi(\widetilde{T}^k (Tx)^*)-\phi(\widetilde{T}^{k+1}x)\Big)\\
&=\phi(x^*)+\sum_{k=0}^{\infty}\Big(\phi(\widetilde{T}^{k+1} x^*)-\phi(\widetilde{T}^{k+1} x)\Big)-\sum_{k=0}^{\infty}\Big(\phi(\widetilde{T}^k (\widetilde{T}x)^*)-\phi(\widetilde{T}^{k+1}x)\Big)\\
&=\phi(x^*)+\sum_{k=0}^{\infty}\Big(\phi[\widetilde{T}^k \widetilde{T}(x^*)]-\phi[\widetilde{T}^k (\widetilde{T}x)^*]\Big).
\end{aligned}$$

The final expression depends only on the non-negative coordinates of x, and can therefore be identified with a function $\phi^*: X\to\mathbb{C}$. Since $\phi^*=\phi+h-h\circ T$, ϕ^* is weakly Hölder continuous with finite first variation. ☐

Remark: The assumption that $\mathrm{var}_1\phi<\infty$ is needed for the boundedness of h. It is not a big restriction, because if one recodes a TMS using the Markov partition of 2–cylinders, then every weakly Hölder continuous function becomes a weakly Hölder continuous function with finite first variation (Problem 1.7).

Problems

1.1. Suppose $T: X\to X$ is a continuous map on a metric space which preserves two different ergodic invariant probability measures μ,ν. Use the pointwise ergodic theorem to show that these measures are mutually singular (i.e. there exists a Borel set E s.t. $\mu(E)=1$, $\nu(E)=0$).

1.2. Suppose X is a TMS with set of states S and transition matrix $\mathbb{A}=(t_{ab})_{S\times S}$ with no rows and columns made only of zeroes.

1. Show that X is separable.
2. Show that the topology of X is given by the metric $d(x,y):=2^{-\min\{i:x_i\neq y_i\}}$ (where $2^{-\min\varnothing}:=0$). Show that with respect to this measure X is complete.
3. Show that X is compact iff $|S|<\infty$.
4. Show that X is locally compact iff for every $a\in S$, $\#\{b\in S: t_{ab}=1\}<\infty$.

1.3. Suppose X is a TMS with set of states S and transition matrix $\mathbb{A}$.

1. Suppose $|S|<\infty$. Show that X is topologically mixing iff there exists an number M s.t. all the entries of $\mathbb{A}^M$ are positive.
2. Construct an example showing that the above could be false when $|S|=\infty$.

1.4. Prove proposition 1.1

1.5. Show that the map $T : x \mapsto \beta x \mod 1$ on $[0,1]$ with $\beta = \frac{1+\sqrt{5}}{2}$ (the "golden mean") is an expanding Markov interval map whose associated TMS is identical to the "hard core lattice gas" model.

1.6 (Rényi's "f–expansions"). Suppose $f : [0,1] \to [0,1]$ is an expanding map of the interval with partition $\{I_a : a \in S\}$. Let $\mathscr{N} := \bigcup_{n\geq 0} f^{-n}[\{0,1\} \cup \bigcup_{a\in S} \partial I_a]$. Let $J_n := (f|_{I_n})^{-1} : f(I_n) \to I_n$, and choose for all $n \in N$ some $\xi_n \in f(J_n)$

1. Show that for every $t \in (0,1) \setminus \mathscr{N}$, if $f^n(t) \in I_{x_n}$ for all n, then
$$x = \lim_{n\to\infty} (J_{x_0} \circ J_{x_1} \circ \cdots \circ J_{x_n})(\xi_n).$$
This is called the f*–expansion* of x.
2. Show that the f–expansion produced by $f(x) = 10x \mod 1$ is the decimal expansion $x = \sum_{n=1}^{\infty} 10^{-n} x_n$
3. Show that the f–expansion produced by $f(x) = \frac{1}{x} \mod 1$ is the continued fraction expansion.
4. What expansion is produced by $f(x) = \beta x \mod 1$ where $\beta = \frac{1+\sqrt{5}}{2}$?

1.7. Suppose X is a TMS with set of states S and transition matrix $\mathbb{A} = (t_{ab})_{S\times S}$. Let $S^* := \{(a,b) \in S \times S : t_{ab} = 1\}$, and define a transition matrix $\mathbb{A}^* = (t_{(a,b),(c,d)})_{S^*\times S^*})$ by $t_{(a,b),(c,d)} = 1 \Leftrightarrow b = c$. Let X^* denote the resulting TMS.

1. Construct a continuous bijection $\pi : X^* \to X$ s.t. $\pi \circ T = T \circ \pi$.
2. Show that for every $\phi : X \to \mathbb{R}$, $\mathrm{var}_n[\phi \circ \pi] = \mathrm{var}_{n+1}\phi$ for all n. In particular any weakly Hölder function is mapped to a weakly Hölder function with finite first variation.
3. Check that this construction also works for two sided shifts.

1.8. Let $\overline{T} : \overline{X} \to \overline{X}$ be the result of inducing a TMS on one of its states a. Let $\varphi_A(x) := 1_{[a]}(x) \inf\{n \geq 1 : x_n = a\}$ and let $\pi : \overline{X} \to [a]$ be the conjugacy described in §1.3.2. Suppose $\phi : X \to \mathbb{R}$ is a function, and let

$$\overline{\phi} := \left(\sum_{k=0}^{\varphi_A - 1} \phi \circ T^k \right) \circ \pi.$$

Prove that:

1. if ϕ is weakly Hölder, then $\overline{\phi}$ is weakly Hölder, and $\mathrm{var}_1 \overline{\phi} < \infty$;
2. if $\sum_{n\geq 2} n \mathrm{var}_n \phi < \infty$, then $\overline{\phi}$ has summable variations and $\mathrm{var}_1 \phi < \infty$;
3. if ϕ has summable variations, then $\overline{\phi}$ satisfies the Walters condition and $\mathrm{var}_1 \phi < \infty$;
4. if ϕ has the Walters property, then $\overline{\phi}$ has the Walters property, and $\mathrm{var}_1 \phi < \infty$.

1.9. Prove the Livsic theorem under the assumption that ϕ satisfies the Walters condition.

1.10. Let X be a topologically transitive TMS and suppose ϕ, ψ are two weakly Hölder continuous functions.

1. Suppose $\phi - \psi = h - h \circ T$. Show that if h is continuous, then h is weakly Hölder continuous;
2. Suppose $\phi - \psi = h_i - h_i \circ T \ (i = 1, 2)$. Show that if h_1 and h_2 are continuous, then h_1, h_2 differ by a constant.

References

1. Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *2nd revised edition, edited by J.-R. Chazottes, LNM* **490***, Springer, 2008, viii+73pp*.
2. Gallavotti, G.: Statistical Mechanics, A short Treatise. *Texts and Monographs in Physics, Springer–Verlag*, 1999, xiv+339pp.
3. Ruelle, D.: Thermodynamic Formalism: The Mathematical Structures of Equilibrium Statistical Mechanics. 2nd edition. *Cambridge Mathematical Library, Cambridge University Press, 2004,* xx+174pp.
4. Sinai, Ja. G.: *Gibbs measures in ergodic theory*. Uspehi Mat. Nauk. **27**, 21—64 (1972). English transl.: Russian Math. Surveys **27**, 21–69 (1972).

Chapter 2
DLR measures, conformal measures, and their ergodic theoretic properties

2.1 DLR measures and conformal measures

2.1.1 DLR measures and the Gibbs cocycle

Let X be a topological Markov shift with set of states S. The Borel σ–algebra of X is the σ–algebra $\mathscr{B} = \mathscr{B}(X)$ which is generated by all cylinders. This is also the minimal σ–algebra with respect to which all coordinate functions $x \mapsto x_k$ are measurable. Let $\sigma(x_n, x_{n+1}, \dots)$ denote the smallest sigma algebra with respect to which all the coordinate maps $x \mapsto x_k$ with $k \geq n$ are measurable, then (Problem 2.1)

$$\sigma(x_n, x_{n+1}, \dots) = T^{-n}\mathscr{B}.$$

Given a probability measure m on X, we write for any $[a_0, \dots, a_{n-1}]$

$$m(a_0, \dots, a_{n-1} | x_n, x_{n+1}, \dots)(x) := \mathbb{E}_m(1_{[a_0,\dots,a_{n-1}]} | T^{-n}\mathscr{B})(x).$$

Armed with this notation, we proceed to give the formal definition of a DLR measure on a general TMS:

Definition 2.1 (DLR measures). Suppose X is a TMS, $\beta > 0$, and $U : X \to \mathbb{R}$ is a measurable function. A probability measure m on X is called a *Dobrushin–Lanford–Ruelle (DLR) measure* for $\phi = -\beta U$ if for all $N \geq 1$ and a.e. $x \in X$

$$m(x_0, \dots, x_{N-1} | x_N, x_{N+1}, \dots) = \frac{\exp \phi_n(x)}{\sum_{T^n y = T^n x} \exp \phi_n(y)} \qquad m\text{–a.s.} \tag{2.1}$$

where $\phi_n := \phi + \phi \circ T + \cdots \phi \circ T^{n-1}$. In particular the sum in the denominator must converge m–a.e.

Equations (2.1) are called the *DLR equations*.

DLR measures can be characterized by the way they transform under the maps $(a_0, \dots, a_{n-1}, x) \mapsto (b_0, \dots, b_{n-1}, x)$. To do this we need the following definition:

Definition 2.2 (Gibbs cocycle). Suppose X is a TMS and $\phi : X \to \mathbb{R}$ is a function.

1. The *tail relation* of a TMS is $x \sim y \Leftrightarrow \exists n$ s.t. $x_n^\infty = y_n^\infty$. This is an equivalence relation. Set $\mathfrak{T} := \{(x,y) \in X \times X : x \sim y\}$.
2. The *Gibbs cocycle* of a function $\phi : X \to \mathbb{R}$ is $\Phi : \mathfrak{T} \to \mathbb{R}$ given by

$$\Phi(x,y) := \sum_{k=0}^{\infty} [\phi(T^k y) - \phi(T^k x)].$$

The sum converges because if $(x,y) \in \mathfrak{T}$, then there exists N s.t. $T^k(x) = T^k(y)$ for all $k \geq N$. The reason Φ is called a "cocycle" is because of the following identity which is reminiscent of identities appearing in the theory of cocycles for group actions: $x \sim y \sim z \Rightarrow \Phi(x,y) + \Phi(y,z) = \Phi(x,z)$.

Proposition 2.1. *A probability measure m is a DLR measure iff for every pair of cylinders $[a_0, \ldots, a_{n-1}], [b_0, \ldots, b_{n-1}]$ such that $a_{n-1} = b_{n-1}$ and $m[\underline{a}] \neq 0$, the map*

$$\vartheta_{\underline{ab}} : [\underline{a}] \to [\underline{b}], \ \ \vartheta_{\underline{ab}} : (\underline{a}x_n^\infty) \mapsto (\underline{b}x_n^\infty)$$

satisfies $\frac{dm \circ \vartheta_{ab}}{dm} = \exp \Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty)$ *a.e. on* $[\underline{a}]$.

Proof. We prove $(\Rightarrow)$. Suppose m is a DLR measure for ϕ. In order to see that $dm \circ \vartheta_{\underline{ab}} = \exp \Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty) dm$, it is enough to show for every cylinder $[\underline{c}]$ which can follow $[\underline{a}]$ that $m[\underline{b}, \underline{c}] = \int_{[\underline{a},\underline{c}]} \exp \Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty) dm$. Here's the proof:

$$\begin{aligned} m[\underline{b}, \underline{c}] &= \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot 1_{[\underline{b}]}\right) = \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot \mathbb{E}(1_{[\underline{b}]} | T^{-n}\mathscr{B})\right) \\ &= \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot m(\underline{b} | x_n, x_{n+1,\ldots})\right) \\ &= \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot e^{\Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty)} m(\underline{a} | x_n, x_{n+1,\ldots})\right) \ \ \text{(by the DLR eqns)} \\ &= \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot e^{\Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty)} \cdot \mathbb{E}(1_{[\underline{a}]} | T^{-n}\mathscr{B})\right) \\ &= \mathbb{E}\left(1_{[\underline{c}]} \circ T^n \cdot e^{\Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty)} \cdot 1_{[\underline{a}]}\right) = \int_{[\underline{a},\underline{c}]} e^{\Phi(\underline{a}x_n^\infty, \underline{b}x_n^\infty)} dm. \end{aligned}$$

Next we prove $(\Leftarrow)$. Fix $c \in S$, $n \geq 1$ and two cylinders $[\underline{a}], [\underline{b}]$ of length n s.t. $\underline{a}c$ and $\underline{b}c$ are admissible. Assume $m[\underline{a}] \neq 0$. We calculate the a.s. limit of $m[\underline{b}, x_n^{n+k}]/m[\underline{a}, x_n^{n+k}]$ as $k \to \infty$ on $T^{-n}[c]$ in two ways:

$$\begin{aligned}
\lim_{k\to\infty}\frac{m[\underline{b},x_n^{n+k}]}{m[\underline{a},x_n^{n+k}]} &= \lim_{k\to\infty}\frac{m[\underline{b},x_n^{n+k}]/m(T^{-n}[x_n^{n+k}])}{m[\underline{a},x_n^{n+k}]/m(T^{-n}[x_n^{n+k}])} = \lim_{k\to\infty}\frac{m(\underline{b}|x_n^{n+k})}{m(\underline{a}|x_n^{n+k})}\\
&= \frac{m(\underline{b}|x_n^{\infty})}{m(\underline{a}|x_n^{\infty})} \text{ a.s., by the Martingale Convergence Theorem;}\\
\lim_{k\to\infty}\frac{m[\underline{b},x_n^{n+k}]}{m[\underline{a},x_n^{n+k}]} &= \lim_{k\to\infty}\frac{(m\circ\vartheta_{\underline{a}c,\underline{b}c})[\underline{a},x_n^{n+k}]}{m[\underline{a},x_n^{n+k}]} \quad\text{because } x\in T^{-n}[c]\\
&= \lim_{k\to\infty}\frac{1}{m[\underline{a},x_n^{n+k}]}\int_{[\underline{a},x_n^{n+k}]}\exp\Phi(\underline{a}x_n^{\infty},\underline{b}x_n^{\infty})dm\\
&= \lim_{k\to\infty}\mathbb{E}\left(\exp\Phi(\underline{a}x_n^{\infty},\underline{b}x_n^{\infty})|x_n^{n+k}\right)(\underline{a}x_n^{\infty})\\
&= \mathbb{E}\left[\exp\Phi(\underline{a}x_n^{\infty},\underline{b}x_n^{\infty})|x_n^{\infty}\right](\underline{a}x_n^{\infty}) \text{ a.s. by martingale convergence}\\
&= \exp\Phi(\underline{a}x_n^{\infty},\underline{b}x_n^{\infty}) \text{ a.s., because this is } T^{-n}\mathscr{B}\text{–measurable.}
\end{aligned}$$

Thus $m(\underline{b}|x_n^{\infty}) = e^{\phi_n(\underline{b}x_n^{\infty})}\left[e^{-\phi_n(\underline{a}x_n^{\infty})}m(\underline{a}|x_n^{\infty})\right]$.

Now fix $\underline{a}$ and vary $\underline{b}$. Since there are at most countably many $[\underline{b}]$'s, for a.e. x s.t. $x_n = c$, for all $[\underline{b}]$ s.t. $\underline{b}c$ is admissible,

$$m(\underline{b}|x_n^{\infty}) = \exp\Phi(\underline{a}x,\underline{b}x)m(\underline{a}|x_n^{\infty}).$$

Thus there is a constant $Z_n(x_n^{\infty})$ which only depends on x_n^{∞} s.t. $m(\underline{b}|x_n^{\infty}) = \frac{\exp\phi_n(\underline{b}x_n^{\infty})}{Z_n(x_n^{\infty})}$. The value of the constant is determined by the condition that the sum over all the possible $\underline{b}$'s is one. We obtain the DLR equations for n and a.s. for all x s.t. $x_n = c$. Since there only countably many c's, the DLR equations hold a.e. in X. ☐

2.1.2 Jacobians and conformal measures

The definition of a DLR measure is expressed in terms of infinitely many equations. As it turns out, there is a stronger condition than the DLR equations, which can be captured in a single equation. To state it, we need the following definitions:

Definition 2.3. A measurable map T on a measure space $(X,\mathscr{B},\mu)$ is called *non–singular*, if $\mu\circ T^{-1}\sim\mu$, i.e., $\mu(T^{-1}E)=0 \Leftrightarrow \mu(E)=0$.

Definition 2.4. Suppose ν is a non–singular measure on a TMS X with set of states S. We let $\nu\circ T$ denote the measure on X given by

$$(\nu\circ T)(E) := \sum_{a\in S}\nu[T(E\cap[a])].$$

Since $\{T(E\cap[a]) : a\in S\}$ is usually not pairwise disjoint, in general $(\nu\circ T)(E)\neq \nu[T(E)]$. It is useful to think of $(\nu\circ T)(E)$ as the measuring $T(E)$ with "multiplicities".

Lemma 2.1. *Suppose ν is a non-singular measure on a TMS X.*

1. *For all non-negative Borel functions $f : X \to \mathbb{R}$:*

$$\int_X f d\nu \circ T = \sum_{a \in S} \int_{T[a]} f(ax) d\nu(x). \tag{2.2}$$

2. *$\nu \ll \nu \circ T$;*
3. *If for every $a \in S$, $T : [a] \to T[a]$ is non–singular in the sense that for every Borel set $E \subset T[a]$, $\nu(E) = 0 \Leftrightarrow \nu(TE) = 0$, then $\nu \circ T \sim \nu$.*

The proof is left as an exercise (Problem 2.2).

Definition 2.5 (Jacobian). Let ν be a non–singular Borel measure on a TMS X.

1. The *Jacobian* of ν is the function $g_\mu := \frac{d\nu}{d\nu \circ T}$.
2. If $\nu \sim \nu \circ T$, then the *log Jacobian* of ν is $\log \frac{d\nu}{d\nu \circ T}$.

Proposition 2.2. *Suppose X is a TMS and $\phi : X \to \mathbb{R}$ is Borel, then any non-singular measure ν s.t. $\nu(X) = 1$ and s.t. for some constant λ*

$$\frac{d\nu}{d\nu \circ T} = \lambda^{-1} \exp \phi,$$

is a DLR measure for ϕ.

Proof. Suppose ν is a non-singular measure such that $d\nu / d\nu \circ T = \lambda^{-1} \exp \phi$. We use proposition 2.1 to show that ν is a DLR measure.

Suppose $\underline{a}, \underline{b}$ are two admissible words of length n which terminate in the same symbol, and suppose $\nu[\underline{a}] \neq 0$. We calculate the Radon–Nikodym derivative of the map $\vartheta_{\underline{a},\underline{b}} : (\underline{a} x_n^\infty) \mapsto (\underline{b} x_n^\infty)$ by noticing that

$$\vartheta_{\underline{a},\underline{b}} := I_{\underline{b}} \circ T^n|_{[\underline{a}]},$$

where $I_{\underline{b}} : T^n[\underline{b}] \to [\underline{b}]$, $I_{\underline{b}} : x \mapsto (\underline{b}, x)$ is the inverse of $T^n : [\underline{b}] \to T^n[\underline{b}]$, and calculating the Radon–Nikodym derivatives of $I_{\underline{b}} : T^n[\underline{b}] \to [\underline{b}]$ and $T^n : [\underline{a}] \to T^n[\underline{a}]$.

Some general facts on Radon–Nikodym derivatives: Suppose $\mu_1 \sim \mu_2 \sim \mu_3$, then

1. $\frac{d\mu_1}{d\mu_2} = (\frac{d\mu_2}{d\mu_1})^{-1}$ a.e.;
2. if V is an *invertible* map s.t. V, V^{-1} are measurable, then $\frac{d\mu_1 \circ V}{d\mu_2 \circ V} = \frac{d\mu_1}{d\mu_2} \circ V$ a.e.;
3. $\frac{d\mu_1}{d\mu_2} \cdot \frac{d\mu_2}{d\mu_3} = \frac{d\mu_1}{d\mu_3}$ a.e.

It follows that

$$\begin{aligned} \left.\frac{d\nu \circ \vartheta_{\underline{a},\underline{b}}}{d\nu}\right|_{[\underline{a}]} &= \left.\frac{d\nu \circ I_{\underline{b}} \circ T^n}{d\nu}\right|_{[\underline{a}]} = \left.\frac{d\nu \circ I_{\underline{b}} \circ T^n}{d\nu \circ T^n} \cdot \frac{d\nu \circ T^n}{d\nu}\right|_{[\underline{a}]} \\ &= \left.\frac{d\nu \circ I_{\underline{b}}}{d\nu}\right|_{T^n[\underline{a}]} \circ T^n \cdot \left.\frac{d\nu \circ T^n}{d\nu}\right|_{[\underline{a}]}. \end{aligned} \tag{2.3}$$

Calculation of $\left.\frac{d\nu\circ I_{\underline{b}}}{d\nu}\right|_{T^n[\underline{a}]}\circ T^n$:

$$\left.\frac{d\nu\circ I_b}{d\nu}\right|_{T[b]}=\left.\frac{d\nu\circ I_b}{d\nu}\right|_{T[b]}=\left.\frac{d\nu\circ I_b}{d\nu\circ T\circ I_b}\right|_{T[b]}=\lambda^{-1}e^{\phi(bx)}|_{T[b]}.$$

Next consider the map $I_{(b_0,b_1)}$. Using the identity $I_{b_0b_1}=I_{b_0}\circ I_{b_1}$, we see that

$$\left.\frac{d\nu\circ I_{b_0,b_1}}{d\nu}\right|_{T[b_1]}=\left.\frac{d\nu\circ I_{b_0}\circ I_{b_1}}{d\nu\circ I_{b_1}}\cdot\frac{d\nu\circ I_{b_1}}{d\nu}\right|_{T[b_1]}=\left.\frac{d\nu\circ I_{b_0}}{d\nu}\right|_{T[b_0]}\circ I_{b_1}\cdot\left.\frac{d\nu\circ I_{b_1}}{d\nu}\right|_{T[b_1]}$$
$$=\lambda^{-2}e^{\phi_2(b_0,b_1,x)}|_{T[b_1]}.$$

Continuing by induction, we get that

$$\left.\frac{d\nu\circ I_{\underline{b}}}{d\nu}\right|_{[\underline{b}]}=\lambda^{-n}\exp\phi_n(\underline{b}x).$$

Calculation of $\frac{d\nu\circ T^n}{d\nu}|_{[\underline{a}]}$:

$$\left.\frac{d\nu\circ T^n}{d\nu}\right|_{[\underline{a}]}=\left.\frac{d\nu}{d\nu\circ I_{\underline{a}}}\circ T^n\right|_{[\underline{a}]}=\left.\left(\frac{d\nu\circ I_{\underline{a}}}{d\nu}\right)^{-1}\circ T^n\right|_{[\underline{a}]}$$
$$=\lambda^n\exp(-\phi_n(\underline{a}T^nx))|_{[\underline{a}]}$$

Thus by (2.3), $\frac{d\nu\circ\vartheta_{\underline{a},\underline{b}}}{d\nu}=\lambda^{-n}e^{\phi_n(\underline{b}T^nx)}\cdot\lambda^n e^{-\phi_n(\underline{a}T^nx)}=\exp\Phi(\underline{a}x_n^\infty,\underline{b}x_n^\infty)$. By proposition 2.1, ν is a DLR measure. □

Definition 2.6 (Conformal measure). Suppose X is a TMS and $\phi:X\to\mathbb{R}$ is Borel. A (possibly infinite) Borel measure ν is called ϕ*–conformal*, if it is finite on cylinders, and if there is a $\lambda>0$ s.t. $\frac{d\nu}{d\nu\circ T}=\lambda^{-1}\exp\phi$ a.e.

Thus any ϕ–conformal probability measure m is a DLR measure for ϕ. Later we shall see that in the case of topologically mixing compact TMS (subshifts of finite type) and functions ϕ with summable variations, the converse is also true: every DLR measure is ϕ–conformal.

2.1.3 Why is the thermodynamic formalism relevant do dynamics?

We are finally ready to address the question which was raised in the introduction: Why should the selection principles of Gibbsian statistical mechanics be relevant to the study of dynamical systems which arise in contexts which have nothing to do with statistical physics?

Before giving the answer let's look again at proposition 2.2. This proposition says that *any* non–singular measure ν with non-vanishing Jacobian is a DLR measure for some Borel measurable "potential" — its log–Jacobian $\phi := \log \frac{d\nu}{d\nu\circ T}$. Thus there nothing too special about DLR measures for Borel measurable potentials.

But of course the potentials $\phi = -\beta U$ in mathematical physics are almost always much better than measurable.[1]

Similarly the measures which appear in the study of smooth dynamical systems, geometric measure theory, and so on often have highly regular log–Jacobians.

As we shall see below, the regularity of the log–Jacobian of measure has many ergodic theoretic consequences. The power of thermodynamic formalism is that enables one to use analogies with statistical physics to guess what these consequences are. In most cases the resulting theorems are valid for all measures with sufficiently regular log–Jacobians.

The reason thermodynamic formalism finds applications outside the realm of statistical physics, is that measures with regular log Jacobians are common in all areas of dynamics.

2.2 Ergodic theory of measures with regular log Jacobians

In this section we start the general study of the ergodic theoretic properties of non–singular measures ν, under the assumption that their log Jacobian $\log \frac{d\nu}{d\nu\circ T}$ has a version with summable variations.

2.2.1 The transfer operator

Before giving the definition, it is useful to perform the following mental experiment: Imagine a mass density $f(x)d\mu(x)$ on a measure space $(\Omega,\mathscr{B},\mu)$, and suppose μ is T–non-singular. If we apply T and "move" every x to $T(x)$ then we will get a new mass density $(\widehat{T}f)(x)d\mu(x)$. The extension of the map $f \mapsto \widehat{T}f$ to signed densities is called the *transfer operator* of T (w.r.t. μ). Formally:

Definition 2.7. The *transfer operator* of a non-singular map T on a sigma finite measure space $(\Omega,\mathscr{B},\mu)$ is the operator $\widehat{T} : L^1(\Omega,\mathscr{B},\mu) \to L^1(\Omega,\mathscr{B},\mu)$ given by

$$\widehat{T}f := \frac{d\mu_f \circ T^{-1}}{d\mu}, \text{ where } d\mu_f := f d\mu.$$

Remark: The transfer operator also called the *dual operator*, or the *Perron–Frobenius operator*.

[1] At the risk of pushing the point too far, we recall that in classical mechanics, a potential U of a conservative force field F is by definition a function U s.t. $F = -\nabla U$. So ∇U exists.

It is perhaps not clear from the definition that $\widehat{T}$ is a well defined operator on L^1. Here is the reason:

1. The Radon-Nikodym derivative which defines $\widehat{T}f$ exists because of the non-singularity of T: $\mu(E)=0$ implies that $\mu(T^{-1}E)=0$, and this entails $\mu_f(T^{-1}E)=\int_{T^{-1}E} f d\mu = 0$, whence $\mu_f \circ T^{-1} \ll \mu$.
2. $\widehat{T}$ maps L^1 into L^1, because

$$\|\widehat{T}f\|_1 = \int \mathrm{sgn}(\widehat{T}f)\cdot\widehat{T}f d\mu = \int \mathrm{sgn}(\widehat{T}f) d\mu_f \circ T^{-1} = \int \mathrm{sgn}(\widehat{T}f)\circ T \cdot f d\mu \le \|f\|_1.$$

Proposition 2.3 (Properties of the transfer operator). *Suppose T is a non-singular map on the σ–finite measure space $(\Omega, \mathscr{F}, \nu)$.*

1. *If $f \in L^1$, then $\widehat{T}f$ is the unique L^1–function s.t. for every $\varphi \in L^\infty$,*

$$\int \varphi \widehat{T} f d\mu = \int (\varphi \circ T) f d\mu. \tag{2.4}$$

2. *$\widehat{T}$ is* positive*: $f \ge 0$ a.e. $\Rightarrow \widehat{T}f \ge 0$ a.e.*
3. *$\widehat{T}$ is a bounded linear operator on L^1, and $\|\widehat{T}\| = 1$.*
4. *$\widehat{T}^*\mu = \mu$ in the sense that for all $f \in L^1$, $\int \widehat{T} f d\mu = \int f d\mu$.*
5. *Suppose $f \in L^1$ is non-negative with integral one, then $\widehat{T}f = f \Leftrightarrow dm = f d\mu$ is a T–invariant probability measure.*
6. *Suppose μ is T–invariant, then $(\widehat{T}f)\circ T = \mathbb{E}(f|T^{-1}\mathscr{B})$.*

Proof. Suppose $f \in L^1$, then for every $\varphi \in L^\infty$,

$$\int \varphi \widehat{T} f d\mu = \int \varphi d\mu_f \circ T^{-1} = \int \varphi \circ T d\mu_f = \int \varphi \circ T \cdot f d\mu.$$

The relation $\int \varphi \widehat{T} f d\mu = \int \varphi \circ T \cdot f d\mu$ characterizes $\widehat{T}f$, because elements in L^1 are characterized by their inner products with all elements in L^∞.

(It is tempting to *define* $\widehat{T}f$ using (2.4). But this identity only defines $\widehat{T}f$ as an element of $(L^\infty)^*$, and not all elements in $(L^\infty)^*$ are realizable as L^1–elements.)

The linearity of $\widehat{T}$ is an immediate consequence of (2.4).

The positivity of $\widehat{T}$ is proved as follows: Suppose $f \ge 0$ a.e., then

$$0 \ge \int 1_{[\widehat{T}f \le 0]} \widehat{T} f d\mu = \int 1_{[\widehat{T}f \le 0]} \circ T f d\mu \ge 0,$$

which means that $\int 1_{[\widehat{T}f \le 0]} \widehat{T} f d\mu = 0$. It follows that $\mu[\widehat{T}f \le 0] = 0$.

The boundedness of $\widehat{T}$ and the fact that $\|\widehat{T}\|_1 \le 1$ was shown above. To see that $\|\widehat{T}\|_1 = 1$, take a non-negative f in L^1, and use the positivity of $\widehat{T}$ to see that

$$\|\widehat{T}\|_1 = \int \widehat{T} f d\mu = \int (1 \circ T) f d\mu = \int f d\mu = \|f\|_1.$$

Part 4 is because $\int \widehat{T} f d\mu = \int 1 \cdot \widehat{T} f d\mu = \int (1 \circ T) f d\mu = \int f d\mu$.

Part 5 is proved as follows: Suppose $f \in L^1$ is non-negative, and $\int f d\mu = 1$. By the definition of $\widehat{T}$, $\widehat{T}f = f \Leftrightarrow \frac{d\mu_f \circ T^{-1}}{d\mu} = f \Leftrightarrow \mu_f \circ T^{-1} = \mu_f \Leftrightarrow f d\mu$ is T–invariant.

For part 6, suppose μ is T–invariant. $(\widehat{T}f) \circ T$ is $T^{-1}\mathscr{B}$ measurable, so it is enough to check that $\int \varphi (\widehat{T}f) \circ T d\mu = \int \varphi f d\mu$ for every $T^{-1}\mathscr{B}$–measurable $\varphi \in L^\infty$. Observe that a function φ is $T^{-1}\mathscr{B}$–measurable iff it can be written in the form $\varphi = \psi \circ T$ with ψ $\mathscr{B}$–measurable.[2] Thus

$$\int \varphi (\widehat{T}f) \circ T d\mu = \int (\psi \circ T)(\widehat{T}f) \circ T d\mu = \int \psi \widehat{T} f d\mu \quad (\because \mu \circ T^{-1} = \mu)$$
$$= \int (\psi \circ T) f d\mu = \int \varphi f d\mu.$$

□

The following proposition gives the formula of the transfer operator of a non–singular measure on a TMS:

Proposition 2.4. *Suppose X is a TMS with set of states S, and ν is T–nonsingular with Jacobian $g_\nu = \frac{d\nu}{d\nu \circ T}$, then the transfer operator of ν is given by*

$$(\widehat{T}f)(x) = \sum_{a \in S} 1_{T[a]}(x) g_\nu(ax) f(ax) = \sum_{Ty = x} g_\nu(y) f(y).$$

Proof. Suppose $\varphi \in L^\infty$, then by (2.2)

$$\int \varphi \sum_{a \in S} 1_{[Ta]}(x) g_\nu(ax) f(ax) d\nu = \sum_{a \in S} \int_{[Ta]} (\varphi \circ T)(ax) g_\nu(ax) f(ax) d\nu$$
$$= \int (\varphi \circ T) g_\nu f d\nu \circ T = \int (\varphi \circ T) f \frac{d\nu}{d\nu \circ T} d\nu \circ T$$
$$= \int (\varphi \circ T) f d\nu.$$

The proposition follows from proposition 2.3, part 1. □

Definition 2.8 (Ruelle Operator). Suppose X is a TMS, and $\phi : X \to \mathbb{R}$ is a function. The *Ruelle operator* associated to ϕ is the operator $(L_\phi f)(x) = \sum_{Ty=x} e^{\phi(y)} f(y)$.

Corollary 2.1. *The transfer operator of a non-singular measure is the Ruelle operator of its log Jacobian.*

The definition of the Ruelle operator is not proper, because we have not specified the domain. The problem is particularly acute in the case of TMS with infinite alphabets, as in such cases the set $\{y : Ty = x\}$ can be infinite, and the sum defining L_ϕ may diverge. But if we know that $\phi = \log \frac{d\nu}{d\nu \circ T}$, then for every $f \in L^1(\nu)$

[2] $(\Leftarrow)$ is obvious. $(\Rightarrow)$ is because if φ is $T^{-1}\mathscr{B}$–measurable, then $\forall t \in \mathbb{R}$ $[\varphi > t] \in T^{-1}\mathscr{B}$, so $\forall t \in \mathbb{R} \exists E_t \in \mathscr{B}$ s.t. $[\varphi > t] = T^{-1}E_t$. We have $\varphi(x) := \sup\{t \in \mathbb{Q} : T(x) \in E_t\} = \psi(Tx)$, $\psi(x) := \sup\{t \in \mathbb{Q} : x \in E_t\}$.

$\int |L_\phi f| d\nu = \|\widehat{T} f\|_1 \leq \|f\|_1 < \infty$, so it is guaranteed that $\sum_{Ty=x} e^{\phi(y)} f(y)$ converges absolutely for a.e. x.

In cases where it is not clear apriori that ϕ is the log–Jacobian of some measure, the domain of L_ϕ will have to be specified.

2.2.2 Conservativity

Definition 2.9. A non-singular map T on a sigma finite measure space $(\Omega, \mathscr{B}, \nu)$ is called *conservative* if every set $W \in \mathscr{B}$ s.t. $\{T^{-n}W\}_{n\geq 0}$ are pairwise disjoint satisfies $W = \varnothing$ or $X \mod \mu$ (such a set W is called a *wandering set*).

The motivation is the following variant of the Poincaré recurrence theorem for non–singular measures:

Theorem 2.1 (Halmos). *Suppose T is a non-singular map on σ–finite measure space $(\Omega, \mathscr{B}, \nu)$. T is conservative iff the following holds for every measurable set E of positive measure: for a.e. $x \in E$, $T^n(x) \in E$ for infinitely many positive n's.*

Proof. ($\Leftarrow$): Suppose for every E measurable of positive measure, a.e. $x \in E$ visits E infinitely many times. Then there are no wandering sets of positive measure.

($\Rightarrow$): Suppose T is conservative, and assume by way of contradiction that there is a measurable set E s.t. $\mu\{x \in E : \#\{n \geq 0 : T^n(x) \in E\} < \infty\} \neq 0$. Set

$$E_N := \{x \in E : \#\{n \geq 0 : T^n(x) \in E\} = N\}.$$

There exists an $N \in \mathbb{N}$ s.t. $\mu(E_N) \neq 0$.

For every $k \geq 1$, $T^{-k}E_N \cap E_N = \varnothing$ because if there were a point $x \in T^{-k}E_N \cap E_N$, then this point would visit E at least $N+1$ times (once at time zero, then N times at times k or larger). But this contradicts the definition of E_N. It follows that E_N is a wandering set of positive measure, in contradiction to the conservativity of T. $\square$

Here is a useful criterion for the conservativity of a non-singular map:

Proposition 2.5. *Let T be a non-singular map of a σ–finite measure space.*

1. *If there is a non–negative $f \in L^1$ s.t. $\sum_{n\geq 1} \widehat{T}^n f = \infty$ a.e., then T is conservative;*
2. *If there is a strictly positive $f \in L^1$ s.t. $\sum_{n\geq 1} \widehat{T}^n f < \infty$ on a set of positive measure, then T is not conservative.*

Proof. Denote the underlying measure space by $(\Omega, \mathscr{B}, \mu)$.

Suppose f is a non-negative integrable function s.t. $\sum \widehat{T}^n f = \infty$ almost everywhere. We show that every wandering set W has measure zero. Since W is wandering, $\sum 1_W \circ T^n \leq 1$ everywhere, and by the Monotone Convergence Theorem

$$\|f\|_1 \geq \int f\Big(\sum_{n=1}^{\infty} 1_W \circ T^n\Big) d\mu = \sum_{n=1}^{\infty} \int f 1_W \circ T^n d\mu = \sum_{n=1}^{\infty} \int_W \widehat{T}^n f d\mu = \int_W \Big(\sum_{n=1}^{\infty} \widehat{T}^n f\Big) d\mu.$$

But $\sum_{n\geq 1}\widehat{T}^n f=\infty$ a.e., so W must have measure zero. This proves part 1.

For part 2, suppose that f is a strictly positive integrable function such that $A:=[\sum\widehat{T}^n f<\infty]$ has positive measure, and show that T cannot be conservative.

Since $\sum\widehat{T}^n f<\infty$ on A, $\exists B\subseteq A$ of positive measure such that $\int_B\sum\widehat{T}^n f d\mu<\infty$: Take $B:=[\sum\widehat{T}^n f<M]$ for M large enough. Arguing as before, we see that

$$\int f\Big(\sum_{n=1}^{\infty}1_B\circ T^n\Big)d\mu<\infty,$$

whence, since f is strictly positive, $\sum_{n=1}^{\infty}1_B\circ T^n d\mu$ almost everywhere in X, whence a.e. in B. It follows that for a.e. $x\in B$, $T^n(x)\in B$ only finitely many times. By theorem 2.1, T is not conservative. □

See Problem 2.8 for an development of this result.

We turn to the special case of non-singular measures with regular log Jacobians:

Theorem 2.2. *Let X be a topologically transitive TMS, and suppose ν is a non-singular measure which is finite on cylinders and s.t. $\frac{d\nu}{d\nu\circ T}=\lambda^{-1}e^{\phi}$. If ϕ has summable variations, then ν is conservative iff for some $a\in S$ (whence all $a\in S$),*

$$\sum_{n=1}^{\infty}\lambda^{-n}Z_n(\phi,a)=\infty,\ \textit{where}\ Z_n(\phi,a):=\sum_{T^n x=x}e^{\phi_n(x)}.$$

Proof. We first show that the if the series $\sum\lambda^{-n}Z_n(\phi,a)$ converges for some $a\in S$, then it converges for all $a\in S$.

Fix two states a,b and pick, using the topological transitivity of X, two admissible words $\underline{u},\underline{v}$ of lengths u,v s.t. $(a,\underline{u},b)$ and $(b,\underline{v},a)$ are admissible. Let $k:=u+v+2$, and define a map $\vartheta:\{x\in[b]:T^n x=x\}\to\{x\in[a]:T^{n+k}x=x\}$ by

$$x\mapsto \text{the periodic point with period }(a,\underline{u},x_0,\dots,x_n,\underline{v}).$$

Since ϕ has summable variations, it has the Walters property (lemma 1.1), so for all $x\in[b]$ s.t. $T^n x=x$,

$$\begin{aligned}|\phi_{n+k}(\vartheta(x))-\phi_n(x)|&\leq|\phi_{u+1}(a,\underline{u},b,*,\dots)|+|\phi_n(x_0,\dots,x_{n-1},b,*,\dots)-\phi_n(x)|\\&\quad+|\phi_{v+1}(b,\underline{v},a,*)|\\&\leq\sup_{[a,\underline{u},b]}|\phi_{u+1}|+\sup_{[b,\underline{v},a]}|\phi_{v+1}|+\sup_n[\mathrm{var}_{n+1}\phi_n]=:\log M<\infty,\end{aligned}$$

where $M=M(a,b,\underline{u},\underline{v})$ is independent of n. Since ϑ is one-to-one, for all n

$$Z_n(\phi,b)\leq M\sum_{T^n x=x,x_0=b}e^{\phi_{n+k}(\vartheta(x))}\leq M\sum_{T^{n+k}y=y,y_0=a}e^{\phi_{n+k}(y)}=MZ_{n+k}(\phi,a).$$

Thus if $\sum\lambda^{-n}Z_n(\phi,b)$ converges, then $\sum\lambda^{-n}Z_n(\phi,a)$ converges.

We saw above that the transfer operator of ϕ is $(\widehat{T}f)(x) = \lambda^{-1}\sum_{Ty=x} e^{\phi(y)}f(y)$. Iterating, it is easy to verify by induction that

$$(\widehat{T}^n f)(x) = \sum_{T^n y=x} e^{\phi_n(y)}f(y), \text{ where } \phi_n := \phi + \phi\circ T + \cdots + \phi\circ T^{n-1}.$$

We claim that for every state b, there are constants $C_1, C_2 > 0$ and $k_1, k_2 \in \mathbb{N}$ s.t.

$$C_1\lambda^{-(n-k_1)}Z_{n-k_1}(\phi,a) \le \widehat{T}^n 1_{[a]}(x) \le C_2\lambda^{-(n+k_2)}Z_{n+k_2}(\phi,a) \text{ for all } x\in[b]. \quad (2.5)$$

The proof is similar to what we did above, so we omit the details.

We can now use proposition 2.5 to prove the theorem.

Suppose $\sum\lambda^{-n}Z_n(\phi,a) = \infty$, then (2.5) implies that $\sum\widehat{T}^n 1_{[a]} = \infty$ everywhere on X. By Proposition 2.5, part 1, T is conservative.

Suppose $\sum\lambda^{-n}Z_n(\phi,a) < \infty$ for some state a. Then $\sum\lambda^{-n}Z_n(\phi,a) < \infty$ for all states a. By (2.5) there are constants C_{ab} s.t. $\sum\widehat{T}^n 1_{[a]} \le C_{ab}$ on $[b]$. Choose b s.t. $\nu[b] \ne 0$, and let $\{\varepsilon_a\}_{a\in S}$ be positive numbers s.t. $\sum\varepsilon_a C_{ab} < \infty$ and $\sum\varepsilon_a\nu[a] < \infty$. (Recall that conformal measures, by definition, are finite on cylinders.) The function

$$f := \sum_{a\in S}\varepsilon_a 1_{[a]}$$

is positive, integrable, and for a.e. $x\in[b]$, $\sum_{n=1}^{\infty}\widehat{T}^n f = \sum_n\sum_a \varepsilon_a\widehat{T}^n 1_{[a]} \le \sum_a\varepsilon_a C_{ab} < \infty$. Since $\nu[b]\ne 0$, proposition 2.5 says that ν is not conservative. □

Remark: The proof shows that the theorem holds for all functions ϕ for which $\sup_{n\ge1}[\text{var}_{n+1}\phi_n] < \infty$, e.g. all functions with the Walters property.

Corollary 2.2. *Suppose X is a* compact *topologically mixing TMS, then any non–singular probability measure on X whose log-Jacobian has summable variations is conservative.*

Proof. Suppose $\frac{d\nu}{d\nu\circ T} = \lambda^{-1}\exp\phi$ where ϕ has summable variations. We claim that there exists a constant M s.t. for every cylinder $[\underline{a}] = [a_0,\dots,a_{n-1}]$,

$$e^{\phi_n(\underline{a}x)} = M^{\pm1}\lambda^n\nu[\underline{a}] \text{ for all } x\in T[a_{n-1}]. \quad (2.6)$$

To see this note, first, that for all $x,y\in T[a_{n-1}]$,

$$\begin{aligned}|\phi_n(\underline{a}x)-\phi_n(\underline{a}y)| &\le |\phi_{n-1}(\underline{a}x)-\phi_{n-1}(\underline{a}y)| + |\phi(a_{n-1}x)-\phi(a_{n-1}y)|\\ &\le \sup_{k\ge1}\text{var}_{k+1}\phi_k + (\max\phi - \min\phi) =: \log M_1,\end{aligned}$$

where $M_1 < \infty$ by the Walters property. Integrating the estimate $e^{\phi_n(\underline{a}x)} = M_1^{\pm1}e^{\phi_n(\underline{a}x)}$ over $T[a_{n-1}]$ we obtain

$$e^{\phi_n(\underline{a}x)}\nu(T[a_{n-1}]) = M_1^{\pm 1}\int_{T[a_{n-1}]} e^{\phi_n(\underline{a}y)}d\nu(y) = M_1^{\pm 1}\lambda^n \int \lambda^{-n} L_\phi^n 1_{[\underline{a}]}d\nu = M_1^{\pm 1}\nu[\underline{a}].$$

The result is (2.6) with $M := M_1/\min\{\nu(T[a]) : a \in S\}$ (the denominator is non-zero because $|S| < \infty$ and because conformal measures on topologically transitive TMS give any cylinder positive measure, see problem 2.9).

Let $\mathscr{W}_n$ denote the collection of admissible words of length n. Fix $a \in S$. Since X is topologically mixing and $|S| < \infty$, there exists n_0 s.t. for every $\xi \in S$ there are words $w_{a\xi}, w_{\xi a} \in \mathscr{W}_{n_0}$ s.t. $(a, w_{a\xi}, \xi), (\xi, w_{\xi a}, a)$ are admissible. It follows that any $\underline{\xi} = (\xi_1, \ldots, \xi_n) \in \mathscr{W}_n$ can be extended to a word

$$w(\underline{\xi}) := (a, w_{a\xi_1}, \underline{\xi}, w_{\xi_n a}, a) \in \mathscr{W}_{n+k_0}, \text{ where } k_0 := 2(n_0+1).$$

Every periodic point with period $w(\underline{\xi})$ contributes to $Z_{n+k_0}(\phi, a) = \sum_{T^{n+k_0}x=x} e^{\phi_{n+k_0}(x)}$. As a result there is a constant C s.t. for all n,

$$\begin{aligned}\lambda^{-(n+k_0)}Z_{n+k_0}(\phi,a) &\geq C\lambda^{-(n+k_0)}\sum_{\underline{\xi}\in\mathscr{W}_n} e^{\max_{[\underline{\xi}]}\phi_n}\\ &\geq CM^{-1}\lambda^{-k_0}\sum_{\underline{\xi}\in\mathscr{W}_n}\nu[\underline{\xi}] = CM^{-1}\lambda^{-k_0} \quad \text{(by (2.6))}.\end{aligned}$$

Thus $\sum \lambda^{-n}Z_n(\phi,a) = \infty$. By theorem 2.2, ν is conservative. □

2.2.3 Ergodicity

Definition 2.10 (Ergodicity). A non–singular map T on a σ–finite measure space $(X, \mathscr{B}, \mu)$ is called *ergodic* if for every $E \in \mathscr{B}$ s.t. $T^{-1}(E) = E$, $\mu(E) = 0$ or $\mu(E^c) = 0$.

The purpose of this section is to prove the following:

Theorem 2.3 (Aaronson, Denker & Urbański). *Suppose X is a topologically transitive TMS and ν is a measure which is finite on cylinders and whose log-Jacobian has summable variations. If ν is conservative, then ν is ergodic.*

Lemma 2.2 (Bounded Distortion). *Suppose ν is a non-singular measure on a TMS s.t. $\phi := \log d\nu/d\nu\circ T$ has summable variations. If $\mathrm{var}_1\phi < \infty$, then $\exists M > 1$ s.t. for all non-empty cylinders $[\underline{a}, \underline{b}] = [a_0, \ldots, a_{n-1}; b_0, \ldots, b_{m-1}]$,*

$$M^{-1}/\nu(T[a_{n-1}]) \leq \frac{\nu[\underline{a},\underline{b}]}{\nu[\underline{a}]\nu[\underline{b}]} \leq M/\nu(T[a_{n-1}]).$$

Proof. Let $\phi := \log d\nu/d\nu\circ T$, then $L_\phi^*\nu = \nu$ and so

$$\nu[\underline{a},\underline{b}] = \int L_\phi^n 1_{[\underline{a},\underline{b}]} d\nu = \int e^{\phi_n(\underline{a}x)} 1_{[\underline{b}]}(x) d\nu(x).$$

We show that $e^{\phi_n(\underline{a}x)} = B^{\pm 1}\nu[\underline{a}]/\nu(T[a_{n-1}])$ for all $x \in [\underline{b}]$, where $B := \exp\sum_{n=1}^\infty \mathrm{var}_n\phi$.

To see this observe that for every $y \in T[a_{n-1}]$, $e^{\phi_n(\underline{a}x)} = B^{\pm 1} e^{\phi_n(\underline{a}y)}$ (it is essential for this estimate that $\mathrm{var}_1\phi < \infty$). Integrating the double inequality over $y \in T[a_{n-1}]$, we obtain that $e^{\phi_n(\underline{a}x)}\nu(T[a_{n-1}]) = B^{\pm 1}\int_{T[a_{n-1}]} e^{\phi_n(\underline{a}y)} d\nu(y) = B^{\pm 1}\int (L_\phi^n 1_{[\underline{a}]})(y) = B^{\pm 1}\nu[\underline{a}]$, as required. ☐

Corollary 2.3. *Under the assumptions of lemma 2.2, $\exists M > 1$ s.t. for all Borel sets E, if $[\underline{a}] = [a_0, \ldots, a_{n-1}]$ has positive measure, then*

$$\nu(T^{-n}E|[\underline{a}]) = M^{\pm 1}\nu(E|T[a_{n-1}]). \tag{2.7}$$

Proof. Let M be as in lemma 2.2, and let $\mathscr{M}$ denote the collection of all sets E which satisfy (2.7). $\mathscr{M}$ is a monotone class, and $\mathscr{M}$ contains all the cylinders (Lemma 2.2). By the monotone class theorem, $\mathscr{M}$ contains all Borel sets. ☐

Proof of theorem 2.3. We begin with a couple of reductions.

First we claim that it is enough to prove the theorem in the case when $\nu(X) = 1$. To see this construct positive constants $\{\varepsilon_a\}_{a\in S}$ s.t. $h = \sum_{a\in S}\varepsilon_a 1_{[a]}$ satisfies $\int h d\nu = 1$. Let $d\nu^* := h d\nu$. It is easy to check that

$$\log\frac{d\nu^*}{d\nu^*\circ T} = \log\frac{d\nu}{d\nu\circ T} + \log h - \log h\circ T.$$

By construction $\mathrm{var}_2[\log h - \log h\circ T] = 0$, so the log–Jacobian of ν^* has summable variations. Since $h > 0$, ν and ν^* are equivalent, so one is ergodic iff the other is ergodic.

