Thermodynamic formalism for countable Markov shifts

Omri M. Sarig

Abstract. We give an overview of the thermodynamic formalism for countable Markov shifts, and indicate applications to surface diffeomorphisms.

1. What is "thermodynamic formalism"?

1.1. Introduction. The ergodic properties of a dynamical system depend on the choice of invariant measure. If a dynamical system has several different ergodic invariant measures, which is often the case, then the question arises which measure to choose to analyze the system.

Physicists working in statistical mechanics encounter a similar problem. Thermodynamic quantities are weighted averages of microscopically defined quantities. Which weighting scheme to use?

Gibbs and Boltzmann found the weighting schemes ("ensembles") which reproduce empirical observations. Sinai and Ruelle imported these schemes to dynamics.[1] They showed that some of the most important invariant measures in smooth ergodic theory can be viewed as analogues of the Gibbs canonical ensemble, and that certain laws in statistical physics translate to mathematical theorems on the properties of these measures.

Sinai called such measures *Gibbs measures*, and Ruelle called the entire program *thermodynamic formalism.*

The word "formalism" is appropriate, since it is just the *formal* aspects of equilibrium statistical physics that is important. The physics itself does not play a role. True enough, thermodynamic formalism studies "Gibbs measures" associated to "interaction potentials." But the theory usually makes no assumptions on the functional form of the potential, which is where the physical content lies. As a result, the theory applies to many problems in geometric measure theory, Riemannian geometry, and number theory, whose natural "interaction potentials" have nothing to do with real physical interactions.

The beauty of thermodynamic formalism is that its language allows for making conjectures (which often turn out to be correct) in contexts far removed from physics, by following analogies with the physical world.

2000 *Mathematics Subject Classification.* Primary .

O.S. is supported by the European Science Foundation, ERC-2009-StG 239885.

[1]Notes and historical references are collected at the end of each section.

In this section we will review some of the basic concepts of equilibrium statistical physics and explain how to translate them to language of ergodic theory.

1.2. The canonical ensemble. As a rough approximation, *thermodynamics* is a collection of empirical laws which govern the behavior of large systems (e.g. a litre of gas) on the large ("macroscopic") scale, and *statistical physics* is a theoretic attempt to derive these laws from the assumption that matter is made from molecules which follow the laws of mechanics (quantum or classical).

The basic idea is to model the thermodynamic quantities q of the system by functions $q(\vec{x}_1(t), \ldots, \vec{x}_N(t))$ of the individual state $\vec{x}_i(t)$ of the molecules which constitute it ("microscopic state"). Because of the enormous number of particles ($N \sim 10^{24}$), the chaotic nature of their motion, and the fact that $q(\cdot)$ is usually not very sensitive to changes in an exponentially small fraction of its coordinates, one expects $q(\vec{x}_1(t), \ldots, \vec{x}_N(t))$ to fluctuate near a constant. Statistical mechanics interprets this constant as the thermodynamic quantity q we measure in the lab.

To find q, we need to average $q(\vec{x}_1(t), \ldots, \vec{x}_N(t))$ in time. A central working assumption (the "ergodic hypothesis") is that the average $\frac{1}{T}\int_0^T q(\vec{x}_1(t), \ldots, \vec{x}_N(t))dt$ can be approximated as $T \to \infty$ by the space average $\int q(\vec{x}_1, \ldots, \vec{x}_N) d\mu(\vec{x}_1, \ldots, \vec{x}_N)$ over all possible microscopic states $(\vec{x}_1, \ldots, \vec{x}_N)$, with respect to some measure μ on the space of configurations. The question is which measure to use.

One of the important choices is called the *canonical ensemble*. Imagine a small closed non-thermally isolated gas container ("system A") sitting in a large open room ("system B"). We assume that A can exchange energy with B; A cannot exchange particles with B; B is at equilibrium; and B is so large that the energy it exchanges with A does not affect its thermodynamic properties.

Assume for simplicity that the list of possible microscopic states $\xi = (\vec{x}_1, \ldots, \vec{x}_N)$ of the container is finite. Gibbs's rule for the probabilistic weight of state ξ at equilibrium is

$$\Pr(\xi) := \frac{1}{Z(\beta)} e^{-\beta U(\xi)} \tag{1.1}$$

where $U(\xi)$:=total energy of system A when in state ξ, $\beta := \frac{1}{k_B T}$ where k_B is a physical constant, T is the temperature of B, and $Z(\beta) := \sum_{\xi} e^{-\beta U(\xi)}$ is the normalizing constant, called the *partition function*. This probability distribution is called the *canonical ensemble*.