Next we claim that it is enough to prove the theorem under the assumption that the first variation of $\log(d\nu/d\nu\circ T)$ is finite. To see this we recode the TMS using the following Markov partition:

$$S^* := \{(a,b) : (a,b) \text{ is admissible}\}.$$

The result is a TMS with alphabet S^* and transition matrix $(t_{(a_1,b_1),(a_2,b_2)})_{S^*\times S^*}$ where $t_{(a_1,b_1),(a_2,b_2)} = 1 \Leftrightarrow b_1 = a_2$. This TMS is conjugate to the original TMS via $\pi(x_0,x_1,x_2,\ldots) = ((x_0,x_1),(x_1,x_2),(x_2,x_3),\ldots)$, and the log Jacobian of $\nu\circ\pi^{-1}$ (equal to $\phi\circ\pi$) satisfies $\mathrm{var}_n\phi\circ\pi = \mathrm{var}_{n+1}\phi$ (Problem 1.7). In particular, $\mathrm{var}_1\phi < \infty$.

Henceforth we assume that ν is a conservative non-singular probability measure whose log–Jacobian has summable variations and finite first variation.

Suppose E is a T–invariant set with positive measure, and let α_n denote the σ–algebra generated by n–cylinders, then for a.e. x and n s.t. $\nu(E\cap T[x_{n-1}]) \neq 0$,

$$\nu(E|\alpha_n)(x) = \nu(T^{-n}E|[x_0,\ldots,x_{n-1}]) = M^{\pm 1}\nu(E|T[x_{n-1}]).$$

Suppose $\nu(E\cap T[a])\neq 0$, then $\nu[a]=\int L_\phi 1_{[a]}d\nu=\int_{T[a]}e^{\phi(ax)}d\nu\neq 0$. Since ν is conservative, for almost every $x\in[a]$, $x_{n-1}=a$ infinitely often, and so

$$\limsup_{n\to\infty}\nu(E|\alpha_n)(x)\geq M^{-1}\nu(E|T[a])>0.$$

But by the Martingale convergence theorem, $\limsup\nu(E|\alpha_n)(x)=1_E(x)$, so $1_E>0$ a.e. on $[a]$, whence $E\supseteq[a] \mod \nu$.

Now suppose p is a state such that $T[p]\supseteq[a]$, then $\nu(E\cap T[p])\geq\nu(E\cap[a])=\nu[a]\neq 0$, and by the previous paragraph $E\supseteq[p] \mod \nu$. It follows by induction that $E\supseteq[p] \mod \nu$ for every $p\in S$ s.t. $T^n[p]\supseteq[a]$ for some n. Since T is topologically transitive, $E=X \mod \nu$. □

Remark: A simple modification of the proof shows that theorem 2.3 holds for all functions ϕ s.t. $\sup_{n\geq 1}[\mathrm{var}_{n+1}\phi]<\infty$, e.g. all functions satisfying the Walters property.

Corollary 2.4. *Suppose X is a compact TMS, and ν is a Borel probability measure whose log–Jacobian has summable variations. If X is topologically transitive, then ν is ergodic.*

Proof. Corollary 2.2. □

2.2.4 Exactness and mixing

Suppose T is a measure preserving map on a probability space $(\Omega,\mathscr{F},\mu)$. The transfer operator of T is the action of T on mass densities: if we move every "mass element" by the dynamics of T, the mass density $f(x)d\mu(x)$ will transform into $(\widehat{T}f)(x)d\mu(x)$. It is reasonable to speculate that if T is "chaotic" enough, then the dynamics of T will tend to flatten the density so that in the limit $\widehat{T}^nf\to\text{const}$. The question is what is the minimal assumption on T which guarantees this. Here it is:

Definition 2.11 (Exactness). Let T be a non–singular map of a σ–finite measure space $(\Omega,\mathscr{F},\mu)$. T is called *exact* if the *tail σ–algebra* $\bigcap_{n\geq 0}T^{-n}\mathscr{F}$ is trivial, i.e. $E\in\bigcap_{n\geq 0}T^{-n}\mathscr{F}\Rightarrow\mu(E)=0$ or $\mu(E^c)=0$.

The definition makes sense for only non–invertible maps (there is an invertible version of the definition called the "K–property"). It is clear that exactness implies ergodicity, because every invariant set belongs to the tail σ–algebra, and must therefore be trivial.

Theorem 2.4 (Lin). *Let T be a measure preserving map on a σ–finite measure space $(\Omega,\mathscr{F},\mu)$ s.t. $L^1(\Omega,\mathscr{F},\mu)$ is separable (e.g. a TMS with a Borel measure), then T is exact iff $\|\widehat{T}^nf\|_1\xrightarrow[n\to\infty]{}0$ for all $f\in L^1$ s.t. $\int fd\mu=0$.*

Proof. We give two proofs, the first when $\mu(\Omega) < \infty$, and the second in general.

Suppose μ is invariant and finite, and assume w.l.o.g. that $\mu(\Omega) = 1$. Suppose T is exact, $f \in L^1$, and $\int f d\mu = 0$. Fix $\varepsilon > 0$ and construct $g \in L^\infty \cap L^1$ s.t. $\|f - g\|_1 < \varepsilon$. Then $|\int g d\mu| < \varepsilon$, and

$$\begin{aligned}
\|\widehat{T}^n f\|_1 &\leq \|\widehat{T}^n g\|_1 + \varepsilon \qquad (\because \|\widehat{T}\| = 1),\\
&= \|(\widehat{T}^n g) \circ T^n\|_1 + \varepsilon \quad (\because T \text{ is measure preserving}),\\
&= \|\mathbb{E}(g|T^{-n}\mathscr{F})\|_1 + \varepsilon \quad (\text{by proposition 2.3}).
\end{aligned}$$

By the Martingale Convergence Theorem and the exactness of T,

$$\mathbb{E}(g|T^{-n}\mathscr{F}) \xrightarrow[n\to\infty]{} \mathbb{E}(g) \text{ almost everywhere.}$$

Now $|\mathbb{E}(g|T^{-n}\mathscr{F})| \leq \|g\|_\infty$ and $\mu(\Omega) = 1$, so $\|\mathbb{E}(g|T^{-n}\mathscr{F})\|_1 \to |\int g d\mu| < \varepsilon$ (bounded convergence theorem). We see that $\limsup \|\widehat{T}^n f\|_1 < 2\varepsilon$, whence since ε was arbitrary, $\|\widehat{T}^n f\|_1 \to 0$.

Now assume that $\|\widehat{T}^n f\|_1 \to 0$ for all $f \in L^1$ s.t. $\int f d\mu = 0$. Suppose $A \in \bigcap_{n\geq 0} T^{-n}\mathscr{F}$, and set $f := 1_A - \mu(A)$. (This makes sense since μ is finite.) Then

$$\begin{aligned}
\|\widehat{T}^n f\|_1 &= \|(\widehat{T}^n f) \circ T^n\|_1 = \|\mathbb{E}(f|T^{-n}\mathscr{F})\|_1\\
&= \|f\|_1, \text{ because } f \text{ is } T^{-n}\mathscr{F}\text{–measurable for all } n.
\end{aligned}$$

But $\|\widehat{T}^n f\|_1 \to 0$ by assumption. This means that $f = 0$ a.e., whence A is trivial.

We now consider the non–singular σ–finite case, under the additional assumption that L^1 is separable. We follow [1].

Suppose T is exact and $f \in L^1$ satisfies $\int f d\mu = 0$. Since $\|\widehat{T}\| = 1$, the sequence $\{\|\widehat{T}^n f\|_1\}_{n\geq 0}$ is bounded. We show that its limit superior is equal to zero. Choose a subsequence $n_k \to \infty$ s.t. $\|\widehat{T}^{n_k} f\|_1 \to \limsup \|\widehat{T}^n f\|_1$. Let

$$g_n := \operatorname{sgn}[\widehat{T}^n f].$$

Then $\|\widehat{T}^n f\|_1 = \int g_n \widehat{T}^n f d\mu = \int (g_n \circ T^n) f d\mu$.

The functions $g_n \circ T^n$ are bounded (by one), so they can be identified with bounded linear functionals on L^1 of norm less than or equal to one. Since L^1 is separable, the closed unit ball in $(L^1)^*$ is sequentially compact. Therefore there is a subsequence $\{n_{k_\ell}\}$ s.t. $g_{n_{k_\ell}} \circ T^n \xrightarrow[\ell\to\infty]{w^*} g$, where g is a bounded linear functional on L^1. By the Riesz representation theorem, $(L^1)^* = L^\infty$ (but caution: $(L^\infty)^* \neq L^1$). Thus g can be identified with an L^∞ function g via $g(h) = \int g h d\mu$ ($h \in L^1$).

We claim that g is $T^{-n}\mathscr{F}$–measurable for all n. This is because, given n, one can choose L s.t. for all $\ell > L$, $n_{k_\ell} > n$. Starting from L all elements of the sequence $g_{n_{k_\ell}} \circ T^{n_{k_\ell}}$ are $T^{-n}\mathscr{F}$–measurable, therefore their limit g lies in $L^1(\Omega, T^{-n}\mathscr{F}, \mu)^* = L^\infty(\Omega, T^{-n}\mathscr{F}, \mu)$.

Since g is $T^{-n}\mathscr{F}$–measurable for all n, it is measurable w.r.t the tail σ–algebra. By exactness, g is a constant, say equal to c. By the choices of $\{n_k\}$ and $\{n_{k_\ell}\}$,

$$\begin{aligned}\limsup_{n\to\infty}\|\widehat{T}^n f\|_1 &= \lim_{k\to\infty}\|\widehat{T}^{n_k} f\|_1 = \lim_{k\to\infty}\int (g_{n_k}\circ T^{n_k})fd\mu \\ &= \lim_{\ell\to\infty}\int (g_{n_{k_\ell}}\circ T^{n_{k_\ell}})fd\mu = \int gfd\mu = c\int fd\mu = 0.\end{aligned}$$

Thus $\|\widehat{T}^n f\|_1 \to 0$.

Next we show that if T is not exact, then there exists some $f\in L^1$ s.t. $\int fd\mu = 0$ but for which $\|\widehat{T}^n f\|_1 \not\to 0$. Choose some $A\in\bigcap_{n\geq 0}T^{-n}\mathscr{F}$ s.t. $A\neq\varnothing,\Omega \mod \mu$. Construct, using the σ–finiteness of $(\Omega,\mathscr{F},\mu)$, $f\in L^1$ s.t. $\int_A fd\mu > 0$ and $\int_\Omega fd\mu = 0$. Since $A\in\bigcap_{n\geq 0}T^{-n}\mathscr{F}$, there are $\mathscr{F}$–measurable sets A_n s.t. $A = T^{-n}A_n$, so

$$\begin{aligned}\|\widehat{T}^n f\|_1 &\geq \int_{A_n}|\widehat{T}^n f|d\mu \geq \int_{A_n}\widehat{T}^n fd\mu \\ &= \int 1_{A_n}\widehat{T}^n fd\mu = \int (1_{A_n}\circ T^n)fd\mu = \int_A fd\mu > 0.\end{aligned}$$

Thus $\|\widehat{T}^n f\|_1 \not\to 0$. ☐

Definition 2.12 (Mixing). Let T be a measure preserving map on a *probability* space $(\Omega,\mathscr{F},\mu)$. T is called *mixing* if $\forall A,B\in\mathscr{F},\ \mu(A\cap T^{-n}B)\xrightarrow[n\to\infty]{}\mu(A)\mu(B)$.

Corollary 2.5. *Suppose T is an exact* probability *preserving map, then*

1. *for every $f\in L^1$, $\|\widehat{T}^n f - \int fd\mu\|_1 \xrightarrow[n\to\infty]{} 0$;*
2. *T is mixing.*

Proof. If T is probability preserving, then $\widehat{T}1 = 1$ (proposition 2.3 part 5), so $\widehat{T}^n(f - \int fd\mu) = \widehat{T}^n f - \int fd\mu$. Since $f - \int fd\mu$ is absolutely integrable with intergal zero, part (1) follows from Lin's theorem. (This argument does not work in the infinite measure setting, because if $\mu(\Omega)=\infty$ and $\int fd\mu\neq 0$, then $f-\int fd\mu$ is not in L^1.)

For part (2), suppose $A,B\in\mathscr{F}$. Since $\widehat{T}^n(1_A - \mu(A)) \xrightarrow[n\to\infty]{L^1} 0$,

$$\begin{aligned}\mu(A\cap T^{-n}B) &= \int 1_A(1_B\circ T^n)d\mu = \int (\widehat{T}^n 1_A)1_B d\mu \\ &= \int [\widehat{T}^n(1_A - \mu(A))1_B d\mu + \mu(A)\mu(B) \ \ (\because \widehat{T}1 = 1),\end{aligned}$$

$$\therefore\ |\mu(A\cap T^{-n}B) - \mu(A)\mu(B)| \leq \|\widehat{T}^n(1_A - \mu(A))\|_1. \tag{2.8}$$

By Lin's theorem, the term on the right side tends to zero. ☐

Remark: Equation (2.8) shows that the rate of mixing can be studied by analyzing the rate of convergence in Lin's theorem. This is the standard approach for analyzing rates of mixing.

We turn to the special case of TMS and measures with regular log Jacobians.

Theorem 2.5 (Aaronson, Denker & Urbański). *Suppose X is a topologically mixing TMS and ν is a non-singular measure which is finite on cylinders and whose log Jacobian has summable variations. If ν is conservative, then ν is exact.*

Proof. As in the proof of Theorem 2.3, it is enough to prove the theorem in the case when $\nu(X) = 1$ and $\log \frac{d\nu}{d\nu\circ T}$ has summable variations and finite first variation. These reductions allow us to use the bounded distortion lemma.

Let $E \in \bigcap_{n\geq 1} T^{-n}\mathscr{B}$ be a tail event of positive measure. We have to show that $E = X \mod \nu$. Since E is a tail event, it can be written in the form $E = T^{-n}E_n$ with $E_n \in \mathscr{B}$. Let α_n denote the σ–algebra generated by n–cylinders. By the corollary to the bounded distortion lemma, there is a constant $M > 1$ s.t. for all n

$$\nu(E|\alpha_n)(x) = \nu(T^{-n}E_n|[x_0,\dots,x_{n-1}]) \geq M^{-1}\nu(E_n|[x_n]). \tag{2.9}$$

If we can show that $\limsup \nu(E_n|T[x_{n-1}]) > 0$ a.e., then it will follow by the Martingale Convergence Theorem that $1_E = \lim \nu(E|\alpha_n)(x) > 0$ a.e., whence $E = X \mod \nu$.

We begin with a couple of observations which will be useful for this purpose.

Fact 1.

$$T^{-k}E_n = E_{n-k}. \tag{2.10}$$

Proof: Suppose $x \in T^{-k}E_n$, and let $y \in T^{-(n-k)}\{x\}$, then $y \in T^{-n}E_n = E$, and so $x = T^{n-k}(y) \in T^{n-k}(E) = T^{n-k}T^{-(n-k)}E_{n-k} \subseteq E_{n-k}$. Next suppose $x \in E_{n-k}$. Then $x = T^{n-k}(y)$ with $y \in E$ (any $y \in T^{-(n-k)}\{x\}$ works), and so $T^k(x) = T^n(y) \in T^nE = T^nT^{-n}E_n \subseteq E_n$, whence $x \in T^{-k}E_n$. ♢

Fact 2. Suppose $S : \Omega \to \Omega$ is a non-singular piecewise invertible transformation on a *finite* measure space $(\Omega, \mathscr{F}, \mu)$, then for every $\varepsilon > 0$ there exists a $\delta(S,\varepsilon) > 0$ with the following property: $\mu(B) > \varepsilon \Longrightarrow \mu[S(B)] > \delta(S,\varepsilon)$.

Proof: Suppose $\Omega = \biguplus_{i=1}^{\infty}\Omega_i$ where $S : \Omega_i \to S(\Omega_i)$ is invertible. Choose N so large that $\mu(\bigcup_{i>N}\Omega_i) < \varepsilon/4$. If $\mu(B) > \varepsilon$, then $B \cap \bigcup_{i=1}^{N}\Omega_i$ has measure at least $\varepsilon/2$. Thus there must exist $1 \leq i_0 \leq N$ s.t.

$$\mu[B \cap \Omega_{i_0}] \geq \frac{\varepsilon}{2N}.$$

For every i, let δ_i be a positive number so small that $\mu[\frac{d\mu\circ S|_{\Omega_i}}{d\mu} < \delta_i] < \varepsilon/4N$, and let $\delta_0 := \min\{\delta_1,\dots,\delta_N\}$. Then

$$\mu\left(B \cap \Omega_{i_0} \cap \left[\frac{d\mu\circ S}{d\mu} \geq \delta_0\right]\right) \geq \frac{\varepsilon}{2N} - \frac{\varepsilon}{4N} = \frac{\varepsilon}{4N},$$

whence $\mu[S(B)] \geq \varepsilon\delta_0/4N =: \delta(S,\varepsilon)$. ♢

We return to the proof of the theorem. Fix a state $[a]$ s.t. $\nu(E \cap [a]) \neq 0$. Since ν is conservative (by assumption) and ergodic (by Theorem 2.3), the symbol a appears infinitely often in a.e. $x \in X$. Let

$$\varphi_0(x) < \varphi_1(x) < \varphi_2(x) < \varphi_3(x) < \cdots$$

be the complete list of indices n s.t. $x_n = a$ for such $x \in X$.

Step 1. $\liminf_{k\to\infty} \nu(E_{\varphi_k(x)}) > 0$ almost everywhere.

This is certainly the case almost everywhere on E, because the corollary to the bounded distortion lemma, and the Martingale Convergence Theorem, for a.e. $x \in E$

$$1 = 1_E(x) \leftarrow \nu(E|\alpha_{\varphi_k(x)})(x) \le M\nu(E_{\varphi_k(x)}|T[x_{\varphi_k(x)-1}]) \le \frac{M}{\nu[a]}\nu(E_{\varphi_k(x)}).$$

Now suppose $x \notin E$. Let $S := T^{\varphi_1(x)}$, then S maps almost all X into $[a]$, and $S|_{[a]}$ is equal to $T_{[a]}$, the *induced transformation on* $[a]$: the map defined a.e. on $[a]$ which maps x to $T^{N(x)}(x)$ where $N(x)$ is the first positive time the orbit enters $[a]$. If a transformation is ergodic and conservative, then its induced version is ergodic and conservative.[3] Thus $S^n(x)$ enters the (positive measure set) $E \cap [a]$ infinitely often, whence at least once. Let $\tau = \tau(x) := \min\{n \ge 1 : S^n(x) \in E \cap [a]\}$, then

$$\nu(E_{\varphi_{k+\tau}(x)}) = \nu(E_{\varphi_\tau(x)+\varphi_k(S^\tau x)}) \ge \nu(T^{\varphi_\tau(x)}T^{-\varphi_\tau(x)}E_{\varphi_\tau(x)+\varphi_k(S^\tau x)}) = \nu(T^{\varphi_\tau(x)}E_{\varphi_k(S^\tau x)}),$$

where the last equality is because of (2.10).

By construction, $S^\tau(x) \in E$, so $\liminf \nu(E_{\varphi_k(S^\tau x)}) > 0$, whence $\nu(E_{\varphi_k(S^\tau x)}) > \varepsilon$ for some ε and all k large enough. The transformation $S^* := x \mapsto T^{\tau(x)}(x)$ is piecewise invertible and non-singular. It follows that $\nu(T^{\varphi_\tau(x)}E_{\varphi_k(S^\tau x)}) > \delta(S^*, \varepsilon)$ for all k large, where $\delta(S^*, \varepsilon)$ is given by the "second fact" above. It follows that $\nu(E_{\varphi_{k+\tau}(x)})$ is eventually bounded from below. $\diamondsuit$

Step 2. $\exists p_0$ s.t. $\liminf_{k\to\infty} \nu(E_{\varphi_k(x)-p} \cap [a]) > 0$ a.e. for all $p \ge p_0$.

Fix $x \in X$ which satisfies the conclusion of step 1, then there exists $\varepsilon > 0$ s.t. $\nu(E_{\varphi_k(x)}) > \varepsilon$ for all k large enough.

The topological mixing of T implies that $\nu(T^p[a]) \xrightarrow[p\to\infty]{} \nu(X)$ (because $T^p[a]$ eventually contains any given finite union of partition sets). Recall that we are assuming w.l.o.g. that $\nu(X) = 1$, and choose p_0 s.t.

$$\nu(T^p[a]) > 1 - \varepsilon \text{ for all } p \ge p_0.$$

From now on, suppose $p \ge p_0$. Find a finite collection Λ of cylinders $[\underline{a}] \subset [a]$ of length p s.t. $\nu(\bigcup_{[\underline{a}]\in\Lambda} T^p[\underline{a}]) > 1 - \varepsilon/2$. For every k large enough, $\exists [\underline{a}] \in \Lambda$ s.t. [4]

$$\nu(E_{\varphi_k(x)} \cap T^p[\underline{a}]) > \frac{\varepsilon}{2|\Lambda|}.$$

---

[3] Conservativity is clear. To see ergodicity, observe that every function which is invariant for an induced transformation has an (obvious) extension to an invariant function for the original map.

[4] If the other inequality holds for all $[\underline{a}] \in \Lambda$, then the summation over $[\underline{a}] \in \Lambda$ produces the inequality $\nu(E_{\varphi_k(x)}) \le \varepsilon/2$. But this contradicts the choice of ε.

Let $I_{\underline{a}} : T^p[\underline{a}] \to [\underline{a}]$ denote the map $I_{\underline{a}}(x) = (\underline{a}, x)$. It is invertible and non-singular, therefore there exists a constant δ s.t.

$$\nu[I_{\underline{a}}(E_{\varphi_k(x)} \cap T^p[\underline{a}])] > \delta$$

(take $\delta := \min\{\delta(I_{\underline{a}}, \varepsilon/2|\Lambda|) : [\underline{a}] \in \Lambda\}$). Thus

$$\delta < \nu[I_{\underline{a}}(E_{\varphi_k(x)} \cap T^p[\underline{a}])] = \nu(T^{-p}E_{\varphi_k(x)} \cap [\underline{a}]) \leq \nu(E_{\varphi_k(x)-p} \cap [a]), \text{ by (2.10).}$$

We see that $\nu(E_{\varphi_k(x)-p} \cap [a]) \geq \delta$ for all k large enough. ♢

Step 3. $\liminf_{k\to\infty} \nu(E_{\varphi_k(x)} \cap [a]) > 0$ almost everywhere.

Suppose $p \geq p_0$ where p_0 is as in step 2. Write for some ℓ to be determined later,

$$\varphi_k(x) - p = \varphi_{k-\ell}(x) + \varphi_\ell(S^{k-\ell}x) - p.$$

The idea is to choose ℓ and p so that, almost surely, for infinitely many k $\varphi_\ell(S^{k-\ell}x) = p$. For such k, $\nu(E_{\varphi_{k-\ell}(x)} \cap [a])$ is bounded below for all k large eneough, proving step 3.

Here is how to choose p and ℓ. The map $S|_{[a]} : [a] \to [a]$ is conservative and ergodic, because it is equal to the induced shift on $[a]$. Therefore for a.e. x, any value p attained by φ_ℓ with positive measure on $[a]$ will be attained by $\varphi_\ell(T^jx)$ for infinitely many j's. By definition $\varphi_\ell \geq \ell$, therefore we may choose any $\ell > p_0$, and any p which is attained by φ_ℓ with positive measure. ♢

We can now complete the proof of the theorem:

$$\begin{aligned} 1_E(x) &= \lim_{n\to\infty} \nu(E|\alpha_n(x))(x) \geq M^{-1} \limsup_{k\to\infty} \nu(E_{\varphi_k(x)}|T[x_{\varphi_k(x)-1}]) && \text{by (2.9)} \\ &\geq M^{-1} \limsup_{k\to\infty} \nu(E_{\varphi_k(x)} \cap [a]) > 0 && \text{(step 2).} \end{aligned}$$

Thus $1_E > 0$ a.e., whence $E = X \mod \nu$. □

Remark: A small modification of the proof shows that the theorem holds for all functions ϕ s.t. $\sup_{n\geq 1}[\mathrm{var}_{n+1}\phi_n] < \infty$, e.g. all functions with the Walters property.

2.2.5 Absolutely continuous invariant densities

Definition 2.13 (ACIM). Suppose T is a non–singular map of a sigma finite measure space $(\Omega, \mathscr{B}, \nu)$. A non-negative measurable function h is called an *absolutely continuous invariant density* if $dm = hd\nu$ is T–invariant. The measure m is called an *absolutely continuous invariant measure (acim)*. If h is integrable and $\int hd\nu = 1$, then m is called an *absolutely continuous invariant probability (acip)*.

There are conservative non–singular measures which do not admit acims. But conservative non-singular measures with regular log Jacobians do.

Theorem 2.6 (Aaronson, Denker & Urbański). *Let X be a topologically mixing TMS, and suppose ν is a conservative non–singular measure which is finite on cylinders. If the log Jacobian of ν has summable variations, then ν has a continuous acim which is bounded away from zero and infinity on cylinders.*

We will derive this theorem from results in the next chapter.

2.3 Existence of conformal measures

2.3.1 Existence of conformal measures on compact TMS

Since conformal measures are given by a single equation, rather than the infinite collection needed to define a DLR measure, they are easier to construct:

Theorem 2.7 (Ruelle). *Suppose X is a* compact *topologically transitive TMS and $\phi : X \to \mathbb{R}$ is continuous, then ϕ possesses a conformal (whence a DLR) probability measure.*

Proof. Let $C(X)$ denote the space of continuous functions on X, equipped with the supremum norm $\|f\| := \max|f|$ (this is well defined because X is compact). Let $C(X)^*$ denote the dual of $C(X)$, equipped with the weak star topology:

1. The elements of $C(X)^*$ are linear maps $\varphi : C(X) \to \mathbb{R}$ which possess a constant C_φ s.t. $|\varphi(f)| \le C_\varphi \|f\|$ for all $f \in C(X)$ ("bounded linear functionals");
2. The weak star topology on $C(X)^*$ is the topology generated by the following basis of open sets: $\{\varphi \in C(X)^* : |\varphi(f_i) - a_i| < b_i \ (i = 1, \ldots, N)\}$ where $N \in \mathbb{N}$, $a_i \in \mathbb{R}$, $b_i > 0$, and $f_i \in C(X)$. Note that these sets are convex, and that the maps $(\varphi_1, \varphi_2) \mapsto \alpha\varphi_1 + \beta\varphi_2$ are continuous for all $\alpha, \beta \in \mathbb{R}$. In the language of functional analysis, $C(X)^*$ is a "locally convex topological vector space".
3. Let $\mathfrak{P}(X) := \{\varphi \in C(X)^* : \varphi(1) = 1, f \ge 0 \Rightarrow \varphi(f) \ge 0\}$. Then
   a. The *Banach–Alaoglu Theorem* implies that $\mathfrak{P}(X)$ is compact in the weak star topology. Here we use the compactness of X in an essential way.
   b. The *Schauder–Tychonoff Theorem*, and the compactness of $\mathfrak{P}(X)$ guarantee that every weak star continuous map $V : \mathfrak{P}(X) \to \mathfrak{P}(X)$ has a fixed point.
   c. The *Riesz Representation Theorem* implies that every $\varphi \in \mathfrak{P}(X)$ can be represented in the form $\varphi(f) = \int f d\mu_\varphi$ where μ_ϕ is a Borel probability measure.

The idea of the proof is to find a weak star continuous map $V : \mathfrak{P}(X) \to \mathfrak{P}(X)$ whose fixed points are ϕ–conformal measures.

A measure ν is ϕ–conformal, if $\frac{d\nu}{d\nu\circ T} = \lambda^{-1} \exp\phi$. This is the case iff for all continuous functions $f : X \to \mathbb{R}$

$$\begin{aligned}\int f d\nu &= \int f\lambda^{-1}e^{\phi}d\nu\circ T\\ &\equiv \sum_{a\in S}\int_{T[a]}\lambda^{-1}e^{\phi(ax)}f(ax)d\nu(x), \text{ by (2.2)}\\ &\equiv \int \lambda^{-1}\sum_{Ty=x}e^{\phi(y)}f(y)d\nu(x) =: \int \lambda^{-1}L_{\phi}f d\nu,\end{aligned}$$

where $L_\phi : C(X) \to C(X)$ is the operator $(L_\phi f)(x) = \sum_{Ty=x} e^{\phi(y)} f(y)$.

Thus ν is ϕ–conformal iff $L_\phi^* \nu = \lambda\nu$, where $L_\phi^* : C(X)^* \to C(X)^*$ is the dual operator to $L_\phi : C(X) \to C(X)$: $[L_\phi^*\varphi](f) := \varphi(L_\phi f)$.

We still need to get rid of the unknown factor λ. This is easy to do, since if $L_\phi^*\nu = \lambda\nu$, then $\lambda = (L_\phi^*\nu)(1) = \nu(L_\phi 1)$. We see that ν is ϕ–conformal iff ν is a fixed point of

$$V : C(X)^* \to C(X)^*\ ,\ V\varphi := \frac{L_\phi^*\varphi}{(L_\phi^*\varphi)(1)}.$$

V is well defined, because by compactness there exists $\varepsilon_0 > 0$ s.t. $L_\phi 1 \geq \varepsilon_0$, so for every $\varphi \in \mathfrak{P}(X)$, $(L_\phi^*\varphi)(1) = \varphi[L_\phi 1] \geq \varepsilon_0\varphi(1) = \varepsilon_0$. Clearly $V[\mathfrak{P}(X)] \subseteq \mathfrak{P}(X)$, and it is routine to check that V is continuous in the weak star topology (Problem 2.3).

By the Schauder-Tychonoff theorem V has a fixed point. By the discussion above this fixed point is represented by a conformal measure. □

2.3.2 Convergence of measures on non-compact TMS

The arguments in the previous section do not extend to non-compact TMS, because if X is not compact, then the set of probability measures on X is not compact, and we cannot use the Schauder–Tychonoff Theorem.

We will construct conformal measures by means of a limiting procedure. To do this, it is important to first understand under what conditions does a sequence of Borel probability measures on X have a convergent subsequence in case X is not compact. This is the purpose of this section.

We use the following notation:

1. (Y,d) is a general metric space;
2. $\mathfrak{P}(Y)$ is the set of Borel probability measures on Y, $\mathfrak{M}(Y)$ is the set of finite Borel measures on Y;
3. $C_B(Y)$ denotes the Banach space of bounded continuous functions on Y, equipped with the metric $\|f\| := \sup_Y |f|$
4. $\mu(f) := \int f d\mu$ $(\mu \in \mathfrak{M}(Y), f \in C_B(Y))$.

Definition 2.14. A sequence $\mu_n \in \mathfrak{M}(Y)$ is said to *converge weak star* to $\mu \in \mathfrak{M}(Y)$, if $\mu_n(f) \xrightarrow[n\to\infty]{} \mu(f)$ for all $f \in C_B(Y)$. In this case we write $\mu_n \xrightarrow[n\to\infty]{w^*} \mu$.

Definition 2.15 (Tightness). A sequence of measures $\mu_n \in \mathfrak{M}(Y)$ is called *tight*, if for every $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subseteq Y$ s.t. $\mu_n(Y \setminus K_\varepsilon) < \varepsilon$ for all n.

Theorem 2.8 (Helly–Prohorov). *Suppose Y is a complete separable metric space, and $\mu_n \in \mathfrak{M}(Y)$ $(n \geq 1)$. If $0 < \liminf\limits_{n\to\infty} \mu_n(X) \leq \limsup\limits_{n\to\infty} \mu_n(X) < \infty$ and μ_n is tight, then $\{\mu_n\}_{n\geq 1}$ has a subsequence which converges weak star to some $0 \not\equiv \mu \in \mathfrak{M}(Y)$.*

Proof. It is enough to prove the theorem in the case when μ_n are *probability measures*. To reduce to this case, pass to a subsequence n_k s.t. $\mu_{n_k}(X) \to c \neq 0$, and work with $\mu_{n_k}/\mu_{n_k}(X)$.

There exists a compact metric space $\widehat{Y}$ such that $\widehat{Y} \supseteq Y$, and such that the topology on Y is the induced topology from $\widehat{Y}$. In the general case, the existence of $\widehat{Y}$ follows from Urysohn's Metrization Theorem (see e.g. [7]). In the particular case of a TMS with set of states S, assume w.l.o.g. that $S = \mathbb{N}$ and take $\widehat{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, and $\widehat{Y} := \widehat{\mathbb{N}}^{\mathbb{N}_0}$ with the metric

$$\widehat{d}(x,y) := \sum_{k=0}^{\infty} 2^{-k} \left| \frac{1}{x_k} - \frac{1}{y_k} \right|, \text{ where } 1/\infty := 0.$$

The metric structure on $(Y, \widehat{d}|_{Y\times Y})$ is different from the standard one, but the topology is the same, and this is all we care about.

Each μ extends to a measure $\widehat{\mu}_n$ on $\widehat{Y}$ given by $\widehat{\mu}_n(E) := \mu_n(E \cap Y)$. Since $\widehat{Y}$ is a *compact* metric space, there is a subsequence $n_k \uparrow \infty$ s.t. $\widehat{\mu}_{n_k} \to \widehat{\mu}$ weak star in $\mathfrak{P}(\widehat{Y})$. This follow from the Banach–Alaoglu theorem, but it is useful to recall the reason:

1. $\widehat{Y}$ is compact, so $C_B(Y)$ contains a countable dense set $\{g_k\}_{k\geq 1}$.[5]
2. Since g_k are bounded, $(\widehat{\mu}_n(g_k))_{n\geq 1}$ are bounded sequences. Using Cantor's diagonal argument, it is possible to extract a subsequence $n_k \uparrow \infty$ s.t. $\widehat{\mu}_{n_k}(g_i)$ converges for each i to some constant $\widehat{\mu}(\cdot)$. Automatically $\widehat{\mu}_{n_k}(g)$ converges for all $g \in C_B(\widehat{Y})$ to some constant $\widehat{\mu}(g)$.
3. $\widehat{\mu}(\cdot)$ is linear on $C_B(\widehat{Y})$, $\widehat{\mu}(1) = 1$, and for every $g \in V$, $|\widehat{\mu}(g)| \leq \|g\|$ and $g \geq 0 \Rightarrow \widehat{\mu}(g) \geq 0$. Thus $\widehat{\mu}$ extends to a bounded positive linear functional $\widehat{\mu}$ on $C_B(\widehat{Y})$.

By the Riesz representation theorem $\widehat{\mu}$ can be identified with a Borel probability measure on $\widehat{Y}$. By construction, $\widehat{\mu}_n \to \widehat{\mu}$ weak star on $\widehat{Y}$.

Claim 1. $\widehat{\mu}(\widehat{Y} \setminus Y) = 0$, thus $\mu := \widehat{\mu}|_Y$ is a Borel probability measure on Y.

Fix $\varepsilon > 0$. The tightness condition says that there exists a compact set $K_\varepsilon \subset Y$ s.t. $\mu_n(K_\varepsilon) > 1 - \varepsilon$ for all n. Since K_ε is compact in Y, it is closed in $\widehat{Y}$.

Let $G_k := \{\widehat{y} \in \widehat{Y} : \mathrm{dist}(\widehat{y}, K_\varepsilon) < \frac{1}{k}\}$. This is an open set which contains K_ε. Construct using Urysohn's Lemma a function $g \in C_B(\widehat{Y})$ s.t. (a) $0 \leq g \leq 1$, (b) $g = 1$ on K_ε, and (c) $g = 0$ outside G_k.

[5] Proof: Compact metric spaces are separable, so let $\{x_n : n \in \mathbb{N}\} \subseteq \widehat{Y}$ be a countable dense set. The functions $g_n(\cdot) = \widehat{d}(\cdot, x_n)$ separate points. Let $\mathscr{A}$ denote the algebra of functions they span together with $g_0 \equiv 1$ over $\mathbb{Q}$. This is a countable collection. Since $\widehat{Y}$ is compact $\mathscr{A}$ is dense in $C(\widehat{Y})$ (Stone–Weierstrass theorem).

We have $\widehat{\mu}(G_k) \geq \widehat{\mu}(g) = \lim \widehat{\mu}_{n_k}(g) \geq \limsup \widehat{\mu}_{n_k}(K_\varepsilon) \geq 1-\varepsilon$. Thus $\widehat{\mu}(G_k) \geq 1-\varepsilon$ for all k. Since $G_k \downarrow K_\varepsilon$, $\widehat{\mu}(K_\varepsilon) > 1-\varepsilon$. Thus

$$\widehat{\mu}(\widehat{Y} \setminus Y) \leq \widehat{\mu}(\widehat{Y} \setminus K_\varepsilon) < \varepsilon. \tag{2.11}$$

Since ε is arbitrary, $\widehat{\mu}(\widehat{Y} \setminus Y) = 0$. $\diamondsuit$

Claim 2. Let $\mu := \widehat{\mu}|_{\widehat{Y}}$, then $\mu_{n_k} \to \mu$ weak star on Y.

It is clear that $\mu_{n_k}(f) \to \mu(f)$ for all $g \in C_B(Y)$ which can be continuously extended to $\widehat{Y}$. We have to show that this is the case for all $f \in C_B(Y)$.

Suppose $f \in C_B(Y)$, fix $\varepsilon > 0$, and let K_ε be a compact set in Y s.t. $\mu_{n_k}(K_\varepsilon) > 1-\varepsilon$. The set K_ε is a compact subset of $\widehat{Y}$. By Tietze's Extension Theorem, $\exists f_\varepsilon \in C_B(\widehat{Y})$ s.t. $f_\varepsilon|_{K_\varepsilon} = f$ and $\|f_\varepsilon\| \leq \|f\|$.

$$\begin{aligned} |\mu_{n_k}(f) - \widehat{\mu}(f_\varepsilon)| &\leq |\mu_{n_k}(f) - \mu_{n_k}(f_\varepsilon|_Y)| + |\mu_{n_k}(f_\varepsilon|_Y) - \widehat{\mu}_{n_k}(f_\varepsilon)| + |\widehat{\mu}_{n_k}(f_\varepsilon) - \widehat{\mu}(f_\varepsilon)| \\ &\leq 2\|f\|\mu_{n_k}(Y \setminus K_\varepsilon) + \|f_\varepsilon\|\widehat{\mu}(\widehat{Y} \setminus Y) + o(1) \leq 3\varepsilon\|f\| + o(1). \end{aligned}$$

Since ε was arbitrary, it follows that $\mu_{n_k}(f) \to \mu(f)$. $\square$

2.3.3 Existence of conformal measures on non-compact TMS

Suppose X is a TMS with set of states S and $\phi : X \to \mathbb{R}$ has the Walters property. Define for $a \in S$

$$Z_n(\phi, a) := \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x), \text{ where } \phi_n := \sum_{k=0}^{n-1} \phi \circ T^k.$$

Theorem 2.9. *Suppose X is a topologically transitive TMS, and $\phi : X \to \mathbb{R}$ is a function with summable variations, then ϕ has a conservative conformal measure which is finite on cylinders iff for some $a \in S$,*

1. $\log \lambda := \limsup_{n\to\infty} \frac{1}{n} \log Z_n(\phi, a) < \infty$;
2. $\sum \lambda^{-n} Z_n(\phi, a) = \infty$.

Proof of necessity. Suppose there is a conservative Borel measure ν which is finite on cylinders and which satisfies $\frac{d\nu}{d\nu\circ T} = \lambda^{-1} \exp \phi$. Let L_ϕ denote the operator $(L_\phi f)(x) = \sum_{Ty=x} e^{\phi(y)} f(y)$. As we saw in the proof of theorem 2.7, the equation $\frac{d\nu}{d\nu\circ T} = \lambda^{-1} e^\phi$ implies that $L_\phi^* \nu = \lambda \nu$ in the sense that $\nu(\lambda^{-1} L_\phi f) = \nu(f)$ for all positive continuous functions f.

We claim that there exists a constant $M > 1$ s.t. $Z_n(\phi, a) = M^{\pm 1} (L_\phi^n 1_{[a]})(x)$ for all $n \in \mathbb{N}$ and $x \in [a]$. To see this we first observe by induction that

$$(L_\phi^n 1_{[a]})(x) = \sum_{T^n y = x} e^{\phi_n(y)} 1_{[a]}(y).$$

Let $\vartheta : \{y \in [a] : T^n y = x\} \to \{z \in [a] : T^n z = z\}$ denote the map which associates to $(y_0, \dots, y_{n-1}, x_0^\infty)$ the periodic point with period $(y_0, \dots, y_{n-1})$. This is a bijection and for every y, $\vartheta(y)$ and y agree on the first $n+1$ coordinates (because $x_0 = y_0 = a$). Since ϕ has the Walters property, $|\phi_n(y) - \phi_n(z)| \le \sup_{n\ge 1} \operatorname{var}_{n+1}\phi_n =: M$ with M finite. It follows that

$$Z_n(\phi,a) = \sum_{T^n z = z} e^{\phi_n(z)} 1_{[a]}(z) = M^{\pm 1} \sum_{T^n y = x} e^{\phi_n(\vartheta(y))} 1_{[a]}(y) = M^{\pm 1}(L_\phi^n 1_{[a]})(x). \quad (2.12)$$

Choose $a \in S$ s.t. $\nu[a] \neq 0$, and average both sides of the inequality $Z_n(\phi,a) \le ML_\phi^n 1_{[a]}(x)$ on $[a]$. The result is

$$\begin{aligned} Z_n(\phi,a) &\le M\left(\frac{1}{\nu[a]}\int 1_{[a]} L_\phi^n 1_{[a]} d\nu\right) \le \frac{M\lambda^n}{\nu[a]}\int \lambda^{-n} L_\phi^n 1_{[a]} d\nu \\ &= \frac{M\lambda^n}{\nu[a]}\int 1_{[a]} d\nu = M\lambda^n \ \ (\because L_\phi^*\nu = \lambda\nu). \end{aligned}$$

It follows that $\limsup \frac{1}{n}\log Z_n(\phi,a) \le \log\lambda < \infty$.

$\sum \lambda^{-n} Z_n(\phi,a) = \infty$ because of the the conservativity of ν and theorem 2.2.

Proof of sufficiency. Let $\lambda := \exp[\limsup\limits_{n\to\infty} \frac{1}{n}\log Z_n(\phi,a)]$. We assume that $\lambda < \infty$ and that $\sum \lambda^{-n} Z_n(\phi,a)$ diverges, and construct a measure ν s.t. (a) ν is finite on cylinders, and (b) $L_\phi^*\nu = \lambda\nu$ in the sense that for every bounded continuous function f, $\nu[L_\phi f] = \lambda\nu[f]$. Such a measure must be conservative, because of Theorem 2.2, and conformal, because by (2.2) for every $f \in L^1(\nu)$

$$\int f\lambda^{-1}e^\phi \mathrm{d}\nu\circ T = \int \sum_{a\in S} 1_{T[a]}(x)\lambda^{-1}e^{\phi(ax)} f(ax)\mathrm{d}\nu(x) = \int \lambda^{-1} L_\phi f \mathrm{d}\nu = \int f\mathrm{d}\nu,$$

whence $\lambda^{-1}e^\phi = \frac{d\nu}{d\nu\circ T}$.

We construct a solution to $L_\phi^*\nu = \lambda\nu$ using a limiting procedure. To set it up, fix once and for all a state $a \in S$, a periodic point $x_a \in [a]$, and define

$$\begin{aligned} a_n &:= \sum_{k=1}^n \lambda^{-k} Z_k(\phi,a); \\ \nu_n^b &:= \frac{1}{a_n}\sum_{k=1}^n \lambda^{-k} \sum_{T^k y = x_a} e^{\phi_k(y)} 1_{[b]}(y)\,\delta_y \qquad (b \in S). \end{aligned}$$

Equivalently, ν_n^b is the measure s.t. $\int f \mathrm{d}\nu_n^b = \frac{1}{a_n}\sum\limits_{k=1}^n \lambda^{-k}(L_\phi^k(f1_{[b]}))(x_a)$. Note that $a_n \uparrow \infty$ and that ν_n^b are supported inside $[b]$.

We use the Helly–Prohorov theorem to construct a subsequence $(n_k)_{k\ge 1}$ s.t. $\nu_{n_k}^b \xrightarrow[k\to\infty]{w^*} \nu^b$ for all $b \in S$, and show that ν given by $\nu|_{[b]} := \nu^b$ satisfies $L_\phi^*\nu = \lambda\nu$.

Part 1: The measures ν_n^b are finite, and $0 < \liminf_{n\to\infty} \nu_n^b(X) \leq \limsup_{n\to\infty} \nu_n^b(X) < \infty$ for all $b \in S$. (This is the reason why we are working with ν_n^b rather than with $\nu_n = \sum_{b\in S} \nu_n^b$, which could be infinite.)

By definition, $\nu_n^b(X) = \nu_n^b[b] = \frac{1}{a_n}\sum_{k=1}^n \lambda^{-k}\left(L_\phi^k 1_{[b]}\right)(x_a)$, so we need to control $(L_\phi^k 1_{[b]})(x_a)$. Construct for constants C_b, n_0, C_b', n_0' such that for all $n > n_0'$,

$$C_b' Z_{n-n_0'}(\phi,a) \leq L_\phi^n 1_{[b]}(x_a) < C_b Z_{n+n_0}(\phi,a). \tag{2.13}$$

We start with the right hand side of (2.13). Since X is topologically transitive, there exists an admissible word of the form $(a, w_1, \ldots, w_{n_0-1}, b)$. Set

$$M := \exp\big(\sup_{n\in\mathbb{N}} \mathrm{var}_{n+1}\phi_n\big) \text{ and } C_b^* := \exp\big(\sup_{[a,w_1,\ldots,w_{n_0-1},b]} |\phi_{n_0}|\big).$$

C_b^* and M are finite because ϕ has the Walters property. For each k, $(L_\phi^k 1_{[b]})(x_a)$ is the sum of all possible expressions of the form $\exp\phi_k(b,\xi_1,\ldots,\xi_{k-1},a,Tx_a)$. Clearly

$$\begin{aligned}\exp\phi_k(b,\xi_1,\ldots,\xi_{k-1},a,Tx_a) &\leq C_b^*\exp\phi_{n_0+k}(a,w_1,\ldots,w_{n_0-1},b,\xi_1,\ldots,\xi_{k-1},a,Tx_a)\\ &\leq C_b^* M \exp\phi_{n_0+k}(\overline{a,w_1,\ldots,w_{n_0-1},b,\xi_1,\ldots,\xi_{k-1}}),\end{aligned}$$

where an over bar over a word indicates the periodic point with this word as a period. Summing over all possible $\underline{\xi}$ we obtain the right side of (2.13) with $C_b := C_b^* M$. The left had side of (2.13) is proved in a similar way, by fixing an admissible word $(b,u_1,\ldots,u_{n_0'-1},a)$ and replacing the terms $\exp\phi_{n-n_0'}(\overline{a,\xi_1,\ldots,\xi_{n-n_0'-1}})$ in $Z_{n-n_0'}(\phi,a)$ by the terms $\exp\phi_n(b,\underline{u},a,\underline{\xi},x_a)$ in $(L_\phi^n 1_{[b]})(x_a)$.

By (2.13), for all $n > n_0'$, $C_b'\lambda^{-n_0'}(a_{n-n_0'}/a_n) \leq \nu_n^b(X) \leq C_b\lambda^{n_0}(a_{n+n_0}/a_n)$. To conclude the proof it is enough to show that $a_{n+1}/a_n \to 1$. By assumption $a_n \to \infty$, so it is enough to show that $|a_{n+1} - a_n| \leq M$, equivalently that

$$\lambda^{-n} Z_n(\phi,a) \leq M \text{ for all } n. \tag{2.14}$$

Suppose by way of contradiction that $Z_k(\phi,a) > (1+\varepsilon)M\lambda^k$ for some $k \in \mathbb{N}$ and $\varepsilon > 0$. Let $\Xi_k := \{\underline{\xi} \in S^{k-1} : [a,\underline{\xi},a] \neq \varnothing\}$, then for every ℓ

$$\begin{aligned}Z_{k\ell}(\phi,a) &\geq \sum_{\underline{\xi}_1,\ldots,\underline{\xi}_\ell\in\Xi_k} \exp\phi_{kl}(\overline{a,\underline{\xi_1},a,\underline{\xi}_2,\cdots,a,\underline{\xi}_\ell})\\ &\geq M^{-\ell}\sum_{\underline{\xi}_1,\ldots,\underline{\xi}_\ell\in\Xi_k}\left(\prod_{i=1}^{\ell}\exp\phi_k(\overline{a,\underline{\xi_i}})\right) = \left(M^{-1}Z_k(\phi,a)\right)^\ell \geq (1+\varepsilon)^\ell\lambda^{k\ell},\end{aligned}$$

whence $\limsup \frac{1}{n}\log Z_n(\phi,a) > \lambda$. But this contradicts the definition of λ. $\diamondsuit$

Part 2 (tightness): $\forall b\, \forall \varepsilon > 0\, \exists F = F_{b,\varepsilon}$ compact such that $\forall n \quad \nu_n^b(F^c) < \varepsilon$.

To construct F, it is useful to identify the set of states S with $\mathbb{N}$ and to notice that for any sequence of natural numbers $\{n_i\}_{i\in\mathbb{N}}$, $\{x \in X : x_i \leq n_i \ (i \in \mathbb{N})\}$ is compact. We construct a sequence $\{n_i\}_{i\geq 1}$ which grows so fast that $\nu_n^b(X \setminus F) < \varepsilon$ for all n.

It is enough to do this in the special case $b = a$. To see why, suppose we can find for every ε a sequence $\{n_i\}_{i\geq 0}$ s.t. $\nu_n^a\{x \in X : \exists i(x_i > n_i)\} < \varepsilon$ for all n. Choose some admissible word of the form $(a, w_1, \ldots, w_{n_0-1}, b)$ (X is topologically transitive). Increase the n_i, if necessary, to ensure that $n_i \geq \max\{a, w_1, \ldots, w_{n_0-1}\}$ for all i, and let $I_{a\underline{w}} : [b] \to [a]$ denote the map $I_{a\underline{w}}(x) = (a, \underline{w}, x)$. Since ϕ has the Walters property, there is a constant C_{ab} s.t. for all n, $\nu_n^b \leq C_{ab}\nu_{n+n_0}^a \circ I_{a\underline{w}}$. Thus $\nu_n^b\{x \in X : \exists i(x_i > n_{i+n_0})\} < C_{ab}\varepsilon$ for all n. Since ε is arbitrary, $\{\nu_n^b\}_{n\geq 1}$ is tight.

Henceforth we stick to case $b = a$, and let $\nu_n := \nu_n^a$. We also fix $\varepsilon > 0$.

The measures ν_n are all carried by the set $\{x \in [a] : x_n = a$ infinitely may times$\}$ (because x_a contains the symbol a infinitely many times). For every x in this set, let $\tau_i(x)$ denote the distance between the i-th a and the $(i+1)$–th a:

1. $\tau_1(x) := \min\{n \geq 1 : T^n(x) \in [a]\}$;
2. and by induction $\tau_{n+1}(x) := \tau_1(T^{\tau_1(x)+\cdots+\tau_n(x)}x)$.

Step 1. Stochastic bound for the growth of τ_i: There exists a sequence of natural numbers $\{T_i\}_{i\geq 0}$ s.t. the set $R(T_0, T_1, T_2, \ldots) := \{x \in [a] : \forall i \ \ \tau_i(x) \leq T_i\}$ satisfies

$$\nu_n [X \setminus R(T_0, T_1, T_2, \ldots)] < \varepsilon \text{ for all } n. \tag{2.15}$$

Proof: Set $\Lambda_a(k_1, \ldots, k_m) := \left\{x \in [a] : \forall j \leq m \quad \tau_j(x) = k_j\right\}$ and

$$Z^*_{k_1,\ldots,k_m} := \sum_{T^{k_1+\cdots+k_m}x = x} e^{\phi_{k_1+\cdots+k_m}(x)} 1_{\Lambda_a(k_1,\cdots,k_m)}(x).$$

Let p_a denote the period of x_a (from the definition of ν_n^b), and assume that $\{T_i\}_{i\geq 1}$ is an increasing sequence of natural numbers s.t. $T_i > p_a$ for all $i \geq 1$. Define, as always, $M := \sup_{n\geq 1} \mathrm{var}_{n+1}\phi_n$.

$$\nu_n [X \setminus R(T_0, T_1, T_2, \ldots)] \leq \sum_{i=1}^{\infty} \nu_n [\tau_i > T_i] = \sum_{i=1}^{\infty} \frac{1}{a_n} \sum_{k=1}^{n} \lambda^{-k} \sum_{\substack{T^k y = x_a \\ y_0 = a}} e^{\phi_k(y)} 1_{[\tau_i > T_i]}(y)$$

$$= \sum_{i=1}^{\infty} \frac{1}{a_n} \sum_{k=T_i+1}^{n} \lambda^{-k} \sum_{\substack{k_1+\cdots+k_N = k \\ k_i > T_i,\, i \leq N \leq k}} \left(\sum_{\substack{T^k y = x_a \\ y_0 = a}} e^{\phi_k(y)} 1_{\left[\forall j \leq N \ \ \tau_j(y) = k_j\right]}(y) \right).$$

Explanation: for any $y \in T^{-k}\{x_a\}$, $(y_0, \ldots, y_k) = (a, \underline{w}^1, a, \underline{w}^2, a, \ldots, a, \underline{w}^N, a)$ where $a\underline{w}^j$ is a word of length $k_j = \tau_j(y)$, $\underline{w}^j$ contains no a's, and $k_1 + \cdots + k_N = k$. Of course $N \leq k$. If $\tau_i(y) > T_i$, then necessarily $N \geq i$ because for all $j > N$, $\tau_j(y) = \tau_{i-N}(x_a) \leq$period of $x_a \leq \min\{T_1, T_2, \ldots\}$.

Thus $\nu_n[X \setminus R(T_0, T_1, T_2, \dots)] \le$

$$\le M^3 \sum_{i=1}^{\infty} \frac{1}{a_n} \sum_{k=T_i+1}^{n} \lambda^{-k} \sum_{\substack{k_1+\cdots+k_N=k \\ k_i>T_i, N\le k}} Z^*_{k_1,\dots,k_{i-1}} Z^*_{k_i} Z^*_{k_{i+1},\dots,k_N}$$
$$\le M^3 \sum_{i=1}^{\infty} \sum_{k_i=T_i+1}^{\infty} \lambda^{-k_i} Z^*_{k_i} \frac{1}{a_n} \sum_{k=k_i}^{n} \lambda^{-(k-k_i)} \sum_{\substack{k_1+\cdots+k_N=k \\ N\le k}} Z^*_{k_1,\dots,k_{i-1}} Z^*_{k_{i+1},\dots,k_N}$$
$$\le M^5 \sum_{i=1}^{\infty} \sum_{k_i=T_i+1}^{\infty} \lambda^{-k_i} Z^*_{k_i} \left(\frac{1}{a_n} \sum_{k=k_i}^{n} \lambda^{-(k-k_i)} Z_{k-k_i}(\phi, a) \right)$$
$$\le M^5 \sum_{i=1}^{\infty} \left(\sum_{k_i=T_i+1}^{\infty} \lambda^{-k_i} Z^*_{k_i} \right).$$

We show below that $\sum \lambda^{-k} Z^*_k < \infty$. Once done, we choose T_i so large that $\sum_{k>T_i} \lambda^{-k} Z^*_k < \varepsilon M^{-5} 2^{-i}$ and get $\nu[X \setminus R(T_0, T_1, T_2, \dots)] < \varepsilon$.

To see that $\sum_{k\ge 1} \lambda^{-k} Z^*_k < \infty$. , set $Z_n := Z_n(\phi, a)$, and note that

$$Z_n = \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x) = \sum_{k=1}^{n} \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x) 1_{[\tau_1=k]}(x)$$
$$= \sum_{k=1}^{n} \sum_{T^n x = x} e^{\phi_k(x) + \phi_{n-k}(T^k x)} 1_{[a]}(x) 1_{[\tau_1=k]}(x)$$
$$= M^{\pm 1} \sum_{k=1}^{n} \sum_{T^n x = x} e^{\phi_k(\overline{a, x_1, \dots, x_{k-1}})} e^{\phi_{n-k}(\overline{a, x_{k+1}, \dots, x_{n-1}})} 1_{[a]}(x) 1_{[\tau_1=k]}(x)$$
$$= M^{\pm 1} \sum_{k=1}^{n} \left(\sum_{\substack{[a, x_1, \dots, x_{k-1}, a] \ne \varnothing \\ x_1, \dots, x_{k-1} \ne a}} e^{\phi_k(\overline{a, x_1, \dots, x_{k-1}})} \right) \left(\sum_{[a, y_1, \dots, y_{k-1}, a] \ne \varnothing} e^{\phi_{n-k}(\overline{a, y_1, \dots, y_{n-k-1}})} \right)$$
$$= M^{\pm 1} \left(Z^*_n + \sum_{k=1}^{n-1} Z^*_k Z_{n-k} \right).$$

The result is $Z_n = M^{\pm 1}(Z^*_1 Z_{n-1} + Z^*_2 Z_{n-2} + \cdots + Z^*_{n-1} Z_1 + Z^*_n)$, an approximate "renewal equation" (see §3.1.2 below). We pass to generating functions:

$$t(x) := 1 + \sum_{k=1}^{\infty} Z_k x^k \quad \text{and} \quad r(x) := \sum_{k=1}^{\infty} Z^*_k x^k.$$

The radius of convergence of $t(x)$ is λ^{-1}. The radius of convergence of $r(x)$ is at least as large. If we expand $r(x)t(x)$ and compare the result to $t(x) - 1$ then we see that $t(x) - 1 = M^{\pm 2} r(x) t(x)$, whence $r(x) \le M^2$ for all $0 < x < \lambda^{-1}$. It follows that $r(\lambda^{-1}) < \infty$, which is the same as saying that $\sum \lambda^{-k} Z^*_k < \infty$.

Step 2. Stochastic bound for the growth of x_i, and the tightness of ν_n.

Any x in the support of ν_n contains infinitely many a's. These a's divide $(x_0, x_1, \ldots)$ into infinitely many segments of lengths τ_i $(i \geq 1)$. In the previous step we bounded the growth of τ_i outside a set with ν_n–measure less than ε for all n. The key to this step is to observe that *given that* $\tau_i < T_i$, the size of x_j for j in the i–th gap can be uniformly controlled outside some other set of x's which has small ν_n–measure for all n.

Every sequence of natural numbers $\{k_i\}_{i\geq 1}$ define a *skeleton*

$$\mathfrak{s}(k_1, k_2, \ldots) := \{x \in [a] : \forall i \quad \tau_i(x) = k_i\}.$$

Of the uncountably many possible skeletons, only countably many have positive ν_n–measure for some n. We call these the "relevant" skeletons.

We construct a compact set $F \subseteq [a]$ such that for every relevant skeleton $\mathfrak{s}\{k_i\}$,

$$\forall i \ (k_i \leq T_i) \ \Rightarrow \ \forall n \ \nu_n(F^c \cap \mathfrak{s}\{k_i\}) \leq \varepsilon\, \nu_n(\mathfrak{s}\{k_i\}). \tag{2.16}$$

This is enough to prove tightness, because (2.16) implies that for every n,

$$\begin{aligned}\nu_n(F^c) &\leq \nu_n(F^c \cap R(T_1, T_2, \ldots)^c) + \sum_{\substack{\text{relevant skeletons} \\ \text{s.t. } \forall i (k_i \leq T_i)}} \nu_n(F^c \cap \mathfrak{s}\{k_i\}) \\ &\leq \varepsilon[1 + \sup_n \nu_n(X)] = O(\varepsilon).\end{aligned}$$

The F we construct is of the form $F = \{x \in [a] : \forall i \ (x_i \leq N_i)\}$ where $N_i \in \mathbb{N}$ need to be chosen. Set $\Theta_a(k; N) := \{x \in [a] : \tau_1(x) = k \text{ and } \exists 1 \leq i \leq k \ (x_i > N)\}$ and

$$Z_k^*(N) = \sum_{T^k x = x} e^{\phi_k(x)} 1_{\Theta_a(k;N)}(x)$$

Obviously, $Z_k^*(N) \downarrow 0$ as $N \uparrow \infty$. For every i, we choose N_i in a way such that

1. $\forall k \leq T_i, \quad Z_k^*(N_i) \leq \varepsilon 2^{-i} M^{-7} Z_k^*$;
2. $N_1 < N_2 < N_3 < \cdots$;
3. $N_1 >$ all the coordinates of x_a (possible, because x_a is periodic).

To prove (2.16), fix some relevant skeleton $\mathfrak{s}\{k_i\}$ such that $\forall i (k_i \leq N_i)$ and $\nu_n(\mathfrak{s}\{k_i\}) > 0$. Fix $N = N(n, \{k_i\})$ s.t. $k_1 + \ldots + k_N \geq n$, then

$$\nu_n(F^c \cap \mathfrak{s}\{k_i\}) \leq \sum_{i=1}^{\infty} \nu_n \left\{ x \in \mathfrak{s}\{k_i\} : \exists j \in \left[\sum_{m=1}^{i-1} k_m, \sum_{m=1}^{i} k_m \right) \ x_j > N_j \right\}.$$

We can replace the half open intervals by open intervals, since $x_j = a < N_j$ at the edges. We can replace $\sum_{i=1}^{\infty}$ by $\sum_{i=1}^{N}$, because for $j > k_1 + \cdots + k_N \geq n$, if $x \in \text{supp}(\nu_n)$ then x_j appears in x_a, whence $x_j < N_1 < N_j$. We can replace "$x_j > N_j$" by "$x_j > N_i$", because $j > k_1 + \cdots + k_{i-1} \geq i - 1$ so $N_j \geq N_i$ and the condition "$x_j > N_i$" is weaker. These modifications give

$$\nu_n\big(F^c\cap\mathfrak{s}\{k_i\}\big)\le\sum_{i=1}^N\nu_n\Big\{x\in\mathfrak{s}\{k_i\}:\exists j\in\big(\sum_{m=1}^{i-1}k_m,\sum_{m=1}^{i}k_m\big)\ \ x_j>N_i\Big\}.$$

If x belongs to the set on the right and $\nu_n\{x\}\neq 0$, then necessarily

- $T^k x = x_a$ for some $k\le n$ (by the definition of ν_n);
- $k=k_1+\cdots+k_\ell$ for some ℓ (because $x_k=(x_a)_0=a$, $x\in\mathfrak{s}\{k_i\}$, and $\ell\le N$);
- $\exists i\le\ell$ and $j\in\left(\sum\limits_{m=1}^{i-1}k_m,\sum\limits_{m=1}^{i}k_m\right)$ s.t. $x_j>N_i$ ($i\le\ell$ otherwise $j\ge\sum_{m=1}^{\ell}k_m=k$, and then $x_j=(x_a)_{j-k}\le N_i$).

Manipulating as before, we see that

$$\begin{aligned}
&\nu_n\left(F^c\cap\mathfrak{s}\{k_i\}\right)\le\\
&\le M^3\sum_{i=1}^N\left(\frac{1}{a_n}\sum_{\ell=i}^N\lambda^{-(k_1+\cdots+k_\ell)}Z^*_{k_1,\dots,k_{i-1}}Z^*_{k_i}(N_i)Z^*_{k_{i+1},\dots,k_\ell}1_{\mathfrak{s}\{k_m\}_{m>\ell}}(x_a)\right)\\
&\le M^3\sum_{i=1}^N\frac{\varepsilon}{2^iM^7a_n}\left(\sum_{\ell=i}^N\lambda^{-(k_1+\cdots+k_\ell)}Z^*_{k_1,\dots,k_{i-1}}Z^*_{k_i}Z^*_{k_{i+1},\dots,k_\ell}1_{\mathfrak{s}\{k_m\}_{m>\ell}}(x_a)\right)\\
&\le M^6\sum_{i=1}^N\frac{\varepsilon}{2^iM^7}\left(\frac{1}{a_n}\sum_{\ell=1}^N\lambda^{-(k_1+\cdots+k_\ell)}Z^*_{k_1,\dots,k_\ell}1_{\mathfrak{s}\{k_m\}_{m>\ell}}(x_a)\right)\\
&\le\varepsilon\,\nu_n\left(\mathfrak{s}\{k_i\}\right).
\end{aligned}$$

This proves (2.16). By the discussion at the beginning of the step, this proves that ν_n is tight.

Part 3: Construction of ν.

Part 1 and 2 show that the conditions of the Helly–Prohorov theorem hold, therefore there exists a subsequence m_k such that $\forall b\in S$, $\left\{\nu^b_{m_k}\right\}_{k\ge1}$ converges weak star to some limit ν^b. Let $\nu=\sum_{b\in S}\nu^b$.

We show that $L^*_\phi\nu=\lambda\nu$. For every cylinder $[\underline{b}]$ and $N\in\mathbb{N}$,

$$\begin{aligned}
\nu\left(1_{[x_0<N]}L_\phi 1_{[\underline{b}]}\right)&=\lim_{k\to\infty}\frac{1}{a_{m_k}}\sum_{j=1}^{m_k}\lambda^{-j}L^j_\phi(1_{[x_0<N]}L_\phi 1_{[\underline{b}]})(x_a)\\
&=\lim_{k\to\infty}\frac{1}{a_{m_k}}\sum_{j=1}^{m_k}\lambda^{-j}L^{j+1}_\phi(1_{[x_1<N]}1_{[\underline{b}]})(x_a)\\
&=\lambda\lim_{k\to\infty}\frac{1}{a_{m_k}}\sum_{j=2}^{m_k+1}\lambda^{-j}L^j_\phi(1_{[x_1<N]}1_{[\underline{b}]})(x_a)\\
&=\lambda\lim_{k\to\infty}\frac{1}{a_{m_k}}\sum_{j=1}^{m_k}\lambda^{-j}L^j_\phi(1_{[x_1<N]}1_{[\underline{b}]})(x_a),
\end{aligned}$$

because $a_n \to \infty$ (by assumption), and $\lambda^{-m_k+1}(L_\phi^{m_k+1}(1_{[x_1<N]}1_{[\underline{b}]})(x_a) = O(1)$ (by (2.13),(2.14)). Thus $\nu\left(1_{[x_0<N]}L_\phi 1_{[\underline{b}]}\right) = \lambda\nu\left(1_{[x_1<N]}1_{[\underline{b}]}\right)$ for all N. By the monotone convergence theorem, $\nu\left(L_\phi 1_{[\underline{b}]}\right) = \lambda\nu[\underline{b}]$. Since $[\underline{b}]$ was arbitrary, we have that $L_\phi^*\nu = \lambda\nu$. □

Problems

2.1. Let $\mathscr{B}$ denote the Borel sigma algebra of a TMS X, and let $\sigma(x_n, x_{n+1}, \ldots)$ denote the smallest sigma algebra with respect to which $x \mapsto x_k$ is measurable for all $k \geq n$. Show that $\sigma(x_n, x_{n+1}, \ldots) = T^{-n}\mathscr{B}$.

2.2. Suppose ν is a sigma finite Borel measure on a TMS X with set of states S, and let $(\nu \circ T)(E) := \sum_{a\in S} \nu[T(E\cap[a])]$. Show that

1. $\nu \circ T$ is sigma finite.
2. for all non-negative Borel functions $f : X \to \mathbb{R}$, $\int f d\nu \circ T = \sum_{a\in S}\int_{T[a]} f(ax)d\nu(x)$.
3. If $\nu \circ T^{-1} \ll \nu$, then $\nu \ll \nu \circ T$.
4. Suppose for every $a \in S$ and $E \subset [a]$ Borel measurable, $\mu(E) = 0 \Leftrightarrow \mu(T[a]) = 0$. Then $\nu \circ T \sim \nu$.

2.3. Show that the map V in the proof of theorem 2.7 is continuous.

2.4. Suppose X is a topologically transitive TMS, and suppose $\phi : X \to \mathbb{R}$ is a function. Let L_ϕ denote the Ruelle operator $(L_\phi f)(x) = \sum_{Ty=x} e^{\phi(y)} f(y)$. Calculate L_ϕ^n, and show that $\limsup \frac{1}{n}\log Z_n(\phi, a) \leq \log \sup_X [L_\phi 1]$.

2.5. Suppose μ is a Markov measure with initial distribution (π_i) and stochastic matrix (p_{ij}). Find the transfer operator of μ.

2.6. Suppose T is a non–singular map on $(\Omega, \mathscr{F}, \mu)$, and $h \geq 0$ is measurable. Express the transfer operator of $hd\mu$ in terms of the transfer operator of ν. In the special case of a TMS, if the transfer operator of μ is the Ruelle operator of ϕ, and the transfer operator of $hd\mu$ is the Ruelle operator of ψ, what can you say on ϕ and ψ?

2.7. Suppose T is a non-singular conservative map of a sigma finite measure space. Show that for every $f \in L^1$, if $\widehat{T}f \leq f$ then $\widehat{T}f = f$.

2.8 (The Hopf Decomposition). Suppose T is a non-singular of a general sigma finite measure space $(\Omega, \mathscr{B}, \mu)$. Let $f \in L^1$ be a positive function, and set

$$\mathfrak{C} := [\sum \widehat{T}^k f = \infty] \text{ and } \mathfrak{D} := [\sum \widehat{T}^k f < \infty].$$

1. Let E be a measurable subset of $\mathfrak{C}$. Show that for a.e. $x \in E$, $T^n(x) \in E$ for infinitely many $n \in \mathbb{N}$.

2. Use the following steps to show that $\mathfrak{D}$ is equal, up to a set of measure zero, to a union of wandering sets.
   a. Suppose $A \subset \mathfrak{D}$ has positive measure. Use the arguments of the proofs of theorem 2.1 and proposition 2.5 to show that A contains a wandering set of positive measure.
   b. Let $X = \biguplus_{N\in\mathbb{N}} X_N$ be a decomposition of X into pairwise disjoint sets of finite measure. Fix $N \in \mathbb{N}$ and define by induction the following objects:
      i. $D_0 := \mathfrak{D} \cap X_N$, $\varepsilon_0 := \sup\{\mu(W) : W \subset D_0 \text{ is wandering}\}$, $W_0 \subset D_0$ a wandering set of measure at least $\varepsilon_0/2$.
      ii. $D_{n+1} := D_n \setminus W_n$, $\varepsilon_{n+1} := \sup\{\mu(W) : W \subset D_n \text{ is wandering}\}$, $W_{n+1} \subset D_{n+1}$ a wandering set of measure at least $\varepsilon_{n+1}/2$.

      Show that $\mathfrak{D} \cap X_N = \biguplus_{n\geq 1} W_n \mod \mu$.
   c. Show that $\mathfrak{D}$ is equal, up to a set of measure zero, to a countable union of wandering sets.
3. Suppose $X = \mathfrak{C}' \uplus \mathfrak{D}'$ is another decomposition of X with the properties described in parts 1 and 2. Show that $\mu(\mathfrak{C} \triangle \mathfrak{C}') = 0 = \mu(\mathfrak{D} \triangle \mathfrak{D}')$.

$\mathfrak{C}$ and $\mathfrak{D}$ are called, respectively, the *conservative* and *dissipative* parts of X.

2.9. Let X be a topologically transitive TMS, and suppose ν is a non–singular measure on X s.t. $\frac{d\nu}{d\nu\circ T} > 0$ almost everywhere. Show that every cylinder has positive measure.

2.10. Suppose T is a conservative non-singular map, and let T_A be the induced map on some set of positive finite measure A. Show that if T is ergodic, then T_A is ergodic. (Hint: show that every T_A–invariant function on A extends to a T–invariant measure on A)

2.11 (Uniqueness of ACIM). Suppose T is a non–singular, conservative, and ergodic map of a sigma finite measure space $(\Omega, \mathscr{B}, \nu)$. Show that any two acim's of ν are proportional, using the following steps:

1. Suppose μ_1, μ_2 are two ergodic invariant *probability* measures s.t. $\mu_1 \ll \mu_2$. Show that $\mu_1 = \mu_2$.
2. Now extend this result to σ–finite invariant measure, by considering the induced map on some set of finite measure.

References

1. Aaronson, J.: An introduction to infinite ergodic theory. *Mathematical Surveys and Monographs* **50**, *American Mathematical Society, Providence RI (1997). xii+284 pp.*
2. Aaronson, J., Denker, M., and Urbański, M.: *Ergodic theory for Markov fibred systems and parabolic rational maps.* Trans. Amer. Math. Soc. **337** (1993), no. 2, 495–548.
3. Bowen, R.: *Invariant measures for Markov maps of the interval.* With an afterword by Roy L. Adler and additional comments by Caroline Series. Comm. Math. Phys. **69** (1979), no. 1, 1–17.

4. Dobrusin, R.L.: *Description of a random field by means of conditional probabilities and conditions for its regularity.* Teor. Verojatnost. i Primenen *13* (1968) 201–229. Engl. Transl.: *Theory of Prob. and Appl.* **13** 197–223 (1968).
5. Halmos, P.R.: Lectures on ergodic theory. *Chelsea Publishing Company (1956), vii+99pp.*
6. Hopf, E.: Ergodentheorie. *Ergeb. Mat.* **5** *Springer, Berlin* (1937).
7. Kelley, John L.: General topology. *Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.]. Graduate Texts in Mathematics, No.* **27**. *Springer-Verlag, New York-Berlin, 1975. xiv+298 pp.*
8. Lanford, O. E., III; Ruelle, D.: *Observables at infinity and states with short range correlations in statistical mechanics.* Comm. Math. Phys. **13** (1969) 194–215.
9. Petersen, K., and Schmidt, K.: *Symmetric Gibbs measures.* Trans. Amer. Math. Soc. **349** (1997), no. 7, 2775–2811.
10. Kac, M.: *On the notion of recurrence in discrete stochastic processes.* Bull. Amer. Math. Soc. (1947), 1002–1010.
11. Kakutani, S.: *Induced measure preserving transformations*, Proc. Imperial Acad. Sci. Tokyo **19** (1943), 635–641.
12. Lin, M.: *Mixing for Markov operators*, Z. Wahrscheinlikeit u.v. Gebiete **19** (1971), 231–243.
13. Rényi, A.: *Representations for real numbers and their ergodic properties*, Acta. Math. Acad. Sci. Hungar. **8** (1957), 477–493.
14. Parthasarathy, K. R.: Probability measures on metric spaces. *Reprint of the 1967 original. AMS Chelsea Publishing, Providence, RI, 2005. xii+276 pp.*
15. Ruelle, D.: Thermodynamic Formalism: The Mathematical Structures of Equilibrium Statistical Mechanics. 2nd edition. *Cambridge Mathematical Library, Cambridge University Press, 2004,* xx+174pp.
16. Schweiger, F.: Ergodic theory of fibred systems and metric number theory. *Clarendon Press, Oxford* (1995),
17. Thaler, M.: *Transformations on* $[0,1]$ *with infinite invariant meaures*, Israel J. Math **46** (1983) 67–96.

Chapter 3
Thermodynamic limits

In the previous chapter we discussed the Dobrushin–Lanford–Ruelle approach for making sense of the impossible relation $\mathbb{P}(\text{configuration}) \propto \exp(-\beta \times \text{Energy})$ in the case of an uncountable configuration space. That approach used systems of conditional measures, and led to the definition of a DLR measure.

In this chapter we discuss a different approach to this problem: Approximate the configuration space by a finite or countable set, define the canonical ensemble there, and pass to the limit. The resulting measures are called "thermodynamic limits".

3.1 Thermodynamic Limits

3.1.1 The Definition of a Thermodynamic Limit

Let X be a *compact* TMS with set of states S, and suppose $\beta > 0$ and $U : X \to \mathbb{R}$ is continuous. Fix some $x \in X$, and consider the following finite approximation to X:

$$X_n^x := T^{-n}\{x\} = \{(y_0, \ldots, y_{n-1}; x_0, x_1, \ldots) \in X : y_0, \ldots, y_{n-1} \in S\}.$$

The Gibbs distribution (canonical ensemble) on X_n^x is the measure $\frac{1}{Z_n^x}\mu_n^x$, where

$$\mu_n^x := \sum_{T^n y = x} e^{\phi_n(y)} \delta_y \text{ and } Z_n^x := \mu_n^x(X) = \sum_{T^n y = x} e^{\phi_n(y)}.$$

As always, $\phi_n = \phi + \phi \circ T + \cdots + \phi \circ T^{n-1}$, and δ_y is the point mass measure at y, defined by $\delta_y(E) = 1_E(y)$ for all measurable sets E. Since X is compact, $|X_n^x| \leq |S|^n < \infty$, so Z_n^x is finite, and $\frac{1}{Z_n^x}\mu_n^x$ is a probability measure.

Definition 3.1 (Thermodynamic Limits for compact TMS). A probability measure ν on a compact TMS is called a *thermodynamic limit* for $\phi = -\beta U$ with *boundary condition* $x \in X$, if $\exists n_k \uparrow \infty$ s.t. $\frac{1}{Z_{n_k}^x}\mu_{n_k}^x[\underline{a}] \xrightarrow[k\to\infty]{} \nu[\underline{a}]$ for all cylinders $[\underline{a}]$.

(Convergence on cylinders can be replaced by weak star convergence, because any continuous function on a compact TMS can be approximated uniformly by a finite linear combination of indicators of cylinders.)

In the *non–compact* case, $|S| = \infty$, and Z_n^x could be infinite. If this happens, then we cannot normalize μ_n^x and we have a problem in defining the canonical ensemble. The requirement that $Z_n^x = \mu_n^x(X)$ be finite is too strong for some applications. The next best thing is to ask for $\mu_n^x(F) < \infty$ for all *finite* unions of partition sets. We are led to the following definition:

Definition 3.2 (Thermodynamic Limits for General TMS). Suppose X is a TMS, $\beta > 0$, and $U : X \to X$ is continuous. Assume that $\mu_n^x(F) < \infty$ for all finite unions of partition sets F. A probability measure ν is called a *thermodynamic limit* for $\phi = -\beta U$ with *boundary condition* $x \in X$, if $\exists n_k \uparrow \infty$ s.t. for every finite union of partition sets F s.t. $\nu(F) \neq 0$, and for every cylinder $[\underline{a}]$,

$$\frac{1}{Z_n^x(F)} \mu_{n_k}^x([\underline{a}] \cap F) \xrightarrow[k\to\infty]{} \nu([\underline{a}]|F), \text{ where } Z_n^x(F) := \mu_n^x(F) = \sum_{T^n y = x} e^{\phi_n(y)} 1_F(y).$$

Remark 1: It is not difficult to check that in this case, there exists sequence of constants $A_n(x)$ s.t. $\frac{1}{A_{n_k}(x)} \mu_{n_k}^x[\underline{a}] \xrightarrow[k\to\infty]{} \nu[\underline{a}]$ for all cylinders $[\underline{a}]$ (Problem 3.1).

Remark 2: For the subsequence $n_k \uparrow \infty$ above, $\frac{1}{A_{n_k}(x)} \int f d\mu_{n_k}^x \to \int f d\nu$ for all uniformly continuous $f : X \to \mathbb{R}$ supported inside a finite union of cylinders.

The sequence μ_n^x can be written concisely using the *Ruelle Operator* of $\phi : X \to \mathbb{R}$

$$(L_\phi f)(x) = \sum_{Ty = x} e^{\phi(y)} f(y).$$

The following holds (Problem 3.2):

$$(L_\phi^n f)(x) = \sum_{T^n y = x} e^{\phi_n(y)} f(y); \tag{3.1}$$

$$\mu_n^x[\underline{a}] = (L_\phi^n 1_{[\underline{a}]})(x); \tag{3.2}$$

$$Z_n^x(F) = (L_\phi^n 1_F)(x). \tag{3.3}$$

Thus the thermodynamic limits of ϕ can be studied by analyzing L_ϕ^n as $n \to \infty$.

3.1.2 A Special Case: Countable Markov Chains

It is instructive to start by considering the following special case:

(a) X is a topologically mixing TMS with set of states S and transition matrix (t_{ij});
(b) $\phi(x) = f(x_0, x_1)$ for some function f;
(c) $L_\phi 1 = 1$.

In this case the problem can be solved using the theory of Markov chains as follows. Define a matrix $P = (p_{ab})_{S\times S}$ by

$$p_{ab} = \begin{cases} \exp f(b,a) & t_{ba} = 1, \\ 0 & t_{ba} = 0. \end{cases}$$

Note that we have inverted the order of the indices. P is a stochastic matrix, because $\sum_b p_{ab} = \sum_{b:t_{ba}=1} e^{f(b,a)} = (L_\phi 1)(a,*,*,\dots) = 1$. Since we have inverted the order of indices, the directed graph associated to P is the the directed graph associated with X with the direction of the edges inverted.

The *Markov chain started at state a* is the stochastic process $\{X_n\}_{n\geq 0}, X_n : X \to S$, $X_n(x) := x_n$, together with the probability measure $\mathbb{P}_a$ given on cylinders by

$$\mathbb{P}_a[\xi_0,\dots,\xi_{n-1}] = \delta_{a\xi_0} p_{\xi_0\xi_1} p_{\xi_1\xi_2} \cdot \dots \cdot p_{\xi_{n-2}\xi_{n-1}}.$$

Here δ_{ij} equals one when $i = j$ and zero otherwise. Of course $\mathbb{P}_a[a] = 1$.

The powers of L_ϕ are related to the powers of P as follows: If $P^n =: (p^{(n)}_{ab})_{S\times S}$, $x_a \in [a]$, and $[\underline{b}] = [b_0,\dots,b_{m-1}]$, then

$$\begin{aligned}
(L_\phi^{n+m} 1_{[\underline{b}]})(x_a) &= \sum_{\substack{(b_0,\dots,b_{m-1},\xi_{n-1},\dots,\xi_1,a)\\ \text{admissible}}} e^{\phi(b_0,b_1)} \cdots e^{\phi(b_{m-2},b_{m-1})} \times \\
&\qquad \times e^{\phi(b_{m-1},\xi_{n-1})} e^{\phi(\xi_{n-1},\xi_{n-2})} \cdots e^{\phi(\xi_2,\xi_1)} e^{\phi(\xi_1,a)} \\
&= \sum_{\substack{(b_{m-1},\xi_{n-1},\dots,\xi_1,a)\\ \text{admissible}}} p_{a\xi_1} \cdots p_{\xi_{n-1}b_{m-1}} \cdot p_{b_{m-1}b_{m-2}} \cdots p_{b_1 b_0} \\
&= p^{(n)}_{ab_{m-1}} \cdot p_{b_{m-1}b_{m-2}} \cdots p_{b_1 b_0},
\end{aligned}$$

where we have used the identity $p^{(n)}_{ab} = \sum_{\eta_1,\dots,\eta_{n-1}\in S} p_{a\eta_1} p_{\eta_1\eta_2} \cdots p_{\eta_{n-1}b}$.

We see that $(L_\phi^{n+m} 1_{[\underline{b}]})(x_a) = p^{(n)}_{ab_{m-1}} p_{b_{m-1}b_{m-2}} \cdots p_{b_1 b_0}$.

If $|S| < \infty$, then the Perron–Frobenius Theorem says that $p^{(n)}_{ab_{m-1}} \xrightarrow[n\to\infty]{} \pi_{b_{m-1}}$, where $(\pi_b)_{b\in S}$ is the unique probability vector s.t. $\pi P = \pi$. In this case we get that $(L_\phi^{n+m} 1_{[\underline{b}]})(x) \to \nu[\underline{b}]$, where $\nu[b_0,\dots,b_{m-1}] = \pi_{b_{m-1}} p_{b_{m-1}b_{m-2}} \cdots p_{b_1,b_0}$ is the Markov measure of (P,π) "ran backwords".

If $|S| = \infty$, then the asymptotic behavior of $p^{(n)}_{ab}$ depends on whether P is positive recurrent, null recurrent, or transient. We recall what this means.

The *taboo probabilities* for states are defined by

$${}_c p^{(n)}_{ab} := \sum_{\xi_1,\dots,\xi_{n-1}\neq c} p_{a\xi_1} p_{\xi_1\xi_2} \cdots p_{\xi_{n-1}b} \qquad (a,b,c \in S).$$

This is the probability that the Markov chain started at a will reach b after n–steps, without having passed through c (after leaving a and before arriving to b). In the special case $a = b = c$ we have the following useful interpretations:

$$p_{aa}^{(n)} = \text{The probability that the Markov chain started at } a \text{ will return to } a \text{ after } n \text{ steps}$$

$$ {}_a p_{aa}^{(n)} = \text{The probability that the Markov chain started at } a \text{ will return to } a \text{ for the } \textit{first time} \text{ after } n \text{ steps}$$

$$\sum_{n=1}^{\infty} {}_a p_{aa}^{(n)} = \text{The probability that the Markov chain started at } a \text{ will return to } a;$$

$$\sum_{n=1}^{\infty} {}_a p_{aa}^{(n)} n = \text{The mean time it takes the Markov chain started at } a \text{ to return to } a.$$

Proposition 3.1. $\sum p_{aa}^{(n)} = \infty \Leftrightarrow \sum {}_a p_{aa}^{(n)} = 1$. *Thus* $\sum p_{aa}^{(n)} = \infty$ *iff a.e. path which starts at a eventually returns to a.*

Proof. Write $u_n := p_{aa}^{(n)}$, $u_0 := 1$ and $f_n :=_a p_{aa}^{(n)}$, $f_0 := 0$. The following holds:

$$u_0 = 1 \text{ and } u_n = f_n u_0 + f_{n-1} u_1 + f_{n-2} u_2 + \cdots + f_1 u_{n-1}. \tag{3.4}$$

To see this decompose the event "return to a after n steps" into the disjoint events "return to a for the first time after k steps, and then return to a again after $n-k$ steps".

Let $U(x) := 1 + \sum_{n \geq 1} u_n x^n$, $F(x) := \sum_{n \geq 1} f_n x^n$. The renewal equation implies that $U(x)F(x) = U(x) - 1$, whence $U(x) = 1/1 - F(x)$. Thus

$$1 + \sum_{n=1}^{\infty} u_n = \lim_{x \to 1^-} U(x) = \begin{cases} \infty & F(1) = 1 \\ \frac{1}{1-F(1)} & F(1) < 1. \end{cases}$$

The proposition follows from this and the interpretation of $F(1) = \sum_a p_{aa}^{(n)}$ as the probability of returning to a from a at least once. □

Terminology: Equation (3.4) is called the *renewal equation*. A sequence $\{u_n\}_{n \geq 0}$ which satisfied (3.4) with $f_n \geq 0$ is called a *renewal sequence*.

Definition 3.3. Suppose $P = (p_{ab})_{S \times S}$ is a stochastic matrix.

1. A state a is called *positive recurrent* ("PR"), if $\sum p_{aa}^{(n)} = \infty$ and $\sum {}_a p_{aa}^{(n)} n < \infty$;
2. A state a is called *null recurrent* ("NR"), if $\sum p_{aa}^{(n)} = \infty$ and $\sum {}_a p_{aa}^{(n)} n = \infty$;
3. A state a is called *transient* ("T"), if $\sum p_{aa}^{(n)} < \infty$.

In other words, a is PR if the Markov chain started at a returns to a with probability one and finite expected first return time; It is NR if the chain returns to a with probability one but with infinite expected return time; And it is T if there is positive probability that the chain will never return to a.

A state which is positive recurrent or null recurrent is called *recurrent*. If a is recurrent, let

$$\mu_a := \sum_a p_{aa}^{(n)} n \in [0,\infty]$$

denote the expectation of the first return time. A stochastic matrix is called *irreducible aperiodic* if its associated TMS is topologically mixing.

Theorem 3.1 (Kolmogorov). *Let $P = (p_{ij})_{S\times S}$ be an irreducible aperiodic stochastic matrix.*

1. *If one state is PR then all states are PR, and $\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} \neq 0$;*
2. *If one state is NR, then all states are NR, and $\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} = 0$;*
3. *If one state is T state, then all states are T, and $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ (trivially).*

We deduce this from the following important theorem of Erdös, Feller, and Pollard:[1]

Theorem 3.2 (Renewal Theorem). *Suppose u_n and f_n satisfy the renewal equation (3.4). If $f_n \geq 0$, $\sum f_n = 1$ and $\gcd\{n : u_n \neq 0\} = 1$, then $u_n \xrightarrow[n\to\infty]{} 1/\sum n f_n$.*

Proof. [2] The renewal equation says that $u_0 = 1$ and $u_n = \sum_{k=1}^{n} f_k u_{n-k}$ for $n \geq 1$. It is easy to see by induction that $u_n \in [0,1]$ for all n. Choose a subsequence $n_k \uparrow \infty$ s.t. $u_{n_k} \to L_0 := \limsup u_n$. Using Cantor's diagonal argument it is no problem to choose n_k so that for every $m \in \mathbb{Z}$,

$$u_{n_k+m} \xrightarrow[k\to\infty]{} L_m \text{ for some } L_m \in [0,1].$$

Step 1. $L_m = L_0 = \limsup_{n\to\infty} u_n$ for all $m \in \mathbb{Z}$.

We prove this in the special case when $f_n \neq 0$ for all n, and delegate the general case $\gcd\{n : f_n \neq 0\} = 1$ to the exercises (Problem 3.4).

For all k, $u_{n_k} = \sum_{j=1}^{n_k} f_j u_{n_k-j}$. In the limit $k \to \infty$, we obtain $L_0 = \sum_{j=1}^{\infty} f_j L_{-j}$. By definition, $L_{-j} \leq \limsup u_n \equiv L_0$, so we get

$$L_0 = \sum_{j=1}^{\infty} f_j L_{-j} \leq \sum_{j=1}^{\infty} f_j L_0 = L_0.$$

It follows that $L_{-j} = L_0$ for all j s.t. $f_j \neq 0$. In other words, $L_{-j} = L_0$ for all $j \geq 0$.

Now consider the identity $u_{n_k+1} = \sum_{j=1}^{n_k+1} f_j u_{n_k+1-j}$. Passing to the limit as before, we get $L_1 = \sum_{j=1}^{\infty} f_j L_{1-j} = \sum_{j=1}^{\infty} f_j L_0 = L_0$. Thus $L_1 = L_0$. The same argument for u_{n_k+2} gives that $L_2 = L_0$. By induction, $L_j = L_0$ for all $j \geq 0$, whence for all j. ◇

Step 2. $L_0 = 1/\sum n f_n$.

Set $F_j := \sum_{i>j} f_i$, we claim that for all n

[1] This approach to Kolmogorov's theorem is due to K.-L. Chung.

[2] We follow Feller's book [5].

$$u_nF_0 + u_{n-1}F_1 + \cdots + u_1F_{n-1} + u_0F_n = 1. \tag{3.5}$$

To see this sum the identities $f_1u_{k-1}+f_2u_{k-2}+\cdots+f_{k-1}u_1+f_k=u_k$ over $1 \le k \le n$:

$$f_1u_{n-1} + (f_1+f_2)u_{n-2} + \cdots + (f_1+\cdots+f_{n-1})u_1 + (f_1+\cdots+f_n) = (u_n+\cdots+u_1).$$

Moving all the u_i's to the right hand side, we find (since $1-(f_1+\cdots+f_j)=F_j$) that $1-F_n = u_nF_0+u_{n-1}F_1+\cdots+u_1F_{n-1}$. This is equivalent to (3.5), because $u_0=1$.

We now apply (3.5) in the particular case $n=n_k$: $\sum_{j=0}^{n_k} F_j u_{n_k-j}=1$. We want to pass to the limit $k\to\infty$. Since $u_{n_k-j}\to L_{-j}=L_0$ (step 2),

1. either $\sum F_j=\infty$, and then we must have $L_0=0$ (otherwise the limit of the left hand side in $\sum_{j=0}^{n_k} F_j u_{n_k-j}=1$ is infinite);
2. or $\sum F_j<\infty$, and then we get the identity $\sum F_jL_0=1$, whence $L_0=1/\sum F_j$.

It is easy to see that $\sum F_j=\sum nf_n$, so $L_0=1/\sum nf_n$. ♢

Step 3. $u_n \xrightarrow[n\to\infty]{} L_0$.

If $\sum nf_n=\infty$, then $L_0=1/\sum nf_n=0$ and $u_n\to 0$ (a non–negative sequence whose limsup is zero, converges to zero). Suppose that $\sum nf_n<\infty$, and assume by way of contradiction that $\exists m_k\uparrow\infty$ s.t. $u_{m_k}\to L\neq L_0$. Since $L_0=\limsup u_n$, $L<L_0$.

Fix $\varepsilon>0$ and let $N(\varepsilon)$ be so large that for all $n\ge N(\varepsilon)$, $u_n\le L_0+\varepsilon$. Given ℓ, choose $K(\varepsilon,\ell)$ s.t. for all $k>K(\varepsilon,\ell)$, $m_k\ge N(\varepsilon)+\ell$, then (3.5) implies that

$$\begin{aligned}1 &= F_0u_{m_k} + (F_1u_{m_k-1}+\cdots+F_\ell u_{m_k-\ell}) + (F_{\ell+1}u_{m_k-\ell-1}+\cdots+F_{m_k}u_0)\\ &\le F_0u_{m_k} + (F_1+\cdots+F_\ell)(L_0+\varepsilon) + \sum_{j=\ell+1}^{\infty} F_j.\end{aligned}$$

As $\ell\to\infty$, $u_{m_k}\to L$ and $(F_1+\cdots+F_\ell)\to\sum_{j\ge 0}F_j-1=\sum nf_n-1=L_0^{-1}-1$, so

$$1 \le L + \left(\frac{1}{L_0}-1\right)(L_0+\varepsilon).$$

But this inequality cannot hold for all $\varepsilon>0$, because the limit of the right hand side as $\varepsilon\to 0^+$ is $1+(L-L_0)<1$. □

Proof of Kolmogorov's Theorem (K.-L. Chung). Recall from the proof of proposition 3.1 that $u_n=p_{aa}^{(n)}$ is the renewal sequence with $f_n={}_a\,p_{aa}^{(n)}$.

By the renewal theorem, if a state is PR then $p_{aa}^{(n)} \xrightarrow[n\to\infty]{} 1/\mu_a\neq 0$, and if it is NR then $p_{aa}^{(n)} \xrightarrow[n\to\infty]{} 1/\mu_a=0$. It clear that if a state is transient, then $p_{aa}^{(n)} \xrightarrow[n\to\infty]{} 0$.

Step 1. Either all states are PR, or all states are NR, or all states are T. (In Markov chain theory, such a statement is called a "solidarity theorem".)

Suppose a is PR, and let b be some other state. It is easy to see, using the irreducibility of P, that there are constants C_1,C_2,k_1,k_2 (which depend on a,b) s.t.

$$C_1 p_{bb}^{(n-k_1)} \le p_{aa}^{(n)} \le C_2 p_{bb}^{(n+k_2)} \text{ for all } n.$$

Since $p_{aa}^{(n)} \not\to 0$, $p_{bb}^{(n)} \not\to 0$, so b is PR. Thus, if one state is PR, all states are PR.

Suppose a is NR, and let b be some other state. Since a is NR, $\sum p_{aa}^{(n)} = \infty$ and $p_{aa}^{(n)} \to 0$. Using the same method as above, we see that $\sum p_{bb}^{(n)} = \infty$ and $p_{bb}^{(n)} \to 0$. The divergence of $\sum p_{bb}^{(n)}$ means that b is PR or NR. The convergence of $p_{bb}^{(n)}$ to zero means that b is not PR. So b is NR. Thus, if one state is NR, all states are NR.

It follows that if one state is T, then all states are T. ♢

Step 2. Calculation of $\lim_{n\to\infty} p_{ij}^{(n)}$.

At the beginning of the proof we saw that if $i = j$, then the limit is $1/\mu_j$ in the PR case, and zero otherwise. Suppose $i \neq j$.

Assume first that all states are PR. For the Markov chain started at i, the event "arriving to j in n steps" is the disjoint union of the events "arriving to j for the first time at time k, then returning to j after additional $n-k$ steps", thus

$$\begin{aligned}
p_{ij}^{(n)} &= \sum_{k=1}^{n} {}_j p_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\lfloor n/2\rfloor -1} + \sum_{k=\lfloor n/2\rfloor}^{n} \\
&= \sum_{k=1}^{\lfloor n/2\rfloor -1} {}_j p_{ij}^{(k)} [\tfrac{1}{\mu_j} + o(1)] + O\left(\sum_{k=\lfloor n/2\rfloor}^{\infty} {}_j p_{ij}^{(k)}\right), \quad \text{because } j \text{ is PR} \\
&= \sum_{k=1}^{\lfloor n/2\rfloor -1} {}_j p_{ij}^{(k)} [\tfrac{1}{\mu_j} + o(1)] + o(1), \text{ because } \sum {}_j p_{ij}^{(k)} = \mathbb{P}_i[\text{arriving to } j] < \infty \\
&= \frac{1}{\mu_j} \sum_{k=1}^{\lfloor n/2\rfloor -1} {}_j p_{ij}^{(k)} + o\left(\sum_{k=1}^{\lfloor n/2\rfloor -1} {}_j p_{ij}^{(k)}\right) + o(1) \\
&= \frac{1}{\mu_j} \sum_{k=1}^{\lfloor n/2\rfloor -1} {}_j p_{ij}^{(k)} + o(1) \xrightarrow[n\to\infty]{} \frac{1}{\mu_j}.
\end{aligned}$$

Next suppose that all states are NR or T, then $p_{jj}^{(n)} \to 0$. The same calculation, with $\frac{1}{\mu_j}$ replaced by zero, shows that $p_{ij}^{(n)} \to 0$. □

Returning to L_ϕ we see that $P_G(\phi) = \lim \frac{1}{n} \log L_\phi^n 1_{[a]} = \lim \frac{1}{n} \log p_{aa}^{(n)} \le 0$. In the positive recurrent or null recurrent cases $\sum p_{aa}^{(n)} = \infty$, so $P_G(\phi) = 0$. If P is positive recurrent, then $(\lambda^{-n} L_\phi^n 1_{[\underline{b}]})(x)$ converges for all x to a positive limit which only depends on $[\underline{b}]$. If P is null recurrent then $(\lambda^{-n} L_\phi^n 1_{[\underline{b}]})(x) \to 0$. We do not know what happens in the transient case.

We also see that the null recurrent and transient cases are only possible when $|S| = \infty$. In the compact case, $p_{aa}^{(n)}$ converges to a positive limit by the Perron–Frobenius theorem, so P must be positive recurrent.

We end this section with a famous example of null recurrence and transience. Let $\mathbf{e}_1,\dots,\mathbf{e}_d$ denote the standard basis of $\mathbb{R}^d$, and define the stochastic matrix $P = (p_{\mathbf{u},\mathbf{v}})_{\mathbb{Z}^d\times\mathbb{Z}^d}$ by

$$p_{\mathbf{u},\mathbf{v}} = \begin{cases}(2d)^{-1} & \mathbf{v}=\mathbf{u}+\mathbf{e}_i \text{ for some } i\\ 0 & \text{otherwise.}\end{cases}$$

The Markov chain started at $\mathbf{0}\in\mathbb{Z}^d$ with stochastic matrix Q is called the *simple random walk on* $\mathbb{Z}^d$, because it describes a "walk" on $\mathbb{Z}^d$ which starts at the origin and proceeds by independent random steps with possible values $\pm\mathbf{e}_1,\dots,\pm\mathbf{e}_d$, each taken with the same probability.

Theorem 3.3 (Polya). *The simple random walk on $\mathbb{Z}^d$ is null recurrent for $d=1,2$ and transient for $d\geq 2$.*

Proof. We will determine the asymptotic behavior of $p^{(2n)}_{\mathbf{0},\mathbf{0}}$ (the probability of returning to zero at time $2n$). We work with even times because the simple random walk cannot return to zero at odd times.

The first step is to realize that

$$p^{(2n)}_{\mathbf{0},\mathbf{0}} = \mathbb{P}[\mathbf{X}_1+\cdots+\mathbf{X}_{2n}=\mathbf{0}]$$

where $\mathbf{X}_i$, called the "jumps of the random walk", are identically distributed independent random variables which take the values $\pm\mathbf{e}_1,\dots\pm\mathbf{e}_d$ with probability $(2d)^{-1}$.

Next we use a trick of Stone, and write the following identity for the Kronecker delta function: $\delta_{\mathbf{0},\eta} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{i\langle\xi,\eta\rangle}d\xi$, where $d\xi$ is the Lebesgue measure on $\mathbb{R}^d$. We have

$$\begin{aligned}
p^{(2n)}_{\mathbf{0},\mathbf{0}} &= \mathbb{E}[1_{[\mathbf{X}_1+\cdots+\mathbf{X}_{2n}=\mathbf{0}]}] = \mathbb{E}\left(\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} e^{i\langle\xi,\mathbf{X}_1+\cdots+\mathbf{X}_{2n}\rangle}d\xi\right)\\
&= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\mathbb{E}\left(e^{i\langle\xi,\mathbf{X}_1+\cdots+\mathbf{X}_{2n}\rangle}\right)d\xi \quad \text{(Fubini)}\\
&= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\mathbb{E}\left(e^{i\langle\xi,\mathbf{X}_1\rangle}\right)^{2n}d\xi \quad (\mathbf{X}_i \text{ are independent, identically distributed})\\
&= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\left(\frac{1}{d}\sum_{j=1}^d \cos\xi_j\right)^{2n}d\xi \quad \text{(check!)}.
\end{aligned}$$

The integrand is bounded by 1 and tends to zero almost everywhere. By the bounded convergence theorem, $p^{(2n)}_{\mathbf{0},\mathbf{0}}\to 0$, so the random walk is either null recurrent, or transient.