In more complicated situations, e.g. when the container is open and can also exchange particles with the room, the function $U(\xi)$ is replaced by a more complicated expression which leads to another weighting scheme, called the *grand canonical ensemble*. The details are not important. What is important is the form

$$\Pr(\text{state}) \propto \exp[-\text{parameteric function of the state}]$$

where the choice of function and parameters depends on the physics of the problem.

If the space of all possible configurations ξ is uncountable, then (1.1) does not make sense, because we cannot perform the normalization. In such cases, we have to use other methods to define the canonical ensemble. Mathematical physicists have come up with several ways to do this, which we now review. For the sake of concreteness we limit ourselves to a particular example: the *one-sided one-dimensional lattice gas model*.

1.3. One-dimensional lattice gas. In this model, we have a one-sided one-dimensional array of sites indexed by 0,1,2, etc. Each site can be either empty ("0") or full ("1"). Thus the space of all possible microscopic states is

$$X := \{(x_0, x_1, x_2, \ldots) : x_i = 0 \text{ or } 1\}.$$

Suppose the particles at the full sites interact with each other and/or with an external force field. This gives each (occupied) site "potential energy". Let

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

denote the energy content of site zero due to its interaction with the world around it. One can think of $U(\underline{x})$ as minus the energy required to "break" site zero and push it to infinity. Breaking sites $1, \ldots, n-1$ successively, we find that the energy content of the first n sites due to their interaction with the world is

$$U(\underline{x}) + U(\sigma\underline{x}) + \cdots + U(\sigma^{n-1}\underline{x}),$$

where $\sigma : X \to X$ is the *left shift map* defined by $\sigma(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots)$. The "total energy" is then $\sum_{k\geq 0} U(\sigma^k \underline{x})$.

Now try to apply (1.1): $\Pr(\underline{x}) = \frac{\text{“}\exp[-\beta\sum_{k\geq 0}U(\sigma^k\underline{x})]\text{”}}{\text{“}\sum_{\underline{y}\in X}\exp[-\beta\sum_{k\geq 0}U(\sigma^k\underline{y})]\text{”}}$. The problems are immediately apparent: the expressions within quotation marks are not necessarily finite, and the sum in the denominator has uncountably many terms.

Mathematical physicists came up with several alternatives to (1.1) which do make sense. We will review these alternatives, paying special attention to the heuristics which motivate them.

1.4. Dobrushin-Lanford-Ruelle states. Instead of giving a formula for Pr directly, we will give formulas for the conditional probabilities of $x_0, \ldots, x_{n-1}$ given $(x_n, x_{n+1}, \ldots)$.

Write $z_n^m = (z_n, z_{n+1}, \ldots, z_m)$ and $\underline{x} = (y_0^{n-1}, x_n^\infty)$. (1.1) and the identity $\sum_{k\geq 0} U[\sigma^k(y_0^{n-1}, x_n^\infty)] = \sum_{k=0}^{n-1} U(y_k^{n-1}, x_n^\infty) + \sum_{k\geq n} U(\sigma^k\underline{x})$ give

$$\frac{\text{“}\Pr(y_0^{n-1}|x_n^\infty)\text{”}}{\text{“}\Pr(z_0^{n-1}|x_n^\infty)\text{”}} = \frac{\text{“}\frac{1}{Z(\beta)}\exp\left(-\beta\sum_{k=0}^{n-1}U(y_k^{n-1},x_n^\infty)\right)\exp\left(-\beta\sum_{k=n}^{\infty}U(\sigma^k\underline{x})\right)\text{”}}{\text{“}\frac{1}{Z(\beta)}\exp\left(-\beta\sum_{k=0}^{n-1}U(z_k^{n-1},x_n^\infty)\right)\exp\left(-\beta\sum_{k=n}^{\infty}U(\sigma^k\underline{x})\right)\text{”}}$$