To determine whether it is NR or T we have to see whether $\sum p^{(2n)}_{\mathbf{0},\mathbf{0}}$ diverges or not. By the monotone convergence theorem, this amounts to determining the convergence properties of the integral

$$\int_{[-\pi,\pi]^d} \sum_{n=0}^{\infty} \left(\frac{1}{d} \sum_{j=1}^{d} \cos \xi_j \right)^{2n} d\xi = \int_{[-\pi,\pi]^d} \frac{d\xi}{1 - \left(\frac{1}{d} \sum_{j=1}^{d} \cos \xi_j \right)^2}.$$

There are three singularities, at $\xi_0 = (0,\dots,0), (\pi,\dots,\pi)$, and $(-\pi,\dots,-\pi)$. Taylor's expansion reveals the singularity to be of the form $[d+o(1)]\|\xi - \xi_0\|^{-2}$. The integral on a ball of radius ε centered at ξ_0 can be written using a spherical shell decomposition as

$$(\text{The surface area of the unit sphere in } \mathbb{R}^d) \times \int_0^{\varepsilon} \frac{1}{r^2} r^{d-1} dr.$$

This is infinite for $d = 1,2$, and finite for $d \geq 3$.

The conclusion is that if $d = 1,2$ then $\sum p_{\mathbf{0},\mathbf{0}}^{(2n)} = \infty$ and the random walk is null recurrent, and that if $d \geq 3$ then $\sum p_{\mathbf{0},\mathbf{0}}^{(2n)} < \infty$, and the random walk is transient. □

To get an example of a null recurrent or transient potential on a *topologically mixing* TMS, let X denote the TMS with set of states $S = \mathbb{Z}^d$ and transition matrix $\mathbb{A} = (t_{\mathbf{u},\mathbf{v}})_{\mathbb{Z}^d \times \mathbb{Z}^d}$, where $t_{\mathbf{u},\mathbf{v}} = 1 \Leftrightarrow \mathbf{v} = \mathbf{u}$, or $\mathbf{v} = \mathbf{u} \pm \mathbf{e}_i$ for some $i = 1,\dots,d$, and take $\phi(x) = \log p_{x_0,x_1}^{(2)}$ where $(p_{\mathbf{u},\mathbf{v}})_{\mathbb{Z}^d \times \mathbb{Z}^d}$ is as above. (This matrix is symmetric, so there is no need to invert the order of the indices.)

3.1.3 The Gurevich pressure

Our aim is to find the asymptotic behavior of L_ϕ^n for general potentials $\phi(x)$. We will model the analysis on the theory of countable Markov chains sketched in the previous section, but there are differences. The most obvious of these is that in general one should expect $L_\phi^n f$ to grow or decrease exponentially. To see why, observe that for every constant c, $L_{\phi+c}^n f = e^{nc} L_\phi^n$.

Thus the first thing to do is to determine the exponential rate of growth of L_ϕ^n and "neutralize" it by subtracting a suitable constant from ϕ.

Proposition 3.2. *Suppose X is a topologically mixing TMS, and suppose $\phi : X \to \mathbb{R}$ has the Walters property. For every state a, let $Z_n(\phi,a) := \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x)$.*

1. *$P_G(\phi) := \lim\limits_{n\to\infty} \frac{1}{n} \log Z_n(\phi,a)$ exists for all $a \in S$, and is independent of a.*
2. *$-\infty < P_G(\phi) \leq \infty$. If $\|L_\phi 1\|_\infty < \infty$, then $P_G(\phi) < \infty$.*
3. *For every bounded continuous function f which is (a) non-negative, (b) not identically equal to zero, and (c) supported inside a finite union of cylinders,*

$$\lim_{n\to\infty} \frac{1}{n} \log(L_\phi^n f)(x) = P_G(\phi) \textit{ for all } x \in X.$$

4. *If $P_G(\phi) < \infty$, then $L_\phi^n 1_{[\underline{a}]}$ is finite for all n and $[\underline{a}]$.*

Terminology: $P_G(\phi)$ is called the *Gurevich pressure* of ϕ. It plays a central role in the theory of equilibrium measures. We will study its properties in detail in the following chapters.

Proof of Proposition 3.2. Fix $a \in S$. Since X is topologically mixing, there exists some $n_0 = n_0(a)$ s.t. $a \xrightarrow{n} a$ for all $n \geq n_0$. We have: $Z_n(\phi, a) > 0$ for all $n \geq n_0$.

Define for $n \geq n_0$, $\zeta_n := \log Z_n(\phi, a)$. If x is an n–periodic point in $[a]$, and y is an m–periodic point in $[a]$, then the periods of x, y can be joined to form an $(n+m)$–periodic point in $[a]$. Using this construction and the Walters property of ϕ, it is easy to check that $\{\zeta_n\}_{n \geq n_0}$ is almost super-additive, in the sense that for some positive constant c, and all $m, n \geq n_0$

$$\zeta_{n+m} \geq \zeta_n + \zeta_m - c.$$

Thus either $\zeta_n = \infty$ for all $n \geq 2n_0$, or $-\infty < \zeta_n < \infty$ for all $n \geq n_0$. In the first case $\zeta_n/n \to \infty$ (trivially). Suppose the second.

Fix $m > n_0$. Any $n > m$ can be written in the form $n = q_n m + r_n$, $0 \leq r_n \leq m-1$. Since $\{\zeta_k\}_{k \geq n_0}$ is almost super additive, if we fix m and send $n \to \infty$, then

$$\frac{\zeta_n}{n} \geq \frac{(q_n - 1)\zeta_m + \zeta_{m+r_n} - q_n c}{q_n m + r_n} \geq \frac{(q_n - 1)\zeta_m + O(1)}{(q_n + 1)m} \xrightarrow[n\to\infty]{} \frac{\zeta_m}{m}.$$

Thus $\liminf(\zeta_n/n) \geq \sup\{\zeta_m/m : m \geq n_0\} \geq \limsup(\zeta_n/n)$, and ζ_n/n converges.

The limit is independent of the choice of a, because for any two states a, b there are constants C_1, C_2, k_1, k_2 s.t. $C_1 Z_{n-k_1}(\phi, a) \leq Z_n(\phi, b) \leq C_2 Z_{n+k_2}(\phi, a)$ for all n (c.f. the proof of Theorem 2.2).

The limit is never $-\infty$ because if $x \in [a]$ and $T^{n_0}(x) = x$, then $Z_{kn_0}(\phi, a) \geq \exp \phi_{kn_0}(x) = \exp[k\phi_{n_0}(x)]$, so $P_G(\phi) \geq \phi_{n_0}(x)/n_0 > -\infty$.

The operator L_ϕ is positive, so $\|L_\phi f\|_\infty \leq \|L_\phi 1\|_\infty \|f\|_\infty$. Thus

$$Z_n(\phi, a) \leq \|L_\phi^n 1\|_\infty \leq \|L_\phi 1\|_\infty^n.$$

If $\|L_\phi 1\|_\infty < \infty$, then $P_G(\phi) \leq \log \|L_\phi 1\|_\infty < \infty$.

Suppose $[a_0, \ldots, a_{N-1}]$ is a non–empty cylinder and b is a state. It is easy to see, using the topological mixing of X and the Walters property of X that there are constants C_1, C_1 and k_1, k_2 such that for all $n > k$ and every $x \in [b]$,

$$C_1 Z_{n-k_1}(\phi, a_{N-1}) \leq (L_\phi^n 1_{[\underline{a}]})(x) \leq C_2 Z_{n+k_2}(\phi, a_0). \tag{3.6}$$

It follows that $\frac{1}{n} \log L_\phi^n 1 \xrightarrow[n\to\infty]{} P_G(\phi)$ on $[b]$. Since $[b]$ was arbitrary, the limit holds on X.

Any non-negative, non–identically zero, bounded continuous function f whose support lies inside finitely many cylinders can be sandwiched between a function of the form $\varepsilon 1_{[\underline{a}]}$ and a function of the form $M \sum_{b \in F} 1_{[b]}$ with $F \subseteq S$ finite. It follows that $\frac{1}{n} \log L_\phi^n f \xrightarrow[n\to\infty]{} P_G(\phi)$.

Finally, (3.6) says that if $L_\phi^n 1_{[\underline{a}]} = \infty$ for some x, then $Z_N(\phi, a_0) = \infty$ for some N. This implies, as before, that $Z_{kN}(\phi, a_0) = \infty$ for all k, and $P_G(\phi) = \infty$. □

3.1.4 Modes of Recurrence and the Generalized Ruelle's Perron–Frobenius Theorem

We now ask for the asymptotic behavior of $\lambda^{-n} L_\phi^n$, where $\lambda = \exp P_G(\phi)$. Our aim is to establish a picture similar to that available for countable Markov chains, as presented in §3.1.2. The trick is to find combinatorial sums which make sense for general ϕ, and behave like $p_{aa}^{(n)}$ and ${}_a p_{aa}^{(n)}$ when $\phi = f(x_0, x_1)$. Here they are: given a state a, let

- $\varphi_a(x) := 1_{[a]}(x) \inf\{n \geq 1 : T^n(x) \in [a]\}$ (first return time);
- $Z_n^*(\phi, a) := \sum_{T^n x = x} e^{\phi_n(x)} 1_{[\varphi_a = n]}(x)$;
- $Z_n(\phi, a) := \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x)$;
- $\lambda := \exp P_G(\phi) = \lim [Z_n(\phi, a)]^{1/n}$.

We treat $\lambda^{-n} Z_n(\phi, a)$ as $p_{aa}^{(n)}$ and $\lambda^{-n} Z_n^*(\phi, a)$ as ${}_a p_{aa}^{(n)}$, and define

Definition 3.4 (Modes of Recurrence). Suppose X is a topologically mixing TMS, $\phi : X \to \mathbb{R}$ has summable variations, and $P_G(\phi) < \infty$. Fix some state a, then

1. ϕ is called *recurrent*, if $\sum \lambda^{-n} Z_n(\phi, a) = \infty$ and *transient* , if $\sum \lambda^{-n} Z_n(\phi, a) < \infty$;
2. ϕ is called *positive recurrent*, if it is recurrent and $\sum n \lambda^{-n} Z_n^*(\phi, a) < \infty$;
3. ϕ is called *null recurrent*, if it is recurrent and $\sum n \lambda^{-n} Z_n^*(\phi, a) = \infty$.

We sometime abbreviate and write *R, T, PR, NR* for recurrence, transience etc.

We shall see below that these properties are independent of the choice of state a.

Remark 1: If $P = (p_{ij})$ is a stochastic matrix, then P is PR (resp. NR) iff $\phi := \log p_{x_1 x_0}$ is PR (resp. NR). This is also the case for transient stochastic matrices, provided their *spectral radius* $\rho := \limsup \sqrt[n]{p_{aa}^{(n)}}$ equals one. If the spectral radius is less than one, then a transient stochastic matrix could lead to a PR or NR potential (which is good, because these modes of recurrence are much better understood).

Remark 2: We have already met the condition of "recurrence" as a necessary and sufficient condition for the conservativity of a ϕ–conformal measure (Theorem 2.2).

Theorem 3.4 (Generalized Ruelle's Perron–Frobenius Theorem). *Let X be a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ a function with summable variations and finite Gurevich pressure.*

1. ϕ **is positive recurrent** *iff there are a $\lambda > 0$, a positive continuous function h, and a conservative measure ν which is finite on cylinders, s.t. $L_\phi h = \lambda h$, $L_\phi^* \nu = \lambda \nu$, and $\int h d\nu = 1$. In this case $\lambda = \exp P_G(\phi)$, and for every cylinder $[\underline{a}]$, $\lambda^{-n} L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}] / \int h d\nu$ uniformly on compacts.*

2. ϕ **is null recurrent** *iff there are a* $\lambda > 0$*, a positive continuous function h, and a conservative measure* ν *which is finite on cylinders, s.t.* $L_\phi h = \lambda h$*,* $L_\phi^* \nu = \lambda \nu$*, and* $\int h d\nu = \infty$*. In this case* $\lambda = \exp P_G(\phi)$*, and for every cylinder* $[\underline{a}]$*,* $\lambda^{-n} L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} 0$ *uniformly on compacts.*
3. ϕ **is transient** *iff there are no conservative measures* ν *which are finite on cylinders s.t.* $L_\phi^* \nu = \lambda \nu$ *for some* $\lambda > 0$*.*

If X is compact, then ϕ *is positive recurrent.*

Remark 1: The theorem holds if "summable variations" is replaced by the Walters property.

Remark 2: As we shall prove later, the following limit theorem holds in the null recurrent case: Let $a_n := (\int_{[a]} h d\nu)^{-1} \sum_{k=1}^n \lambda^{-k} Z_k(\phi, a)$, then for all cylinders $[\underline{a}]$, $\frac{1}{a_n} \sum_{k=1}^n \lambda^{-k} L_\phi^k 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}]$ uniformly on compacts.

Corollary 3.1 (Solidarity of States). *The modes of recurrence are independent of the choice of the state a used to define them.*

Corollary 3.2 (Ruelle's Perron-Frobenius Theorem). *Suppose X is a* compact *topologically mixing TMS and* $\phi : X \to \mathbb{R}$ *has summable variations, then there exist* $\lambda > 0$*, h positive continuous, and* ν *a positive probability measure s.t.* $L_\phi h = \lambda h$*,* $L_\phi^* \nu = \lambda \nu$*,* $\int h d\nu = 1$*, and s.t.*

$$\lambda^{-n} L_\phi^n f \xrightarrow[n\to\infty]{} h \int f d\nu \quad \textit{uniformly}$$

for all continuous functions $f : X \to \mathbb{R}$*.*

Corollary 3.3. *Suppose X is a topologically mixing TMS, and* ϕ *has summable variations and finite Gurevich pressure. If* ϕ *is positive recurrent, then* ϕ *has a unique thermodynamic limit, and this limit is a DLR measure for* ϕ*.*

3.2 Proof of the GRPF Theorem

Throughout this section, let X be a topologically mixing TMS with set of states S and transition matrix $\mathbb{A} = (t_{ab})_{S\times S}$, and let $\phi : X \to \mathbb{R}$ be some function with the Walters property (e.g. a function with summable variations). We assume that $P_G(\phi) < \infty$.

3.2.1 Eigenmeasures and Eigenfunctions

Proposition 3.3. ϕ *is recurrent iff there exists a conservative measure* ν *which is finite on cylinders such that for some* $\lambda > 0$*,* $L_\phi^* \nu = \lambda \nu$*. In this case* $\lambda = \exp P_G(\phi)$*, and* ν *gives any cylinder positive measure.*

Proof. ($\Rightarrow$): If ϕ is recurrent, then Theorem 2.9 says that there is a conservative measure ν s.t. $\frac{d\nu}{d\nu\circ T}=\lambda^{-1}\exp\phi$. This measure satisfies $L_\phi^*\nu=\lambda\nu$.

To see ($\Leftarrow$), suppose $L_\phi^*\nu=\lambda\nu$ for some conservative measure ν which is finite on cylinders. We claim that $\lambda=\exp P_G(\phi)$. The transfer operator of ν is $\lambda^{-1}L_\phi$:

$$\int\varphi\lambda^{-1}L_\phi\psi d\nu=\int\lambda^{-1}L_\phi(\varphi\circ T\psi)d\nu=\int\varphi\circ T\psi d\nu$$

for all φ,ψ indicators of cylinders, whence for all $\varphi\in L^\infty$ and $\psi\in L^1$. It follows that $\frac{d\nu}{d\nu\circ T}=\lambda^{-1}e^\phi$. By Theorem 2.2, $\sum\lambda^{-n}Z_n(\phi,a)=\infty$. As a result $\lambda\le e^{P_G(\phi)}$, because the radius of convergence of the series $\sum_{k\ge1}Z_k(\phi,b_0)x^k$ is $e^{-P_G(\phi)}$.

On the other hand, since ν is finite on cylinders and $M:=\exp\sup_n\mathrm{var}_{n+1}\phi_n<\infty$, $Z_n(\phi,a)\le M(L_\phi^n 1_{[a]})(y)$ for all $y\in[a]$. Integrating both sides of the inequality over $[a]$, we see that $Z_n(\phi,a)\nu[a]\le M\lambda^n\nu[a]$, whence (assuming w.l.o.g. that $\nu[a]\neq0$)

$$\lambda^{-n}Z_n(\phi,a)=O(1). \tag{3.7}$$

It follows that $\lambda\ge\exp P_G(\phi)$.

Thus $\lambda=\exp P_G(\phi)$. But $\frac{d\nu}{d\nu\circ T}=\lambda^{-1}\exp\phi$, and ν is conservative, so Theorem 2.2 says that $\sum e^{-nP_G(\phi)}Z_n(\phi,a)=\infty$. This means that ϕ is recurrent. $\square$

Proposition 3.4. *If ϕ is recurrent, and ν is as in the previous proposition then there exists a positive continuous function h s.t.*

1. $L_\phi h=e^{P_G(\phi)}h$;
2. $\mathrm{var}_k[\log h]\le\sup_n[\mathrm{var}_{n+k}\phi_n]\le\sum_{\ell\ge k+1}\mathrm{var}_\ell\phi$;
3. $\log h$ *is uniformly continuous, and* $\mathrm{var}_1[\log h]<\infty$.

Proof. Let $\lambda:=\exp P_G(\phi)$, and fix some state a. Since ϕ is recurrent,

$$a_n:=\sum_{k=1}^n\lambda^{-k}Z_n(\phi,a)\xrightarrow[n\to\infty]{}\infty.$$

By (3.7), $|a_{n+1}-a_n|=O(1)$. Let $f_N:=(1/a_N)\sum_{k=1}^N\lambda^{-k}L_\phi^k 1_{[a]}$. We show that f_N has a subsequence which converges uniformly on compacts to a non–zero limit. This will be h.

We need the following observation (Problem 3.6): For any two states a,b and any cylinder $[\underline{c}]$, there are constants C_1,C_2,k_1,k_2,N s.t. for all $x_b\in[b]$ and $n\ge N$,

$$C_1Z_{n-k_1}(\phi,a)\le(L_\phi^n 1_{[\underline{c}]})(x_b)\le C_2Z_{n+k_2}(\phi,a). \tag{3.8}$$

It follows that for every state b, there exists $N_b\in\mathbb{N}$ such that $\{\log f_N\}_{N\ge N_b}$ is uniformly bounded on $[b]$.

It is also easy to see that, using the Walters property of ϕ, that $\mathrm{var}_k[\log f_N]\le\sup_n\mathrm{var}_{n+k}\phi_n\xrightarrow[k\to\infty]{}0$, so $\{\log f_N\}_{N\ge N_b}$ is equicontinuous on $[b]$. By the Arzela–Ascoli theorem, $\{f_N\}_{N\ge1}$ has a subsequence which converges uniformly on compact subsets of $[b]$ to a positive continuous limit.

Since there are only countably many possible $b \in S$, we can obtain a subsequence $N_k \uparrow \infty$ s.t. $f_{N_k} \xrightarrow[k\to\infty]{} h$ uniformly on compacts in X for some limiting function h.

The limit h must also be bounded away from zero and infinity on partition sets, and it must satisfy $\mathrm{var}_k[\log h] \leq \sup_n[\mathrm{var}_{n+k}\phi_n] \leq \sum_{j=k+1}^{\infty} \mathrm{var}_k \phi$. In particular $\mathrm{var}_1[\log h] < \infty$.

We show that $L_\phi h \leq \lambda h$. Choose some increasing sequence of finite sets $S_m \subseteq S$ s.t. $S_m \uparrow S$, then for every m

$$\begin{aligned}
\lambda^{-1} \sum_{\substack{Ty=x\\ y\in S_m}} e^{\phi(y)} h(y) &= \lim_{k\to\infty} \frac{1}{a_{N_k}} \sum_{n=1}^{N_k} \lambda^{-(n+1)} \sum_{\substack{Ty=x\\ y\in S_m}} e^{\phi(y)} L_\phi^n 1_{[a]}(y) \quad (\because |S_m| < \infty)\\
&\leq \lim_{k\to\infty} \frac{1}{a_{N_k}} \sum_{n=1}^{N_k} \lambda^{-(n+1)} L_\phi^{n+1} 1_{[a]}(x) \quad (\because S_m \subseteq S)\\
&= \lim_{k\to\infty} \frac{1}{a_{N_k}} \sum_{n=1}^{N_k} \lambda^{-n} L_\phi^n 1_{[a]}(x) + \\
&\qquad + \lim_{k\to\infty} \frac{1}{a_{N_k}} \left(\lambda^{-(N_k+1)} (L_\phi^{N_k+1} 1_{[a]})(x) - \lambda^{-1} (L_\phi 1_{[a]})(x) \right).
\end{aligned}$$

The term in the brackets is uniformly bounded on partition sets, because of (3.8) and (3.7). It follows that $\lambda^{-1} \sum_{\substack{Ty=x\\ y\in S_m}} e^{\phi(y)} h(y) \leq h(x)$ for all m. In the limit $m \to \infty$, we obtain $\lambda^{-1} L_\phi h \leq h$.

We claim that $\lambda^{-1} L_\phi h = h$. Otherwise, $f := h - \lambda^{-1} L_\phi h > \varepsilon > 0$ on some cylinder $[\underline{a}]$, and we have for all N,

$$\sum_{k=1}^{N} \lambda^{-k} L_\phi^k 1_{[\underline{a}]} \leq \frac{1}{\varepsilon} \sum_{k=1}^{N} \lambda^{-k} L_\phi^k f \leq \frac{1}{\varepsilon} h \qquad \text{(telescopic sum)}.$$

Passing to the limit $N \to \infty$ we see that $\sum \lambda^{-k} L_\phi^k 1_{[\underline{a}]} < \infty$ everywhere. This implies by (3.8) that $\sum \lambda^{-k} Z_k(\phi, a) < \infty$, in contradiction to the recurrence of ϕ. □

Proposition 3.5. *Suppose ϕ is recurrent, and λ, h and ν are as in the previous propositions. If ϕ is PR, then $\int h d\nu < \infty$, and if ϕ is NR, then $\int h d\nu = \infty$.*

Proof. Let $dm = h d\nu$, then the transfer operator of m is $\widehat{T} f = \lambda^{-1} h^{-1} L_\phi(hf)$ (check) and m is T–invariant:

$$\begin{aligned}
\int f \circ T dm &= \int \lambda^{-1} L_\phi(hf \circ T) d\nu \ \ (\because L_\phi^* \nu = \lambda \nu)\\
&= \int f \lambda^{-1} L_\phi h d\nu = \int f h d\nu = \int f dm.
\end{aligned}$$

Recall that $\varphi_a(x) := 1_{[a]}(x) \inf\{n \geq 1 : T^n(x) \in [a]\}$. Since $\mathrm{var}_1 h < \infty$, h is uniformly bounded away from zero and infinity on states. Choose some constant C s.t. $C^{-1} < h(x) < C$ on $[a]$, and let $M := \exp[\sup_n(\mathrm{var}_{n+1}\phi_n)]$ (c.f. lemma 1.1), then

$$
\begin{aligned}
m[\varphi_a = N] &= \int \widehat{T}^N 1_{[\varphi_a=N]}\,dm = \int \lambda^{-N} h^{-1} L_\phi^n (h 1_{[\varphi_a=N]}) dm \\
&= C^{\pm 2}\lambda^{-N}\int_{[a]} L_\phi^N 1_{[\varphi_a=N]} dm \quad (\because L_\phi^N 1_{[\varphi_a=N]} \text{ is supported in } [a]) \\
&= C^{\pm 2} M^{\pm 1}\lambda^{-N} Z_N^*(\phi,a) m[a].
\end{aligned}
$$

It follows that $\lambda^{-N}Z_N^*(\phi,a) = (C^2M)^{\pm 1} m[\varphi_a = N]/m[a]$.

Thus $\sum n\lambda^{-n}Z_n^*(\phi,a)$ converges iff $\int_{[a]}\varphi_a dm < \infty$. By the Kac formula, this happens iff $m(X) < \infty$. ☐

3.2.2 The limit of $\lambda^{-n}L_\phi^n$

Proposition 3.6. *Suppose ϕ is NR and $\lambda = \exp P_G(\phi)$, then $\lambda^{-n}L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} 0$ uniformly on partition sets.*

Proof. Let h and ν be the eigenfunction and eigenmeasure of L_ϕ which we constructed in the previous section, and set $dm = hd\nu$. Recall that m is invariant, and that every cylinder has positive finite measure.

Since h is bounded away from zero and infinity on partition sets, it is enough to show that for every $a \in S$, $\lambda^{-n}h^{-1}L_\phi^n\left(h1_{[a]}\right) \xrightarrow[n\to\infty]{} 0$ uniformly on partition sets.

Choose finite unions of partition sets F_n, such that $F_n \uparrow X$ and $0 < m(F_n) < \infty$. Since ϕ is null recurrent, $m(F_N) \uparrow m(X) = \infty$. Set $f_N = 1_{[a]} - 1_{F_N}\cdot m[a]/m(F_N)$.

Recall that the transfer operator of m is $f \mapsto \lambda^{-1}h^{-1}L_\phi(hf)$, and define as always $M := \exp[\sup_n(\mathrm{var}_{n+1}\phi_n)]$. For every $b \in S$ the usual estimations yield (for $\|\cdot\|_1 = \|\cdot\|_{L^1(m)}$)

$$
\begin{aligned}
\left\|1_{[b]}\widehat{T}^n 1_{[a]}\right\|_\infty &\le \frac{M}{m[b]}\left\|1_{[b]}\widehat{T}^n 1_{[a]}\right\|_1 \le \frac{M}{m[b]}\left(\left\|1_{[b]}\widehat{T}^n f_N\right\|_1 + \frac{m[a]}{m(F_N)}\left\|1_{[b]}\widehat{T}^n 1_{F_N}\right\|_1\right) \\
&\le \frac{M}{m[b]}\left(\left\|\widehat{T}^n f_N\right\|_1 + \frac{m[a]m[b]}{m(F_N)}\right) \quad (\because m\circ T^{-1} = m).
\end{aligned}
$$

The measure m is exact, because it is equivalent to ν and ν is a conservative measure whose log–Jacobian satisfies the Walters property (theorem 2.5). Since $\int f_N dm = 0$, Lin's theorem (theorem 2.4) says that $\left\|\widehat{T}^n f_N\right\|_{L^1(m)} \to 0$. This and $m(F_N)\uparrow\infty$ imply that $\left\|1_{[b]}\widehat{T}^n 1_{[a]}\right\|_\infty \xrightarrow[n\to\infty]{} 0$ as required. ☐

Proposition 3.7. *Suppose ϕ is positive recurrent and $\lambda = \exp P_G(\phi)$, and let h,ν be the eigenfunction and eigenmeasure constructed above. For every cylinder $[\underline{a}]$, $\lambda^{-n}L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}]/\int hd\nu$ uniformly on compacts and in $L^1(\nu)$.*

Proof. Normalize the eigenfunction h so that $\int h\,d\nu = 1$. We saw above that $dm = hd\nu$ is a shift invariant probability measure. Since m is equivalent to ν, and ν is exact (Theorem 2.5), m is exact — whence mixing.

Step 1. For every state a, $\exists N, C > 1$ s.t. $C^{-1} \le \lambda^{-n} Z_n(\phi, a) \le C$ for all $n \ge N$.

Fix a state a, and let $\widehat{T}$ denote the transfer operator of $dm = hd\nu$. Recall that this is the operator $\widehat{T} f = \lambda^{-1} h^{-1} L_\phi(hf)$.

Since $\mathrm{var}_1 h < \infty$, h is uniformly bounded away from zero and infinity on $[a]$. Choose some $H > 1$ s.t. $H^{-1} \le h(x) \le H$ for all $x \in [a]$, and denote as always $M := \exp[\sup_{n\ge 1} \mathrm{var}_{n+1}\phi_n]$. For all $x, y \in [a]$ and $n \in \mathbb{N}$, we have

$$\begin{aligned}\lambda^{-n}(L_\phi^n 1_{[a]})(x) &= M^{\pm 1}\lambda^{-n}(L_\phi^n 1_{[a]})(y) = M^{\pm 1}H^{\pm 2}(\widehat{T}^n 1_{[a]})(y)\\ &= (MH^2)^{\pm 1}\frac{1}{m[a]}\int_{[a]}(\widehat{T}^n 1_{[a]})(y)dm(y) = \frac{(MH^2)^{\pm 1}}{m[a]} m([a]\cap T^{-n}[a]).\end{aligned}$$

But m is mixing, so $m([a]\cap T^{-n}[a]) \xrightarrow[n\to\infty]{} m[a]$. Since $m[a] \ne 0$ (Problem 2.9), we see that there are constants C, N s.t. $C^{-1} \le \lambda^{-n} L_\phi^n 1_{[a]} \le C$ on $[a]$ for all $n \ge N$.

We now invoke (3.8) in the particular case $[a] = [b] = [\underline{c}]$ to see that there are constants C_0, N_0 s.t. $C_0^{-1} \le \lambda^{-n} Z_n(\phi, a) \le C_0$ on $[a]$ for all $n \ge N_0$. $\diamondsuit$

Step 2. For every state b, $\exists N_b$ s.t. $\{\lambda^{-n} L_\phi^n 1_{[\underline{a}]}\}_{n\ge N_b}$ is equicontinuous and uniformly bounded away from zero and infinity on $[b]$.

Fix a state b. Equation (3.8) provides constants C_1, C_2, k_1, k_2 and N_b s.t.

$$C_1 \lambda^{n-k_1} Z_{n-k_1}(\phi, a) \le \lambda^{-n}(L_\phi^n 1_{[\underline{a}]})(x) \le C_2 \lambda^{n+k_2} Z_{n+k_2}(\phi, a)$$

for all $x \in [b]$ and $n \ge N_b$. By step 1, $\lambda^{-n}(L_\phi^n 1_{[\underline{a}]})(x)$ is bounded away from zero and infinity on $[b]$.

Since ϕ has the Walters property, it is clear that $\{\log[\lambda^{-n} L_\phi^n 1_{[\underline{a}]}]\}_{n\ge N_b}$ is equicontinuous on $[b]$. For any sequence of positive functions $\varphi_n(x)$, if $\sup\|\varphi_n\|_\infty < \infty$ and $\{\log\varphi_n\}_{n\ge N}$ is equicontinuous, the $\{\varphi_n\}_{n\ge N}$ is equicontinuous. It follows that $\{\log[\lambda^{-n} L_\phi^n 1_{[\underline{a}]}]\}_{n\ge N_b}$ is equicontinuous on $[b]$. $\diamondsuit$

Step 3. $\lambda^{-n} L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}]$ pointwise.

Suppose by way of contradiction that $\varphi_n := \lambda^{-n} L_\phi^n 1_{[\underline{a}]} \not\to h\nu[\underline{a}]$ on some point x^*. By step 2 and the Arzela–Ascoli Theorem, there is a subsequence $\{\varphi_{n_k}\}_{k\ge 1}$ which converges uniformly on compacts to a continuous limit φ s.t. $\varphi(x^*) \ne h(x^*)\nu[\underline{a}]$.

For every k, $\varphi_{n_k} = \lambda^{-n_k} L_\phi^{n_k} 1_{[\underline{a}]} \le C^* h$ where $C^* = 1/\inf\{h(x) : x \in [\underline{a}]\}$. Since $C^* h$ is integrable, we have by the dominated convergence theorem that

$$\begin{aligned}\int |\varphi - h\nu[\underline{a}]|\, d\nu &= \lim_{k\to\infty}\int \left|\lambda^{-n_k} L_\phi^{n_k} 1_{[\underline{a}]} - h\nu[\underline{a}]\right| d\nu\\ &= \lim_{k\to\infty}\int \left|\widehat{T}^{n_k}\left(h^{-1}1_{[\underline{a}]} - \nu[\underline{a}]\right)\right| dm = 0,\end{aligned}$$

by Lin's Theorem and the exactness of m.

We have that $\varphi = h\nu[\underline{a}]$ almost everywhere. Since φ is continuous, $\varphi = h\nu[\underline{a}]$ everywhere, in particular at x^*. $\diamondsuit$

Since $\varphi_n := \lambda^{-n} L_\phi^n 1_{[\underline{a}]}$ converges to $h\nu[\underline{a}]$ pointwise and $\{\varphi_n\}_{n\geq 1}$ is equicontinuous, $\varphi_n \to h\nu[\underline{a}]$ uniformly on compacts. The convergence also holds in the sense of $L^1(\nu)$, because as we saw in the proof of step 3, $|\varphi_n| \leq C^* h$ and $h \in L^1(\nu)$. $\square$

3.2.3 Compact TMS and shifts with the BIP Property

Definition 3.5. A TMS with a set of states S and a transition matrix $\mathbb{A} = (t_{ij})_{S\times S}$ is said to have the *Big Images and Prepimages (BIP) property* if there is a finite set of states $b_1, \ldots, b_N$ s.t. $\forall a \in S\ \exists 1 \leq i, j \leq N$ s.t. $t_{b_i a} t_{a b_j} = 1$.

For example, any compact TMS has the BIP property (take $\{b_1, \ldots, b_N\} = S$). Non–compact examples include the full shift. The following result is implicit in the work of Mauldin & Urbański:

Proposition 3.8. *If X has the BIP property, then any $\phi : X \to \mathbb{R}$ with the Walters property s.t.* $\mathrm{var}_1\phi < \infty$ *and* $P_G(\phi) < \infty$ *is positive recurrent.*

Proof. [3] Fix some $a \in S$ and set $\lambda := e^{P_G(\phi)}$. We construct $N_1 \in \mathbb{N}$, such that

$$\inf\left\{\lambda^{-n} Z_n(\phi, a) : n \geq N_1\right\} > 0 \tag{3.9}$$

This implies recurrence, and rules out null recurrence, since for every $x \in [a]$, $\lambda^{-n} L_\phi^n 1_{[a]}(x) \asymp \lambda^{-n} Z_n(\phi, a) \neq o(1)$, whereas Proposition 3.6 says that in the NR case, $\lambda^{-n} L_\phi^n 1_{[a]} \to 0$.

Let $\mathscr{W}_n := \{\underline{a} \in S^n : [\underline{a}] \neq \varnothing\}$ and set for every $\underline{a} \in \mathscr{W}_n$,

$$\phi_n(a_0, \ldots, a_{n-1}) := \inf\{\phi_n(x) : x \in [\underline{a}]\}\ ,\ \phi_n[a_0, \ldots, a_{n-1}] := \sup\{\phi_n(x) : x \in [\underline{a}]\}.$$

These numbers are finite, because if we define the constant M_0 by

$$\log M_0 := \sup_n [\mathrm{var}_{n+1}\phi_n] + \mathrm{var}_1\phi,$$

then $|\phi_n(\underline{a}) - \phi(x)|, |\phi[\underline{a}] - \phi(x)| \leq \log M_0$ for all $x \in [\underline{a}]$. This also shows that $|\phi_n(\underline{w}) - \phi_n[\underline{w}]| \leq \log M_0$.

Let $\{b_1, \ldots, b_N\}$ be the finite set of states given by the BIP property. The topological mixing of X guarantees the existence of some $n_1 \in \mathbb{N}$ and $\underline{w}_{ab_i}, \underline{w}_{b_j a} \in \mathscr{W}_{n_1}$ $(1 \leq i, j \leq N)$ such that $(a, \underline{w}_{ab_i}, b_i)$ and $(b_j, \underline{w}_{b_j a}, a)$ are admissible. We call these words "bridge words".

By virtue of the defining property of $\{b_1, \ldots, b_N\}$ and the bridge words, for every n and $\underline{w} \in \mathscr{W}_n$ there are $1 \leq i, j \leq N$ s.t.

[3] The proof we give is from [9].

$$(a, \underline{w}_{ab_i}, b_i, \underline{w}, b_j, \underline{w}_{b_j a}, a) \in \mathscr{W}_{n+k_1+1}, \text{ where } k_1 := 2(n_1+2)-1 = 2n_1+3.$$

Set $C := \min\{e^{\phi_{n_1+2}(a, \underline{w}_{ab_i}, b_i)} \cdot e^{\phi_{n_1+1}(b_j, \underline{w}_{b_j a})} : i, j = 1, \dots, N\}$, then for $C_1 := C\lambda^{-N_1}$,

$$\lambda^{-(n+N_1)} Z_{n+N_1}(\phi, a) \geq C_1 \lambda^{-n} \sum_{\underline{w} \in \mathscr{W}_n} e^{\phi_n(\underline{w})} \geq \frac{C_1}{M_0} \lambda^{-n} \sum_{\underline{w} \in \mathscr{W}_n} e^{\phi_n[\underline{w}]}.$$

If we can show that the last expression is bounded below, then we will have (3.9) and be done.

Indeed, $\lambda^{-n} \sum_{\underline{w} \in \mathscr{W}_n} e^{\phi_n[\underline{w}]} \geq 1$ for all n, otherwise $\exists n_0$ such that

$$\lambda^{-n_0} \sum_{\underline{w} \in \mathscr{W}_{n_0}} e^{\phi_{n_0}[\underline{w}]} < r < 1,$$

and then for all k $\lambda^{-kn_0} Z_{kn_0}(\phi, a) \leq \left(\lambda^{-n_0} \sum_{\underline{w} \in \mathscr{W}_{n_0}} e^{\phi_{n_0}[\underline{w}]}\right)^k < r^k$. But this is impossible, since by the definition of the Gurevich pressure $\lambda^{-kn} Z_{kn}(\phi, a)$ does not decay exponentially fast. □

Corollary 3.4. *Any potential with summable variations on a topologically mixing compact TMS is positive recurrent.*

3.3 Applications

3.3.1 Absolutely Continuous Invariant Densities

We can now prove the following theorem which was mentioned in the previous chapter:

Theorem 2.6 [Aaronson, Denker, and Urbański] *Let X be a topologically mixing TMS, and suppose ν is a conservative non–singular measure which is finite on cylinders. If the log Jacobian of ν has summable variations, then ν has a continuous acim which is bounded away from zero and infinity on cylinders.*

Proof. Let ϕ be a continuous version of the log–Jacobian of ν, then the transfer operator of ν is L_ϕ (Corollary 2.1) and $L_\phi^* \nu = \nu$.

Since ν is conservative, the Generalized Ruelle's Perron–Frobenius Theorem says that ϕ is recurrent with pressure zero. Since ϕ is recurrent, L_ϕ has a positive continuous eigenfunction h. This is an invariant density for ν, because $m = h d\nu$ satisfies $m(T^{-1}[\underline{a}]) = m[\underline{a}]$ for all cylinders $[\underline{a}]$:

$$m(T^{-1}[\underline{a}]) = \int 1_{[\underline{a}]} \circ T h d\nu = \int 1_{[\underline{a}]} L_\phi h d\nu = \int 1_{[\underline{a}]} h d\nu = m[\underline{a}].$$

This invariant density is bounded away from zero and infinity on cylinders, because as we saw above $\mathrm{var}_1[\log h] < \infty$. □

Theorem 3.5 ("Folklore Theorem"). *Suppose $f : [0,1] \to [0,1]$ is a topologically mixing piecewise expanding Markov map with Markov partition $\{I_k : k \in \mathbb{N}\}$. Assume that*

1. *"Finite images": $\{f(I_k) : k \in \mathbb{N}\}$ is a finite set;*
2. *"Adler's condition": $\sup|f''|/|f'|^2 < \infty$.*

The f has an integrable Hölder continuous ACIP, and this ACIP is mixing.

3.3.2 Uniqueness of DLR measures

The following result was proved independently by Ruelle and by Dobrushin.

Theorem 3.6. *Let X be a* compact *topologically mixing TMS, and suppose $\phi : X \to \mathbb{R}$ has the Walters property, then ϕ has a unique DLR measure, and this measure is the ϕ–conformal measure.*

Proof. [4] Every ϕ with summable variations on a compact TMS has finite pressure. Assume w.l.o.g. that $P_G(\phi) = 0$ (otherwise work with $\phi - P_G(\phi)$ and note that every DLR measure of ϕ is a DLR measure of $\phi - P_G(\phi)$ and vice verse).

Existence: Theorem 2.7.

Uniqueness: Since X is compact, ϕ is positive recurrent and there exist h, ν s.t. $L_\phi^* \nu = \nu$, $L_\phi h = h$, and $\int h d\nu = 1$. The function h is continuous, and X is compact, so h is bounded below. It follows that $\nu(X) < \infty$. Re–normalizing ν and h (if necessary), we may assume w.l.o.g. that $\nu(X) = 1$. The idea of the proof is to show that if ν' is a DLR measure for ϕ, then $\nu' = \nu$.

Let $m := h d\nu$. We need the following estimate on m: There exists a constant $C > 1$ s.t. for every $x \in X$,

$$m[x_0, \ldots, x_{n-1}] = C^{\pm 1} e^{\phi_n(x)}. \tag{3.10}$$

To see this use the compactness of X and the continuity of h, ϕ to find constants $H > 1$ s.t. $H^{-1} \le h \le H$ and $M := \exp[\mathrm{var}_1 \phi + \sup_{n \ge 1} \mathrm{var}_n \phi_{n+1}] < \infty$, then

$$\begin{aligned} m[x_0, \ldots, x_{n-1}] &= H^{\pm 1} \nu[x_0, \ldots, x_{n-1}] = H^{\pm 1} \int 1_{[x_0, \ldots, x_{n-1}]} d\nu(y) \\ &= H^{\pm 1} \int (L_\phi^n 1_{[x_0, \ldots, x_{n-1}]})(y) d\nu(y) = (HM)^{\pm 1} e^{\phi_n(x)} m(T[x_{n-1}]). \end{aligned}$$

Since $|S| < \infty$, $m(T[x_{n-1}])$ is bounded away from zero and infinity, and (3.10) follows.

Now let ν' be some DLR measure for ϕ. For every state $s \in S$ and n–cylinders $[\underline{a}], [\underline{b}]$ which terminate at s, let $\vartheta_{\underline{b},\underline{a}} : [\underline{b}] \to [\underline{a}]$ be the map $(\underline{b}, x) \mapsto (\underline{a}, x)$. By Proposition 2.1 and (3.10), for every $x \in T[s]$,

[4] We follow the proof in [1]

$$\nu'[\underline{a}] = \int_{[\underline{b}]} \frac{d\nu' \circ \vartheta_{\underline{ba}}}{d\nu'} d\nu' = M^{\pm 1} e^{\phi_n(\underline{a}x) - \phi_n(\underline{b}x)} \nu'[\underline{b}] = (CM)^{\pm 1} m[\underline{a}] \cdot e^{-\phi_n(\underline{b}x)} \nu'[\underline{b}].$$

Fixing $[\underline{b}]$ and varying $[\underline{a}]$, we see that there are constants $K_n(s)$ $(s \in S)$ s.t. for every n–cylinder $[\underline{a}]$ which terminates in the symbol s,

$$\nu'[\underline{a}] = (CM)^{\pm 1} K_n(s) m[\underline{a}] \quad (x \in T[s]). \tag{3.11}$$

We claim that $\sup_n K_n(s) < \infty$. Sum the inequalities (3.11) over all N–cylinders $[\underline{a}]$ which terminate at s, then

$$1 \geq \nu'(T^{-n}[s]) \geq (CM)^{-2} K_n(s) m(T^{-n}[s]) = (CM)^{-2} K_n(s) m[s]$$

(because $m \circ T^{-1} = m$). It follows that $\sup_n K_n(s) \leq (CM)^2 \max\{1/m[s] : s \in S\} < \infty$.

By (3.11), $\nu' \ll m$, whence $\nu' \ll \nu$.

To see that $\nu' = \nu$, set $F := \frac{d\nu'}{d\nu}$, then $F d\nu = d\nu'$ is a DLR measure. So is ν (because it is conformal). It follows that the transformations $(\underline{a}x) \mapsto (\underline{b}x)$ have the same Radon Nikodym derivatives w.r.t. ν and $F d\nu$. Equating these derivatives we see that

$$F(\underline{a}x) = F(\underline{b}x)$$

for all pairs of cylinders of the same length $\underline{a}, \underline{b}$ which end at the same state. It follows that for every n there is a Borel function F_n s.t. $F(x) = F_n(T^n x)$, and so F is $T^{-n}\mathscr{B}$ measurable. This holds for all n, so F is measurable w.r.t. the tail σ–algebra $\bigcap_{n \geq 1} T^{-n}\mathscr{B}$. Since m is exact, F is constant. The value of the constant must be one, because ν, ν' are probability measures. We conclude that $\nu' = \nu$. □

3.3.3 *g–functions*

In section 3.1.2 we determined the asymptotic behavior of L_ϕ^n for potentials ϕ of the form $\phi(x) = f(x_0, x_1)$, subject to the condition $L_\phi 1 = 1$, which allowed us create a stochastic matrix out of L_ϕ. The condition $L_\phi 1 = 1$ appears in many other situations. The purpose of this section is to show that under fairly general situations, it can always be assumed "without loss of generality". But first we give it a name:

Definition 3.6 (*g*–functions). Suppose X is a TMS. A continuous function $g : X \to (0, 1]$ is called a *g–function*, if $\sum_{Ty=x} g(y) = 1$ for all x.

The terminology is due to Mike Keane.

Proposition 3.9. *Suppose ν is a non–singular measure on a TMS X s.t. $g = \log \frac{d\nu}{d\nu \circ T}$ is continuous. The following conditions are equivalent:*

1. *g is a g–function;*
2. *ν is T–invariant.*

The proof is left as an exercise.

Theorem 3.7. *Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ has summable variations. Suppose $P_G(\phi) < \infty$.*

1. *If ϕ is recurrent, then $\phi - P_G(\phi)$ is cohomologous via a continuous transfer function φ to $\log g$, where g is a g–function and $\log g$ has summable variations.*
2. *If ϕ is transient, then $\phi - P_G(\phi)$ is cohomologous via a continuous transfer function φ to $\log g$, where $\log g$ has summable variations and $\sum_{Ty=x} g(y) \leq 1$ everywhere. We call g a* sub g–function.

In both cases the cohomology can be done s.t. $\mathrm{var}_1 \varphi < \infty$.

Proof. Let $\lambda := \exp P_G(\phi)$.

In the recurrent case, let h be positive function s.t. $L_\phi h = \lambda h$. Set

$$g := \frac{e^{\phi} h}{\lambda h \circ T},$$

then $\log g = \phi + \log h - \log h \circ T - P_G(\phi)$ and $\log g$ is a g–function:

$$\sum_{Ty=x} g(y) = (\lambda h(x))^{-1} (L_\phi h)(x) = 1.$$

Since $\mathrm{var}_1 [\log h] < \infty$, $\log g$ has summable variations.

If ϕ is transient, let $h := \sum_{n \geq 1} \lambda^{-n} L_\phi^n 1_{[a]}$. This sum converges, because of the transience of ϕ and Problem 3.6. It is easy to verify that $\mathrm{var}_1 [\log h] < \infty$ and that $L_\phi h \leq \lambda h$. The proof follows, as before, by setting $g := e^{\phi} h / (\lambda h \circ T)$. ☐

Corollary 3.5. *Suppose X is a topologically mixing TMS, then any $\phi : X \to \mathbb{R}$ with summable variations is cohomologous to a function $\psi : X \to \mathbb{R}$ with summable variations s.t. $\psi \leq P_G(\psi)$.*

Remark: Theorem 3.7 and Corollary 3.5 hold, with the same proof, if "summable variations" is replaced by "the Walters property".

Problems

3.1. Suppose ν is a thermodynamic limit, and $\{n_k\}_{k \geq 1}$ is a sequence as in definition 3.2.

1. Suppose $\nu[a] \neq 0$. Calculate the limit of $Z_n^x([a] \cup [b]) / Z_n^x([a])$ as $n \to \infty$.
2. Construct a sequence $A_n(x)$ s.t. $\frac{1}{A_{n_k}(x)} \mu_{n_k}^x [\underline{a}] \xrightarrow[k \to \infty]{} \nu[\underline{a}]$ for all cylinders $[\underline{a}]$.

3.2. Check identities (3.1),(3.2), and (3.3).

3.3. Prove (3.6).

3.4. The purpose of this problem is to complete the proof of step 1 in the proof the renewal theorem.

1. Read the first step of the proof of the Renewal Theorem. The following uses the notation of that proof.
2. Show, using induction on k, that $L_{-(j_1+\cdots+j_k)} = L_0$ for all $j_1, \ldots, j_k$ s.t. $f_{j_i} \neq 0$.
3. Show that the set $\{m_1 j_1 + \cdots + m_k j_k : m_i \in \mathbb{N}, f_{j_i} \neq 0\}$ contains a set of the form $\{N, N+1, \ldots\}$. (Hint: $\gcd\{j : f_j \neq 0\} = 1$.)
4. Prove that $L_j = L_0$ for all $j \in \mathbb{Z}$.

3.5. Suppose X is a topologically transitive TMS, and $\phi : X \to \mathbb{R}$ has the Walters property. Show that any measure ν s.t. $L_\phi^* \nu = \lambda \nu$ with $\lambda > 0$ gives positive measure to every cylinder.

3.6. Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ has the Walters property. Show that for any two states a, b and a non–empty cylinder $[\underline{c}]$, there are positive constants C_1, C_2, k_1, k_2, N s.t. $C_1 Z_{n-k_1}(\phi, a) \leq (L_\phi^n 1_{[\underline{c}]})(x_b) \leq C_2 Z_{n+k_2}(\phi, a)$ for all $x_b \in [c]$ and $n \geq N$.

3.7. Suppose X is a topologically mixing TMS and $\phi : X \to \mathbb{R}$ has summable variations. Show that ϕ is positive recurrent iff there are constants $M, N > 1$ and $\mu > 0$ s.t. for some state $a \in S$, $M^{-1} \leq \mu^{-n} Z_n(\phi, a) \leq M$ for all $n > N$.

3.8. Let X be a topologically mixing TMS with set of states S. Suppose X has *finitely many images*: the set $\{T[a] : a \in S\}$ is finite, and let $\phi : X \to \mathbb{R}$ be a function with summable variations and finite pressure. Show that

1. $\exists h > 0$ s.t. $L_\phi h = e^{P_G(\phi)} h$, and
2. $0 < \inf_X h \leq \sup_X h < \infty$.

3.9. Prove the Folklore theorem 3.5. (Hint: Prove first that if $\pi : X \to [0,1]$ is the Markov coding of $f : [0,1] \to [0,1]$, then $-\log |f'|$ is Hölder continuous).

3.10. Prove proposition 3.9.

References

1. J. Aaronson, H. Nakada, O. Sarig, R. Solomyak: *Invariant measures and asymptotics for some skew–products*, Israel J. Math. **121**, 285–311 (2001)
2. R.L. Adler: *f–expansions revisited*, Recent Advances in topological dynamics (Proc. Conf. Yale Univ. New–Haven, Conn, 1972; in honor of Gustav Hedlund), 1–5, Springer Lecture Notes in Math. **318** (1973)
3. K.-L. Chung: Markov chains with stationary transition probabilities. *Die Grund. Math. Wiss* **104** *Springer* (1960), x+278pp.
4. R.L. Dobrushin: *The problem of uniqueness of a Gibbsian random field and the problem of phase transitions*, Functional Anal. & Appl. **2**, 302–312 (1968).
5. W. Feller: An Introduction to Probability Theory and Its Applications, Vol. 1 (Volume 1), *Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc. (1971)*
6. R.D. Mauldin & M. Urbański: *Gibbs states on the symbolic space over an infinite alphabet*, Israel J. Math. **125** (2001), 93–130.

7. D. Ruelle: *Statistical mechanics of a one–dimensional lattice gas*, Commun. Math. Phys. **9**, 267–268 (1968).
8. O. Sarig: *Thermodynamic formalism for countable Markov shifts*, Ergodic Th. & Dynam. Syst. **19** (1999), 1565–1593.
9. O. Sarig: *Existence of Gibbs measures for countable Markov shifts*, Proc. AMS **131** (2003), 1751–1758.
10. D. Vere–Jones: *Ergodic properties of nonnegative matrices. I& II* Pacific J. Math. **22** (1967) 361–386, and Pacific J. Math. **26** (1968) 601–620.
11. P. Walters: *Ruelle's operator theorem and g–measures*. Trans. AMS **214** (1975), 375–387.

Chapter 4
Pressure, Equilibrium measures, and Gibbs measures

The Gibbs probability vector $p_i = e^{-\beta U_i} / \sum_{i=1}^{N} e^{-\beta U_i}$ $(i = 1, \ldots, N)$ can be characterized as the (unique) probability vector which maximizes $-\sum p_i \log p_i + \sum(-\beta U_i) p_i$. The variational approach to the canonical ensemble on TMS is summarized by the following definition:

Definition 4.1 (Equilibrium Measure). Let X be a TMS, $\beta > 0$, and $U : X \to \mathbb{R}$ be a measurable function. A shift invariant probability measure m is called an *equilibrium measure* for $\phi = -\beta U$, if

$$h_m(T) + \int \phi dm = \sup\{h_\mu(T) + \int \phi d\mu\} \tag{4.1}$$

where the supremum ranges over all invariant Borel probability measures μ for which $h_\mu(T) + \int \phi d\mu$ is well defined.

"Well–defined" means that ϕ is μ one–sided integrable,[1] and $h_\mu(T) + \int \phi d\mu \neq \infty - \infty$.

Convention: Throughout this chapter, $h_\mu(T)$ signifies the metric entropy defined using natural logarithms (and not logarithms to base two).

4.1 Entropy and information

4.1.1 Entropy

Suppose $(\Omega, \mathscr{F}, \mu)$ is a probability space, and β is a finite or countable partition of Ω into measurable sets (*"measurable partition"*). The *entropy* of β is the number

[1] A measurable function f is called *one sided integrable* if at least one of the functions $f 1_{[f>0]}$, $f 1_{[f<0]}$ is absolutely integrable.

$$H_\mu(\beta) := -\sum_{B\in\beta} \mu(B)\log\mu(B), \text{ where } 0\log 0 := 0.$$

The *join* of two partitions β,γ is $\beta\vee\gamma := \{B\cap C : B\in\beta, C\in\gamma\}$. The concavity of $f(t) = -t\log t$ implies that

$$H_\mu(\beta\vee\gamma) \le H_\mu(\beta) + H_\mu(\gamma). \tag{4.2}$$

Now suppose $\tau : \Omega \to \Omega$ is a probability preserving map. Given a measurable partition β, we let

- $\tau^{-i}\beta := \{\tau^{-i}B : B\in\beta\}$ (since $\mu\circ\tau^{-1} = \mu$, $H_\mu(\tau^{-i}\beta) = H_\mu(\beta)$);
- $\beta_n^k := \tau^{-n}\beta\vee\tau^{-(n+1)}\beta\vee\cdots\vee\tau^{-k}\beta$;
- $h_\mu(\tau,\beta) := \lim\limits_{n\to\infty}\frac{1}{n}H_\mu(\beta_0^n)$.

The limit exists and equals $\inf_n \frac{1}{n}H_\mu(\beta_0^{n-1})$, because $H_\mu(\beta_0^{n+m}) \le H_\mu(\beta_0^n) + H_\mu(\beta_0^m)$ for all $m,n > 0$, by (4.2).

Definition 4.2 (Metric Entropy). The *metric entropy* of a probability preserving map τ on a probability space $(\Omega,\mathscr{F},\mu)$ is

$$h_\mu(\tau) := \sup\{h_\mu(\tau,\beta) : \beta \text{ is a countable measurable partition s.t. } H_\mu(\beta) < \infty\}.$$

The following theorems provide the main tool we shall need to calculate $h_\mu(T)$:

Theorem 4.1 (Sinai's Generator Theorem). *Suppose β is a countable measurable partition s.t. the smallest σ–algebra containing $\tau^k\beta$ for all $k\in\mathbb{Z}$ is $\mathscr{F}$. If $H_\mu(\beta) < \infty$, then $h_\mu(\tau) = h_\mu(\tau,\beta)$.*

Theorem 4.2 (Rokhlin Formula). *Suppose X is a TMS with set of states S. Let $\alpha := \{[a] : a\in S\}$ ("natural partition"). For every invariant Borel probability measure μ,*

1. *If $H_\mu(\alpha) < \infty$, then $h_\mu(T) = -\int\log\frac{d\mu}{d\mu\circ T}d\mu$;*
2. *If $H_\mu(\alpha) = \infty$, then $h_\mu(T) \ge -\int\log\frac{d\mu}{d\mu\circ T}d\mu$.*

The proof is given in the next sections.

4.1.2 The information function

Suppose $(\Omega,\mathscr{F},\mu)$ is a probability space, β is a finite or countable measurable partition, and $\mathscr{G}\subseteq\mathscr{F}$ a σ–algebra.

Definition 4.3 (Information function). The *information function* of β given $\mathscr{G}$ is

$$I_\mu(\beta|\mathscr{G})(x) := -\sum_{B\in\beta} 1_B(x)\log\mu(B|\mathscr{G})(x), \text{ where } \mu(B|\mathscr{G})(x) := \mathbb{E}_\mu(1_B|\mathscr{G})(x).$$

The *conditional entropy* of β given $\mathscr{G}$ is $H_\mu(\beta|\mathscr{G}) := \int I_\mu(\beta|\mathscr{G})d\mu$.

Suppose β, γ are two finite or countable measurable partitions of a probability space smallest σ–algebra containing γ). We write

- $I_\mu(\beta|\gamma) := I_\mu(\beta|\sigma(\gamma))$;
- $H_\mu(\beta|\gamma) := \int I_\mu(\beta|\gamma) d\mu$.

If $H_\mu(\beta), H_\mu(\gamma) < \infty$, this turns out to be equivalent to the more friendly

$$H_\mu(\beta|\gamma) = H_\mu(\beta \vee \gamma) - H_\mu(\gamma)$$

(see equation (4.3) below). But we emphasize that $H_\mu(\beta|\gamma)$ makes sense (and could be finite) even when $H_\mu(\beta), H_\mu(\gamma) = \infty$.

Proposition 4.1 (Properties of the information function). *Let $(\Omega, \mathscr{F}, \mu)$ be a probability space, and suppose α, β, γ are finite or countable measurable partitions.*

1. *For every σ–algebra $\mathscr{G} \subseteq \mathscr{F}$,*
$$I_\mu(\alpha \vee \beta|\mathscr{G}) = I_\mu(\alpha|\mathscr{G}) + I_\mu(\beta|\mathscr{G} \vee \alpha), \tag{4.3}$$
where $\mathscr{G} \vee \alpha$ is the smallest σ–algebra which contains $\mathscr{G} \cup \alpha$.
2. *For every σ–algebra $\mathscr{G} \subset \mathscr{F}$, If $\sigma(\beta) \subset \sigma(\gamma)$, then $I_\mu(\beta|\mathscr{G}) \leq I_\mu(\gamma|\mathscr{G})$ and $H_\mu(\beta|\mathscr{G}) \leq H_\mu(\gamma|\mathscr{G})$.*
3. *If $\sigma(\beta) \subset \sigma(\gamma)$, then $H_\mu(\alpha|\beta) \geq H_\mu(\alpha|\gamma)$.*

Proof.

Part 1. $\mathscr{G} \vee \alpha = \{\biguplus_{A \in \alpha} A \cap G_A : G_A \in \mathscr{G}\}$ (this is the minimal σ–algebra which contains α and $\mathscr{G}$). Thus every $\mathscr{G} \vee \alpha$–measurable function is of the form $\varphi = \sum_{A \in \alpha} 1_A \varphi_A$ with φ_A $\mathscr{G}$–measurable. If $B \in \beta$, then

$$\begin{aligned}
\int 1_B \varphi d\mu &= \sum_{A \in \alpha} \int 1_B 1_A \varphi_A d\mu = \sum_{A \in \alpha} \int 1_{B \cap A} \varphi_A d\mu \\
&= \sum_{A \in \alpha} \int \mathbb{E}_\mu(1_{B \cap A}|\mathscr{G}) \varphi_A d\mu = \sum_{A \in \alpha} \int \frac{\mathbb{E}_\mu(1_{B \cap A}|\mathscr{G})}{\mathbb{E}_\mu(1_A|\mathscr{G})} \mathbb{E}_\mu(1_A|\mathscr{G}) \varphi_A d\mu \\
&= \sum_{A \in \alpha} \int \frac{\mathbb{E}_\mu(1_{B \cap A}|\mathscr{G})}{\mathbb{E}_\mu(1_A|\mathscr{G})} 1_A \varphi_A d\mu \quad (\because \frac{\mathbb{E}_\mu(1_{B \cap A}|\mathscr{G})}{\mathbb{E}_\mu(1_A|\mathscr{G})} \varphi_A \text{ is } \mathscr{G}\text{–measurable}) \\
&= \int \left(\sum_{A \in \alpha} 1_A \frac{\mathbb{E}_\mu(1_{B \cap A}|\mathscr{G})}{\mathbb{E}_\mu(1_A|\mathscr{G})} \right) \varphi d\mu.
\end{aligned}$$

The term in the brackets is $\mathscr{G} \vee \alpha$–measurable. Since φ was an arbitrary $\mathscr{G} \vee \alpha$–measurable L^∞ function, we must have $\mathbb{E}_\mu(1_B|\mathscr{G} \vee \alpha) = \sum_{A \in \alpha} 1_A \frac{\mu(B \cap A|\mathscr{G})}{\mu(A|\mathscr{G})}$. Thus

$$I_\mu(\alpha) + I_\mu(\beta|\mathscr{G} \vee \alpha) = -\sum_{A \in \alpha} 1_A \log \mu(A|\mathscr{G}) - \sum_{B \in \beta} 1_B \log \sum_{A \in \alpha} 1_A \frac{\mu(B \cap A|\mathscr{G})}{\mu(A|\mathscr{G})}$$

$$= -\left(\sum_{A\in\alpha}\sum_{B\in\beta} 1_{A\cap B}\log\mu(A|\mathscr{G}) + \sum_{B\in\beta}\sum_{A\in\alpha} 1_{A\cap B}\log\frac{\mu(B\cap A|\mathscr{G})}{\mu(A|\mathscr{G})}\right)$$
$$= -\sum_{A\in\alpha}\sum_{B\in\beta} 1_{A\cap B}\log\mu(A\cap B|\mathscr{G}) = I_\mu(\alpha\vee\beta|\mathscr{G}). \qquad \diamondsuit$$

Part 2. If $\sigma(\beta)\subset\sigma(\gamma)$, then every element of β is a union of (disjoint) elements of γ, and so $\gamma=\gamma\vee\beta$. By (4.3),

$$I_\mu(\gamma|\mathscr{G}) = I_\mu(\gamma\vee\beta|\mathscr{G}) = I_\mu(\beta|\mathscr{G}) + I_\mu(\gamma|\mathscr{G}\vee\beta).$$

Since the information function is non–negative $I_\mu(\gamma|\mathscr{G})\geq I_\mu(\beta|\mathscr{G})$. Integrating, we see that $H_\mu(\gamma|\mathscr{G})\geq H_\mu(\beta|\mathscr{G})$. $\diamondsuit$

Part 3. Suppose $\sigma(\beta)\subset\sigma(\gamma)$, then

$$\begin{aligned}
H_\mu(\alpha|\sigma(\beta)) &= \int I_\mu(\alpha|\sigma(\beta))d\mu = -\int\left(\sum_{A\in\alpha} 1_A\log\mathbb{E}_\mu(1_A|\sigma(\beta))\right)d\mu\\
&= -\int\left(\sum_{A\in\alpha}\mathbb{E}_\mu(1_A|\sigma(\beta))\log\mathbb{E}_\mu(1_A|\sigma(\beta))\right)d\mu\\
&= \sum_{A\in\alpha}\int f\left(\mathbb{E}_\mu(1_A|\sigma(\beta))\right)d\mu, \quad \text{where } f(t) := -t\log t\\
&= \sum_{A\in\alpha}\int f\left(\mathbb{E}_\mu[\mathbb{E}_\mu(1_A|\sigma(\gamma))|\sigma(\beta))]\right)d\mu, \quad (\because \sigma(\beta)\subset\sigma(\gamma))\\
&\geq \sum_{A\in\alpha}\int \mathbb{E}_\mu[f(\mathbb{E}_\mu(1_A|\sigma(\gamma)))|\sigma(\beta))]d\mu, \quad \text{(Jensen's inequality)}\\
&= \sum_{A\in\alpha}\int f(\mathbb{E}_\mu(1_A|\sigma(\gamma)))d\mu = H_\mu(\alpha|\sigma(\gamma)). \qquad \square
\end{aligned}$$

Proposition 4.2 (Chung–Neveau). *Suppose β is a finite or countable measurable partition of a probability space $(\Omega,\mathscr{F},\mu)$, and $\{\mathscr{G}_n\}_{n\geq 1}$ is an increasing sequence of σ–algebras contained in $\mathscr{F}$. If $f^* := \sup_{n\geq 1} I(\beta|\mathscr{G}_n)$, then $\int f^* d\mu\leq H_\mu(\beta)+1$.*

Proof. Fix $B\in\beta$, and decompose $B\cap[f^*>t] = \biguplus_{m\geq 1} A\cap B_m(t)$, where

$$B_m(t) := \{x\in X:\ m \text{ is the minimal natural number s.t. } -\log\mu(B|\mathscr{G}_m)>t\}.$$

This is an element of $\mathscr{G}_m$, and

$$\begin{aligned}
\mu[B\cap B_m(t)] &= \mathbb{E}_\mu\left(1_B 1_{B_m(t)}\right) = \mathbb{E}_\mu\left(\mathbb{E}_\mu\left(1_B 1_{B_m(t)}|\mathscr{G}_m\right)\right)\\
&= \mathbb{E}_\mu\left(1_{B_m(t)}\mathbb{E}_\mu\left(1_B|\mathscr{G}_m\right)\right), \text{ because } B_m(t)\in\mathscr{G}_m\\
&\equiv \mathbb{E}_\mu\left(1_{B_m(t)} e^{\log\mu(B|\mathscr{G}_m)}\right)\leq\mathbb{E}_\mu\left(1_{B_m(t)}e^{-t}\right) = e^{-t}\mu[B_m(t)].
\end{aligned}$$

Summing over m we see that $\mu(B\cap[f^*>t])\leq e^{-t}$. Of course we have the stronger inequality $\mu(B\cap[f^*>t])\leq\min\{\mu(B),e^{-t}\}$.

We now use the identity $\int g d\mu = \int_0^\infty \mu[g>t]dt$ for all $g \geq 0$ measurable:[2]

$$\begin{aligned}\int_B f^* d\mu &= \int_0^\infty \mu(B\cap[f^*>t])dt \leq \int_0^\infty \min\{\mu(B), e^{-t}\}dt \\ &\leq \int_0^{-\log\mu(B)} \mu(B)dt + \int_{-\log\mu(B)}^\infty e^{-t}dt = -\mu(B)\log\mu(B) - e^{-t}\Big]_{-\log\mu(B)}^\infty \\ &= -\mu(B)\log\mu(B) + \mu(B).\end{aligned}$$

Summing over $B \in \beta$ we get that $\int f^* d\mu \leq H_\mu(\beta)+1$. □

4.1.3 Ledrappier's formula and proof of Rokhlin's formula

Proposition 4.3 (Ledrappier). *Let μ be a shift invariant probability measure on a TMS with set of states S, and let $\alpha := \{[a] : a \in S\}$, then*

$$I_\mu(\alpha|T^{-1}\mathscr{B}) = -\log\frac{d\mu}{d\mu\circ T}.$$

Proof. The transfer operator of μ is $\widehat{T}f = \sum_{Ty=x} g_\mu(y)f(y)$, where $g_\mu := \frac{d\mu}{d\mu\circ T}$. By proposition 2.3, for every $f \in L^1(\mu)$, $\mathbb{E}_\mu(f|T^{-1}\mathscr{B}) = (\widehat{T}f)\circ T$, so $I_\mu(\alpha|T^{-1}\mathscr{B}) = -\sum\limits_{A\in\alpha} 1_A(x)\log(\widehat{T}1_A)(Tx) = -\sum\limits_{A\in\alpha} 1_A(x)\log\sum\limits_{Ty=Tx} g_\mu(y)1_A(y)$. If $x \in A$, then the only preimage of Tx in A is x, so the inner sum equals $-\log g_\mu(x)$. □

Proof of Rokhlin's formula (Theorem 4.2). Suppose X is a TMS with set of states S, α is the natural partition $\{[a] : a \in S\}$, and $\mu \in \mathfrak{P}_T(X)$. Fix once and for all a finite sets of states S_n s.t. $S_n \uparrow S = \{\text{states}\}$, and let

$$\beta^{(n)} := \{[a] : a \in S_n\} \cup \{\bigcup_{b\notin S_n}[b]\}.$$

These are finite partitions, so $H_\mu(\beta^{(n)}) < \infty$ for all n and

$$\begin{aligned}h_\mu(T,\beta^{(n)}) &= \lim_{k\to\infty}\frac{1}{k}H_\mu((\beta^{(n)})_0^{k-1}) = \lim_{k\to\infty}\frac{1}{k}\sum_{\ell=1}^{k-1}[H_\mu((\beta^{(n)})_0^\ell) - H_\mu((\beta^{(n)})_0^{\ell-1})] \\ &= \lim_{k\to\infty}\frac{1}{k}\sum_{\ell=1}^{k-1}[H_\mu((\beta^{(n)})_0^\ell) - H_\mu(T^{-1}(\beta^{(n)})_0^{\ell-1})] \quad (\because \mu\circ T^{-1} = \mu) \\ &= \lim_{k\to\infty}\frac{1}{k}\sum_{\ell=1}^{k-1}H_\mu(\beta^{(n)}|(\beta^{(n)})_1^\ell) \geq \lim_{k\to\infty}\frac{1}{k}\sum_{\ell=1}^{k-1}H_\mu(\beta^{(n)}|\alpha_1^\ell), \qquad (4.4)\end{aligned}$$

because $\sigma(\alpha) \supset \sigma(\beta^{(n)})$ (Proposition 4.1).

[2] Proof: $\int_X g d\mu = \int_X\int_0^\infty 1_{[0\leq t<g(x)]}(x,t)dtd\mu(x) = \int_0^\infty\int_X 1_{[g>t]}(x,t)d\mu(x)dt = \int_0^\infty \mu[g>t]dt$.

We claim that

$$H_\mu(\beta^{(n)}|\alpha_1^\ell) \xrightarrow[\ell\to\infty]{} H_\mu(\beta^{(n)}|\alpha_1^\infty). \tag{4.5}$$

This is because

- $I_\mu(\beta^{(n)}|\alpha_1^\ell) \xrightarrow[\ell\to\infty]{} I_\mu(\beta^{(n)}|T^{-1}\mathscr{B})$ μ–a.e. (Martingale convergence theorem);
- $\int \sup_{\ell\geq 1} I_\mu(\beta^{(n)}|\alpha_1^\ell)d\mu < \infty$, by the Chung–Neveu Lemma (Proposition 4.2);
- The dominated convergence theorem.

(4.4) and (4.5) say that $h_\mu(T,\beta^{(n)}) \geq H_\mu(\beta^{(n)}|T^{-1}\mathscr{B}) \equiv \int I_\mu(\beta^{(n)}|T^{-1}\mathscr{B})d\mu$.

By the definition of $\beta^{(n)}$, $I_\mu(\beta^{(n)}|T^{-1}\mathscr{B}) \xrightarrow[n\to\infty]{} I_\mu(\alpha|T^{-1}\mathscr{B})$. Since $\{\sigma(\beta_n)\}_{n\geq 1}$ is increasing, $\{I_\mu(\beta^{(n)}|\alpha_1^\infty)\}_{n\geq 1}$ is increasing (Lemma 4.1). By the monotone convergence theorem $H_\mu(\beta^{(n)}|T^{-1}\mathscr{B}) \uparrow H_\mu(\alpha|T^{-1}\mathscr{B})$, so

$$h_\mu(T,\beta^{(n)}) \geq H_\mu(\beta^{(n)}|T^{-1}\mathscr{B}) \xrightarrow[n\to\infty]{} H_\mu(\alpha|T^{-1}\mathscr{B}).$$

Since $h_\mu(T) \geq h_\mu(T,\beta^{(n)})$, $h_\mu(T) \geq H_\mu(\alpha|T^{-1}\mathscr{B}) = -\int \log \frac{d\mu}{d\mu\circ T}d\mu$ (Proposition 4.3). This proves Rokhlin's formula, when $H_\mu(\alpha) = \infty$. $\diamondsuit$

Suppose $H_\mu(\alpha) < \infty$. In this case we can repeat the calculation leading to (4.4) with α replacing $\beta^{(n)}$, and get

$$h_\mu(T,\alpha) = \lim_{k\to\infty} \frac{1}{k}\sum_{\ell=1}^{\infty} H_\mu(\alpha|\alpha_1^\ell).$$

Next we use the argument leading to (4.5) to obtain

$$H_\mu(\alpha|\alpha_1^\ell) \xrightarrow[\ell\to\infty]{} H_\mu(\alpha|\alpha_1^\infty).$$

The result is that $h_\mu(T,\alpha) = H_\mu(\alpha|\alpha_1^\infty) = -\int \log \frac{d\mu}{d\mu\circ T}d\mu$.

Since $H_\mu(\alpha) < \infty$ and $\alpha_0^\infty = \mathscr{B}$, $h_\mu(T) = h_\mu(T,\alpha)$ (Sinai's Generator Theorem). Rokhlin's formula follows. $\square$

4.2 Pressure and the Variational Principle

Let $T: Y \to Y$ be a continuous map on a complete separable metric space Y, and let $\phi: Y \to \mathbb{R}$ be a continuous function. The *variational topological pressure* of ϕ is

$$P_{top}(\phi) = \sup\{h_\mu(T) + \int \phi d\mu\},$$

where the supremum ranges over all invariant Borel probability measures μ s.t. $h_\mu(T) + \int \phi d\mu$ is well–defined.

The first step in the construction an equilibrium measure it to calculate the variational topological pressure. The purpose of this section is to do this for TMS.

4.2.1 Gurevich Pressure

We encountered the Gurevich pressure when we determined the exponential rate of growth of $L_\phi^n 1_{[\underline{a}]}$. We recall the definition:

Definition 4.4 (Gurevich pressure). Let X be a topologically mixing TMS, and ϕ some function with summable variations. The *Gurevich Pressure* of ϕ is the number

$$P_G(\phi) = \lim_{n\to\infty} \frac{1}{n}\log Z_n(\phi,a), \text{ where } Z_n(\phi,a) = \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x).$$

The *Gurevich entropy* of X is $P_G(0)$.

Proposition 3.2 says that the limit exists and is independent of the choice of a. Here are some of its properties:

Proposition 4.4. *Let X be topologically transitive, and assume ϕ and ψ have summable variations. Then,*

1. Addition of constants*: For every $c \in \mathbb{R}$, $P_G(\phi + c) = P_G(\phi) + c$;*
2. Convexity*: For every $t \in [0,1]$, $P_G(t\phi + (1-t)\psi) \le tP_G(\phi) + (1-t)P_G(\psi)$;*
3. Cohomology*: If for some f, $\phi - \psi = f - f\circ T$, then $P_G(\phi) = P_G(\psi)$.*

Proof. The first part is obvious. The second follows Hölder's inequality for sums: $Z_n(t\phi + (1-t)\psi, a) \le Z_n(\phi,a)^t Z_n(\psi,a)^{1-t}$. The third is because if ϕ,ψ are cohomologous, then $\phi_n(x) = \psi_n(x)$ whenever $T^n x = x$. □

We show that the Gurevich pressure is captured by the restrictions of ϕ to a compact TMS lying inside X. To make the statement unambiguous, we introduce the following terminology:

Definition 4.5 (sub systems). Suppose X is a TMS with set of states S and transition matrix $\mathbb{A} = (t_{ij})_{S\times S}$. A *sub system* of X is a TMS with set of states $S' \subseteq S$ and transition matrix $\mathbb{A}' = (t'_{ij})_{S'\times S'}$ s.t. $t'_{ij} = 1 \Rightarrow t_{ij} = 1$.

The following proposition was proved in the special case $\phi = \phi(x_0,x_1)$ by Gurevich.

Theorem 4.3. *If X is topologically mixing and ϕ has summable variations, then $P_G(\phi) = \sup\{P_G(\phi|_Y) : Y$ is a topologically mixing compact sub system of $X\}$.*

Proof. The $(\ge)$ inequality is because for every sub system Y of X, if $Z_n(Y,\phi,a) = \sum_{T^n x = x} e^{\phi_n(x)} 1_Y(x) 1_{[a]}(x)$ then $Z_n(\phi,a) \ge Z_n(Y,\phi,a)$ for all n.

We prove the $(\le)$–inequality under the assumption that $P_G(\phi) < \infty$, and leave the case $P_G(\phi) = \infty$ is left as an exercise. Let $M := \sup_n \mathrm{var}_{n+1}\phi_n$, fix $\varepsilon > 0$ and let $m > \frac{2}{\varepsilon}M$ be large enough so that

$$P_G(\phi) < \frac{1}{m}\log Z_m(\phi,a) + \varepsilon. \tag{4.6}$$

Identifying S with $\mathbb{N}$, we choose N large enough so that

$$\frac{1}{m}\log Z_m(\phi,a) \leq \frac{1}{m}\log Z_m\left(\{1,\ldots,N\}^{\mathbb{N}_0}\cap X,\phi,a\right) + \varepsilon.$$

Adding a finite number of states to $\{1,\ldots,N\}$, one can construct a topologically mixing finite Markov shift $Y \subseteq X$, such that

$$\frac{1}{m}\log Z_m(\phi,a) < \frac{1}{m}\log Z_m(Y,\phi,a) + \varepsilon. \tag{4.7}$$

Set $\zeta_n = \log Z_n(Y,\phi,a)$. By the definition of M, $\zeta_n + \zeta_m \leq \zeta_{n+m} + 2M$, whence for $n = km + r \ \ (r = 0,\ldots,k-1)$

$$\frac{k\zeta_m + \zeta_r}{km+r} \leq \frac{\zeta_{km+r} + 2(k+1)M}{km+r} \leq \frac{\zeta_n}{n} + \frac{k+1}{k}\varepsilon.$$

Fixing m and passing to the limit as $n \to \infty$, we have

$$\frac{1}{m}\log Z_m(Y,\phi,a) \leq P_G(\phi|_Y) + \varepsilon. \tag{4.8}$$

By (4.6),(4.7), (4.8), $P_G(\phi) \leq P_G(\phi|_Y) + 3\varepsilon$. ☐

4.2.2 The Variational Principle

Denote by $\mathfrak{P}_T(X)$ the set of all shift invariant Borel probability measures. The following theorem was proved for compact TMS by Lanford, Ruelle, and Bowen, and for non– compact TMS in the particular case $\phi \equiv 0$ by Gurevich.

Theorem 4.4 (Variational Principle). *Let X be a topologically mixing countable Markov shift and ϕ have summable variations. If* $\sup\phi < \infty$ *then*

$$P_G(\phi) = \sup\left\{ h_\mu(T) + \int \phi\, d\mu \,\middle|\, \mu \in \mathfrak{P}_T(X) \text{ s.t. } h_\mu(T) + \int \phi d\mu \text{ is well defined}\right\}.$$

Proof. We break the equality into two inequalities.

Part 1. $P_G(\phi) \leq \sup\{h_\mu(T) + \int \phi\, d\mu\}$.

Fix $\varepsilon > 0$ and a topologically mixing compact sub system $Y \subseteq X$ s.t. $P_G(\phi) \leq P_G(\phi|_Y) + \varepsilon$ (Theorem 4.3). Let $\psi := \phi|_Y$.

Since Y is compact, the Generalized Ruelle–Perron–Frobenius theorem says that ψ is positive recurrent, and that there is a positive continuous function $h : Y \to \mathbb{R}$

and a probability measure ν on Y s.t. $L_\psi h = e^{P_G(\psi)}h$,$L_\psi^* \nu = e^{P_G(\psi)}\nu$,$\int_Y h d\nu = 1$. Set $m := hd\nu$. This is a shift invariant probability measure, because

$$m(T^{-1}[\underline{a}]) = \nu(h1_{[\underline{a}]} \circ T) = \nu(e^{-P_G(\psi)} L_\psi h1_{[\underline{a}]}) = m[\underline{a}].$$

We claim that $h_m(T|_Y) + \int_Y \psi dm = P_G(\psi)$.

To see this, let $\alpha_Y := \{[a] : a \in S, [a] \subset Y\}$. Since Y is compact, α_Y is finite, so $H_m(\alpha_Y) < \infty$. Rokhlin's formula says that in this case

$$\begin{aligned} h_m(T|_Y) &= \int_Y \log \frac{dm}{dm \circ T} dm \\ &= \int_Y \log \left(\frac{h}{h \circ T} \frac{d\nu}{d\nu \circ T} \right) dm = \int_Y [\psi + \log h - \log h \circ T - P_G(\psi)] d\nu \\ &= \int_Y \psi dm - P_G(\psi), \end{aligned}$$

because m is T–invariant and $\log h$ is absolutely integrable (a continuous function on a compact space).

Thus $P_G(\psi) = h_m(T|_Y) + \int_Y \psi dm \leq \sup\{h_\mu(T) + \int \phi d\mu\}$. Since by construction, $P_G(\phi) \leq P_G(\psi) + \varepsilon$, $P_G(\phi) \leq \sup\{h_\mu(T) + \int \phi d\mu\} + \varepsilon$. ♢

Part 2. $P_G(\phi) \geq \sup\{h_\mu(T) + \int \phi d\mu\}$.

This is trivial when $P_G(\phi) = \infty$, so assume $P_G(\phi) < \infty$. We have to show that $h_\mu(T) + \int \phi d\mu \leq P_G(\phi)$ for every $\mu \in \mathfrak{P}_T(X)$ s.t. $h_\mu(T) + \int \phi d\mu$ makes sense.

If $\int \phi\, d\mu = -\infty$, then $h_\mu(T) < \infty$ (otherwise $h_\mu(T) + \int \phi d\mu = \infty - \infty$), and $h_\mu(T) + \int \phi d\mu = -\infty < P_G(\phi)$. Suppose $-\int \phi d\mu < \infty$.

Assume w.l.o.g. that $S = \mathbb{N}$, and set $\alpha_m := \{[1], \ldots, [m-1], [\geq m]\}$ where $[\geq m] := \{x : x_0 \geq m\}$. Let $\mathscr{B}_m$ denote the σ–algebra generated by α_m. As $m \uparrow \infty$, $\mathscr{B}_m \uparrow \bigcup_m \mathscr{B}_m \subseteq \sigma(\cup_m \mathscr{B}_m) = \mathscr{B}$, whence (Problem 4.7)

$$h_\mu(T, \alpha_m) + \int \phi\, d\mu \xrightarrow[m\to\infty]{} h_\mu(T) + \int \phi\, d\mu.$$

Fix m and set $\beta = \alpha_m$. For every $\underline{a} = (a_0, \ldots, a_n)$ where $\forall i\ a_i \in \beta$ set

$$\langle \underline{a} \rangle = \langle a_0, \ldots, a_n \rangle := \bigcap_{k=0}^{n} T^{-k} a_i.$$

Set $\phi_n \langle \underline{a} \rangle := \sup\{\phi_n(x) : x \in \langle \underline{a} \rangle\}$.

Since $\mu \circ T^{-1} = \mu$,

$$\frac{1}{n} H_\mu(\beta_0^n) + \int \phi\, d\mu = \frac{1}{n}\left(H_\mu(\beta_0^n) + \int \phi_n\, d\mu \right) \leq \frac{1}{n} \sum_{\langle \underline{a} \rangle \in \beta_0^n} \mu \langle \underline{a} \rangle \log \frac{e^{\phi_n \langle \underline{a} \rangle}}{\mu \langle \underline{a} \rangle}$$

$$= \frac{1}{n}\sum_{a,b\in\beta}\mu(a\cap T^{-n}b)\sum_{\substack{\langle\underline{a}\rangle\subseteq a\cap T^{-n}b\\ \langle\underline{a}\rangle\in\beta_0^n}}\mu\left(\langle\underline{a}\rangle|a\cap T^{-n}b\right)\log\frac{e^{\phi_n\langle\underline{a}\rangle}}{\mu\langle\underline{a}\rangle}$$

$$\leq \frac{1}{n}\sum_{a,b\in\beta}\mu(a\cap T^{-n}b)\log\sum_{\substack{\langle\underline{a}\rangle\subseteq a\cap T^{-n}b\\ \langle\underline{a}\rangle\in\beta_0^n}}e^{\phi_n\langle\underline{a}\rangle}+\frac{1}{n}H_\mu(\beta\vee T^{-n}\beta)$$

(Jensen's inequality for sums)

$$=: \sum_{a,b\in\beta}\mu(a\cap T^{-n}b)P_n(a,b)+O(\frac{2}{n}H_\mu(\beta)),$$

where $$P_n(a,b):=\frac{1}{n}\log\sum_{\substack{\langle\underline{a}\rangle\subseteq a\cap T^{-n}b\\ \langle\underline{a}\rangle\in\beta_0^n}}e^{\phi_n\langle\underline{a}\rangle}.$$

Passing to the limit $n\to\infty$, and recalling that $\beta=\alpha_m$,

$$h_\mu(T,\alpha_m)+\int\phi\,d\mu\leq\limsup_{n\to\infty}\left\{\sum_{a,b\in\beta}\mu\left(a\cap T^{-n}b\right)P_n(a,b)\right\} \tag{4.9}$$

It remains to estimate $P_n(a,b)$.

Case 1. $a,b\neq[\geq m]$.

In this case, since α_0^n is finer than β_0^n,[3]

$$P_n(a,b)\leq\frac{1}{n}\log\sum_{\substack{[\underline{a}]\subseteq[a]\cap T^{-n}[b]\\ [\underline{a}]\in\alpha_0^n}}e^{\sup\{\phi_n(x):x\in[\underline{a}]\}}.$$

By the summable variations of ϕ and the topological mixing of X, the right side tends to $P_G(\phi)$, whence

$$a,b\neq[\geq m]\Longrightarrow\limsup_{n\to\infty}P_n(a,b)\leq P_G(\phi). \tag{4.10}$$

Case 2. $a=[\geq m]$ or $b=[\geq m]$.

For every $\langle\underline{a}\rangle\in\beta_0^n$ s.t. $\langle\underline{a}\rangle\subseteq a\cap T^{-n}b$, either $\langle\underline{a}\rangle=\langle\geq m,\dots,\geq m\rangle$ or $\exists i,j,k\geq 1$ such that $i+j+k=n+1$ and

$$\langle\underline{a}\rangle=\langle\underbrace{\geq m,\dots,\geq m}_{i},\underbrace{\xi_0,\dots,\xi_j}_{j},\underbrace{\geq m,\dots,\geq m}_{k}\rangle\quad\text{where }\xi_1,\xi_j\neq[\geq m].$$

[3] Either $\phi\langle\underline{a}'\rangle=\sup\{\phi_n(x):x\in[\underline{a}]\}$ for some $\langle\underline{a}'\rangle\supset[\underline{a}]$, or there are at least two $[\underline{a}'],[\underline{a}'']\subset\langle\underline{a}\rangle$ s.t. $\sup\{\phi_n(x):x\in[\underline{a}']\},\sup\{\phi_n(x):x\in[\underline{a}'']\}>\frac{1}{2}\phi_n\langle\underline{a}\rangle$.

For such i,j,k, $\phi_n\langle\underline{a}\rangle \leq (i+k)\sup\phi + \phi_j\langle\xi_0,\ldots,\xi_j\rangle$. Summing over all possibilities we have

$$P_n(a,b) \leq \frac{1}{n}\log\left(e^{\phi_n\langle\geq m,\cdots,\geq m\rangle} + \sum_{i=1}^{n}\sum_{j=1}^{n+1-i}\sum_{\xi_1=1}^{m-1}\sum_{\xi_j=1}^{m-1} e^{jP_j(\xi_0,\xi_j)+(n+1-j)\sup\phi}\right)$$
$$\leq \sup\phi + \frac{1}{n}\log\left(1+n\sum_{\xi,\eta\neq[\geq m]}\sum_{j=1}^{n} e^{jP_j(\xi,\eta)}\right).$$

We saw in case 1 that $\limsup P_n(\xi,\eta) \leq P_G(\phi)$ when $\xi,\eta \neq [\geq m]$, so there exists a constant C s.t.

$$a = [\geq m] \text{ or } b = [\geq m] \Longrightarrow \limsup_{n\to\infty} P_n(a,b) \leq C. \tag{4.11}$$

We now use (4.9), (4.10), and (4.11) to finish the proof:

$$\begin{aligned}
h_\mu(T,\alpha_m) + \int\phi\,d\mu &\leq \limsup_{n\to\infty}\left\{\sum_{a,b\in\beta}\mu(a\cap T^{-n}b)P_n(a,b)\right\}\\
&\leq \limsup_{n\to\infty}\left[P_G(\phi)\sum_{a,b\neq[\geq m]}\mu(a\cap T^{-n}b) + C\sum_{\neg(a,b\neq[\geq m])}\mu(a\cap T^{-n}b)\right]\\
&\leq P_G(\phi)\mu(X) + C\mu[\geq m] + C\mu(T^{-n}[\geq m])\\
&= P_G(\phi) + 2C\mu[\geq m] = P_G(\phi) + o(1), \quad \text{as } m\to\infty.
\end{aligned}$$

Thus $h_\mu(T) + \int\phi\,d\mu = \lim_{m\to\infty}\left(h_\mu(T,\alpha_m) + \int\phi\,d\mu\right) \leq P_G(\phi)$ as required. □

4.3 Equilibrium measures

4.3.1 Existence

Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is positive recurrent with finite Gurevich pressure and summable variations. By the Generalized Ruelle's Perron–Frobenius Theorem (Theorem 3.4), there are $\lambda > 0$, h positive continuous, and ν conservative finite on cylinders s.t. $L_\phi h = \lambda h$, $L_\phi^*\nu = \lambda\nu$, and $\int h d\nu = 1$. Moreover $\lambda = \exp P_G(\phi)$ and $m := hd\nu$ is a shift invariant probability measure.

Terminology: We call m the *RPF measure* of ϕ.

Theorem 4.5. *Let m be the RPF measure of a positive recurrent function ϕ with summable variations s.t.* $\sup\phi < \infty$ *and $P_G(\phi) < \infty$. If m has finite entropy, then m is an equilibrium measure for ϕ.*

Proof. Let $I_m := I_m(\alpha|T^{-1}\mathscr{B})$. Ledrappier's formula says that $I_m = -\log \frac{dm}{dm\circ T}$. Since $m = h\nu$ and $\frac{d\nu}{d\nu\circ T} = \lambda^{-1}e^{\phi}$ (because $L_\phi^*\nu = \lambda\nu$),

$$I_m = -[\phi + \log h - \log h\circ T - P_G(\phi)]. \tag{4.12}$$

Since $h_m(T) < \infty$ and $\sup\phi < \infty$, $h_m(T) + \int\phi dm$ is well defined. The proof that m is an equilibrium measures consists of a justification of the following calculation (the non–trivial steps are tagged by a question mark):

$$\begin{aligned} h_m(T) + \int \phi d\mu &\geq \int I_m dm + \int \phi dm \quad \text{(Rokhlin's formula)} \\ &\stackrel{?}{=} \int (I_m + \phi) dm = \int [\log h\circ T - \log h + P_G(\phi)] dm \stackrel{?}{=} P_G(\phi). \end{aligned}$$

By the variational principle, m must be the equilibrium measure of ϕ.

To justify these steps we need to show that $I_m, \phi, \log h\circ T - \log h$ are absolutely integrable, and that $\int(\log h\circ T - \log h)dm = 0$. We do this.

I_m is non-negative, by definition. By Rokhlin's formula, $\int I_m dm \leq h_m(T) < \infty$, so I_m is absolutely integrable.

By (4.12), $\phi + \log h - \log h\circ T$ is absolutely integrable. Adding its integral to $\int I_m dm$ we see (again by Rokhlin's formula) that

$$h_m(T) + \int [\phi + \log h - \log h\circ T] dm \geq \int [I_m + \phi + \log h - \log h\circ T] dm = P_G(\phi).$$

Since ϕ is recurrent, ν is ergodic. Therefore m is ergodic. The following holds for almost every $x\in X$:

- $\phi_n(x)/n \xrightarrow[n\to\infty]{} \int\phi dm$ ($\sup\phi < \infty$ and the pointwise convergence in the ergodic theorem holds for all one–sided integrable functions[4]);
- $[\phi_n(x) + \log h(x) - \log h(T^n x)]/n \xrightarrow[n\to\infty]{} \int[\phi + \log h - \log h\circ T]dm$ (because $\phi + \log h - \log h\circ T \in L^1(m)$);
- $\exists n_k(x)\uparrow\infty$ s.t. $|\log h(x) - \log h(T^{n_k(x)}x)| \leq 1$ (because of the Poincaré recurrence theorem, and the continuity of h).

Choose one such x, then

$$\begin{aligned} \int \phi dm &= \lim_{k\to\infty} \frac{1}{n_k(x)} \phi_{n_k(x)}(x) = \lim_{k\to\infty} \frac{1}{n_k(x)} \left(\phi_{n_k(x)} + \log h(x) - \log h(T^{n_k(x)}x)\right) \\ &= \lim_{n\to\infty} \frac{1}{n}\left(\phi_n + \log h(x) - \log h(T^n x)\right) = \int (\phi + \log h - \log h\circ T)\, dm. \end{aligned}$$

It follows that $\int\phi dm = \int(\phi + \log h - \log h\circ T)dm \neq -\infty$. Since $\sup\phi < \infty$, $\phi\in L^1$.

---

[4] Proof: Suppose f is one–sided integrable and, say $\int f = +\infty$. For every M, $f1_{[f<M]}$ is absolutely integrable, and the Birkhoff averages of f are at least as large as the Birkhoff averages of $f1_{[f<M]}$. Thus the liminf of the Birkhoff averages of f is at least $\int_{[f<M]} f$. Since M was arbitrary, the liminf (whence the limit) is ∞.

Since $I_m, \phi, \phi + \log h - \log h \circ T \in L^1$, $\log h \circ T - \log h \in L^1$. Since ϕ and $\phi + \log h \circ T - \log h$ have equal integrals, $\int(\log h - \log h \circ T)dm = 0$. □

Here is an example of a positive recurrent potential whose RPF measure has infinite entropy, and is therefore *not* an equilibrium measures:

Example: Let $X = \mathbb{N}^{\mathbb{N}_0}$ denote the full shift with set of states $\mathbb{N}$, and let $\phi(x) = \log p_{x_0}$ where $(p_i)_{i\geq 1}$ is a probability vector which satisfies $-\sum p_i \log p_i = \infty$. It is easy to verify that $L_\phi 1 = 1$ and $L_\phi^* \mu = \mu$ where μ is the Bernoulli measure corresponding to (p_i), so ϕ is positive recurrent with RPF measure μ. But this is not an equilibrium measure, because $h_\mu(T) + \int \phi d\mu = \infty - \infty$ is not well defined.

4.3.2 Uniqueness

The following theorem can be found in [4].

Theorem 4.6. *Suppose X is a topologically mixing TMS, $\phi : X \to \mathbb{R}$ has summable variations,* $\sup \phi < \infty$ *and* $P_G(\phi) < \infty$.

1. *ϕ has at most one equilibrium measure;*
2. *this equilibrium measure, if it exists, is equal to the RPF measure of ϕ;*
3. *in particular, if ϕ has an equilibrium measure, then ϕ is positive recurrent and the RPF measure of ϕ has finite entropy.*

Before proving this theorem, we need a couple of remarks on the ergodic theory of invariant measures which are not ergodic.

Suppose τ is a probability preserving map on a probability space $(\Omega, \mathscr{F}, \mu)$. A set $E \in \mathscr{F}$ is called a *sweep–out set* for τ if for a.e. $x \in \Omega$, $\tau^n(x) \in E$ for infinitely many positive n.

If A is a sweep–out set, then $\mu(A) \neq 0$. If μ is ergodic, then any measurable set of positive measure is a sweep–out set. More generally, A is a sweep–out set for μ iff almost all the ergodic components of μ give A positive measure (Problem 4.8).

If A is a sweep–out set for T, then the induced transformation $\tau_A := \tau^{\varphi_A}$, $\varphi_A(x) := 1_A(x)\inf\{n \geq 1 : \tau^n(x) \in A\}$ is well defined on a set of full measure of A, and

1. **Kac formula**: for all $f \in L^1(\Omega, \mathscr{F}, \mu)$,

$$\int_\Omega f d\mu = \frac{1}{\mu(A)} \int_A \left(\sum_{k=0}^{\varphi_A - 1} f \circ \tau^k \right) d\mu.$$

2. **Abramov's formula**: if $\mu_A := \mu(\cdot|A)$, then

$$h_\mu(\tau) = \mu(A) h_{\mu_A}(\tau_A).$$

This was proved in M506 under the assumption that μ is ergodic and that $\mu(A) > 0$, but the proofs given there work verbatim in the non-ergodic case provided A is a sweep–out set.

Lemma 4.1 (Affinity of the entropy map). *Suppose μ_1,μ_2 are two invariant probability measures for some measurable map τ on a measurable space $(\Omega,\mathscr{F})$. Suppose $\mu = t\mu_1+(1-t)\mu_2$ where $0<t<1$, then $h_\mu(\tau)=th_{\mu_1}(\tau)+(1-t)h_{\mu_2}(\tau)$.*

Proof. Problem 4.2. □

Proof of Theorem 4.6 Suppose μ is an equilibrium measure for ϕ, and assume w.l.o.g. that $P_G(\phi)=0$ (else pass to $\phi-P_G(\phi)$). By Theorem 3.7 there exists a sub g-function g, and a continuous function φ s.t. $\mathrm{var}_1(\varphi)<\infty$ and

$$\phi = \log g + \varphi - \varphi\circ T. \tag{4.13}$$

We will show that

$$L_\phi e^{-\varphi} = e^{-\varphi} \text{ and } L_\phi^*(e^{\varphi}\mu) = e^{\varphi}\mu. \tag{4.14}$$

This means that $\mu = e^{\varphi}(e^{-\varphi}\mu)$ is an RPF measure of ϕ. Since ϕ can have at most one RPF measure (Problem 4.10), μ is uniquely determined.

Step 1. ϕ, $\log g$, and $\varphi-\varphi\circ T$ are absolutely integrable, and $\int(\varphi-\varphi\circ T)d\mu=0$. Moreover, $\varphi-\varphi\circ T$ has zero integral with respect to μ and with respect to almost every ergodic component of μ.

Proof. We are assuming that $\sup\phi<\infty$, so to prove that $\phi\in L^1$ it is enough to show that $\int\phi d\mu>-\infty$. This must be the case, because $h_\mu(T)+\int\phi d\mu$ is well defined and equal to zero.

Since $\phi\in L^1$ and $\log g\le 0$ (g is a sub–g–function), $\varphi-\varphi\circ T=\phi-\log g$ is one–sided integrable, with integral in $(-\infty,\infty]$. We will show that $\int(\varphi-\varphi\circ T)d\mu=0$, and deduce that $\varphi-\varphi\circ T$ and $\log g=\phi-(\varphi-\varphi\circ T)$ are absolutely integrable.

Let $\mu=\int_X\mu_x d\mu(x)$ denote the ergodic decomposition of μ. Since $\phi\in L^1(\mu)$, $\phi\in L^1(\mu_x)$ for μ–a.e. x. For such x, $\varphi-\varphi\circ T=\phi-\log g$ is one–sided integrable, whence by the Birkhoff ergodic theorem (see footnote on page 90)

$$\int(\varphi-\varphi\circ T)d\mu_x = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(\varphi-\varphi\circ T)\circ T^k \quad \mu_x\text{–almost surely.}$$

The ergodic sum on the right is telescopic, and equal to $\frac{1}{n}(\varphi-\varphi\circ T^n)$. By the Poincaré Recurrence Theorem, $|\varphi-\varphi\circ T^n|\le 1$ infinitely often μ_x–almost surely. It follows that $\int(\varphi-\varphi\circ T)d\mu_x=0$ for μ–a.e. x. Integrating over x with respect to μ, we see that $\int(\varphi-\varphi\circ T)d\mu=0$. ◇

Step 2. μ can be written as a finite or countable convex combination of equilibrium measures μ_i, such that for each i there is a state a_i s.t. $[a_i]$ is a sweep–out set for μ_i.

Proof. Let $\mu=\int_X\mu_x d\mu(x)$ be the ergodic decomposition of μ. Let $\{a_1,a_2,\ldots\}$ be a list of the states a s.t. $\mu[a]\neq 0$, and set

$$E_i := \{x\in X : \mu_x[a_1],\ldots,\mu_x[a_{i-1}]=0, \text{ and } \mu_x[a_i]\neq 0\}.$$

E_i are pairwise disjoint measurable sets, and $\bigcup E_i = X$. Thus $p_i:=\mu(E_i)$ satisfy $\sum p_i=1$. Assume all the indices i s.t. $p_i=0$ have been removed, then

$$\mu = \sum p_i \mu_i, \text{ where } \mu_i := \frac{1}{\mu(E_i)} \int_{E_i} \mu_x d\mu(x) \equiv \mu(\cdot|E_i).$$

By construction, all the ergodic components of μ_i charge $[a_i]$, therefore $[a_i]$ is a sweep–out set for μ_i. We claim that μ_i are all equilibrium measures.

Suppose first $\{p_i\}$ is a finite collection. Since μ is an equilibrium measure and $P_G(\phi) < \infty$, $h_\mu(T) < \infty$. By the affinity of the entropy map, $\sum p_i h_{\mu_i}(T) = h_\mu(T) < \infty$, so $h_{\mu_i}(T) < \infty$ for all i and we have

$$\sum p_i \left(h_{\mu_i}(T) + \int \phi d\mu_i \right) = h_\mu(T) + \int \phi d\mu = P_G(\phi) = 0.$$

By the variational principle, each summand on the left is non–positive. Since $p_i \neq 0$, all the summands are equal to zero, and so μ_i are all equilibrium measures.