(terms within quotation marks are not well-defined)

$$= \frac{\exp\left(-\beta\sum_{k=0}^{n-1}U(y_k^{n-1},x_n^\infty)\right)}{\exp\left(-\beta\sum_{k=0}^{n-1}U(z_k^{n-1},x_n^\infty)\right)}.$$

The meaningless quantities cancelled out. Fixing z_0^{n-1} and varying y_0^{n-1} we get "$\Pr(y_0^{n-1}|x_n^\infty)$" $\propto \exp[-\beta\sum_{k=0}^{n-1}(U\circ\sigma^k)(y_0^{n-1}, x_n^\infty)]$. Now there is no problem to normalize, because $(y_0, \ldots, y_{n-1})$ ranges over a finite set, with 2^n elements.

To make this precise, let $\mathscr{B}$ denote the smallest σ–algebra which makes the coordinate functions $X_i(x_0, x_1, x_2, \ldots) = x_i$ measurable. $\mathscr{B}$ is the Borel σ–algebra for the metric $d(\underline{x}, \underline{y}) = \exp[-\min\{i \geq 0 : x_i \neq y_i\}]$, and $\sigma^{-n}(\mathscr{B})$ is generated by X_i, $i \geq n$. Given a Borel probability measure m, we define

$$m(x_0, \ldots, x_{n-1}|x_n, x_{n+1}, \ldots) := \mathbb{E}_m(1_{[x_0,\ldots,x_{n-1}]}|\sigma^{-n}(\mathscr{B}))(\underline{x}). \tag{1.2}$$

$\mathbb{E}_m$ is the conditional expectation, and $1_{[x_0,\ldots,x_{n-1}]}$ is the indicator function of

$$[x_0,\ldots,x_{n-1}] := \{\underline{y} : y_i = x_i \ (i = 0,\ldots,n-1)\}.$$

The right hand side of (1.2) an L^1 element, not a function, and is only defined a.e.

DEFINITION 1.1. A Borel probability measure m on $(X,\mathscr{B})$ is called a *Dobrushin–Lanford–Ruelle (DLR) state* for the potential U at inverse temperature β, if

$$m(x_0,\ldots,x_{n-1}|x_n,x_{n+1},\ldots) = \frac{1}{Z_n(\beta,x_n^\infty)}\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{x}))] \text{ for } m\text{–a.e. } \underline{x}\in X$$

where $Z_n(\beta,x_n^\infty) := \sum_{\underline{y}\in\sigma^{-n}\{\sigma^n(\underline{x})\}}\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{y}))]$.

The equations for $m(x_0,\ldots,x_{n-1}|x_n,x_{n+1},\ldots)$ are called the *DLR equations*. It is not clear that they can be solved, or that if a solution exists then it is unique. We will discuss these questions later.

The DLR equations appear quite naturally in ergodic theory, even in situations which have nothing to do with physics. To explain how, we need the following definitions. We continue to work in the context of the left shift σ on $X = \{0,1\}^{\mathbb{N}}$, postponing the discussion of the general case to later sections.

A Borel measure ν on X is called *non-singular*, if for every Borel set E, $\nu[\sigma^{-1}(E)] = 0 \Leftrightarrow \nu(E) = 0$.

Let $\nu\circ\sigma$ denote the measure $(\nu\circ\sigma)(E) := \nu[\sigma(E\cap[0])] + \nu[\sigma(E\cap[1])]$, where $[a] := \{\underline{x} : x_0 = a\}$. Notice that $(\nu\circ\sigma)(E) \geq \nu[\sigma(E)]$, with strict inequality possible due to the non-invertibility of σ. It is easy to see that if ν is non-singular, then $\nu \ll \nu\circ\sigma$, and therefore the Radon-Nikodym derivative $\frac{d\nu}{d\nu\circ\sigma}$ is well–defined. We call $\frac{d\nu}{d\nu\circ\sigma}$ the *Jacobian* of ν w.r.t. σ.[2]

Suppose $\phi : X \to \mathbb{R}$ is Borel. A Borel probability measure ν on X is called a *conformal* measure for ϕ if it is non-singular, and if there is a constant λ s.t.

$$\frac{d\nu}{d\nu\circ\sigma} = \lambda^{-1}\exp\phi, \quad \nu\circ\sigma\text{–almost everywhere.}$$

Every non-singular measure ν is conformal for $\phi := \ln\frac{d\nu}{d\nu\circ\sigma}$. The next theorem shows that every non-singular measure is a DLR states for some suitable (measurable) "interaction potential."

THEOREM 1.2. *Any conformal measure for $\phi := -\beta U$ is a DLR state with potential U at inverse temperature β.*

PROOF. Given two $\underline{a},\underline{b}\in\{0,1\}^n$, the *holonomy map* $\vartheta_{\underline{a},\underline{b}} : [\underline{a}] \to [\underline{b}]$ is the bijection $\vartheta_{\underline{a},\underline{b}}(\underline{a},x_n^\infty) = (\underline{b},x_n^\infty)$. It is standard to check that if $\frac{d\nu}{d\nu\circ\sigma} = \lambda^{-1}\exp\phi$, then $\frac{d\nu\circ\vartheta_{\underline{a},\underline{b}}}{d\nu} = \exp\Phi(\underline{a}x_n^\infty,\underline{b}x_n^\infty)$, where $\Phi(\underline{u},\underline{v}) := \sum_{k=0}^\infty[\phi(\sigma^k\underline{v}) - \phi(\sigma^k\underline{u})]$. $\Phi(\underline{u},\underline{v})$ makes sense for all pairs $(\underline{u},\underline{v})$ s.t. $\sigma^n(\underline{u}) = \sigma^n(\underline{v})$ for some n.

[2] If $T : [0,1] \to [0,1]$ is an expanding piecewise smooth interval map with two full branches, and ν is the symbolic coding of Lebesgue's measure, then $\frac{d\nu}{d\nu\circ\sigma}$ is the symbolic coding of $1/|T'|$.

By the martingale convergence theorem, for all words $\underline{a}$ and $\underline{b}$ of length n,

$$\frac{\nu(b_0,\ldots,b_{n-1}|x_n,x_{n+1},\ldots)}{\nu(a_0,\ldots,a_{n-1}|x_n,x_{n+1},\ldots)} = \lim_{k\to\infty}\frac{\nu[\underline{b},x_n^{n+k}]}{\nu[\underline{a},x_n^{n+k}]} = \lim_{k\to\infty}\frac{\nu\circ\vartheta_{\underline{a},\underline{b}}[\underline{a},x_n^{n+k}]}{\nu[\underline{a},x_n^{n+k}]}$$
$$= \lim_{k\to\infty}\frac{1}{\nu[\underline{a},x_n^{n+k}]}\int_{[\underline{a},x_n^{n+k}]}\exp\Phi(\underline{a}x_n^\infty,\underline{b}x_n^\infty)d\nu$$
$$= \lim_{k\to\infty}\mathbb{E}_\nu(\exp\Phi(\underline{a}x_n^\infty,\underline{b}x_n^\infty)|x_0^{n+k})(\underline{a}x_n^\infty) = \exp\Phi(\underline{a}x_n^\infty,\underline{b}x_n^\infty) \text{ a.s.}$$

Treating $\underline{a}$ as a constant and $\underline{b}$ as a variable, we get $\nu(\underline{b}|x_n^\infty)\propto\exp\sum_{i=0}^{n-1}\phi[\sigma^i(\underline{b}x_n^\infty)]$, whence the DLR equations. □

1.5. Thermodynamic limits. Another approach for making sense of (1.1) is to "approximate" X from within by countable or finite $X_n\subset X$, define the canonical ensembles for X_n, and pass to the limit.

One popular approximation scheme is to impose "periodic boundary conditions": $X_n^{per} := \{\underline{x}\in X:\sigma^n(\underline{x})=\underline{x}\}$. Another common approach is to impose a "fixed boundary condition":

$$X_n = X_n^{\underline{x}} := \sigma^{-n}\{\underline{x}\} = \{\underline{y}\in X : y_i = x_{i-n}\quad (i\geq n)\}.$$

We focus on the latter approach, because it leads more quickly to one of the basic tools of the trade, *Ruelle's operator*.