Now suppose $\{p_i\}$ is an infinite collection. For every N, let $q_N := \sum_{i>N} p_i$, then $q_N \neq 0$ and $\mu^*_{N+1} := \frac{1}{q_N} \sum_{i>N} p_i \mu_i$ makes sense. We have $\mu = \sum_{i \leq N} p_i \mu_i + q_N \mu^*_{N+1}$. The same argument as before shows that $\mu_1, \ldots, \mu_N; \mu^*_{N+1}$ are equilibrium measures for all N. Since N was arbitrary, μ_i is an equilibrium measure for all i. $\diamondsuit$

Step 3. For all i, $h_{\mu_i}(T) = -\int \log \frac{d\mu_i}{d\mu_i \circ T} d\mu_i$.

Proof: This would have been obvious, had we know that the natural partition has finite entropy, but we cannot assume this (unless X is compact). The idea is to show that T induces some map which possesses a generator with finite entropy.

Fix i and let a_i be the state s.t. $[a_i]$ is a sweep–out set for μ_i. Let $\overline{T}(x) = T^{\varphi_{a_i}(x)}(x)$, $\varphi_{a_i}(x) := 1_{[a_i]}(x) \inf\{n \geq 1 : T^n(x) \in [a_i]\}$, denote the induced map on $[a_i]$. $\overline{T}$ preserves the measure $\overline{\mu}_i := \mu_i(\cdot|[a_i])$ and admits the Markov partition $\beta := \{[a_i, \xi_1, \ldots, \xi_{n-1}, a_i] : n \geq 1, \xi_i \neq a_i\} \setminus \{\varnothing\}$.

We show that $H_{\mu_i}(\beta) < \infty$, and use Rokhlin's formula to prove the step as follows:

$$\begin{aligned}
\frac{1}{\mu_i[a_i]} h_{\mu_i}(T) &= h_{\overline{\mu}_i}(\overline{T}) = -\int \log \frac{d\overline{\mu}_i}{d\overline{\mu}_i \circ \overline{T}} d\overline{\mu}_i \quad \text{(Rokhlin formula)} \\
&= -\frac{1}{\mu_i[a_i]} \int_{[a_i]} \log \frac{d\mu_i}{d\mu_i \circ T^{\varphi_{a_i}}} d\mu_i = -\frac{1}{\mu_i[a_i]} \int_{[a_i]} \sum_{j=0}^{\varphi_{a_i}-1} \log \frac{d\mu_i}{d\mu_i \circ T} \circ T^j d\mu_i \\
&= -\frac{1}{\mu_i[a_i]} \int_X \log \frac{d\mu_i}{d\mu_i \circ T} d\mu_i \quad \text{(Kac formula)}.
\end{aligned}$$

The step follows.

Thus it is enough to prove that $H_{\mu_i}(\beta) < \infty$. To do this, define a Bernoulli measure $\overline{\mu}^i_B$ on $[a_i]$ by $\overline{\mu}^i_B\left(\bigcap_{j=0}^{n-1} \overline{T}^{-j} B_j\right) = \prod_{j=0}^{n-1} \overline{\mu}_i(B_j)$ whenever $B_j \in \beta$. Since $\overline{\mu}^i_B(B) = \overline{\mu}_i(B)$ for all $B \in \beta$, the entropy of the Bernoulli measure $\overline{\mu}^i_B$ is given by

$$h_{\overline{\mu}^i_B}(\overline{T}) = H_{\overline{\mu}_i}(\beta).$$

We will show that $h_{\overline{\mu}^i_B}(\overline{T}) < \infty$ and deduce that $H_{\overline{\mu}_i}(\beta) < \infty$.

Let μ_B^i be the shift invariant measure on X given by Kac's formula

$$\mu_B^i(E) = \mu[a_i] \int_{[a_i]} \sum_{j=0}^{\varphi_{a_i}-1} 1_E \circ T^j d\overline{\mu}_B^i.$$

This is a probability measure, because

$$\mu_B^i(X) = \mu[a_i] \int_{[a_i]} \varphi_a d\overline{\mu}_B^i = \mu[a_i] \sum_{B\in\beta} \overline{\mu}_B^i(B)\cdot \text{length of } B$$
$$= \mu[a_i] \sum_{B\in\beta} \overline{\mu}_i(B)\cdot \text{length of } B = \mu[a_i]\int_X \varphi_a d\overline{\mu}_i = 1,$$

by Kac formula. Since $\overline{\mu}_B^i$ is $\overline{T}$–ergodic (a Bernoulli measure), μ_B^i is T–ergodic.[5]

We claim that $\phi \in L^1(\mu_B^i)$. Set $M := \sup_{n\geq 1} \mathrm{var}_{n+1}\phi_n$, and define

$$\overline{\phi} := \sum_{j=0}^{\varphi_a - 1} \phi\circ T^j.$$

Every $B \in \beta$ is a cylinder of length $n(B)+1$ where $n(B)$ is the unique value of φ_{a_i} on B. Fix $x_B \in B$. For all $y \in B$, $\big||\overline{\phi}(y)| - |\overline{\phi}(x_B)|\big| \leq |\overline{\phi}(y) - \overline{\phi}(x_B)| \leq \mathrm{var}_{n(B)+1}\phi_{n(B)} \leq M$. It follows that the average value of $|\overline{\phi}|$ on B is $|\overline{\phi}(x_B)| \pm M$ no matter which weights we use. Thus for all $B \in \beta$,

$$\left| \frac{1}{\overline{\mu}_i(B)} \int_B |\overline{\phi}| d\overline{\mu}_i - \frac{1}{\overline{\mu}_B^i(B)} \int_B |\overline{\phi}| d\overline{\mu}_B^i \right| \leq 2M.$$

This, and the fact that $\overline{\mu}_i(B) = \overline{\mu}_B^i(B)$ for all $B \in \beta$, shows that $\overline{\phi} \in L^1(\overline{\mu}_B^i) \Leftrightarrow \overline{\phi} \in L^1(\overline{\mu}_i)$. Since $\int |\overline{\phi}| d\overline{\mu}_i \leq \int \overline{|\phi|} d\overline{\mu}_i = \frac{1}{\mu[a_i]} \int |\phi| d\mu_i < \frac{1}{p_i\mu_i[a_i]} \int |\phi| d\mu < \infty$, $\overline{\phi} \in L^1(\overline{\mu}_B^i)$, whence $\phi \in L^1(\mu_B^i)$.

Since $\phi \in L^1(\mu_B^i)$, $h_{\mu_B^i}(T) + \int \phi d\mu_B^i$ is well-defined, whence by the variational principle, $h_{\mu_B^i}(T) + \int \phi d\mu_B^i \leq P_G(\phi) = 0$. Thus $h_{\mu_B^i}(T) \leq -\int \phi d\mu_B^i < \infty$. By Abramov's formula $h_{\overline{\mu}_B^i}(\overline{T}) = (1/\mu_i[a_i]) h_{\mu_B^i}(T) < \infty$ and we are done. $\diamondsuit$

Step 4. $\frac{d\mu_i}{d\mu_i\circ T} = g$ for all i.

Set $g_i := \frac{d\mu_i}{d\mu_i\circ T}$. The transfer operator of μ_i is $\widehat{T}_{\mu_i} f = \sum_{Ty=x} g_i(y) f(y)$. The invariance of μ_i implies that $\sum_{Ty=x} g_i(y) = \widehat{T}_{\mu_i} 1 = 1$ μ_i-almost everywhere, so g_i is a g-function. By step 1 and the definition of μ_i, $\int(\varphi - \varphi\circ T) d\mu_i = 0$ and so

[5] To see this note if f is a measurable T–invariant function on X, then $f|_{[a_i]}$ is a $\overline{T}$–invariant function on $[a_i]$. Thus f is constant on $[a_i]$. Since $[a_i]$ is a sweep–out set, f is constant on X.

$$
\begin{aligned}
0 &= h_{\mu_i}(T) + \int \phi \, d\mu_i && (\mu_i \text{ is an equilibrium measure})\\
&= h_{\mu_i}(T) + \int \log g \, d\mu_i && (\text{step 1})\\
&= \int \log \frac{g}{g_i} d\mu && (\text{step 3})\\
&= \int \left(\sum_{Ty=x} g_i(y) \log \frac{g(y)}{g_i(y)} \right) d\mu(x) && (\widehat{T}^*_{\mu_i}\mu_i = \mu_i)\\
&\le \int \log \left(\sum_{Ty=x, g_i(y)>0} g(y) \right) d\mu(x) && (\log \text{ is concave})\\
&\le 0 && (g \text{ is a sub } g\text{–function}).
\end{aligned}
$$

The inequality must be an equality, so

$$
\sum_{Ty=x} g_i(y) \log \frac{g(y)}{g_i(y)} = \log \left(\sum_{Ty=x} g_i(y) \cdot \frac{g(y)}{g_i(y)} \right) = 0 \ \text{ for } \mu_i - \text{almost all } x. \tag{4.15}
$$

Since the logarithm function is strictly concave, we must have the following (Problem 4.1): For μ_i-a.e. x there exists $c(x)$ such that

$$
y \in T^{-1}\{x\}, g_i(y) > 0 \text{ implies } g(y) = c(x) g_i(y).
$$

Since for a.e. x, $\sum_{Ty=x} g_i(y) = (\widehat{T}1)(x) = 1$, (4.15) forces $c(x)$ to equal one almost everywhere. In summary, for μ_i-almost every x, $y \in T^{-1}\{x\}$ implies $g(y) = g_i(y)$. $\diamondsuit$

We can now complete the proof. Since μ_i is $\log g$–conformal, the transfer operator of μ_i is $L_{\log g}$. In particular, $L^*_{\log g}\mu_i = \mu_i$, whence by (4.13), $L^*_\phi(e^\varphi \mu_i) = e^\varphi \mu_i$. Since $\mu = \sum p_i \mu_i$, $L^*_\phi(e^\varphi \mu) = e^\varphi \mu$.

We also have that $L_{\log g} 1 = 1$ μ_i–almost everywhere, because $\mu_i \circ T^{-1} = \mu_i$. Thus $L_\phi e^\varphi = e^\varphi$ μ–almost everywhere. The identity $L_\phi(e^\varphi \mu) = e^\varphi \mu$ implies that $e^\varphi \mu$ gives positive measure to every cylinder (Problem 3.5). It follows that $L_\phi e^\varphi = e^\varphi$ on a dense set, whence by continuity, everywhere. This proves (4.14), whence by the discussion at the beginning of the proof, the theorem. $\square$

4.3.3 Ergodic properties of equilibrium measures

Theorem 4.7. *Let X be a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ be some function with the Walters property s.t.* $\sup \phi < \infty$ *and* $P_G(\phi) < \infty$*. If m is an equilibrium measure of ϕ, then m is exact (whence ergodic and strong mixing), and*

$$
h_m(T) = -\int \log \frac{dm}{dm \circ T} dm.
$$

Proof. We saw that $m = h\nu$ where $L_\phi^*\nu = \lambda\nu$, whence $\frac{d\nu}{d\nu\circ T} = \lambda^{-1}\exp\phi$. Since ϕ has the Walters property, ν is exact. Thus m is exact. The formula $h_m(T) = -\int \log\frac{dm}{dm\circ T}dm$ can be extracted from the proof of the uniqueness theorem. □

4.3.4 Equilibrium measures and cohomology

Theorem 4.8. *Suppose X is a topologically mixing TMS, and let ϕ,ψ be two functions with the Walters property s.t. $P_G(\phi), P_G(\psi), \sup\phi, \sup\psi < \infty$. Suppose ϕ has an equilibrium measure m, then m is also an equilibrium measure for ψ iff $\phi-\psi$ is cohomologous to a constant.*

Proof. ($\Rightarrow$): Suppose m is an equilibrium measure for ϕ and ψ, we show that ϕ and ψ are cohomologous. Subtract suitable constants from ϕ and ψ to have $P_G(\phi) = P_G(\psi) = 0$ (this has no effect on the equilibrium measures of these functions).

If m is an equilibrium measure for ϕ, then m is an RPF measure for ϕ, so $m = h\nu$ where $L_\phi h = h$, $L_\phi^*\nu = \nu$. The proof of the generalized RPF theorem shows that $\mathrm{var}_1[\log h] < \infty$, so h is bounded away from zero and infinity on partition sets. Let C_a $(a \in S)$ be a family of constants s.t. $C_a^{-1} < h < C_a$ on $[a]$. Define, as usual, $M := \exp(\sup_n \mathrm{var}_{n+1}\phi_n)$, then for every cylinder $[\underline{a}] = [a_0,\dots,a_{n-1}]$,

$$m[\underline{a}] = C_{a_0}^{\pm1}\nu[\underline{a}] = C_{a_0}^{\pm1}\int L_\phi^n 1_{[\underline{a}]}d\nu = C_{a_0}^{\pm1}\int_{T[a_{n-1}]} e^{\phi_n(\underline{a}x)}d\nu = (MC_{a_0})^{\pm1}\nu(T[a_{n-1}])e^{\phi_n(\underline{a}x)},$$

for all $x \in [\underline{a}]$.

This means that if $T^p(x) = x$, then $\phi_p(x)/p = \lim_{n\to\infty}\frac{1}{n}\log m[x_0,\dots,x_{np-1}]$. In the same way, $\psi_p(x)/p = \lim_{n\to\infty}\frac{1}{n}\log m[x_0,\dots,x_{np-1}]$. It follows that $T^p(x) = x \Rightarrow \phi_p(x) = \psi_p(x)$. By Livsic's Theorem, ϕ and ψ a cohomologous. ♢

($\Leftarrow$): Suppose $\psi = \phi + \varphi - \varphi\circ T + c$, and m is an equilibrium measure for ϕ. Since $\sup\phi < \infty$ and $P_G(\phi) < \infty$, $h_m(T) < \infty$ and $\phi \in L^1(m)$. Since $\sup\psi < \infty$, ψ is one–sided integrable w.r.t m. By the ergodic theorem (see page 90)

$$\int \psi dm = \lim_{n\to\infty}\frac{1}{n}\psi_n = \int \phi dm + \lim_{n\to\infty}\frac{1}{n}(\varphi - \varphi\circ T^n) + c \text{ a.s.}$$

By Poincaré's Recurrence Theorem, the only possible a.s. limit for $\frac{1}{n}(\varphi-\varphi\circ T^n)$ is zero. It follows that $\int\psi dm = \int\phi dm + c$, whence

$$h_m(T) + \int\psi dm = h_m(T) + \int\phi dm + c = P_G(\phi) + c = P_G(\psi).$$

By the variational principle, m is an equilibrium measure for ψ. □

Corollary 4.1. *Suppose X is a compact topologically mixing TMS, then two Walters functions are cohomologous iff they have equal pressure and the same equilibrium measure.*

The following example shows that the condition that $\sup\phi, \sup\psi < \infty$ cannot be removed in the non–compact case.

Example: Let $X = \mathbb{N}^{\mathbb{N}_0}$ denote the full shift with set of states $\mathbb{N}$. Fix some probability vector $(p_i)_{i\geq 1}$ s.t. $-\sum p_i \log p_i < \infty$. Let $\phi(x) := \log p_{x_0}$, then it is easy to check that $L_\phi 1 = 1$ and that $L_\phi^* \mu = \mu$, where μ is the Bernoulli measure with probability vector $(p_i)_{i\geq 1}$. It follows that $P_G(\phi) = 0$. The measure μ is an equilibrium measure for ϕ, because $-\sum p_i \log p_i + \sum p_i \log p_i = 0$.

Now let $(h_i)_{i\geq 1}$ be a vector of positive numbers s.t. $\sum p_i \log h_i = \infty$ and set

$$\psi(x) = \log p_{x_0} + \log h_{x_1} - \log h_{x_0}.$$

This function is cohomologous to ϕ. But μ is not an equilibrium measure for ϕ, because $\int \phi dm$ does not exist:

- for all a, $\int_{[a]} \phi d\mu = p_a (\log p_a + \sum p_i \log h_i - \log h_a) = \infty$;
- for all a, $\int_{T^{-1}[a]} \phi d\mu = \sum p_i \log p_i + p_a \log h_a - p_a \sum p_i \log h_i = -\infty$.

Such a function cannot be one–sided integrable.

4.4 Gibbs measures in the sense of Bowen

In his studies of compact TMS, Bowen observed that if m is the equilibrium measure of a Walters function ϕ, then $m[\underline{a}]$ can be uniformly approximated by $\frac{1}{Z_n}\exp\phi_n(x)$ where n is the length of $\underline{a}$, $x \in [\underline{a}]$ is arbitrary, and $Z_n \asymp \exp(nP)$ for some P. This leads to the following definition:

Definition 4.6. Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ has the Walters property. A Borel probability measure m is called a *Gibbs measure (in the sense of Bowen)* for ϕ if m is shift invariant, and if there are constants $G > 1$ and P s.t. for all cylinders $[\underline{a}]$ and every $x \in [\underline{a}]$,

$$G^{-1} \leq \frac{m[a_0,\ldots,a_{n-1}]}{\exp[\sum_{k=0}^{n-1}\phi(T^k x) - nP]} \leq G. \tag{4.16}$$

Although this characterization of equilibrium measures does not hold in general for non-compact TMS, it does hold in certain important cases. When it does, it is of great use because it is the closest one can hope to get to an explicit formula for the equilibrium measure of ϕ.

Theorem 4.9. *Let X be a topologically mixing TMS with set of states S and transition matrix $(t_{ij})_{S\times S}$. A Walters function ϕ has admits a Gibbs measure m iff*

1. *X satisfies the BIP property: $\exists b_1,\ldots,b_N \in S$ s.t. for all $a \in S$ there are $1 \leq i, j \leq N$ s.t. $t_{b_j a} t_{a b_i} = 1$;*
2. *$P_G(\phi) < \infty$ and $\mathrm{var}_1\phi < \infty$.*

In this case ϕ is positive recurrent, m is equal to the RPF measure of ϕ, and P in (4.16) is equal to $P_G(\phi)$.

Proof. ($\Leftarrow$): Suppose (1)–(2) hold. Proposition 3.8 says that ϕ is positive recurrent. Let $m = h\nu$ be its RPF measure, i.e., $L_\phi h = \lambda h$, $L_\phi^* \nu = \lambda \nu$, and $\int h d\nu = 1$. We claim that m is a Gibbs measure for ϕ.

Step 1: h is bounded away from zero and infinity on X.

Let $\{b_1, \dots, b_N\}$ be the collection of states given by the BIP property, and fix some $z^i \in [b_i]$ $(1 \le i \le N)$. For every point of the form $(\xi_0, \dots, \xi_{n-1}, x)$ one can find $1 \le i \le N$ s.t. $(\xi_0, \dots, \xi_{n-1}, z^i)$ is admissible. Since ϕ has the Walters property and $\mathrm{var}_1 \phi < \infty$, $\exists C > 1$ s.t. $\exp \phi_n(\xi_0, \dots, \xi_{n-1}, x) \le C \exp \phi_n(\xi_0, \dots, \xi_{n-1}, z^i)$. Summing over all possible $\underline{\xi}$ we see that

$$\lambda^{-n}(L_\phi 1_{[a]})(x) \le C \sum_{i=1}^N \lambda^{-n}(L_\phi^n 1_{[a]})(z^i).$$

In the limit $n \to \infty$ we obtain $h(x)\nu[a] \le C \sum_{i=1}^N h(z^i)\nu[a]$, whence $\sup h < \infty$.

Next we claim that $\inf h > 0$. For every $x \in X$ $\exists 1 \le i \le N$, s.t. (b_i, x) is admissible. Then $h(x) = \lambda^{-1}\left(L_\phi h\right)(x) \ge \lambda^{-1} e^{\phi(b_i x)} h(b_i x)$. Since $\mathrm{var}_1[\log h] < \infty$ (by Proposition 3.4), $\mathrm{var}_1 \phi < \infty$ (by assumption), and $\{b_1, \dots, b_N\}$ is finite, $\inf h > 0$. $\diamondsuit$

Step 2: $m = h\nu$ is a Gibbs measure for ϕ.

This measure is shift invariant because for every f continuous and m integrable,

$$\int f \circ T \, dm = \int \lambda^{-1} L_\phi(hf \circ T)\, d\nu = \int f \cdot \lambda^{-1} L_\phi h \, d\nu = \int f \, dm.$$

To see that this measure satisfies (4.16), choose using the previous step a constant $H > 1$ such that $\forall x \in X$, $\frac{1}{H} < h(x) < H$, and let $M_0 := \exp(\sup_n \mathrm{var}_n \phi_n)$ (this is finite by the Walters property and since $\mathrm{var}_1 \phi < \infty$). For every cylinder $[\underline{a}] = [a_0, \dots, a_{n-1}]$,

$$\begin{aligned} m[a_0, \dots, a_{n-1}] &= \int \lambda^{-n} L_\phi^n(h 1_{[\underline{a}]})\, d\nu = H^{\pm 1} \int \lambda^{-n} L_\phi^n 1_{[\underline{a}]}\, d\nu \\ &= (HM_0)^{\pm 1} e^{\phi_n(\underline{a}) - n \log \lambda} \nu(T^n[\underline{a}]) = (H^2 M_0)^{\pm 1} e^{\phi_n(\underline{a}) - n \log \lambda} m(T^n[\underline{a}]). \end{aligned}$$

It is therefore enough to show that $m(T^n[\underline{a}])$ is bounded away from zero and infinity. Boundness from above is clear, since $m(X) = 1$. Boundness from below is because of the BIP property which says that $m(T^n[\underline{a}]) \ge \min\{m[b_i] : i = 1, \dots, N\} > 0$.

A by-product of the proof is that the P in (4.16) is equal to $P_G(\phi)$. $\diamondsuit$

($\Rightarrow$): Assume that ϕ admits an invariant Gibbs measure m.

The first variation of ϕ must be finite, because by (4.16) if $x_0 = y_0$, then $e^{\phi(x)}, e^{\phi(y)} = G^{\pm 1} e^{-P} m[x_0]$ whence $|\phi(x) - \phi(y)| \le 2 \log G$.

The Gurevich pressure of ϕ is finite, because by (4.16),

$$Z_n(\phi, a) = \sum_{T^n x = x} e^{\phi_n(x)} 1_{[a]}(x) \le G e^{nP} \sum m[a, x_0, \dots, x_{n-1}] \le G e^{nP} m[a].$$

We show that X satisfies the BIP property. For every state $a \in S$, let $\phi(a) := \inf\{\phi(x) : x \in [a]\}$. This number is finite because $\mathrm{var}_1\phi < \infty$ so ϕ is bounded on partition sets.

By (4.16), there is some global constant C, such that for all $p,q \in S$ s.t. $p \xrightarrow{1} q$, $m[p,q] = C^{\pm 1} e^{\phi(p)} \cdot e^{\phi(q)}$ If we sum over all possibilities, first for p (fixing q) and then for q (fixing p), then we obtain

$$m(T^{-1}[q]) = C^{\pm 1} e^{\phi(q)} \sum_{p:p \xrightarrow{1} q} e^{\phi(p)}, \text{ and } m[p] = C^{\pm 1} e^{\phi(p)} \sum_{q:\, p \xrightarrow{1} q} e^{\phi(q)}.$$

Since $m(T^{-1}[q]) = m[q] = G^{\pm 1} e^{-P} e^{\phi(q)}$ and $m[p] = G^{\pm 1} e^{-P} e^{\phi(p)}$,

$$\delta := \inf \left\{ \sum_{b:\, b \xrightarrow{1} q} e^{\phi(b)}, \sum_{b:\, p \xrightarrow{1} b} e^{\phi(b)} : p,q \in S \right\} > 0. \tag{4.17}$$

We show that this implies the BIP property.

Assume w.l.o.g. that $S = \mathbb{N}$. The sum $\sum e^{\phi(b)}$ converges, because by (4.16) $e^{\phi(b)} \leq Ge^P m[b]$ and $m(X) < \infty$. Pick N s.t. $\sum_{b>N} e^{\phi(b)} < \delta/2$, then for every $p,q \in S$ there must exist $b, b' \leq N$ s.t. $b \xrightarrow{1} q$, $p \xrightarrow{1} b'$. This is the BIP property. ♢

Uniqueness: We show that every Gibbs measure for ϕ equal the RPF measure of ϕ. Let m denote the RPF measure of ϕ, and let μ be some other Gibbs measure. There is a constant $C > 1$ s.t. for all cylinders $[\underline{a}]$,

$$C^{-1} m[\underline{a}] \leq \mu[\underline{a}] \leq Cm[\underline{a}],$$

because m and μ both satisfy (4.16) with $P = P_G(\phi)$. The collection of sets E s.t. $C^{-1}m(E) \leq \mu(E) \leq m(E)$ forms a monotone class, so these inequalities hold for all Borel sets E. It follows that $m \sim \mu$.

But m is ergodic, because it is conservative (like all invariant probability measures) and its log Jacobian $\phi + \log h - \log h \circ T - P_G(\phi)$ has the Walters property (Theorem 2.3). Thus μ is ergodic. It is easy to see using the pointwise ergodic theorem that any two equivalent ergodic invariant probability measures for the same transformation must be equal. So $\mu = m$. □

Corollary 4.2 (Bowen). *Suppose X is a compact TMS. Any $\phi : X \to \mathbb{R}$ with the Walters property admits a unique equilibrium measure, and this measure is a Gibbs measure in the sense of Bowen.*

In the non–compact case the situation is more complicated. Since equilibrium measures may exist in the absence of the BIP property, there are (many) cases where there are equilibrium measures but no Gibbs measures. The following example that the opposite is also possible.

Example. Let (p_i) be a probability vector with infinite entropy, and let $X := \mathbb{N}_0^{\mathbb{N}}$. The Bernoulli measure of (p_i) is a Gibbs measure for $\phi(x) := \log p_{x_i}$. But ϕ has no equilibrium measure, because the RPF measure of ϕ has infinite entropy.

Problems

4.1. Let p_i, x_i $(i \in \mathbb{N})$ be real numbers s.t. $p_i \geq 0$ and $x_i > 0$ for all i, and $\sum p_i = 1$. If $\sum_{i=1}^{\infty} p_i \log x_i = \log \sum p_i x_i$, then all the x_i with $p_i \neq 0$ are equal.

4.2. (Entropy is affine) Suppose T is a measurable map on $(\Omega, \mathscr{F})$, and μ_1, μ_2 are two T–invariant probability measures. Set $\mu = t\mu_1 + (1-t)\mu_2$ $(0 \leq t \leq 1)$. Prove that $h_\mu(T) = t\mu_1 + (1-t)\mu_2$. *Guidance:* Start by showing that for all $0 \leq x, y, t \leq 1$,

$$0 \leq \varphi(tx + (1-t)y) - [t\varphi(x) + (1-t)\varphi(y)] \leq -tx \log t - (1-t)y \log(1-t)$$

4.3. Let $(X, \mathscr{B}, \mu)$ be a probability space. If α, β are two measurable partitions of X, then we write $\alpha = \beta \mod \mu$ if $\alpha = \{A_1, \ldots, A_n\}$ and $B = \{B_1, \ldots, B_n\}$ where $\mu(A_i \triangle B_i) = 0$ for all i. Let $\mathfrak{P}$ denote the set of all countable measurable partitions of X, modulo the equivalence relation $\alpha = \beta \mod \mu$. Show that

$$\rho(\alpha, \beta) := H_\mu(\alpha|\beta) + H_\mu(\beta|\alpha)$$

induces a metric on $\mathfrak{P}$.

4.4. Let $(X, \mathscr{B}, \mu, T)$ be a ppt. Show that $|h_\mu(T, \alpha) - h_\mu(T, \beta)| \leq H_\mu(\alpha|\beta) + H_\mu(\beta|\alpha)$.

4.5. Use the previous problem to show that the supremum in the definition of metric entropy is attained by *finite* measurable partitions.

4.6. Suppose $\alpha = \{A_1, \ldots, A_n\}$ is a finite measurable partition. Show that for every ε, there exists $\delta = \delta(\varepsilon, n)$ such that if $\beta = \{B_1, \ldots, B_n\}$ is measurable partition s.t. $\mu(A_i \triangle B_i) < \delta$, then $\rho(\alpha, \beta) < \varepsilon$.

4.7. Entropy via generating sequences of partitions
Suppose $(X, \mathscr{B}, \mu)$ is a probability space, and $\mathscr{A}$ is an algebra of $\mathscr{F}$–measurable subsets (namely a collection of sets which contains $\varnothing$ and which is closed under finite unions, finite intersection, and forming complements). Suppose $\mathscr{A}$ generates $\mathscr{B}$ (i.e. $\mathscr{B}$ is the smallest σ–algebra which contains $\mathscr{A}$).

1. For every $F \in \mathscr{F}$ and $\varepsilon > 0$, there exists $A \in \mathscr{A}$ s.t. $\mu(A \triangle F) < \varepsilon$.
2. For every $\mathscr{F}$–measurable finite partition β and $\varepsilon > 0$, there exists an $\mathscr{A}$–measurable finite partition α s.t. $\rho(\alpha, \beta) < \varepsilon$.
3. If $T : X \to X$ is probability preserving, then

$$h_\mu(T) = \sup\{h_\mu(T, \alpha) : \alpha \text{ is an } \mathscr{A}\text{–measurable finite partition}\}.$$

4. Suppose $\alpha_1 \leq \alpha_2 \leq \cdots$ is an increasing sequence of finite measurable partitions such that $\sigma(\bigcup_{n\geq 1} \alpha_n) = \mathscr{B} \mod \mu$, then $h_\mu(T) = \lim_{n\to\infty} h_\mu(T, \alpha_n)$.

4.8. Let $\mu = \int_X \mu_x d\mu(x)$ denote the ergodic decomposition of an invariant probability measure μ, and let E be a measurable set. Prove that E is a sweep–out set for μ iff $\mu_x(E) \neq 0$ for μ–a.e. x.

4.9. Complete the proof of Theorem 4.3 by proving the $(\leq)$–inequality in the case $P_G(\phi) = \infty$.

4.10. Suppose X is a topologically mixing TMS and ϕ has the Walters property. Show that ϕ can have at most one RPF measure. (Hint: Theorem 3.4)

4.11. Much of the work required to prove the existence and uniqueness theorems for equilibrium measures was aimed at showing that the entropy of certain measures is equal to minus the integral of their log–Jacobians. Explain why this is automatic in the compact case, and find simplified proofs for compact TMS.

4.12. Suppose X is a topologically mixing compact TMS. Show that the measure with maximal possible entropy is a Markov measure, and describe this measure.

4.13. Prove corollary 4.2.

4.14. Suppose X is a topologically mixing TMS and $\phi : X \to \mathbb{R}$ is a positive recurrent with finite pressure and the Walters property. Let m denote the RPF measure of ϕ. Show that there are constant $\{G_{ab} : a, b \in S\}$ s.t. for every cylinder $[a_0, \ldots, a_{n-1}]$ and every $x \in [a_0, \ldots, a_{n-1}]$,

$$G_{a_0, a_{n-1}}^{-1} \leq \frac{m[a_0, \ldots, a_{n-1}]}{\exp[\sum_{k=0}^{n-1} \phi(T^k x) - nP_G(\phi)]} \leq G_{a_0 a_{n-1}}.$$

References

1. R. Bowen: *Bernoulli equilibrium states for Axiom A diffeomorphisms*, Math. Systems Theory **8** (1975), 289–294.
2. R. Bowen: Equilibrium states and the ergodic theory of Anosov diffeomorphisms.*Second revised edition. Edited by J.-R. Chazzottes. Lecture Notes in Mathematics* **470***, Springer 2008, viii+75pp.*
3. R. Bowen and D. Ruelle: *The ergodic theory of Axiom A flows*, Invent. Math. **29**, 181–202 (1975).
4. J. Buzzi & O. Sarig: *Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps*, Ergodic Th. & Dynam. Syst. **23**, 1383–1400 (2003).
5. B.M. Gurevich: *Topological entropy of a countable Markov chain*. Dokl. Akad. Nauk. SSSR **187** (1969) 715–718.
6. B.M. Gurevich: *Shift entropy and Markov measures in the space of paths of a countable graph*. Dokl. Akad. Nauk. SSSR **192** (1970) 963–965.

7. B.M. Gurevich: *A variational characterization of one–dimensional countable state Gibbs random fields*, Prob. Theory & Related Fields **68**, 205–242 (1984).
8. F. Ledrappier: *Principe variationnel et systémes dynamiques symboliques*, Z. Wahrscheinlichkeitstheorie und. Verb. Gebiete **30** (1974), 185–202.
9. R.D. Mauldin and M. Urbański:*Gibbs states on the symbolic space over an infinite alphabet.* Israel J. Math. **125** 93–130 (2001).
10. W. Parry: *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** 55–66 (1964).
11. K. Petersen: Ergodic theory, Corrected reprint of the 1983 original,*Cambridge Studies in Advanced Mathematics* **2***, Cambridge Univ. Press 1989. xii+329pp.*
12. D. Ruelle: *A variational formulation of equilibrium statistical mechanics and the Gibbs phase rule*, Commun. Math. Phys. **5**, 324–329 (1967).
13. D. Ruelle: *Statistical mechanics of a one–dimensional lattice gas*, Commun. Math. Phys. **9**, 267–268 (1968).
14. D. Ruelle: *A measure associated with Axiom A attractors.* Amer. J. Math. **98**, 619–654 (1976).
15. O. Sarig: *Thermodynamic formalism for countable Markov shifts*, Ergodic Th. & Dynam. Syst. **19** (1999), 1565–1593.
16. O. Sarig: *Existence of Gibbs measures for countable Markov shifts.* Proc. Amer. Math. Soc. **131** , no. 6, 1751–1758 (2003).
17. P. Walters: A variational principle for the pressure of continuous transformations, Amer. J. Math. **97** 937–971 (1975).
18. P. Walters: Introduction to ergodic theory, *Graduate Texts in Mathematics* **79***, Springer 1982, ix+250pp.*

Chapter 5
Spectral Gap

5.1 Functional analytic background

5.1.1 Banach spaces

A (complex) *normed vector space* is a vector space B over $\mathbb{C}$ together with a function $\|\cdot\| : B \to \mathbb{R}$ satisfying the following axioms:

1. $\|x\| \geq 0$ for all $x \in B$ and $\|x\| = 0 \Rightarrow x = 0$ (the zero vector in B);
2. $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in B$ and $\lambda \in \mathbb{C}$;
3. $\|x+y\| \leq \|x\| + \|y\|$ ("triangle inequality").

We call $\|\cdot\|$ a *norm*. The norm $\|\cdot\|$ defined a metric $d(x,y) := \|x-y\|$ which turns $(B, \|\cdot\|)$ into a metric space.

Definition 5.1. A normed vector space $(B, \|\cdot\|)$ is called a *Banach space*, if it is complete: Any sequence $\{x_n\}_{n\geq 1}$ such that $\|x_m - x_n\| \xrightarrow[m,n\to\infty]{} 0$ possesses a vector $x \in B$ s.t. $\|x_n - x\| \xrightarrow[n\to\infty]{} 0$.

A *bounded linear functional* on a Banach space B is a map $f : B \to \mathbb{C}$ which is linear and continuous. Such maps are called "bounded" because of the simple fact that a linear map $f : B \to \mathbb{C}$ is continuous iff there exists $M > 0$ s.t. $|f(x)| \leq M\|x\|$ for all $x \in B$.[1] The *norm* of a bounded linear functional is

$$\|f\| := \sup_{x \in B \setminus \{0\}} \frac{\|f(x)\|}{\|x\|}.$$

If we endow the collection of bounded linear functionals on B with this norm, then the result is another Banach space, called the *dual space* of B. The dual space of B is denoted by B^*.

[1] If such M exists, then $|f(x) - f(y)| \leq M\|x-y\|$ and f is continuous. Since f is linear, $f(0) = 0$. If f is continuous, then $\exists \delta > 0$ s.t. $\|x\| < \delta \Rightarrow |f(x)| < 1$, and so $|f(x)| = |f(\frac{\delta x}{2\|x\|})| \cdot (2\|x\|/\delta) < \frac{2}{\delta}\|x\|$.

More generally, if B_1, B_2 are two Banach spaces, then a *bounded linear operator* from B_1 to B_2 is a continuous linear map $L : B_1 \to B_2$. Such a map has a constant M s.t. $\|Lx\| \leq M\|x\|$ for all $x \in X$. Its norm is defined to be

$$\|L\| := \sup_{x \in B_1 \setminus \{0\}} \frac{\|Lx\|}{\|x\|}.$$

If we endow the collection of bounded linear operators from B_1 to B_2 with this norm, then the result is a Banach space, which we denote by $\mathrm{Hom}(B_1, B_2)$. We will often use the abbreviation $\mathrm{Hom}(B) := \mathrm{Hom}(B, B)$.

We need the following three fundamental results, which can be found in any textbook on Banach space theory.

Theorem 5.1 (Hahn–Banach). *Suppose B is a normed vector space and $V \subset B$ is a linear subspace of B. Any bounded linear functional f on V can be extended to a bounded linear functional F on B s.t. $\|F\| = \|f\|$.*

Since it is very easy to construct bounded linear functionals on finite dimensional subspaces, the Hahn–Banach theorem guarantees the existence of (many) bounded linear functionals on any infinite dimensional Banach space.

Theorem 5.2 (Banach & Steinhaus). *Suppose $(B, \|\cdot\|)$ is a Banach space. If $\{f_\lambda\}_{\lambda \in \Lambda}$ is a family of bounded linear functionals s.t. $\sup_{\lambda \in \Lambda} |f_\lambda(x)| < \infty$ for all $x \in B$, then $\sup_{\lambda \in \Lambda} \|f_\lambda\| < \infty$.*

(Banach called this phenomenon "the accumulation of singularities", because it is equivalent to saying that if $\sup_{\lambda \in \Lambda} \|f_\lambda\| = \infty$, then the sets where f_λ takes large values must accumulate on a common singularity $x \in B$ where $\sup |f_\lambda(x)| = \infty$.)

Theorem 5.3 (Open Mapping Theorem). *Suppose $L : B_1 \to B_2$ is a linear operator between two Banach spaces. If L is continuous and surjective, then L is open (i.e. if U is an open subset of B_1, then $L(U)$ is an open subset of B_2.*

In particular, if $L : B_1 \to B_2$ is linear invertible map, and L is bounded, then L^{-1} is bounded.

5.1.2 Analytic functions on Banach spaces

Suppose $U \subset \mathbb{C}$ is open, and $x : U \to B$ is a function taking values in a Banach space B. We wish to define what it means for $x(\cdot)$ to be "holomorphic". Two two natural definitions come to mind:

Definition 5.2. $x : U \to B$ is *strongly holomorphic* on U, if for every $\xi \in U$, if there exists $y \in B$ such that $\left\| \frac{x(\xi+h)-x(\xi)}{h} - y \right\| \xrightarrow[h\to 0]{} y$. In this case y is called the *derivative* of x at ξ and is denoted by $y = x'(\xi)$.

Definition 5.3. $x : U \to B$ is *weakly holomorphic* on U, if $z \mapsto f(x(z))$ is holomorphic on U in the usual sense for every $f \in B^*$.

The two definitions turn out to be equivalent:

Theorem 5.4. *Suppose $U \subset \mathbb{C}$ is open. A function $x : U \to B$ is weakly holomorphic on U iff it is strongly holomorphic on U.*

Proof. ($\Leftarrow$): Suppose $x(\cdot)$ is strongly holomorphic on U. If $f \in B^*$, then

$$\left| \frac{f(x(\xi_0+h)) - f(x(\xi_0))}{h} - f(x'(\xi_0)) \right| \le \|f\| \cdot \left\| \frac{x(\xi_0+h) - x(\xi_0)}{h} - x'(\xi_0) \right\| \xrightarrow[h\to 0]{} 0.$$

It follows that $f \circ x$ is differentiable, whence holomorphic, on U. $\diamond$

($\Rightarrow$): Suppose $x(\xi)$ is weakly holomorphic on U, and fix some ξ_0 and $f \in B^*$. We show that $[x(\xi_0+h) - x(\xi_0)]/h$ satisfies the Cauchy criterion as $h \to 0$.

Fix $f \in B^*$. Since x is weakly holomorphic on U, $f \circ x$ is holomorphic on U in the usual sense. Fix $r > 0$ so small that $B_r(\xi_0) := \{z \in \mathbb{C} : |z - \xi_0| < r\} \subset U$. Cauchy's Integral Formula for $f \circ x$ says that for all $h \neq 0$ small enough,

$$\begin{aligned} f\left(\frac{x(\xi_0+h) - x(\xi_0)}{h}\right) &= \frac{f(x(\xi_0+h)) - f(x(\xi_0))}{h} \\ &= \frac{1}{2\pi hi} \oint_{\partial B_r(\xi_0)} \left[\frac{f(x(\xi))}{\xi - (\xi_0+h)} - \frac{f(x(\xi))}{\xi - \xi_0} \right] d\xi \\ &= \frac{1}{2\pi i} \oint_{\partial B_r(\xi_0)} \frac{f(x(\xi))}{(\xi - (\xi_0+h))(\xi - \xi_0)} d\xi. \end{aligned}$$

A similar expression holds for $f[(x(\xi_0+k) - x(\xi_0))/k]$ for all $k \neq 0$ small enough. Subtracting, we see that

$$\begin{aligned} A_f(h,k) &:= f\left(\frac{x(\xi_0+h) - x(\xi_0)}{h} - \frac{x(\xi_0+k) - x(\xi_0)}{k} \right) \\ &= \frac{1}{2\pi i} \oint_{\partial B_r(\xi_0)} \left[\frac{f(x(\xi))}{(\xi - (\xi_0+h))(\xi - \xi_0)} - \frac{f(x(\xi))}{(\xi - (\xi_0+k))(\xi - \xi_0)} \right] d\xi \\ &= \frac{h-k}{2\pi i} \oint_{\partial B_r(\xi_0)} \frac{f(x(\xi))}{(\xi - (\xi_0+h))(\xi - (\xi_0+k))(\xi - \xi_0)} d\xi. \end{aligned}$$

Setting $M_f := \sup\{|f(x(\xi))| : |\xi - \xi_0| = r\}$, we see that if $|h|, |k| < \frac{r}{2}$, then

$$\left| A_f(h,k) \right| \le |h-k| \cdot \frac{2\pi r}{2\pi} \cdot \frac{M_f}{\frac{r^3}{4}} = \frac{4M_f}{r^2} |h-k|.$$

Rearranging terms, we deduce that for all $f \in B^*$,

$$\sup_{0<|h|,|k|<\frac{r}{2}} \left| f\left(\underbrace{\frac{1}{|h-k|}\left[\frac{x(\xi_0+h)-x(\xi_0)}{h} - \frac{x(\xi_0+k)-x(\xi_0)}{k} \right]}_{=:x(h,k)} \right) \right| < \infty. \quad (5.1)$$

We denote the term in the brackets by $x(h,k)$.

Let $x^*_{h,k} : B^* \to \mathbb{C}$ be the function $x^*_{h,k}(f) := f(x_{h,k})$. This is a bounded linear functional on B^*, because $|x^*_{h,k}(f)| = |f(x_{h,k})| \le \|f\|\|x_{h,k}\|$ so $\|x^*_{h,k}\| \le \|x_{h,k}\|$. In fact $\|x^*_{h,k}\| = \|x_{h,k}\|$, because it is no problem to construct, using the Hahn–Banach theorem, a bounded linear functional $f \in B^*$ s.t. $f(x_{h,k}) = \|x_{h,k}\|$ and $\|f\| = 1$.

Equation (5.1) reads $\sup_{0<|h|,|k|<\frac{r}{2}} \|x^*_{h,k}(f)\| < \infty$ for all $f \in B^*$. By the Banach–Steinhaus theorem (applied to bounded linear functionals on B^*), $\sup_{0<|h|,|k|<\frac{r}{2}} \|x^*_{h,k}\| < \infty$. Since $\|x^*_{h,k}\| = \|x_{h,k}\|$, $\sup_{0<|h|,|k|<\frac{r}{2}} \|x_{h,k}\| < \infty$. Recalling the definition of $x_{h,k}$, we see that

$$\left\| \frac{x(\xi_0+h)-x(\xi_0)}{h} - \frac{x(\xi_0+k)-x(\xi_0)}{k} \right\| = O(|h-k|), \text{ as } h,k \to 0.$$

It follows that $[x(\xi_0+h)-x(\xi_0)]/h$ satisfies Cauchy's criterion as $h \to 0$. Since B is a Banach space, $[x(\xi_0+h)-x(\xi_0)]/h$ must converge to a limit $y \in B$. Thus $x(\cdot)$ is holomorphic on U in the strong sense.

Terminology: Henceforth we call $x(\cdot)$ "holomorphic" on U if it is holomorphic in one of the equivalent senses above.

Theorem 5.4 makes it easy to generalize much of the theory of complex holomorphic functions to vector valued holomorphic functions. In what follows, suppose B is a Banach space and $U \subseteq \mathbb{C}$ is open.

Corollary 5.1. *If $x(\xi)$ is holomorphic on U, then $x'(\xi)$ is holomorphic on U.*

Proof. If $x(\xi)$ is holomorphic on U, then $(f \circ x)(\xi)$ is holomorphic on U for all $f \in B^*$. By complex function theory, $\frac{d}{d\xi}(f \circ x)$ is holomorphic on U. If we can show that $\frac{d}{d\xi}(f \circ x) = f(\frac{d}{d\xi}x)$, then it will follow that $\frac{d}{d\xi}x$ is holomorphic on U. This is indeed the case, because $\|\frac{1}{h}(x(\xi+h)-x(\xi)) - x'(\xi)\| \xrightarrow[h\to 0]{} 0$, so for every $f \in B^*$ $\|\frac{1}{h}(f(x(\xi+h)) - f(x(\xi))) - f(x'(\xi))\| \xrightarrow[h\to 0]{} 0$ □

Corollary 5.2. *Suppose X is a Banach space and $D \subseteq \mathbb{C}$ is open and simply connected, then $x : D \to \mathbb{R}$ is holomorphic on D iff for every closed smooth curve γ in D, $\oint_\gamma x(\xi)d\xi = 0$.*

Proof. Integrals of continuous functions $x : [a,b] \to B$ are defined for all continuous functions $x : [a,b] \to B$ via Riemann sums, as in the classical case. In the case of smooth curves $\gamma : [0,1] \to D$,

$$\oint_\gamma x(\xi)d\xi := \int_0^1 x(\gamma(t))\gamma'(t)dt.$$

Working with Riemann sums, it is easy to see that $f(\int_\gamma x(\xi)d\xi) = \int_\gamma f(x(\xi))d\xi$ for every $f \in B^*$.

If $x : D \to B$ is holomorphic, then $x : D \to B$ is continuous (even locally Lipschitz) so we can integrate and see that for every $f \in B^*$,

$$f\left(\oint_\gamma x(\xi)d\xi\right) = \int_0^1 f(x(\gamma(t)))\gamma'(t)dt = \oint_\gamma f(x(\xi))d\xi = 0,$$

where the last equality is because $f \circ x$ holomorphic. It follows that $f(\oint_\gamma x(\xi)d\xi) = 0$ for all $f \in B^*$. By the Hahn–Banach theorem, $\oint_\gamma x(z)dz = 0$.

Next suppose that $\oint_\gamma x(\xi)d\xi$ for all closed smooth curves γ in D, then $\oint_\gamma f[x(\xi)]d\xi = f(\oint_\gamma x(\xi)d\xi) = 0$ for all closed smooth curves γ in D and all bounded linear functionals $f \in B^*$. By Morera's Theorem, $f[x(\xi)]$ is holomorphic on U. Thus $x(\xi)$ is (weakly) holomorphic on U. □

Corollary 5.3 (Cauchy's Integral Formula). *Suppose $D \subset \mathbb{C}$ is open and simply connected. If $x : D \to B$ is holomorphic, then for every $\xi_0 \in D$ and closed smooth curve γ in D,*

$$x(\xi_0) = \frac{1}{2\pi i}\oint_\gamma \frac{x(z)}{z-\xi_0}dz\ ,\ x^{(n)}(\xi_0) = \frac{n!}{2\pi i}\oint_\gamma \frac{x(z)}{(z-\xi_0)^{n+1}}dz.$$

Proof. Using Corollary 5.1, it is easy to see by induction that if $x(z)$ is holomorphic on U, then $x^{(n)}(z) := \frac{d^n}{dz^n}x(z)$ exists and is holomorphic on U, and $f[x^{(n)}(z)] = \frac{d^n}{dz^n}f[x(z)]$ for all $f \in B^*$. Cauchy's formula for $f[x(z)]$ says that

$$f(x^{(n)}(\xi_0)) = \frac{n!}{2\pi i}\oint_\gamma \frac{f[x(z)]}{(z-\xi_0)^{n+1}}dz.$$

Since f is a bounded linear functional, we can pull it out of the integrand:

$$f(x^{(n)}(\xi_0)) = f\left(\frac{n!}{2\pi i}\oint_\gamma \frac{x(z)}{(z-\xi_0)^{n+1}}d\xi\right).$$

This identity holds for all $f \in B^*$, and B^* separates points in B (Hahn–Banach Theorem), so $x^{(n)}(\xi_0) = \frac{n!}{2\pi i}\oint_\gamma \frac{x(z)}{(z-\xi_0)^{n+1}}dz$. □

Corollary 5.4 (Power Series Expansion). *Let $B_r(z_0) := \{z \in \mathbb{C} : |z-z_0| < r\}$. A function $x : B_r(z_0) \to B$ is holomorphic on $B_r(z_0)$ iff it can be put in the form*

$$x(z) = x(z_0) + \sum_{n=1}^{\infty} a_n(z-z_0)^n \quad (|z-z_0| < r),$$

where $a_n \in B$, and the series on the right hand side converges uniformly in norm on compact subsets of $B_r(z_0)$. In this case $a_n = x^{(n)}(z_0)/n!$.

Proof. ($\Rightarrow$): Suppose $x(z)$ is holomorphic on $B_r(z_0)$. By corollary 5.1, $x^{(n)}(z)$ exists for all n on $B_r(z_0)$. By Cauchy's integral formula, the following estimates holds for all $z \in B_\rho(z_0)$ and $0 < \rho < r$:

$$\|x^{(n)}(z)\| \leq \frac{n!}{2\pi} \cdot 2\pi\rho \cdot \frac{\max\{\|x(z)\| : |z| = \rho\}}{\rho^{n+1}} = O(n!\rho^{-n}).$$

It follows that the series

$$y(z) := \sum_{n=0}^{\infty} \frac{x^{(n)}(z_0)}{n!}(z - z_0)^n$$

converges uniformly in norm on compact subsets of $B_r(z_0)$.