If we apply (1.1) blindly to X_n then we get the following formula (expressions within quotation marks are not well-defined):

$$\Pr_{X_n}(\underline{y}) = \frac{\text{“}\exp[-\beta\sum_{i=0}^\infty U(\sigma^i(\underline{y}))]\text{”}\,\delta_{\underline{x}}(\sigma^n\underline{y})}{\text{“}\sum_{\sigma^n(\underline{z})=\underline{x}}\exp[-\beta\sum_{i=0}^\infty U(\sigma^i(\underline{z}))]\text{”}}\qquad(\delta_{\underline{x}} = \text{Dirac's delta function})$$
$$= \frac{\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{y}))]\delta_{\underline{x}}(\sigma^n\underline{y})\cdot\text{“}\exp[-\beta\sum_{i=0}^\infty U(\sigma^i(\underline{x}))]\text{”}}{\sum_{\sigma^n(\underline{z})=\underline{x}}\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{z}))]\cdot\text{“}\exp[-\beta\sum_{i=0}^\infty U(\sigma^i(\underline{x}))]\text{”}}$$
$$= \frac{\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{y}))]\delta_{\underline{x}}(\sigma^n\underline{y})}{\sum_{\sigma^n(\underline{z})=\underline{x}}\exp[-\beta\sum_{i=0}^{n-1}U(\sigma^i(\underline{z}))]}.$$

The idea is now to let $n\to\infty$ and look for weak-star limit points.

The measures $\Pr_{X_n}$ can be expressed very efficiently using *Ruelle's operator* $L_\phi : C(X)\to C(X)$

$$(L_\phi f)(x) = \sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})}f(\underline{y}), \tag{1.3}$$

where $\phi := -\beta U$. To do this, let $\phi_n := \phi+\phi\circ\sigma+\cdots+\phi\circ\sigma^{n-1}$, then it is not difficult to see that $(L_\phi^n f)(\underline{x}) = \sum_{\sigma^n\underline{y}=\underline{x}} e^{\phi_n(\underline{y})}f(\underline{y})$. Thus for every $f : X\to\mathbb{R}$,

$$\int f d\Pr_{X_n} = \frac{\sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})}f(\underline{y})}{\sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})}} = \frac{(L_\phi^n f)(\underline{x})}{(L_\phi^n 1)(\underline{x})}.$$

Equivalently, $\Pr_{X_n} = (L_\phi^n)^*\delta_{\underline{x}}/(L_\phi^n 1)(\underline{x})$.

Definition 1.3. A *thermodynamic limit with potential* U*, inverse temperature* β *and boundary condition* $\underline{x}$ is a weak-star limit point of $\Pr_{X_n} = (L_\phi^n)^*\delta_{\underline{x}}/(L_\phi^n 1)(\underline{x})$ as $n\to\infty$. Here $\phi = -\beta U$.

To understand the weak-star limit points of Pr_{X_n}, one needs to understand the asymptotic behavior of L_ϕ^n as $n \to \infty$.

Again we see the relevance to dynamics: Averaging operators of the form (1.3) appear naturally in dynamics as *transfer operators*, or *dual operators* for non-singular measures. The asymptotic behavior of their powers contains information on the ergodic and stochastic properties of the dynamical system, and any general tool for determining this behavior has potential applications to dynamics.

1.6. Equilibrium measures. A completely different approach to the canonical ensemble is to characterize the probability distribution (1.1) as a solution to a variational problem, and then hope that the variational problem is well–posed when the configuration space is uncountable.

This can be done. Let $(U_1, \ldots, U_N)$ be a vector, representing the energies of a system which can only occupy a finite number of states (1 to N). The canonical ensemble distribution of this system $\underline{p} = (p_1, \ldots, p_N)$, $p_i = e^{-\beta U_i} / \sum e^{-\beta U_i}$ turns out to be the unique minimizer of the quantity

$$F = \sum_{i=1}^{N} p_i U_i - \frac{1}{\beta} H(\underline{p}), \text{ where } H(\underline{p}) = -\sum_{i=1}^{N} p_i \ln p_i.$$

F is called the average *Helmholtz free energy*, $\sum p_i U_i$ is called the average *energy* of $\underline{p}$, and $H(\underline{p})$ is called the *entropy*.[3] The canonical ensemble can thus be thought of as the probability distribution which *minimizes* the average free energy. Ruelle had the idea of using a similar principle for the lattice gas model.