We claim that $y(z) = x(z)$. To see this, note that if $f \in B^*$, then

$$f[y(z)] = \sum_{n=0}^{\infty} \frac{f[x^{(n)}(z_0)]}{n!}(z - z_0)^n.$$

It is easy to see by induction that $f[x^{(n)}(z_0)] = \frac{d^n}{dz^n} f[x(z)]$, thus $f[y(z)]$ is the Taylor series of $f[x(z)]$. Since $f[x(z)]$ is holomorphic on $B_r(z_0)$, the Taylor series of $f[x(z)]$ equals $f[x(z)]$. We see that $x(z) = y(z)$. ♢

($\Leftarrow$): Suppose $x(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$, then for all $f \in B^*$, $f[x(z)] = \sum_{n=0}^{\infty} f(a_n)(z - z_0)^n$, whence $f[x(z)]$ is holomorphic on $B_r(z_0)$. Since f was an arbitrary element of B^*, $x(z)$ is holomorphic on $B_r(z_0)$. □

5.1.3 Resolvents

Suppose B is a Banach space over $\mathbb{C}$.

Definition 5.4. Suppose $L : B \to B$ is a bounded linear operator on B. The *resolvent set* of L is

$$\text{Res}(L) := \{z \in \mathbb{C} : (zI - L) \text{ has a bounded inverse}\}.$$

If $z \in \text{Res}(L)$, then $R(z) := (zI - L)^{-1}$ is called the *resolvent operator* of L at z.[2]

Proposition 5.1 (Properties of the Resolvent).

1. Res(L) *is open and non–empty;*
2. $R(z)$ *is holomorphic on* Res(L)*, and* Res(L) *cannot be holomorphically continued to points outside* Res(L)*;*
3. Resolvent equation*: for all* $z, w \in \text{Res}(L)$*,*

$$R(w) - R(z) = (z - w)R(w)R(z) \tag{5.2}$$

[2] Some people define the resolvent to be $(L - Iz)^{-1} = -(zI - L)^{-1}$.

4. Neumann's Series Expansion*: Suppose* $z_0 \in \mathrm{Res}(L)$*, then*

$$R(z) = \sum_{n=0}^{\infty} (-1)^n (z-z_0)^n R(z_0)^{n+1} \quad (|z-z_0| < \|R(z_0)\|^{-1}). \tag{5.3}$$

Proof. First we claim that if A is a bounded linear operator s.t. $\|A\| < 1$, then $I-A$ has a bounded inverse, and $(I-A)^{-1} = \sum_{n\geq 0} A^n$. The series on the right converges in norm, because of the trivial estimate $\|A^n\| \leq \|A\|^n$, therefore it defines a bounded linear operator B. A direct calculation shows that $B - AB = I$, whence $(I-A)B = I$. Similarly $B(I-A) = I$. Thus $B = (I-A)^{-1}$.

We can now see that $(zI-L)^{-1} = z^{-1}(I - z^{-1}L)^{-1}$ exists and is bounded whenever $|z| > \|L\|$. Thus $\mathrm{Res}(L)$ is non–empty (moreover, $\mathbb{C} \setminus \mathrm{Res}(L)$ is bounded).

We show that $\mathrm{Res}(L)$ is open. Suppose $z \in \mathrm{Res}(L)$ and w is close to z (we see how close below). Formal manipulations show that

$$(wI-L)^{-1} = (zI - L - (zI - wI))^{-1} = [I - (z-w)(zI-L)^{-1}]^{-1}(zI-L)^{-1}.$$

If $|z-w| < \|R(z)\|^{-1}$ then $[I-(z-w)(zI-L)^{-1}]^{-1}$ is bounded, so $w \in \mathrm{Res}(L)$.

Returning to the identity just proved for $(wI-L)^{-1}$ we see that for all w s.t. $|w-z| < \|R(z)\|^{-1}$,

$$R(w) = (I-(z-w)R(z))^{-1}R(z) = \sum_{n=0}^{\infty} (z-w)^n R(z)^{n+1},$$

where the series on the right converges in norm. Viewing z as a constant and w as a variable, we obtain a power series expansion for the resolvent near an arbitrary $z \in \mathrm{Res}(L)$. It follows that the resolvent operator is holomorphic on $\mathrm{Res}(L)$. Incidently, we have proven the Neumann series mentioned is part 4 of the proposition.

We show that $R(z)$ cannot be extended holomorphically outside $\mathrm{Res}(L)$. Otherwise, there exists $z_0 \in \partial\mathrm{Res}(L)$, $r > 0$ and a holomorphic function $\widetilde{R}(z)$ on $B_r(z_0) := \{z : |z-z_0| < r\}$ s.t. $\widetilde{R}(z) = R(z)$ on $B_r(z_0) \cap \mathrm{Res}(L)$. We claim that for every $x \in B$ and for every $z \in B_r(z_0)$,

$$(zI-L)\widetilde{R}(z)x = x \text{ and } \widetilde{R}(z)(zI-L)x = x,$$

and derive a contradiction from the fact that $zI-L$ has no bounded inverse for $z \in B_r(z_0) \setminus \mathrm{Res}(T)$. The functions $(zI-L)\widetilde{R}(z)$, $\widetilde{R}(z)(zI-L)$ are holomorphic on $B_r(z_0)$ (because they can be expanded to a power series there), therefore for all $f \in B^*$ and $x \in B$ $f[(zI-L)\widetilde{R}(z)x]$ and $f[\widetilde{R}(z)(zI-L)x]$ are holomorphic on $B_r(z_0)$ (we are using the fact that $L \mapsto f(Lx)$ is a bounded linear functional on the Banach space of bounded linear operators on B). We know that

$$f[(zI-L)\widetilde{R}(z)] = f[\widetilde{R}(z)(zI-L)] = f(x) \text{ for all } z \in B_r(z_0) \cap \mathrm{Res}(L).$$

A holomorphic function with connected domain which equals a constant on a non–empty open subset is equal to this constant on its entire domain. It follows that

$$f[(zI-L)\widetilde{R}(z)] = f[\widetilde{R}(z)(zI-L)] = f(x) \text{ for all } z \in B_r(z_0).$$

Since this holds for all $f \in B^*$ and $x \in B$, $(zI-L)\widetilde{R}(z) = \widetilde{R}(z)(zI-L)$ as required.

Proof of the resolvent equation: $R(z)-R(w) = R(z)(wI-L)R(w) - R(z)(zI-L)R(w) = R(z)[(w-z)I]R(w) = (w-z)R(z)R(w)$. □

5.1.4 Spectrum

Suppose $(B, \|\cdot\|)$ is a Banach space over the complex numbers.

Definition 5.5. The *spectrum* of a bounded linear operator L is the complement of the resolvent set of L. It is denoted by $\mathrm{Spect}(L)$. The *spectral radius* of L is $\rho(L) := \sup\{|z| : z \in \mathrm{Spect}(L)\}$.

Every eigenvalue of L (i.e. $\lambda \in C$ s.t. $Lx = \lambda x$ for some $0 \neq x \in B$) belongs to $\mathrm{Spect}(L)$, because if $\ker(\lambda I - L) \neq \{0\}$, then $\lambda I - L$ is not invertible. If $\dim B < \infty$, then the spectrum is equal to the collection eigenvalues of L. But if $\dim B = \infty$ then the spectrum is usually much larger than the collection of eigenvalues:

Example: Suppose $B = L^1[0,1]$ and $(Lf)(t) = tf(t)$. The spectrum of L is equal to $[0,1]$, but L has no eigenvalues.[3]

Proposition 5.2. *Let L be a bounded linear operator on a Banach space B. The spectrum of L is compact and non–empty. The spectral radius of L is given by*

$$\rho(L) = \lim_{n\to\infty} \sqrt[n]{\|L^n\|}.$$

Proof. The spectrum is closed, because its complement, $\mathrm{Res}(L)$, is open. The spectrum is bounded, because (as we saw in the previous section), $zI-L$ has a bounded inverse of the form $\sum_{n\geq 0} z^{-(n+1)}L^n$ whenever this series converges in norm. This is the case whenever $|z| > \limsup \sqrt[n]{\|L^n\|} =: \rho'(L)$ (a finite number since $\|L^n\| \leq \|L\|^n$). Thus $\mathrm{Spect}(L) \subset \{z : |z| \leq \rho'(L)\}$, and $\mathrm{Spect}(L)$ is compact.

We also see that $\rho(L) \leq \rho'(L)$. Suppose by way of contradiction that $\rho(L) \lneqq \rho'(L)$. Then $\mathrm{Res}(T)$ contains a neighborhood of $\{z : |z| = \rho'(L)\}$. It follows that the series $\sum z^{-(n+1)}L^n$ admits holomorphic continuation beyond its radius of convergence. But this is impossible, because the summands of $\sum z^{-(n+1)}L^n$ do not converge to zero in norm for such z.

It follows that $\rho(L) = \rho'(L) = \limsup \sqrt[n]{\|L^n\|}$. In fact the limsup is a limit, because $\|L^{n+m}\| \leq \|L^n\|\|L^m\|$, so the sequence $\log \|L^n\|$ is sub additive.

It remains to show that $\mathrm{Spect}(L) \neq \varnothing$. Otherwise, $R(z)$ would be holomorphic on $\mathbb{C}$, and so $f[R(z)]$ is an entire function for every $f \in \mathrm{Hom}(B)^*$. Now

[3] Proof that $\mathrm{Spect}(L) = [0,1]$: The resolvent of L at z is the operator $[R(z)f](t) = (z-t)^{-1}f(t)$. If $z \notin [0,1]$, then this is a bounded linear operator on $L^1[0,1]$. But if $z \in [0,1]$ then $R(z)$ is not well defined on $L^1[0,1]$, because it maps the constant function 1 to the function $(z-t)^{-1}$, which has a non-integrable singularity at $t = z$.

$$\|R(z)\| = \|(zI-L)^{-1}\| = |z|^{-1}\|(I-z^{-1}L)^{-1}\| \leq |z|^{-1}\sum_{n=0}^{\infty}|z|^{-n}\|L\|^n \xrightarrow[z\to\infty]{} 0,$$

so $f[R(z)] \xrightarrow[z\to\infty]{} 0$. In particular, $f[R(z)]$ is bounded. By Liouville's theorem, $f[R(z)]$ is constant, whence $f[R(z)-R(0)] = 0$ for all $f \in \mathrm{Hom}(B)^*$. This means that $R(z) = R(0) \equiv L^{-1}$ for all z, whence $zI - L = L$ for all z. This is certainly false, proving that $\mathrm{Spect}(L)$ cannot be empty. □

5.1.5 Eigenprojections and the Separation of Spectrum Theorem

Let $(B, \|\cdot\|)$ denote a Banach space. A bounded linear operator $P: B \to B$ is called a *projection*, if it satisfies $P^2 = P$. Any projection gives rise to a direct sum decomposition of B:

$$B = \mathrm{Im}(P) \oplus \ker(P).$$

The decomposition is $x = Px + (I-P)x$.

Theorem 5.5 (Separation of Spectrum). *Suppose L is a bounded linear operator s.t.* $\mathrm{Spect}(L) = \sigma_1 \uplus \sigma_2$ *where σ_1, σ_1 are compact and disjoint, and suppose γ is a closed smooth simple curve which does not intersect* $\mathrm{Spect}(L)$ *and which contains σ_1 in its interior, and σ_2 in its exterior. Let*

$$P := \frac{1}{2\pi i}\oint_{\gamma}(zI-L)^{-1}dz.$$

1. *P is a projection;*
2. *$PL = LP$ and $\ker P, \mathrm{Im} P$ are L–invariant: $L(\mathrm{Im}P) \subseteq \mathrm{Im}P, L(\ker P) \subseteq \ker P$;*
3. $\mathrm{Spect}(L|_{\mathrm{Im}P}) = \sigma_1$ *and* $\mathrm{Spect}(L|_{\ker P}) = \sigma_2$.

Proof.

Step 1. P is a projection.

Since γ does not intersect $\mathrm{Spect}(L)$, $R(z)$ is bounded for each $z \in \gamma$, and $z \mapsto R(z)$ is continuous on γ. This implies that $\|P\| \leq (2\pi)^{-1}\ell(\gamma)\max\{\|R(z)\| : z \in \gamma\} < \infty$.

To see that $P^2 = P$, "expand" γ to a slightly larger closed simple smooth curve γ^* which contains γ (and σ_1) in its interior and σ_2 in its exterior. Then $P = \frac{1}{2\pi i}\oint_{\gamma^*}(wI - L)^{-1}dw$ and so

$$\begin{aligned}
P^2 &= \frac{1}{(2\pi i)^2}\oint_{\gamma}R(z)\oint_{\gamma^*}R(w)dw\,dz = \frac{1}{(2\pi i)^2}\oint_{\gamma}\oint_{\gamma^*}R(z)R(w)dwdz \quad (R(z) \in \mathrm{Hom}(B))\\
&= \frac{1}{(2\pi i)^2}\oint_{\gamma}\oint_{\gamma^*}\frac{R(z)-R(w)}{w-z}dwdz \quad \text{(by the resolvent equation)}\\
&= \frac{1}{(2\pi i)^2}\oint_{\gamma}\oint_{\gamma^*}\frac{R(z)}{w-z}dwdz - \frac{1}{(2\pi i)^2}\oint_{\gamma}\oint_{\gamma^*}\frac{R(w)}{w-z}dwdz
\end{aligned}$$

$$= \frac{1}{(2\pi i)^2}\oint_\gamma \left(R(z)\oint_{\gamma^*}\frac{1}{w-z}dw\right)dz - \frac{1}{(2\pi i)^2}\oint_{\gamma^*}\left(R(w)\oint_\gamma \frac{1}{w-z}dz\right)dw$$
$$= \frac{1}{(2\pi i)^2}\oint_\gamma R(z)\cdot 2\pi i\, dz - \frac{1}{(2\pi i)^2}\oint_{\gamma^*} R(zw)\cdot 0\, dw \quad \text{(Cauchy integral formula)}$$
$$= \frac{1}{2\pi i}\oint_\gamma (zI-L)^{-1}dz = P.$$

(We used the fact that γ^* is outside γ.) $\diamondsuit$

Step 3. $PL = LP$, $\text{Spect}(LP) \subseteq \sigma_1$ and $\text{Spect}(L(I-P)) \subseteq \sigma_2$.

It is a general fact that if an invertible operator S commutes with an operator T, then its inverse S^{-1} also commutes with T.[4] Since $(zI-L)$ commutes with L, $R(z) = (zI-L)^{-1}$ commutes with L (wherever it is defined). It follows that $PL = \left(\frac{1}{2\pi i}\oint_\gamma R(z)dz\right)L = \frac{1}{2\pi i}\oint_\gamma R(z)Ldz = \frac{1}{2\pi i}\oint_\gamma LR(z)dz = L\left(\frac{1}{2\pi i}\oint_\gamma R(z)dz\right) = LP$, so P commutes with L. Since $PL = LP$, $\ker P, \text{Im}P$ are L–invariant: $L(\ker P) \subseteq \ker P$, $L(\text{Im}P) \subseteq \text{Im}P$.

Let $\widetilde{R}(z)$ denote the resolvent operator of $L|_{\text{Im}P} : \text{Im}P \to \text{Im}P$. Since $\text{Im}P$ is L–invariant, $\text{Im}P$ is $(zI-L)$–invariant, and so $\text{Res}(L|_{\text{Im}P}) \supseteq \text{Res}(L)$ and

$$\widetilde{R}(z) = ((zI-L)|_{\text{Im}P})^{-1} = (zI-L)^{-1}|_{\text{Im}P} = R(z)|_{\text{Im}P}.$$

We show that $\widetilde{R}(z)$ has a holomorphic continuation to the exterior of γ. Since the resolvent operator of $L|_{\text{Im}P}$ cannot be holomorphically extended outside $\text{Res}(L|_{\text{Im}P})$, this will show that $\text{Res}(L|_{\text{Im}P}) \supseteq$ exterior of $\gamma \supset \sigma_2$, whence

$$\text{Spect}(L|_{\text{Im}P}) \subseteq \text{Spect}(L) \setminus \sigma_2 = \sigma_1.$$

Here is the holomorphic extension: For every $z \in \text{Res}(L)$ outside the exterior of γ,

$$\widetilde{R}(z) = R(z)P \quad \text{(because } P = I \text{ on Im}P)$$
$$= \frac{1}{2\pi i}\oint_\gamma R(z)R(w)dw = \frac{1}{2\pi i}\oint_\gamma \frac{R(z)-R(w)}{w-z}dw \quad \text{(resolvent equation)}$$
$$= \frac{1}{2\pi i}R(z)\left(\oint_\gamma \frac{1}{w-z}dw\right) - \frac{1}{2\pi i}\oint_\gamma \frac{R(w)}{w-z}dw$$
$$= -\frac{1}{2\pi i}\oint_\gamma \frac{R(w)}{w-z}dw.$$

The last expression is holomorphic everywhere outside γ (to find is Taylor series around z_0, expand $(w-z)^{-1}$ into a power series in $(z-z_0)$ and integrate term by term over $w \in \gamma$).

In the same way one shows that the resolvent of $L|_{\ker P}$ extends holomorphically to the interior of γ, with the consequence that $\text{Spect}(L|_{\ker P}) \subseteq \sigma_2$. $\diamondsuit$

We saw that $\text{Spect}(L|_{\text{Im}P}) \subseteq \sigma_1$ and $\text{Spect}(L|_{\ker P}) \subseteq \sigma_2$. On the other hand

[4] Proof: If $ST = TS$, then $T = S^{-1}TS$, whence $TS^{-1} = S^{-1}T$.

$$\text{Spect}(L|_{\text{Im}P}) \cup \text{Spect}(L|_{\ker P}) \supseteq \text{Spect}(L) = \sigma_1 \cup \sigma_2,$$

because if $(zI-L)|_{\text{Im}P}$, $(zI-L)|_{\ker P}$ have bounded inverses, then $zI-L$ has a bounded inverse (equal to $(zI-L)^{-1}|_{\text{Im}P} + (zI-L)^{-1}|_{\ker P}$). It follows that $\text{Spect}(L|_{\text{Im}P}) = \sigma_1$ and $\text{Spect}(L|_{\ker P}) = \sigma_2$. □

There is a particularly important special case. A point $\lambda \in \text{Spect}(L)$ is called *isolated*, if $\text{Spect}(L) \cap \{z : |z-\lambda| < \varepsilon\} = \{\lambda\}$ for some $\varepsilon > 0$. Any such point can be separated from the rest of the spectrum of L by a simple closed smooth curve γ. Let

$$P := \frac{1}{2\pi i}\oint_\gamma (zI-L)^{-1} dz.$$

1. P is called the *eigenprojection* of λ;
2. $\text{Im}P$ is called the *eigenspace* of λ;
3. $\dim \text{Im}P$ is called the *geometric multiplicity* of λ.

It is easy to see that these notions are independent of the choice of γ (subject to the requirement that γ separate the spectrum of L as above).

Suppose P is the eigenprojection of an isolated spectrum point λ. If $\dim \text{Im}P < \infty$, then λ must be an eigenvalue, because $LP : \text{Im}P \to \text{Im}P$ is a linear operator on a finite dimensional space, and the spectrum of any linear operator on a finite dimensional space equals its set of eigenvalues.

Terminology: An eigenvalue with geometric multiplicity equal to one is called a *simple eigenvalue*.

5.1.6 Analytic perturbation theory

Suppose $(B, \|\cdot\|)$ is a Banach space over $\mathbb{C}$, and let U be an open neighborhood of zero in $\mathbb{C}$. A one–parameter family $\{L_z\}_{z\in U} \subset \text{Hom}(B)$ is called *analytic*, if $z \mapsto L_z$ is a holomorphic map from U into $\text{Hom}(B)$. We think of such a family as of an *analytic perturbation* of $L = L_0$.

We are interested in the effects of analytic perturbations on the spectrum. Of principle importance to us is the behavior of simple eigenvalues under analytic perturbations.

We begin with a basic stability result for projections. Recall that an *isomorphism* of a Banach space B is an invertible linear map $\pi : B \to B$ s.t. π, π^{-1} are bounded, and that two maps $T, S : B \to B$ are called *similar* if there is an isomorphism π s.t. $\pi^{-1}T\pi = S$.

Lemma 5.1 (Kato). *Suppose $P : B \to B$ is a projection on a Banach space. Any projection $Q : B \to B$ such that $\|Q-P\| < 1$ is similar to P.*

Proof. First we construct a map $U : B \to B$ which maps $\ker P$ into $\ker Q$ and $\text{Im}P$ into $\text{Im}Q$: $U := (I-Q)(I-P) + QP$.

Since $P^2 = P, Q^2 = Q$, $UP = QP$ and $QU = QP$. Thus

$$UP = QU.$$

The only question is whether U is invertible with bounded inverse.

Let $V : B \to B$ be the map $V := (I-P)(I-Q) + PQ$ (which maps $\ker Q$ onto $\ker P$ and $\mathrm{Im}P$ onto $\mathrm{Im}Q$). This is "almost" an inverse for U:

$$\begin{aligned}
UV &= (I-Q)(I-P)(I-Q) + QPQ = (I-Q)(I-Q-P+PQ) + QPQ \\
&= I-Q-P+PQ-Q+Q+QP-QPQ+QPQ = I-Q-P+PQ+QP \\
VU &= (I-P)(I-Q)(I-P) + PQP = (I-P)(I-P-Q+QP) + PQP \\
&= I-P-Q+QP+PQ-PQP+PQP = I-P-Q+QP+PQ
\end{aligned}$$

So $UV = VU = I - R$, where $R := P + Q - QP - PQ$. The point is that if $\|P-Q\|$ is small, then $\|R\|$ is small, because

$$R = P^2 + Q^2 - QP - PQ = (P-Q)^2.$$

In particular, if $\|P-Q\| < 1$, then $I-R$ is invertible.

Suppose this is the case, then

- Since $VU = I-R$ and $\ker(I-R) = \{0\}$, $\ker U = \{0\}$, so U is one-to-one.
- Since $UV = I-R$ and $\mathrm{Im}(I-R) = B$, $\mathrm{Im}U = B$, so U is onto.

It follows that U is invertible. An invertible map on a Banach space has a bounded inverse (Open Mapping Theorem), so U^{-1} is bounded, and P, Q are similar. $\square$

Theorem 5.6. *Let $\{L_z\}_{z\in U}$ be an analytic family of bounded linear operators. If L_0 has a simple eigenvalue λ which is separated from the rest of the spectrum of L_0 by a smooth closed simple curve γ, then there exists $\varepsilon > 0$ and holomorphic functions $\lambda(z) \in \mathbb{C}, P(z) \in \mathrm{Hom}(B)$ defined on $\{z : |z| < \varepsilon\}$ s.t.*

1. *$\lambda(0) = \lambda$ and $P(0) =$eigenprojection of λ for L_0;*
2. *$\lambda(z)$ is a simple isolated eigenvalue for L_z, with eigenprojection $P(z)$;*
3. *$\lambda(z)$ is separated from $\mathrm{Spect}(L_z) \setminus \{\lambda\}$ by γ.*

(We allow the separating curve to contain λ in its interior or exterior.)

Proof. Let $\Lambda := \{(\xi, z) \in \mathbb{C} \times U : (\xi I - L_z)^{-1}$ has a bounded inverse$\}$. This is an open set, because $z \mapsto L_z$ is continuous and because the property of having a bounded inverse is open in $\mathrm{Hom}(B)$.[5]

[5] More precisely: for every operator A, if A has a bounded inverse, then B has a bounded inverse for all B s.t. $\|B-A\|$ is sufficiently small. Here is why: Formally,

$$B^{-1} = (A-(A-B))^{-1} = A^{-1}(I-(A-B)A^{-1})^{-1} = A^{-1}\sum_{k=0}^{\infty}((A-B)A^{-1})^k.$$

The last expression defined a bounded operator whenever $\|A-B\| < 1/\|A^{-1}\|$, and this operator can be checked to be an inverse for B by direct calculation.

Pick a closed smooth simple curve γ which contains λ_0 in its interior and $\mathrm{Spect}(L_0)\setminus\{\lambda_0\}$ in its exterior. Such γ lies inside $\mathrm{Res}(L_0)$, so $\xi I - L_0$ has a bounded inverse for all ξ on γ. Since γ is compact, and the existence of a bounded inverse is an open property, there exists $0<\varepsilon<1$ so small that

$$(\xi,z)\in\Lambda \text{ for all } \xi\in\gamma, |z|<\varepsilon.$$

Thus $R(\xi,z):=(\xi I-L_z)^{-1}$ is well–defined and bounded for all $\xi\in\gamma$ and $|z|<\varepsilon$.

Step 1. $P(z):=\frac{1}{2\pi i}\oint_\gamma(\xi I-L_z)^{-1}d\xi$ is holomorphic on $\{z:|z|<\varepsilon\}$.

We claim that $(\xi,z)\mapsto R(\xi,z)$ is holomorphic on Λ in the sense that it can be developed into a norm–convergent power series in $(\xi-\xi_0),(z-z_0)$ near any $(\xi_0,z_0)\in\Lambda$.

$$\begin{aligned}R(\xi,z)&=\big(\xi_0 I-L_{z_0}-[(\xi_0-\xi)I+(L_z-L_{z_0})]\big)^{-1}\\&=R(\xi_0,z_0)\big(I-[(\xi_0-\xi)I+(L_z-L_{z_0})]R(\xi_0,z_0)\big)^{-1}\\&=\sum_{n=0}^{\infty}R(\xi_0,z_0)[((\xi_0-\xi)I+(L_z-L_{z_0}))R(\xi_0,z_0)]^n,\end{aligned}$$

provided $\|(\xi_0-\xi)I+L_z-L_{z_0}\|<1/\|R(\xi_0,z_0)\|$. The convergence is uniform in norm on compact subsets of $\{(\xi,z):\|(\xi_0-\xi)I+L_z-L_{z_0}\|<1/\|R(\xi_0,z_0)\|\}$.

Since $z\mapsto L_z$ is holomorphic at z_0, it can be expanded into a power series in $(z-z_0)$, and this power series converges in norm uniformly on some closed disc centered at z_0. We substitute this series above, expand, and collect terms. The result is power series in $(z-z_0)$ and $(\xi-\xi_0)$ which converges in norm uniformly on some compact neighborhood $V(\xi_0,z_0)$ of (ξ_0,z_0).

Collecting the terms which multiply $(z-z_0)^n$, we are led to the following series expansion for $R(\xi,z)$ on $V(\xi_0,z_0)$:

$$R(\xi,z)=R(\xi_0,z_0)+\sum_{n=1}^{\infty}A_n(\xi)(z-z_0)^n, \tag{5.4}$$

where $A_n(\xi)$ are operator valued functions which are holomorphic on $\{z:|z-z_0|<\varepsilon\}$ for some $\varepsilon=\varepsilon(\xi_0,z_0)$. This is sometimes called the "second Neumann series" (the first Neumann series is (5.3)).

Now cover the compact set $\{(\xi,z):\xi\in\gamma,|z|<\varepsilon\}$ by neighborhoods $V(\xi,z)$ as above. By compactness, finitely many neighborhoods suffice, and the result is that there exists $\varepsilon>0$ so small that for all ξ on γ and $|z|<\varepsilon$,

$$R(\xi,z)=R(\xi_0,z_0)+\sum_{n=1}^{\infty}A_n(\xi)(z-z_0)^n,$$

where $A_n(\xi)$ are continuous on γ and the series on the right converges uniformly in norm.

Integrating over γ, we see that $P(z)$ has a norm convergent power series expansion of the form $P(\xi,z) = P(\xi_0,z_0) + (2\pi i)^{-1}\sum_{n\geq 1}(z-z_0)^n \oint_\gamma A_n(\xi)d\xi$. $\diamondsuit$

Claim 2. For all z sufficiently close to zero, $P(z)$ is an eigenprojection of a simple eigenvalue $\lambda(z)$ of L_z.

Since γ and ε were constructed so that $\gamma \subset \mathrm{Res}(L_z)$ for all $|z| < \varepsilon$, γ separates the part of the spectrum of L_z inside γ (denoted by $\sigma_{\mathrm{in}}(z)$) from the part of the spectrum of L_z outside γ (denoted by $\sigma_{\mathrm{out}}(z)$). By Theorem 5.5, $P(z)$ is a projection, $\mathrm{Spect}(L_z|_{\mathrm{Im}P(z)}) = \sigma_{\mathrm{in}}(z)$, and $\mathrm{Spect}(L_z|_{\ker P(z)}) = \sigma_{\mathrm{out}}(z)$.

We claim that if $|z|$ is sufficiently small, then $\sigma_{\mathrm{in}}(z)$ consists of a single point $\lambda(z)$, and that $\lambda(z)$ is a simple eigenvalue of L_z.

Since $z \mapsto P(z)$ is continuous, Lemma 5.1 and a compactness argument imply that $P(z)$ is similar to $P(0)$ for all $|z|$ sufficiently small. Now $P(0)$ is by definition the eigenprojection of L_0, which by assumption, satisfies $\dim \mathrm{Im}P(0) = 1$. It follows that $\dim \mathrm{Im}P(z) = 1$ for all $|z|$ sufficiently small.

Let $\mathrm{Im}P(z) = \mathrm{span}\{x_z\}$. Since $\mathrm{Im}P(z)$ is L_z–invariant, $L_z x_z = \lambda(z)x_z$ for some $\lambda(z) \in \mathbb{C}$. It follows that $\sigma_{\mathrm{in}}(z) = \mathrm{Spect}(L_z|_{\mathrm{Im}P(z)}) = \{\lambda(z)\}$, where $\lambda(z)$ is an eigenvalue. Moreover $P(z)$ is the eigenprojection of $\lambda(z)$. Since $\dim \mathrm{Im}P(z) = 1$, $\lambda(z)$ is simple. $\diamond$

Step 3. $\lambda(z)$ is holomorphic on $\{z : |z| < \varepsilon\}$.

Fix some $|z_0| < \varepsilon$. We saw that $\dim \mathrm{Im}P(z_0) = 1$, so there exists some $v \in B$ s.t. $x_0 := P(z_0)v \neq 0$. Choose some $f \in B^*$ s.t. $f(x_0) \neq 0$. Since $f[P(z)v]$ is holomorphic, it is non-zero on some neighborhood of z_0. The identity

$$\lambda(z) = \frac{f(L_z P(z)v)}{f(P(z)v)}$$

shows that $\lambda(z)$ is holomorphic on a neighborhood of z_0. $\square$

5.2 Quasi-compactness for Ruelle operators

5.2.1 Quasi-compactness and Hennion's Theorem

Definition 5.6. Let B be a Banach space, and $L : B \to B$ a bounded linear operator with spectral radius $\rho(L)$. We say that L is *quasi–compact*, if there exists $0 < \rho < \rho(L)$ and a decomposition $B = F \oplus H$ where

1. F, H are closed and L–invariant: $L(F) \subseteq F$, $L(H) \subseteq H$;
2. $\dim F < \infty$ and all the eigenvalues of $L|_F$ are larger than or equal to ρ;
3. the spectral radius of $L|_H$ is strictly less than ρ.

The origin of the term "quasi–compact" is the following: Let $\pi : B \to F$ denote the projection onto F parallel to H (i.e. $\pi(x) \in F, (I-\pi)(x) \in H$ for all $x \in B$), and set $P := L\pi$, $N := L(I-\pi)$. Then $PN = NP = 0$, so $L^n = P^n + N^n$. Now $\|P^n(x)\| \geq \rho^n \|\pi(x)\|$ and $\|N^n(x)\| = o(\rho^n)\|x\|$, thus if L is quasi compact then $L^n = P^n +$negligible error, where P is a compact operator.[6]

In the special case when L has a unique eigenvalue λ of maximal modulus, and this eigenvalue is simple, we can choose the decomposition $B = F \oplus H$ in such a way that F is the one–dimensional eigenspace of λ and $0 < \rho < |\lambda|$ satisfies $\mathrm{Spect}(L) \subset \{z : |z| \leq \rho\} \cup \{\lambda\}$. In this case π is the eigenprojection of λ and we get the following simple asymptotic formula for L^n:

$$L^n = \lambda^n \pi + O(\rho^n) \text{ as } n \to \infty \ \ (0 < \rho < |\lambda|).$$

This situation is often called *spectral gap*, because in this case there is a gap between the leading eigenvalue of L and the rest of its spectrum.

In general, the quasi-compactness of an operator depends on the Banach space it is acting on. The main tool for constructing a Banach space on which Ruelle's operator L_ϕ acts quasi-compactly is the following theorem which emerged out of the works of Doeblin & Fortet, Ionescu–Tulcea & Marinescu, and (in the version presented below) Hennion.

Recall that a *semi–norm* on a vector space V is a real valued function $v \mapsto \|v\|'$ on V s.t. (a) $\|v\|' \geq 0$, (b)$\|\lambda v\|' = |\lambda|\|v\|'$, and (c) $\|u+v\|' \leq \|u\|' + \|v\|'$. But we do not require that $\|v\|' = 0 \Rightarrow v = 0$.

Theorem 5.7 (Hennion). *Suppose $(B, \|\cdot\|)$ is a Banach space and $L : B \to B$ is a bounded linear operator with spectral radius $\rho(L)$ for which there exists semi–norm $\|\cdot\|'$ s.t.:*

1. *$\|\cdot\|'$ is continuous on B;*
2. *there exists $M > 0$ s.t. $\|Lf\|' \leq M\|f\|'$ for all $f \in B$;*
3. *for any sequence of $f_n \in B$, if $\sup\|f_n\| < \infty$ then there exists a subsequence $\{n_k\}_{k\geq 1}$ and some $g \in B$ s.t. $\|Lf_{n_k} - g\|' \xrightarrow[k\to\infty]{} 0$;*
4. *there are $k \geq 1$, $0 < r < \rho(L)$, and $R > 0$ s.t.*

$$\|L^k f\| \leq r^k \|f\| + R\|f\|'. \tag{5.5}$$

Then L is quasi compact

Before giving the proof, we first show how this theorem can be applied in the context of Ruelle operators.

[6] An operator Q on a Banach space B is called *compact* if for every bounded sequence $\{x_n\}_{n\geq 1}$ of elements of B, $\{Q(x_n)\}_{n_k \geq 1}$ has a convergence subsequence.

5.2.2 Application to Ruelle operators of potentials with Gibbs measures

Let X be a topologically mixing TMS with the BIP property, and let $\phi : X \to \mathbb{R}$ denote a function such that there are constants $0 < \theta < 1$, $A_\phi > 0$ s.t.

$$\mathrm{var}_n \phi \leq A_\phi \theta^n \text{ for all } n \geq 1,$$

and such that $P_G(\phi) < \infty$. Under these assumptions ϕ has a Gibbs measure in the sense of Bowen, which we denote by m.

Let $t(x,y) := \min\{n \geq 0 : x_n \neq y_n\}$, and define for $f : X \to \mathbb{C}$, $\mathrm{Lip}(f) := \sup\{|f(x) - f(y)|/\theta^{t(x,y)} : x_0 = y_0\}$. The following space is a Banach space (check!):

$$B := \{f : X \to \mathbb{C} : \|f\| := \mathrm{Lip}(f) + \|f\|_\infty < \infty\}. \tag{5.6}$$

Theorem 5.8 (Aaronson & Denker). *$L_\phi : B \to B$ is a bounded linear operator, and its spectrum consists of a simple eigenvalue equal to $\exp P_G(\phi)$ and a compact subset of $\{z \in C : |z| \leq \rho\}$ for some $\rho < \exp P_G(\phi)$.*

Proof. We assume without loss of generality that $P_G(\phi) = 0$.

Step 1. $L_\phi : B \to B$ is quasi-compact.

We let $\|\cdot\|' := \|\cdot\|_{L^1(m)}$ where m is the Gibbs measure of ϕ, and check the conditions of Hennion's theorem. It is clear that $\|\cdot\|' \leq \|\cdot\| \equiv \mathrm{Lip}(\cdot) + \|\cdot\|_\infty$, so $\|\cdot\|'$ is continuous on B and condition 1 holds.

To see (2), we use the fact that $m = h\nu$ where $L_\phi^* \nu = \nu$, $L_\phi h = h$, and we recall that if X has the BIP property, then there is a constant $C > 1$ s.t. $C^{-1} < h < C$. We have $\|L_\phi f\|' \leq \int L_\phi |f| h d\nu \leq C \int |f| d\nu \leq C^2 \|f\|_{L^1(m)} = C^2 \|f\|'$.

It is convenient to prove condition 4 before condition 3. Fix some k (to be determined later). We need the following estimate:

$$\sup_{x \in [a_0, \ldots, a_{n-1}]} |f(x)| \leq \frac{1}{m[\underline{a}]} \int_{[\underline{a}]} |f| dm + \theta^n \mathrm{Lip}(f).$$

To see this start from the inequality $|f(x)| \leq |f(y)| + \theta^n \mathrm{Lip}(f)$ valid for every $x, y \in [a_0, \ldots, a_{n-1}]$, fix x and average over $y \in [a]$. Suppose $f \in B$, we estimate $\|L_\phi^k f\|$:

Estimation of $\|L_\phi^k f\|_\infty$: We use the Gibbs property $m[x_0, \ldots, x_{k-1}] = G^{\pm 1} \exp \phi_k(x)$ and the notation $P^k(a) := \{(p_0, \ldots, p_{k-1}) : (\underline{p} x_0) \text{ is admissible}\}$. For all $x \in X$,

$$
\begin{aligned}
|(L_\phi^k f)(x)| &\leq \sum_{\underline{p}\in P^k(x_0)} e^{\phi_k(\underline{p}x)}|f(\underline{p}x)| \\
&\leq \sum_{\underline{p}\in P^k(x_0)} Gm[\underline{p}]\left(\frac{1}{m[\underline{p}]}\int_{[\underline{p}]}|f|dm+\theta^k\mathrm{Lip}(f)\right) \\
&\leq G\|f\|_{L^1(m)}+G\theta^k\|f\|. \qquad (5.7)
\end{aligned}
$$

Estimation of $\mathrm{Lip}(L_\phi^k f)$: Suppose $x_0=y_0=a$,

$$
|(L_\phi^k f)(x)-(L_\phi^k f)(y)| \leq \sum_{\underline{p}\in P^k(a)}\left|e^{\phi_k(\underline{p}x)}f(\underline{p}x)-e^{\phi_k(\underline{p}y)}f(\underline{p}y)\right|
$$

$$
\begin{aligned}
&= \sum_{\underline{p}\in P^k(a)} e^{\phi_k(\underline{p}x)}\left|f(\underline{p}x)-f(\underline{p}y)\right| + \sum_{\underline{p}\in P^k(a)} e^{\phi_k(\underline{p}y)}\left|e^{\phi_k(\underline{p}x)-\phi_k(\underline{p}y)}-1\right||f(\underline{p}y)| \\
&\leq \sum_{\underline{p}\in P^k(a)} Gm[\underline{p}]\mathrm{Lip}(f)\theta^{k+t(x,y)} \\
&\quad + \sum_{\underline{p}\in P^k(a)} Gm[\underline{p}]\left(\frac{1}{m[\underline{p}]}\int_{[\underline{p}]}|f|dm+\theta^k\mathrm{Lip}(f)\right)\left|e^{\phi_k(\underline{p}x)-\phi_k(\underline{p}y)}-1\right|
\end{aligned}
$$

Now $|\phi_k(\underline{p}x)-\phi_k(\underline{p}y)| \leq \sum_{j=0}^{k-1} A_\phi \theta^{k+t(x,y)-j} \leq \frac{A_\phi}{1-\theta}\cdot\theta^{t(x,y)}$. Fix a constant C s.t. $|e^\delta - 1| \leq C|\delta|$ for all $|\delta| \leq A_\phi/(1-\theta)$, then

$$
\left|e^{\phi_k(\underline{p}x)-\phi_k(\underline{p}y)}-1\right| \leq \frac{CA_\phi}{1-\theta}\cdot\theta^{t(x,y)} =: C_\phi\theta^{t(x,y)}.
$$

Thus $|(L_\phi^k f)(x)-(L_\phi^k f)(y)| \leq \theta^{t(x,y)}\left[G\theta^k\mathrm{Lip}(f)+GC_\phi(\|f\|_{L^1(m)}+\theta^k\mathrm{Lip}(f))\right]$, so

$$
\mathrm{Lip}(L_\phi^k f) \leq G(1+C_\phi)\theta^k\mathrm{Lip}(f)+GC_\phi\|f\|_{L^1(m)}. \qquad (5.8)
$$

Putting (5.7) and (5.8) together, we see that

$$
\|L_\phi^k f\| \leq G(2+C_\phi)\theta^k\|f\|+GC_\phi\|f\|'.
$$

Since $\|\cdot\|' \leq \|\cdot\|$, this inequality with $k=1$ shows that $L_\phi : B \to B$ is bounded. We also see that $\|L_\phi^k\| = O(1)$ as $k\to\infty$, whence the spectral radius of L_ϕ is no larger than one. Since $G(2+C_\phi)\theta^k \xrightarrow[k\to\infty]{} 0$, one can choose k so large that $r := G(2+C_\phi)\theta^k < \rho(L_\phi)$, and condition 4 holds.

Next we check condition 3: Suppose $\sup\|f_n\| < \infty$. By what we just proved, $C := \sup\|L_\phi f_n\| < \infty$, so then $\{L_\phi f_n\}_{n\geq 1}$ is uniformly bounded and equicontinuous (even equi-Lipschitz). By the Arzela–Ascoli theorem, there is a subsequence $n_k \uparrow \infty$ s.t.

$L_\phi f_{n_k} \xrightarrow[k\to\infty]{} f$ pointwise. The limit belongs to B and satisfies $\|L_\phi f\| \leq 2C$, because $\mathrm{Lip}(f) \leq \sup \mathrm{Lip}(f_n) \leq C$ and $\|f\|_\infty \leq \sup \|L_\phi f_n\|_\infty$. By the bounded convergence theorem, $\|L_\phi f_{n_k} - f\|' = \int |L_\phi f_{n_k} - f| dm \xrightarrow[k\to\infty]{} 0$.

Having checked all the conditions in Hennion's theorem, we can now deduce that L_ϕ is quasi–compact. ♢

Step 2. $\lambda = 1$ is a simple eigenvalue for L_ϕ.

We know that $L_\phi h = h$ where h is the eigenfunction given by the GRPF theorem. This function belongs to B, because as we have already remarked the BIP property of X forces h to be bounded away from zero an infinity, and because

$$\mathrm{var}_k[\log h] \leq \sup_n \mathrm{var}_{n+k}\phi_n = O(\theta^k),$$

whence (since $\inf h > 0$), $\mathrm{var}_k h = O(\theta^k)$. Thus one is an eigenvalue of L_ϕ.

Recall that $m = h\nu$ where h, ν are eigenvectors of L_ϕ, and let $P : B \to B$ denote the operator

$$Pf := h \int f d\nu.$$

Since $h \in B$ and $\nu(X) < 1/\inf h$, P is a bounded linear operator. It is easy to verify that $P^2 = P$ and that $PL_\phi = L_\phi P$. P is useful for us because of the following limit theorem:

$$\left\| L_\phi^n f - Pf \right\|_{L^1(\nu)} \xrightarrow[n\to\infty]{} 0 \text{ for all } f \in L^1(m).$$

This is a result of the exactness of m, Lin's Theorem, and the fact that the transfer operator of $m = h\nu$ is the operator $f \mapsto h^{-1} L_\phi(hf)$.

To show that $\lambda = 1$ is simple, we first we claim that the geometric multiplicity is equal to one. This is because if $f \in B$ is some other eigenvector, then

$$\|f - Pf\|_{L^1(\nu)} = \|L_\phi^n f - Pf\|_{L^1(\nu)} \xrightarrow[n\to\infty]{} 0,$$

whence $f = Pf$ m–almost everywhere ($\nu \sim m$). Since f, Pf are continuous, and m has global support, $f = Pf$ proving that f is proportional to h.

Next we claim that the algebraic multiplicity of $\lambda = 1$ is equal to one. Let V denote the eigenspace of 1. Since L_ϕ is quasi-compact, V is finite dimensional. If $\dim(V) > 1$, then V has a basis in which $L_\phi : V \to V$ acts like a Jordan block with ones on the diagonal. The powers of such a matrix are not bounded, therefore there must exists some function $f \in V$ s.t. $\|L_\phi^n f\| \to \infty$. But this is not the case because as we have already seen above, $\|L_\phi^k\| = O(1)$. ♢

Step 3. $\lambda = 1$ is the only point of the spectrum of L_ϕ which lies on the unit circle.

Suppose there were some other $e^{i\theta} \in \mathrm{Spect}(L_\phi)$. Since L_ϕ is quasi-compact, and $|e^{i\theta}| = 1 =$spectral radius of L_ϕ, $e^{i\theta}$ is an eigenvalue. Let $f \in B$ be an eigenfunction: $L_\phi f = e^{i\theta} f$, $f \neq 0$. Applying P to both sides, using the identity $PL_\phi = P$, we see that $Pf = e^{i\theta} Pf$ whence, since $e^{i\theta} \neq 1$, $Pf = 0$. It follows that

$$\|L_\phi^n f\|_{L^1(\nu)} = \|L_\phi^n f - Pf\|_{L^1(\nu)} \xrightarrow[n\to\infty]{} 0.$$

But $\|L_\phi^n f\|_{L^1(\nu)} = \|e^{in\theta} f\|_{L^1(\nu)} = \|f\|_{L^1(\nu)}$, so this implies that $\|f\|_{L^1(\nu)} = 0$, whence $f = 0$ a.e., whence $f = 0$ in contradiction to our assumptions. $\square$

Remark: In the absence of the BIP property, it is not true that the Ruelle operator of a general $\phi : X \to \mathbb{R}$ such that $P_G(\phi) < \infty$ and $\mathrm{var}_n\phi \le \mathrm{const.}\,\theta^n$ $(n \ge 1)$ acts with a spectral gap on a (non–trivial) Banach space. But this case be shown to be the case for "most" $\phi : X \to \mathbb{R}$ is some sense which can be made precise.

5.2.3 Proof of Hennion's Theorem

It is enough to prove the theorem in the case $k = 1$ (explain why!), so henceforth we assume that (5.5) holds with $k = 1$.

Fix $r < \rho \le \rho(L)$, and let $A(\rho,\rho(L)) := \{z \in C : \rho \le |z| \le \rho(L)\}$ (an annulus). The plan of the proof is to show that for all $z \in A(\rho,\rho(L))$,

- $K(z) := \bigcup_{\ell\ge 0} \ker(zI - L)^\ell$ is finite dimensional, and $I(z) := \bigcap_{\ell\ge 0} \mathrm{Im}(zI - L)^\ell$ is closed;
- $K(z), I(z)$ are L–invariant and $B = K(z) \oplus I(z)$;
- $(zI - L) : I(z) \to I(z)$ is a bijection with bounded inverse;
- the set of $\lambda \in A(\rho,\rho(L))$ s.t. $K(\lambda) \ne \{0\}$ is finite, non–empty.

In particular, the intersection of the spectrum of L with the annulus $A(\rho,\rho(L))$ is a finite set of eigenvalues $\lambda_1,\dots,\lambda_t$ with finite multiplicity (because if z is not an eigenvalue, then $K(z) = 0$, whence $B = I(z)$, whence $(zI - L) : B \to B$ is a bijection with a bounded inverse).

Once we have this spectral information, we let $\{\lambda_1,\dots,\lambda_t\}$ denote the eigenvalues of L in $A(\rho,\rho(L))$ and form

$$F := \bigoplus_{i=1}^{t} K(\lambda_i)\ ,\ H := \bigcap_{i=1}^{t} I(\lambda_i).$$

By the properties of $K(z), I(z)$ mentioned above, F, H are L–invariant, F is finite dimensional, and H is closed. We will show, using standard linear algebra techniques, that $B = F \oplus H$, that the eigenvalues of $L|_F$ are $\lambda_1,\dots,\lambda_t$, and that the spectral radius of $L|_H$ is less than ρ.

The double norm inequality (5.5) and the semi–norm $\|\cdot\|'$ are utilized through the following statement, which is the main technical tool:

Conditional Closure Lemma: Fix $|z| > r$ and let $\{g_n\}_{n\ge 1}$ be a sequence in B s.t. $g_n = (zI - L)f_n$ has a solution $f_n \in L$ for all n. If $g_n \xrightarrow[n\to\infty]{B} g$ and $\sup\|f_n\| < \infty$, then $\{f_n\}_{n\ge 1}$ has a subsequence which converges in B to a solution f of $g = (zI - L)f$.

Proof. Starting from the equation $(g_n - g_m) = (zI - L)(f_n - f_m)$, we see that

$$|z|\|f_n - f_m\| = \|(g_n - g_m) + L(f_n - f_m)\| \le \|g_n - g_m\| + r\|f_n - f_m\| + R\|f_n - f_m\|'.$$

Rearranging terms, we obtain

$$\|f_n - f_m\| \le \frac{1}{|z| - r}\left[\|g_n - g_m\| + \|f_n - f_m\|'\right]. \tag{5.9}$$

1. $\|g_n - g_m\|$ tends to zero as $n, m \to \infty$, because $g_n \xrightarrow[n\to\infty]{B} g$.
2. To deal with $\|f_n - f_m\|'$ we start again from $g_n = (zI - L)f_n$ and deduce

   $$|z| \cdot \|f_n - f_m\|' \le \|g_n - g_m\|' + \|Lf_n - Lf_m\|'.$$

   Since $\sup\|f_n\| < \infty$, there is a subsequence $\{Lf_{n_k}\}_{k\ge1}$ s.t. $\|Lf_{n_k} - h\|' \to 0$ for some $h \in B$. Since $\|\cdot\|'$ is continuous, $\|g_{n_k} - g\|' \to 0$. Thus $\|f_{n_k} - f_{m_k}\|' \le \frac{1}{|z|}\left(\|g_{n_k} - g_{m_k}\|' + \|Lf_{n_k} - Lf_{m_k}\|'\right) \xrightarrow[k,\ell\to\infty]{} 0$.

Returning to (5.9), we see that $\|f_{n_k} - f_{n_\ell}\| \xrightarrow[k,\ell\to\infty]{} 0$, so $\exists f \in B$ s.t. $f_{n_k} \xrightarrow[k\to\infty]{B} f$. Since $zI - L$ is continuous, $g = (zI - L)f$. ♢

We also need the following well–known fact:

Riesz Lemma: Let $(V, \|\cdot\|)$ be a normed vector space, and suppose $U \subseteq V$ is a subspace. If $\overline{U} \neq V$, then for every $0 < t < 1$ there exists $v \in V$ s.t. $\|v\| = 1$ and $\mathrm{dist}(v, U) \ge t$.

Proof. Fix $v_0 \in V \setminus \overline{U}$, and construct $u_0 \in U$ s.t. $\mathrm{dist}(v_0, U) \le \|v_0 - u_0\| \le \frac{1}{t}\mathrm{dist}(v_0, U)$. Calculating, we see that for every $u \in U$,

$$\left\|\frac{v_0 - u_0}{\|v_0 - u_0\|} - \frac{u}{\|v_0 - u_0\|}\right\| = \frac{\|v_0 - (u_0 + u)\|}{\|v_0 - u_0\|} \ge \frac{\mathrm{dist}(v_0, U)}{\frac{1}{t}\mathrm{dist}(v_0, U)} = t.$$

Since this holds for all $u \in U$, $v := (v_0 - u_0)/\|v_0 - u_0\|$ is as required. ♢

We are now ready for the proof. Define, as before, $K(z) := \bigcup_{\ell>0} \ker(zI - L)^\ell$, $I(z) := \bigcap_{\ell>0} \mathrm{Im}(zI - L)^\ell$.

Step 1. Suppose $|z| > r$, then

1. $\ker(zI - L)^\ell$ is finite dimensional for all ℓ;
2. $\mathrm{Im}(zI - L)^\ell$ is closed for all ℓ;
3. there exists m s.t. $K_\ell(z) = K(z)$ and $I_\ell(z) = I(z)$.

Proof. Fix z s.t. $|z| > r$.

Set $K_\ell := \ker(zI - L)^\ell$. We show by induction that $\dim K_\ell < \infty$ for all ℓ. Suppose by way of contradiction that $\dim K_1 = \infty$. Using the Riesz Lemma with $t = 1/2$, it is not difficult to construct $f_n \in \ker(zI - L)$ s.t. $\|f_n\| = 1$ and $\|f_n - f_m\| \ge 1/2$ for all $n \neq m$. We have for all n, $\sup\|f_n\| < \infty$ and $(zI - L)f_n = 0$, so by the conditional closure lemma $\{f_n\}_{n\ge1}$ contains a convergent sequence. But this cannot be the case, so we get a contradiction which proves that $\dim K_1 < \infty$.

Assume by contradiction that $\dim K_{\ell+1}=\infty$, then $\exists f_n\in\ker(zI-L)^{\ell+1}$ s.t. $\|f_n\|=1$ and $\|f_n-f_m\|\geq 1/2$ for $n\neq m$. By construction $g_n:=(zI-L)f_n\in K_\ell$, and $\|g_n\|\leq |z|+\|L\|$. The unit ball in K_ℓ is compact, because $\dim K_\ell<\infty$ by the induction hypothesis, so $\exists n_k\uparrow\infty$ s.t. g_{n_k} converges in norm. By the conditional closure lemma, $\exists n_{k_\ell}$ s.t. $\{f_{n_{k_\ell}}\}$ converges in norm. But this cannot be the case because $\|f_n-f_m\|\geq 1/2$ when $n\neq m$. So $\dim K_{\ell+1}$ must be finite.

We show that $I_\ell:=\mathrm{Im}(zI-L)^\ell$ is closed for all ℓ. Again we use induction on ℓ, except that this time we start the induction at $\ell=0$, with the understanding that $(zI-L)^0=I$.

$I_0=B$ is closed. Suppose by induction that I_ℓ is closed. We have to show that for every sequence of functions $g_n\in(zI-L)I_\ell$, if $g_n\to g$, then $g\in(zI-L)I_\ell$.

Write

$$g_n=(zI-L)f_n,\quad f_n\in I_\ell.$$

We are free to modify f_n by subtracting arbitrary elements of $K_1\cap I_\ell$. For example, we may subtract the closest element to f_n in $K_1\cap I_\ell$ (the closest element exists since $\dim K_1<\infty$ and I_ℓ is closed). Thus we may assume without loss of generality that $\|f_n\|=\mathrm{dist}\,(f_n,K_1\cap I_\ell)$.

We claim that $\sup\|f_n\|<\infty$. Otherwise, $\exists n_k\uparrow\infty$ s.t. $\|f_{n_k}\|\to\infty$, and then $g_{n_k}/\|f_{n_k}\|\to 0$ (because $g_{n_k}\to g$). But

$$\frac{g_{n_k}}{\|f_{n_k}\|}=(zI-L)\frac{f}{\|f_{n_k}\|}$$

so $\exists n_{k_\ell}\uparrow\infty$ s.t. $f_{n_{k_\ell}}/\|f_{n_{k_\ell}}\|\to h$ where $(zI-L)h=0$ (conditional closure lemma). Since $f_n\in I_\ell$ and I_ℓ is closed, $h\in I_\ell$. Thus $f_{n_{k_\ell}}/\|f_{n_{k_\ell}}\| toh\in K_1\cap I_\ell$. But this is impossible, since we have constructed f_n so that $\mathrm{dist}\,(f_n/\|f_n\|,K_1)=1$ for all n. This contradiction shows that $\sup\|f_n\|<\infty$.

Since $\sup\|f_n\|<\infty$, the conditional closure lemma provides a subsequence $n_k\uparrow\infty$ s.t. $f_{n_k}\to f$ where $g=(zI-L)f$. The limit f belongs to I_ℓ, because $f_n\in I_\ell$ and I_ℓ is closed. Thus $g\in(zI-L)I_\ell\equiv I_{\ell+1}$ as required.

We show that $K(z)=K_\ell$ for some ℓ. By definition, $K_1\subseteq K_2\subseteq\cdots$, so if the assertion is false, then $K_1\subsetneq K_2\subsetneq\cdots$. Construct, using the Riesz lemma a sequence of vectors $f_n\in K_n$ s.t. $\|f_n\|=1$ and $\mathrm{dist}\,(f_n,K_{n-1})\geq\frac{1}{2}$. So $\{f_n\}_{n\geq 1}$ is $\frac{1}{2}$–separated.

We claim that for every $m\in\mathbb{N}$, $\{L^m f_n\}_{n\geq 1}$ is $\frac{1}{2}|z|^{m+1}$–separated. To show this we write

$$z^{-m}L^m f_{n+k}-z^{-m}L^m f_n=f_{n+k}-\left[(I-z^{-m}L^m)f_{n+k}+z^{-m}L^m f_n\right],$$

and show that the term in the brackets belongs to K_{n+k-1}. This means that $\|L^m f_{n+k}-L^m f_n\|\geq|z|^m\mathrm{dist}\,(f_{n+k},K_{n+k-1})\geq|z|^{m+1}/2$.

We begin with two trivial observations on K_ℓ. Firstly, $L(K_\ell)\subseteq K_{\ell-1}$ (because $L(zI-L)^\ell=(zI-L)^\ell L$). Secondly, $(zI-L)K_\ell\subseteq K_{\ell-1}$. The first observation shows that $L^m f_n\in K_n$. The second observation shows that

$$(I - z^{-m}L^m)f_{n+k} = \sum_{j=0}^{m-1} z^{-j}L^j(I - z^{-1}L)f_{n+k} \in \sum_{j=0}^{m-1} L^j K_{n+k-1} \subseteq K_{n+k-1}.$$

Thus the term in the brackets belongs to K_{n+k-1}, and $\|L^m f_{n+k} - L^m f_n\| \geq \frac{1}{2}|z|^{m+1}$.

We obtain a contradiction to this fact as follows. Recall that we are assuming that (5.5) holds with $k = 1$. Iterating, we get for all m and $f \in B$,

$$\|L^m f\| \leq r^m \|f\| + R \sum_{j=1}^{m} r^j \|L^{m-j} f\|'.$$

Applying this to $Lf_k - Lf_\ell$ we get

$$\begin{aligned}\|L^{m+1} f_k - L^{m+1} f_\ell\| &\leq r^m \|Lf_k - Lf_\ell\| + R \sum_{j=1}^{m} r^j \|L^{m-j} Lf_k - L^{m-j} Lf_\ell\|.' \\ &\leq 2\|L\| r^m + R \sum_{j=1}^{m} r^j M^{m-j} \|Lf_k - Lf_\ell\|',\end{aligned}$$

Since $\sup \|f_n\| < \infty$, $\exists n_k \uparrow \infty$ s.t. $\|Lf_{n_k} - h\|' \to 0$ for some $h \in B$. This means that for all $\varepsilon > 0$, we can find $\ell \neq k$ so large that

$$\|L^{m+1} f_{n_k} - L^{m+1} f_{n_\ell}\| \leq 2\|L\| r^m + \varepsilon.$$

Choosing m so large that $2\|L\| r^m < \frac{1}{4}|z|^{m+1}$ and $\varepsilon < \frac{1}{4}|z|^{m+1}$, we obtain $n_k \neq n_\ell$ s.t. $\|L^{m+1} f_{n_k} - L^{m+1} f_{n_\ell}\| < \frac{1}{2}|z|^{m+1}$. But this is impossible, because $\{L^m f_n\}_{n \geq 1}$ is $\frac{1}{2}|z|^{m+1}$–separated.

This proves that the sequence $K_1 \subseteq K_2 \subseteq \cdots$ stabilizes eventually. A similar argument, applied to $I_1 \supseteq I_2 \supseteq \cdots$ shows that that sequence also eventually stabilizes (Problem **??**). ♢

Step 2. $LK(z) \subseteq K(z)$, $LI(z) \subseteq I(z)$, and $B = K(z) \oplus I(z)$.

Proof. The first two statements are obvious, we show the third. The previous step shows that for some m, $K(z) = K_\ell, I(z) = I_\ell$ for all $\ell \geq m$, so it's enough to show that $B = K_m \oplus I_m$.

$B = K_m + I_m$: Suppose $f \in B$, then $(zI - L)^m f \in I_m = I_{2m}$ ($\because 2m > m$), so $\exists g \in B$ s.t. $(zI - L)^m f = (zI - f)^{2m} g$. We have $f = [f - (zI - L)^m g] + (zI - L)^m g \in K_m + I_m$.

$K_m \cap I_m = \{0\}$: Suppose $f \in K_m \cap I_m$, then $f = (zI - L)^m g$ for some $g \in B$. Necessarily $(zI - L)^{2m} g = (zI - L)^m f = 0$, so $g \in K_{2m}$. But $K_{2m} = K_m$, so $g \in K_m$. It follows that $f = (zI - L)^m g = 0$. ♢

Step 3. $(zI - L) : I(z) \to I(z)$ is a bijection with bounded inverse.

Proof. Let m be a number s.t. $I(z) = I_m, K(z) = K_m$. $(zI - L)$ is one-to-one on $I(z)$, because $\ker(zI - L) \cap I(z) \subseteq K_1 \cap I_m \subseteq K_m \cap I_m = \{0\}$. $(zI - L)$ is onto $I(z)$, because $(zI - L)I(z) = (zI - L)I_m = I_{m+1} = I_m = I(z)$. Thus $(zI - L) : I(z) \to I(z)$ is a bijection.

Since $I(z)$ is a closed subset of a Banach space, $(I(z), \|\cdot\|)$ is complete. The open mapping theorem says that any bijection between Banach spaces is open, therefore its inverse is continuous, whence bounded. $\diamondsuit$

Step 4. $K(z) = 0$ for all but at most finitely many $z \in A(\rho, \rho(L))$. $K(z) \neq 0$ for at least one z s.t. $|z| = \rho(L)$.

Proof. Suppose by way of contradiction that $K(z) \neq \{0\}$ for infinitely many different points $z_i \in A(\rho, \rho(L))$ $(i \geq 1)$. Since $A(\rho, \rho(L))$ is compact, we may assume w.l.o.g. that $z_n \xrightarrow[n\to\infty]{} z \in A(\rho, \rho(L))$.

Since $K(z_n) \neq 0$, $\ker(z_n I - L) \neq 0$. Let $F_n := \ker(z_1 I - L) \oplus \cdots \oplus \ker(z_n I - L)$, then $F_1 \subsetneq F_2 \subsetneq F_3 \subsetneq \cdots$. We now argue as in step 1. By the Riesz Lemma, $\exists f_n \in F_n$ s.t. $\|f_n\| = 1$ and $\mathrm{dist}(f_n, F_{n-1}) \geq \frac{1}{2}$. Using the decomposition

$$L^m f_{n+k} - L^m f_n = z_{n+k}^m f_{n+k} - z_n^m f_n$$

we see that $\|L^m f_{n+k} - L^m f_n\| \geq \mathrm{dist}(z_{n+k}^m f_{n+k}, F_n) \geq \frac{1}{2}|z_{n+k}|^m \geq \frac{1}{2}\rho^m$. But this leads to a contradiction exactly as in step 1.

Thus $\{z \in A(\rho, \rho(L)) : K(z) \neq 0\}$ is finite. Next we claim that it contains an element on $\{z : |z| = \rho(L)\}$. Otherwise, $\exists \rho' < \rho(L)$ s.t. $K(z) = 0$ for all $|z| \geq \rho'$. This means that $I(z) = B$ for all $|z| \geq \rho'$, whence by the previous step, $(zI - L)$ has a bounded inverse for all $|z| \geq \rho'$. It follows that the spectral radius of L is less than or equal to ρ'. But this is not the case, because $\rho' < \rho(L)$. $\diamondsuit$

Step 5. Let $\lambda_1, \ldots, \lambda_t$ denote the complete list of different eigenvalues of L in $A(\rho, \rho(L))$, then

$$F := \bigoplus_{i=1}^{t} K(\lambda_i)$$

is a direct sum, $\dim F < \infty$, $L(F) \subseteq F$, and the eigenvalues of $L|_F$ are $\lambda_1, \ldots, \lambda_t$.

Proof. Suppose $v_i \in K(\lambda_i) \setminus \{0\}$ and $\sum \alpha_i v_i = 0$. We have to show that $\alpha_j = 0$ for all j. Suppose by way of contradiction that $\alpha_j \neq 0$ for some j.

Find, using step 1, an $m \geq 1$ s.t. $K(\lambda_i) = \ker(\lambda_i I - L)^m$, and set $p_i(z) := (\lambda_i - z)^m$. For every j, let $q_j(z) := \prod_{i \neq j} p_i(z)$, then $q_j(L) v_i = 0$ for all $i \neq j$, and so

$$0 = q_j(L)\left(\sum_i \alpha_i v_i\right) = \alpha_j q_j(L) v_j.$$

Since $\alpha_j \neq 0$, $q_j(L) v_j = 0$.

The polynomials $q_j(z), p_j(z)$ have no zeroes in common, so they are relatively prime. Find polynomials $a(z), b(z)$ s.t. $a(z)p_j(z) + b(z)q_j(z) = 1$. Evaluating the expression on the right at L and applying it to v_j we get that $v_j = 0$, contrary to our assumptions.

Thus the sum defining F is direct. The remaining statements are obvious consequences of steps 1 and 4.

Step 6. $H := \bigcap_{i=1}^{t} I(\lambda_i)$ is closed, $L(H) \subseteq H$, and $B = F \oplus H$.

Proof. H is closed and L–invariant because $I(\lambda_i)$ are closed and L–invariant (step 1).

For every $i = 1, \ldots, t$ $B = K(\lambda_i) \oplus H(\lambda_i)$ (step 2), so there exist projection operators $\pi_i : B \to K(\lambda_i)$ s.t. for every $f \in B$,

$$\pi_i(f) \in K(\lambda_i) \text{ and } (I - \pi_i)(f) \in I(\lambda_i).$$

We have

1. $\pi_i L = L\pi_i$, because $LK(\lambda_i) \subseteq K(\lambda_i), LI(\lambda_i) \subseteq I(\lambda_i)$;
2. $i \neq j \Rightarrow \pi_i \pi_j = 0$: Suppose $u \in B$, and let $v := \pi_j(u)$. Then $v \in K(\lambda_j)$, so $\exists m$ s.t. $(\lambda_j I - L)^m v = 0$. We have

$$(\lambda_j - \lambda_i)^m v + \sum_{k=1}^{m} \binom{m}{k} (\lambda_j - \lambda_i)^k (\lambda_i I - L)^m v = (\lambda_i I - L)^m v = 0,$$

whence $v = -(\lambda_j - \lambda_i)^{-m} \sum_{k=1}^{m} \binom{m}{k} (\lambda_j - \lambda_i)^k (\lambda_i I - L)^{m-k} v$. Iterating this identity we see that for every ℓ

$$v = \left[-(\lambda_j - \lambda_i)^{-m} \sum_{k=1}^{m} \binom{m}{k} (\lambda_j - \lambda_i)^k (\lambda_i I - L)^{m-k} \right]^{\ell} v \in \mathrm{Im}(\lambda_i I - L)^{\ell},$$

whence $v \in I(\lambda_i) \subseteq \ker \pi_i$. It follows that $(\pi_i \circ \pi_j)(u) = \pi_i(v) = 0$.

We can now show that $B = F \oplus H$. Every $f \in B$ can be decomposed into

$$\sum_{i=1}^{t} \pi_i(f) + \left(f - \sum_{i=1}^{t} \pi_i(f) \right).$$

The left summand is in F, the right summand is in $\bigcap_{i=1}^{t} \ker \pi_i = \bigcap_{i=1}^{n} I(\lambda_i) = H$. Thus $B = F + H$. On the other hand $F \cap H = \{0\}$, because if $f \in F \cap H$, then $\pi_i(f) = 0$ for all i (because $f \in H$), whence $f = 0$ (because $f \in F$) $\diamondsuit$

Step 7. The spectral radius of $L|_H$ is less than or equal to ρ.

Proof. It is enough to show that $(zI - L) : H \to H$ has a bounded inverse for all $|z| \geq \rho$. Fix such a z, and let h be some element of H.

Suppose $z \notin \{\lambda_1, \ldots, \lambda_t\}$, then $K(z) = \{0\}$ so $I(z) = B$. By step 3, $(zI - L) : B \to B$ is invertible with bounded inverse.

Now suppose $z = \lambda_i$ for some i. Recall that $(\lambda_i I - L) : I(\lambda_i) \to I(\lambda_i)$ is an isomorphism, so $\exists! f \in I(\lambda_i)$ s.t. $h = (\lambda_i I - L) f$. We show that f belongs to H, by checking that $\pi_j(f) = 0$ for all j. This is clear for $j = i$ (because $f \in I(\lambda_i) = \ker \pi_i$). If $j \neq i$, then

$$0 = \pi_j(h) = \pi_j(\lambda_i I - L) f = (\lambda_i I - L)\pi_j(f),$$

so $\pi_j(f) \in K(\lambda_i) \cap K(\lambda_j) = \{0\}$. Thus $f \in \bigcap \ker \pi_j = H$. We see that $\exists! f \in H$ s.t. $h = (zI - L) f$. It follows that $(zI - L) : H \to H$ is invertible. Since H is closed, H is a Banach space and the inverse mapping theorem says that the inverse of $(zI - L)$ must be bounded. $\diamondsuit$

In summary, $B = F \oplus H$ where F, H are L–invariant spaces such that (a) F is finite dimensional, (b) H is closed, (c) all the eigenvalues of $L|_F$ have modulus larger than or equal to ρ, and (d) the spectral radius of $L|_H$ is less than or equal to ρ. □

5.3 Implications of spectral gaps

In this section we discuss several applications of spectral gaps to the theory of Gibbs measures.

5.3.1 Exponential decay of correlations

Suppose X, Y are two measurable random variables on a general probability space $(\Omega, \mathscr{F}, \mu)$. Suppose X, Y have finite variance. The *covariance* of X, Y is

$$\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}(x))(Y - \mathbb{E}(Y))] \equiv \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

The *correlation coefficient* of X, Y is

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbb{E}((X - \mathbb{E}(X)^2)\mathbb{E}((Y - \mathbb{E}(Y)^2)}}.$$

Note that if X and Y have a tendency to deviate from their means in the same direction, then $\mathrm{Cov}(X,Y)$ would tend to be positive, if X, Y tend to deviate from their means in opposite directions then $\mathrm{Cov}(X,Y)$ would tend to be negative, and if there is little correlation between the directions of deviations then $|\mathrm{Cov}(X,Y)|$ would tend to be small. Thus $\mathrm{Cov}(X,Y)$ contains information on the correlation between X and Y. The correlation coefficient is just a normalization of the covariance which makes it take values in $[-1,1]$ (Cauchy-Schwarz).

Theorem 5.9 (Ruelle). *Suppose m is a Gibbs measure for a locally Hölder continuous function $\phi : X \to \mathbb{R}$ with finite pressure, on a topologically mixing TMS with the BIP property. Let B denote the Banach space in (5.6). For every $f, g \in B$,*

$$\mathrm{Cov}(f, g \circ T^n) \xrightarrow[n\to\infty]{} 0 \quad \textit{exponentially fast}.$$

Proof. W.l.o.g. that $P_G(\phi) = 0$. Write $m = h\nu$, where h, ν are the eigenfunctions of Ruelle's operator s.t. $\int h d\nu = 1$. Since X has the BIP property, h is bounded away from zero and infinity, so $\nu(X) < \infty$ we normalize ν and h to have $\nu(X) = 1$. The transfer operator of m is $\widehat{T} : f \mapsto h^{-1} L_\phi(fh)$.

It is enough to prove the theorem for f, g with zero intergals. For such functions

$$\begin{aligned}\mathrm{Cov}(f, g\circ T^n) &= \int fg\circ T^n dm = \int h^{-1}L_\phi^n(fh)gdm = \int L_\phi^n(fh)gd\nu \\ &\le \|L_\phi^n(fh)\|_\infty \|g\|_\infty \le \|L_\phi^n(fh)\|\|g\|_\infty.\end{aligned}$$

We saw in the course of the proof of theorem 5.8 that $h \in B$. It is not difficult to see $\|fh\| \le \|f\|\|h\| < \infty$, so $fh \in B$. Since $L_\phi : B \to B$ is quasi–compact with no eigenvalues on $\{z : |z| = 1\}$ other than a simple eigenvalue at one, we can write

$$L_\phi = \pi + N$$

where $\pi : f \mapsto h\int f d\nu$ is the eigenprojection of $\lambda = 1$, and N is an operator s.t. $N\pi = \pi N = 0$ and $\rho(N) < 1$. Fix some $\rho(N) < \kappa < 1$, then $\|N^n\| = O(\kappa^n)$, and so

$$\|L_\phi^n(fh)\| = \|\pi(fh) + N^n(fh)\| = O(\kappa^n)\|fh\|,$$

because $\pi(fh) = h\int fhd\nu = h\int fdm = 0$. We see that $|\mathrm{Cov}(f, g\circ T^n) = O(\kappa^n)$. □

5.3.2 The central limit theorem

The central limit theorem is a statement on convergence in distribution to the normal (or Gaussian) distribution. We review these probabilistic notions.

Suppose $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space. A (real valued) *random variable* on Ω is an $\mathscr{F}$–measurable function $X : \Omega \to \mathbb{R}$. The *distribution function* of X is

$$F_X(t) := \mathbb{P}[X < t].$$

The following properties are automatic:

1. F_X is non–decreasing and takes values in $[0,1]$;
2. F_X is continuous from the left: $F_X(t) \xrightarrow[t\to a^-]{} F_X(a)$;
3. $F_X(t) \xrightarrow[t\to-\infty]{} 0$, $F_X(t) \xrightarrow[t\to\infty]{} 1$

Any function $F : \mathbb{R} \to [0,1]$ like that is called a *distribution function*.[7]

Definition 5.7. Let X_n, Y be real valued random variables (possibly on different probability spaces). We say that X_n *converges in distribution* to Y, and write $X_n \xrightarrow[n\to\infty]{\text{dist}} Y$, if $\mathbb{P}[X_n < t] \xrightarrow[n\to\infty]{} \mathbb{P}[Y < t]$ for all t where $F_Y(\cdot)$ is continuous.

(The reason we only ask for $F_{X_n}(t)$ to converge to $F_Y(t)$ at points where $F_Y(\cdot)$ is continuous, is to deal with examples such as $X_n \equiv 2 - \frac{1}{n}, Y \equiv 2$. We would like to say that $X_n \xrightarrow[n\to\infty]{\text{dist}} Y$, eventhough $F_{X_n}(2) = 1 \not\to 0 = F_Y(2)$.)

[7] Some people prefer to define the distribution function of a random variable by the formula $F_X(t) := \mathbb{P}[X \le t]$, and then replace the requirement of continuity from the left by the requirement of continuity from the right.

The *normal distribution* (or *Gaussian distribution*) with *mean* $\mu \in \mathbb{R}$ and *variance* $\sigma^2 > 0$ is the following distribution function:

$$N(\mu,\sigma^2)(t) := \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} e^{-(t-\mu)^2/2\sigma^2} dt.$$

The *degenerate normal distribution* with mean μ is $N(\mu,0)(t) := 1_{(-\infty,\mu]}(t)$. A random variable with normal distribution is called *Gaussian*. A degenerate Gaussian random variable is constant almost everywhere.

Theorem 5.10 (Central Limit Theorem). *Suppose m_ϕ is the Gibbs measure of a locally Hölder continuous function ϕ on a topologically mixing TMS, and suppose $\psi : X \to \mathbb{R}$ is a (bounded) Hölder continuous function, then*

$$\frac{1}{\sqrt{n}}\left(\sum_{k=0}^{n-1} \psi \circ T^k - n\int \psi dm\right) \xrightarrow[n\to\infty]{dist} N(0,\sigma^2) \text{ with respect to } m_\phi$$

where $\sigma^2 \geq 0$ is given by $\sigma^2 := \frac{d^2}{dt^2}\Big|_{t=0} P_G(\phi + t\psi)$.

To prove such a result, it is convenient to use *characteristic functions*: The characteristic function of a random variable $X : \Omega \to \mathbb{R}$ is

$$f_X(t) := \mathbb{E}(e^{itX}).$$

For example, the characteristic function of $N(\mu,\sigma^2)$ is $g(t) = \exp(i\mu - \frac{1}{2}\sigma^2 t^2)$. Characteristic functions are useful for studying convergence in distribution, because of the following theorem:

Theorem 5.11 (Lévy's Continuity Theorem). *A sequence of random variables X_n converges in distribution to a random variable Y iff the characteristic functions of X_n converge pointwise to the characteristic function of Y.*

Thus $X_n \xrightarrow[n\to\infty]{\text{dist}} N(0,\sigma^2)$ iff $f_{X_n}(t) \xrightarrow[n\to\infty]{} \exp(-\frac{1}{2}\sigma^2 t^2)$ pointwise. We will not give a proof of Lévy's continuity theorem as stated above. Rather, we will prove a more special result, which assumes more (the limiting distribution function needs to have a bounded derivative) and which says more (rate of convergence):

Proposition 5.3. *There exists a constant $C > 0$ s.t. for every pair of real valued random variables X,Y with distribution functions F,G and characteristic functions f,g, if G is differentiable and $\int |F-G|dx < \infty$, then for all $x \in \mathbb{R}$ and $T > 0$,*

$$|F(x) - G(x)| \leq C\left(\frac{1}{2\pi}\int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|} dt + \frac{\sup|G'|}{T}\right).$$

In particular, if X_n are bounded random variables with distribution functions F_n and characteristic functions f_n, and if $f_n \to f$ =characteristic function of the Gaussian random variable $N(0,\sigma^2)$, then for every T

$$|F_n(x) - N(0,\sigma^2)(x)| \leq C\left(\frac{1}{2\pi}\int_{-T}^{T}\frac{|f_n(t)-f(t)|}{|t|}dt + \frac{1}{T}\right).$$

Characteristic functions are uniformly bounded (by one), so the integral term tends to zero by the bounded convergence theorem. Since T can be taken to be arbitrarily large, this is enough to deduce that $F_n(x) \to N(0,\sigma^2)(x)$ for all x.

If one has more information on the convergence $f_n \to f$, then one can hope to construct $T_n \to \infty$ such that $\int_{-T_n}^{T_n} \frac{|f_n(t)-f(t)|}{|t|}dt \to 0$, and then obtain a rate of convergence (Berry–Eseen).

Proposition 5.3 is pure Fourier analysis, so we delegate its proof to a separate section, and proceed to prove the CLT accepting it as given.

Proof of Theorem 5.10. We assume w.l.o.g. that $P_G(\phi) = 0$ and $\sum_{Ty=x} e^{\phi(y)} = 1$. In this case the transfer operator of m_ϕ is L_ϕ, and $L_\phi^* m_\phi = m_\phi$. We may also assume w.l.o.g. that $\int \psi dm_\phi = 0$.

Let $\psi_n := \psi + \psi\circ T + \cdots + \psi \circ T^{n-1}$. The characteristic function of ψ_n is

$$\mathbb{E}(e^{it\psi_n}) = \int e^{it\psi_n} dm_\phi = \int L_\phi^n(e^{it\psi_n})dm_\phi = \int L^n_{\phi+it\psi} 1 dm_\phi.$$

We shall study this expression using the spectral properties of $L_\phi : B \to B$, where $B := \{f : X \to \mathbb{C} : \|f\| := \|f\|_\infty + \mathrm{Lip}_\theta(f) < \infty\}$, and $0 < \theta < 1$ be the number such that $\mathrm{var}_n\phi, \mathrm{var}_n\psi < A\theta^n$ for some $A > 0$ and all $n \geq 1$.

Step 1. Suppose ψ is a (bounded) Hölder continuous function, then $z \mapsto L_z := L_{\phi+iz\psi}$ is holomorphic map from a neighborhood of zero to $\mathrm{Hom}(B)$.

For every $f \in B$, $L_z f = L_\phi(e^{iz\psi}f) = \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} L_\phi(\psi^n f)$, so

$$L_z = \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} L_\phi M^n,$$

where $M : B \to B$ is the operator $Mf = \psi f$.

It is easy to check that $(B, \|\cdot\|)$ is a Banach algebra: $\|fg\| \leq \|f\|\|g\|$. Thus $\|M\| \leq \|\psi\|$ and so $\|L_\phi M^n\| \leq \|L_\phi\|\|M\|^n$. Thus the above series expansion converges on $\{z : |z| < 1/\|\psi\|\}$. Since L_z admits a power series expansion on a neighborhood of zero, it is holomorphic there. ♢

Step 2. Application of spectral gap and analytic perturbation theory.

The space B has the property that $L_\phi : B \to B$ has spectral gap: $\exists 0 < \kappa < 1$ s.t.

$$\mathrm{Spect}(L_\phi) = \{1\} \cup \text{compact subset of } \{z : |z| < \kappa\},$$

where 1 is a simple eigenvalue. By analytic perturbation theory, there exists $\kappa < \kappa_0 < \kappa_1 < 1$ and $\varepsilon > 0$ s.t. for every $|z| < \varepsilon$,

$$\mathrm{Spect}(L_z) = \{\lambda_z\} \cup \text{compact subset of } \{z : |z| < \kappa_0\},$$

where λ_z is a simple eigenvalue s.t. $|\lambda_z| > \kappa_1$. Moreover, λ_z and its eigenprojection P_z are holomorphic on $\{z : |z| < \varepsilon\}$. For such z we can write

$$L_z = \lambda_z P_z + N_z,$$

where $N_z := L_z(I - P_z)$ has spectral radius smaller than κ_0, and $P_z N_z = N_z P_z = 0$.

Note that $\lambda_0 = 1$ (the leading eigenvalue of L_ϕ), and P_0 is the operator $P_0 f = \int f dm_\phi$ (the eigenprojection of this eigenvalue).

We see that

$$\begin{aligned}\mathbb{E}(e^{it\psi_n}) &= \int L_t^n 1 dm_\phi = \int (\lambda_t^n P_t 1 + N_t^n 1) dm_\phi \\ &= \lambda_t^n \int [P_0 1 + (P_t - P_0)1 + \lambda_t^{-n} N_t^n 1] dm_\phi \\ &= \lambda_t^n [1 + \varepsilon_n(t)], \text{ where } \varepsilon_n(t) := \int (P_t - P_0) 1 dm_\phi + \lambda_t^{-n} \int N_t^n 1 dm_\phi.\end{aligned}$$

Since $t \mapsto P_t$ is continuous and $|\lambda_t| > \kappa_1 > \kappa_0 > \rho(N_t)$,

$$\varepsilon_n(t) = O(\|P_t - P_0\|) + O(\kappa_1^{-n}\kappa_0^n) \xrightarrow[n\to\infty, t\to 0]{} 0.$$

Thus $\mathbb{E}(e^{it\psi_n/\sqrt{n}}) = \lambda_{t/\sqrt{n}}^n [1 + \varepsilon_n(\frac{t}{\sqrt{n}})] = \lambda_{t/\sqrt{n}}^n [1 + o(1)]$. $\diamondsuit$

Step 3. Finding the asymptotic behavior of $\lambda_{t/\sqrt{n}}^n$ as $n \to \infty$.

Since $z \mapsto \lambda_z$ is holomorphic on a neighborhood of zero,

$$\lambda_t = a + bt + ct^2 + o(t^2) \text{ as } t \to 0.$$

We find the coefficients a, b, c. To find a, we substitute $t = 0$ to see that $a = \lambda_0 = 1$. To find $b = \frac{d}{dt}|_{t=0}\lambda_t$, we start from the identity $L_t P_t = \lambda_t P_t$ and differentiate:

$$L_t' P_t + L_t P_t' = \lambda_t' P_t + \lambda_t P_t'.$$

Now apply P_t on the left: $P_t L_t' P_t + \lambda_t P_t P_t' = \lambda_t' P_t^2 + \lambda_t P_t P_t'$. Since $P_t^2 = P_t$ this gives upon cancelation and substitution $t = 0$

$$P(L_t'\big|_{t=0})P = bP.$$

The series expansion we found above for L_z shows that $L_t'|_{t=0} = iL_\phi M$, so

$$b = PL_\phi(\psi P1) = \int L_\phi \psi dm_\phi = \int \psi dm_\phi = 0.$$

The coefficient c does not have such a simple formula. But we can at least determine its sign: Let $\Lambda(x) := \lambda_{-ix}$, where x is real. This the leading eigenvalue of $L_{\phi+x\psi}$, so $\Lambda(x) = \exp P_G(\phi + x\psi)$. Since $P_G(\cdot)$ and $\exp(\cdot)$ are convex, $\Lambda(x)$ is convex. Therefore $\Lambda''(0) \geq 0$. But $\Lambda''(0) = (-i)^2 \lambda_t''|_{t=0} = -c$, so $c \leq 0$. We write $c = -\frac{1}{2}\sigma^2$ for some $\sigma \geq 0$.

Putting this altogether we have that $\lambda_t = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$, whence

$$\lambda^n_{t/\sqrt{n}} = \left(1 - \frac{1}{2n}\sigma^2 t^2 + o(\tfrac{1}{n})t^2\right)^n = \exp\left[n \log\left(1 - \frac{1}{2n}\sigma^2 t^2 + o(\tfrac{1}{n})t^2\right)\right]$$
$$= \exp\left[n\left(-\frac{1}{2n}\sigma^2 t^2 + o(\tfrac{1}{n})t^2\right)[1+o(1)]\right] \quad (\because \log(1-x) = x[1+o(1)]).$$

Fixing t and passing to the limit as $n \to \infty$, we see that $\lambda^n_{t/\sqrt{n}} \to \exp(-\frac{1}{2}\sigma^2 t^2)$. $\diamondsuit$

These steps show that $\mathbb{E}(e^{it\psi_n/\sqrt{n}}) \xrightarrow[n\to\infty]{} \exp(-\frac{1}{2}\sigma^2 t^2)$, the characteristic function of $N(0,\sigma^2)$. So the CLT is proved. $\square$

Remark 1: By expanding λ_t to the third degree, we can obtain information on the rate of convergence in the CLT.

Remark 2: It is important to know when is the limit non–degenerate, i.e. when does $\sigma \neq 0$. The following can be shown, but we do not have time to do this:

Theorem 5.12. *Under the assumption of the previous theorem, $\sigma \neq 0$ iff ψ is not cohomologous via a continuous transfer function to a constant.*

Remark 3: There is no explicit formula for σ, but the following identities are known.

Theorem 5.13. *Under the assumptions of the previous theorem,*

1. Asymptotic variance formula*: $\sigma^2 = \lim_{n\to\infty} \frac{1}{n} Var(\psi_n)$ w.r.t. m_ϕ*
2. Linear response formula*: $\sigma^2 = \frac{d^2}{dt^2}\Big|_{t=0} P_G(\phi + t\psi)$*
3. Green–Kubo formula*: Suppose $\int \psi dm_\phi = 0$, then (w.r.t. m_ϕ),*

$$\sigma^2 = Var(\psi) + 2\sum_{k=1}^{\infty} Cov(\psi, \psi \circ T^k)$$

Here $Cov(f,g) = \int fg dm_\phi - \int f dm_\phi, \int g dm_\phi$ and $Var(f) = Cov(f,f)$.

The "linear response formula" originates from statistical physics. There first partial derivatives of the free energy function are interpreted as thermodynamic quantities, and second order partial derivatives are called "linear response functions", because they measure the rate of change in a thermodynamic quantity when one of the parameters of the systems is varied linearly.

The "Green–Kubo formula" originates from the theory of diffusion processes. It is a statement about the relation between the diffusion coefficient and the rate of mixing.

5.3.3 Proof of proposition 5.3

We first recall some standard facts on Lebesgue–Stieltjes integrals. Any distribution function F determines a unique Borel probability measure on $\mathbb{R}$ by $\mu_F([a,b)) := F(b) - F(a)$. This is called the *Lebesgue–Stieltjes measure* of F. It is customary to use the following notation

$$\int_a^b f(x)F(dx) \text{ or } \int_a^b f(x)dF(x) \text{ for } \int_{[a,b)} f d\mu_F.$$

Note that the right endpoint of the interval is not included. (This matters when $F(x)$ has a jump discontinuity at b, because in this case μ_F has an atom at b.)

The *Fourier transform* of a function $f \in L^1(\mathbb{R})$ is

$$\mathfrak{F}(f)(t) = \int e^{itx} f(x)dx$$

This has the following properties:

1. $\mathfrak{F}(\mathfrak{F}(f)) = 2\pi f$
2. $\mathfrak{F}(f * g) = \mathfrak{F}(f) \cdot \mathfrak{F}(g)$, where $(f * g)(x) = \int f(x-y)g(y)dy$ (the *convolution*).

The Fourier transform of a Borel probability measure μ on $\mathbb{R}$ is the function $(\mathfrak{F}\mu)(t) := \int e^{itx} d\mu(x)$.

The reader can check that characteristic function of a random variable X is the Fourier transform of the Stieltjes measure of the distribution function of X:

$$f_X(t) := \int e^{itx} d\mu_{F_X}(x).$$

This only depends on the distribution function of X. Therefore we can safely speak of the characteristic function of a distribution function.

Lemma 5.2. *Suppose $F(x), G(x)$ are two distribution functions with characteristic functions $f(t), g(t)$. If $\int |F(x) - G(x)|dx < \infty$, then*

$$[\mathfrak{F}(F-G)](t) = -\frac{f(t) - g(t)}{it}.$$

Proof. First note that the Fourier transform of $F - G$ exists, because $F - G \in L^1$ by assumption. Let μ_F and μ_G denote the Lebesgue–Stieltjes measures of F, G, then

$$\begin{aligned}
[\mathfrak{F}(F-G)](t) &= \lim_{T\to\infty}\int_{-T}^{T} e^{itx}[F(x)-G(x)]dx \\
&= \lim_{T\to\infty}\left[\int_{-T}^{T}\int_{-\infty}^{T} e^{itx}1_{[\xi<x]}d\mu_F(\xi)dx - \int_{-T}^{T}\int_{-\infty}^{T} e^{itx}1_{[\xi<x]}d\mu_G(\xi)dx\right] \\
&= \lim_{T\to\infty}\int_{-\infty}^{T}\left(\int_{\xi}^{T} e^{itx}dx\right)d\mu_F(\xi) - \lim_{T\to\infty}\int_{-\infty}^{T}\left(\int_{\xi}^{T} e^{itx}dx\right)d\mu_G(\xi) \\
&= \lim_{T\to\infty}\int_{-\infty}^{T}\frac{e^{itT}-e^{it\xi}}{it}d\mu_F(\xi) - \lim_{T\to\infty}\int_{-\infty}^{T}\frac{e^{itT}-e^{it\xi}}{it}d\mu_G(\xi) \\
&= \lim_{T\to\infty}\left[\frac{e^{itT}}{it}[F(T)-G(T)] - \int_{-\infty}^{T}\frac{e^{it\xi}}{it}d(\mu_F-\mu_G)(\xi)\right].
\end{aligned}$$

The first summand tends to zero (because $F(T), G(T) \xrightarrow[T\to\infty]{} 1$), and the second summand tends to $-\frac{f(t)-g(t)}{it}$, where f, g are the characteristic functions of the distributions F, G. □

Lemma 5.3. *There exists a non–negative absolutely integrable function $H(x)$ s.t.*

1. *H is even;*
2. *$\int H(x)dx = 1$;*
3. *$b := \int |x|H(x)dx < \infty$;*
4. *$H(x) \xrightarrow[|x|\to\infty]{} 0$;*
5. *The Fourier transform of $H(x)$ is real–valued, non–negative, and supported inside $[-1,1]$.*

Proof. There are many possible constructions. Here is one. Start with the indicator of of a symmetric interval $[-a,a]$, and take its Fourier transform

$$H_0(y) = \int_{-a}^{a} e^{ity}dt = \frac{2\sin ay}{y}.$$

The Fourier transform of H_0 is $\mathfrak{F}H_0 = 2\pi 1_{[-a,a]}$, so it has compact support. But H_0 is not non-negative, and $\int |x|H_0(x)dx = \infty$. To correct this we let $H_1(x) := (H_0(x))^4$, and observe that $H_1(x) \geq 0$ and $\int |x|H(x)dx < \infty$. The Fourier transform of H_1 still has compact support (in $[-4a, 4a]$), because

$$\mathfrak{F}[(H_0)^4] = \mathfrak{F}[(\mathfrak{F}1_{[-a,a]})^4] = \mathfrak{F}\{\mathfrak{F}[(1_{[-a,a]} * 1_{[-a,a]} * 1_{[-a,a]} * 1_{[-a,a]}]\} = (1_{[-a,a]})^{*4},$$

and the convolution of functions with compact support has compact support. H_1 is even, because it is the convolution of even functions. It remains to normalize H_1 to have integral equal to one. □

Proof of Proposition 5.3 Let $H(x)$ be the function given by the lemma, and let $h := \mathfrak{F}H$. Set $H_T(x) := TH(Tx)$, then $H_T(x)$ is an even non–negative absolutely integrable function s.t.

1. $\int H_T dx = 1$;

2. $\int |x| H_T(x) dx = b/T$;
3. The Fourier transform of H_T is $h_T(t) := h(t/T)$ where $h = \mathfrak{F}H$.

Note that h_T is supported in $[-T,T]$, and $|h_T| \leq \|H_T\|_1 = 1$.

The proof is based on the following heuristic: For T large, $H_T(x)$ has a sharp peak at $x = 0$, and rapid decay for x far from zero. If we average a "nice" function $\varphi(y)$ with weights $H_T(x-y)$, then we expect the result to be close to $\varphi(x)$, in particular we expect $|F(x) - G(x)| \overset{?}{\approx} |\text{average of } F(y) - G(y) \text{ with weights } H_T(x-y)|$. This average is the integral

$$I_T(x) := \int H_T(x-y)[F(y) - G(y)]dy.$$

The proof consists of (a) estimating $I_T(x)$ in terms of $f(t), g(t)$, and (b) relating $M := \sup|F(x) - G(x)|$ to the value of $I_T(\cdot)$ at a point where $|F(x) - G(x)|$ is nearly maximal.

Step 1. $I_T(x) \leq \frac{1}{2\pi} \int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|} dt$.

Proof:

$$\begin{aligned}
I_T(x) &= \left| \int H_T(x-y)[F(y) - G(y)]dy \right| = |H_T * (F-G)| \\
&= (2\pi)^{-1} \left| \mathfrak{F}^2[H_T * (F-G)] \right| = 2\pi |\mathfrak{F}[\mathfrak{F}H_T \cdot \mathfrak{F}(F-G)]| \\
&= (2\pi)^{-1} |\mathfrak{F}[h_T \cdot \mathfrak{F}(F-G)]| \\
&= (2\pi)^{-1} \left| \int_{-\infty}^{\infty} e^{itx} h_T(t) \frac{f(t)-g(t)}{it} dt \right| \quad \text{(lemma 5.2)} \\
&\leq \frac{1}{2\pi} \int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|} dt, \qquad (5.10)
\end{aligned}$$

because $|h_T(t)| \leq \|H_T\|_1 = 1$ and h_T is supported in $[-T,T]$. $\diamondsuit$

Step 2. Relating $\sup|F(x) - G(x)|$ to $I_T(x_0)$ at a point x_0 where $|F(x_0) - G(x_0)|$ is nearly maximal.

Let $A := \sup|G'(x)|$ and $M := \sup|F(x) - G(x)|$. Fix some point $x_0 \in \mathbb{R}$ s.t. $M_0 := |F(x_0) - G(x_0)| > \frac{1}{2}M$. Since we are free to translate the distributions F, G by the same amount, we may assume w.l.o.g. that $x_0 = 0$. So $M_0 = |F(0) - G(0)|$ and

$$I_T(x_0) = I_T(0) = \int H_T(y)[F(y) - G(y)]dy.$$

(we have used the fact that H_T is even).

Suppose first $F(0) > G(0)$, and decompose the integral $I_T(0)$ into $\int_0^{M_0} + \int_{-\infty}^0 + \int_{M_0}^{\infty}$.

1. To analyze $\int_0^{M_0}$ we note that if $y \in [0, M_0]$, then $[F(y) - G(y)] - [F(0) - G(0)] \geq G(0) - G(y) = -\int_0^y G'(y)dy \geq -Ay$. Thus $[F(y) - G(y)] \geq [F(0) - G(0)] - Ay = M_0 - Ay$ (because $F(0) > G(0)$), whence

$$\int_0^{M_0} H_T(y)[F(y)-G(y)]dy \geq \int_0^{M_0} (M_0 - Ay)H_T(y)dy.$$

2. We estimate $\int_{-\infty}^{0}$ from below by replacing $[F(y)-G(y)]$ by $-M > -2M_0$:

$$\int_{-\infty}^{0} H_T(y)[F(y)-G(y)]dy \geq -\int_{-\infty}^{0} H_T(y)\cdot 2M_0 dy$$

3. Similarly, $\int_{M_0}^{\infty} H_T(y)[F(y)-G(y)]dy \geq -\int_{M_0}^{\infty} H_T(y)\cdot 2M_0 dy$.

Putting this all together, we obtain

$$\begin{aligned}
I_T(0) &\geq \int_0^{M_0} (M_0 - Ay)H_T(-y)dy - \int_{-\infty}^{0} 2M_0 H_T(-y)dy - \int_{M_0}^{\infty} 2M_0 H_T(y)dy \\
&= \int_0^{M_0} (3M_0 - Ay)H_T(y)dy - M_0 \\
&\geq 3M_0 \int_0^{M_0} H_T(y)dy - A\int |y|H_T(y)dy - M_0 \\
&= -M_0 + 3M_0 \int_0^{M_0} H_T(y)dy - \frac{Ab}{T} \quad (\because \int |y|H_T(y)dy = \frac{1}{T}\int |y|H(y)dy = \frac{b}{T}) \\
&= -M_0 + \frac{3M_0}{2}\int_{-M_0}^{M_0} H_T(y)dy - \frac{Ab}{T}
\end{aligned}$$

In summary $M_0[\frac{3}{2}\int_{-M_0}^{M_0} H_T(y)dy - 1] \leq I_T(0) + \frac{Ab}{T}$.

Fix some $\sigma > 0$ s.t. $\int_{-\sigma}^{\sigma} H(y)dy = \frac{8}{9}$, then $\int_{-\sigma/T}^{\sigma/T} H_T(y)dy = \frac{8}{9}$. It is no problem to choose H from the beginning in such a way that $\sigma < A$.

There are two cases:

1. $M_0 \leq \frac{\sigma}{T}$, and then $M \leq 2\sigma/T < 2A/T$;
2. $M_0 > \frac{\sigma}{T}$, and then $\frac{3}{2}\int_{-M_0}^{M_0} H(y)dy - 1 > \frac{1}{3}$, so $M_0 \leq 3I_T(0) + \frac{Ab}{T}$.

In both cases, this and step 1 yields

$$\sup|F(x)-G(x)| < 2M_0 \leq 6\left(\frac{1}{2\pi}\int \frac{|f(t)-g(t)|}{|t|}dt + \frac{\max\{b,2\}A}{T}\right),$$

and the proposition is proved.

This proof was done under the assumption that $F(0) > G(0)$. If $F(0) \leq G(0)$, then we repeat the same procedure, but with the decomposition $\int_{-M_0}^{0} + \int_{-\infty}^{-M_0} + \int_0^{\infty}$. This leads to

$$\int H_T(y)[G(y)-F(y)]dy \geq \int_{-M_0}^{0} (3M_0 - A|y|)H_T(y)dy - M_0.$$

From this point onward, the proof continues as before. $\square$

References

1. Aaronson, Jon; Denker, Manfred: *Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps.* Stoch. Dyn. **1** (2001), no. 2, 193237.
2. Baladi, Viviane: Positive transfer operators and decay of correlations. *Advanced Series in Nonlinear Dynamics,* **16***. World Scientific Publishing Co., Inc., River Edge, NJ, 2000. x+314 pp.*
3. Breiman, Leo: Probability. *Corrected reprint of the 1968 original. Classics in Applied Mathematics, 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. xiv+421 pp.*
4. Doeblin, Wolfgang; Fortet, Robert: *Sur des chaînes á liaisons complétes.* (French) Bull. Soc. Math. France **65** (1937), 132148.
5. Dunford, Nelson; Schwartz, Jacob T.: Linear operators. Part I. General theory. With the assistance of William G. Bade and Robert G. Bartle. Reprint of the 1958 original. *Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988. xiv+858 pp.*
6. Gnedenko, B. V.; Kolmogorov, A. N.: Limit distributions for sums of independent random variables. *Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954. ix+264 pp*
7. Guivarc'h, Y.; Hardy, J.: *Théorémes limites pour une classe de chanes de Markov et applications aux difféomorphismes d'Anosov.* Ann. Inst. H. Poincar Probab. Statist. **24** (1988), no. 1, 7398.
8. Hennion, Hubert; Hervé, Loïc: Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness. *Lecture Notes in Mathematics,* **1766***. Springer-Verlag, Berlin, 2001. viii+145 pp.*
9. Hennion, Hubert: *Sur un théoréme spectral et son application aux noyaux lipchitziens.* (French) Proc. Amer. Math. Soc. **118** (1993), no. 2, 627634.
10. Rousseau-Egele, J.: *Un théoréme de la limite locale pour une classe de transformations dilatantes et monotones par morceaux.* Ann. Probab. **11** (1983), no. 3, 772788.
11. Ionescu Tulcea, C. T.; Marinescu, G.: *Théorie ergodique pour des classes d'opérations non complétement continues.* (French) Ann. of Math. (2) **52**, (1950). 140147.
12. Kato, Tosio: Perturbation theory for linear operators. Reprint of the 1980 edition. *Classics in Mathematics. Springer-Verlag, Berlin, 1995. xxii+619 pp.*
13. Parry, William; Pollicott, Mark: *Zeta functions and the periodic orbit structure of hyperbolic dynamics.* Astérisque No. **187-188** (1990), 268 pp.
14. D. Ruelle: Thermodynamic Formalism (The Mathematical Structures of Classical Equilibrium Statistical Mechanics),*Encyclopedia of Mathematics and Its Applications, Vol.* **5** *(Addison-Wesley, 1978).*

Index