Worried by the fact that the total energy $\sum_{k \geq 0} U \circ \sigma^k$ diverges at many configurations, we will replace the total free energy by the *free energy per site*, defined for a Borel probability measure μ by the following limit, when it exists:

$$\lim_{n \to \infty} \frac{1}{n} \left(\int \sum_{k=0}^{n-1} U(\sigma^k \underline{x}) d\mu - \frac{1}{\beta} \Big(- \sum_{\underline{a} \in \{0,1\}^n} \mu[\underline{a}] \ln \mu[\underline{a}] \Big) \right)$$

If μ is shift invariant and U is absolutely integrable w.r.t μ, then the limit does exist, and it equals $\int U d\mu - \frac{1}{\beta} h_\mu(\sigma)$ where $h_\mu(\sigma)$ is the metric entropy of μ.

Thus it makes sense to look for the shift invariant measures μ which minimize $\int U d\mu - \frac{1}{\beta} h_\mu(\sigma)$. The custom in dynamics is to pull the negative constant $-\frac{1}{\beta}$ out, and look for measures which *maximize* $h_\mu(\sigma) + \int \phi d\mu$, where $\phi := -\beta U$.

DEFINITION 1.4. An *equilibrium measure* for U at inverse temperature β is a shift invariant measure μ which maximizes $h_\mu(\sigma) + \int \phi d\mu$ for $\phi = -\beta U$, over all invariant Borel measures.

Notice that equilibrium measures are shift invariant by definition. This is not always the case for DLR states or thermodynamic limits. However, we shall see later that for countable Markov shifts, equilibrium measures only differ from DLR states or thermodynamic limits by a positive density function.

1.7. Gibbs measures in the sense of Bowen. This definition, due to Bowen, originated in dynamics rather than mathematical physics:

---

[3]Notice that we used ln, not $\log_2$, to define $H(\underline{p})$.

Definition 1.5. Let $\phi := -\beta U$. A *Gibbs measure in the sense of Bowen* is a shift invariant Borel probability measure m s.t. that for some $M > 1$ and $P \in \mathbb{R}$, for every finite word $\underline{a} = (a_0, \ldots, a_{n-1})$

$$M^{-1} \leq \frac{m[a_0, \ldots, a_{n-1}]}{\exp(\sum_{k=0}^{n-1} \phi(\sigma^k \underline{x}) - nP)} \leq M \text{ for all } \underline{x} \in [\underline{a}]. \tag{1.4}$$

This is a strong uniform way of saying that $m[\underline{a}] \asymp \frac{1}{Z_n} \exp(-\beta \sum_{k=0}^{n-1} U(\sigma^k \underline{x}))$, where $Z_n = \exp(nP)$.

It can be shown that for the lattice gas model, if U is Hölder continuous, then every equilibrium measure is a Gibbs measure in the sense of Bowen. This extends to topological Markov shifts with a finite alphabet. But when the alphabet is infinite, Gibbs measures in the sense of Bowen cannot exist unless the underlying shift satisfies a very strong combinatorial property, called the *big images and pre-images property*, see Theorem 5.9 below.

Therefore while this definition is central to the thermodynamic formalism of subshifts of finite type, it is less important to the thermodynamic formalism of countable Markov shifts.

1.8. Plan of the survey. We have given several non-equivalent definitions for the canonical ensemble in the case of the lattice gas model $\{0,1\}^{\mathbb{N}}$: DLR states, thermodynamics limits, equilibrium measures, and Gibbs measures in the sense of Bowen. Do these measures exist? Are they unique? What are their ergodic properties? What are their stochastic properties? Do the different approaches to the canonical ensemble lead to the same measures? We will discuss these questions in the case when $\{0,1\}^{\mathbb{N}}$ is replaced by a general topological Markov shift.

The interest in such questions is that topological Markov shifts appear as symbolic models for smooth dynamical systems, and measures with smooth Jacobians (which as we saw above can be viewed to be DLR states for a smooth potentials) appear naturally in dynamical contexts. The interest in countable state topological Markov shifts is that they appear in the non-uniformly hyperbolic setup, see §8.1. At the end of the survey we will mention applications to surface diffeomorphisms.

1.9. Notes and references. The different approaches to the canonical ensemble originate in the mathematical theory of the foundations of equilibrium statistical physics. Equilibrium measures were introduced by Ruelle [**Rue67**]. DLR measures were introduced by Dobrushin [**Dob68a**], [**Dob68b**], [**Dob68c**] and Lanford & Ruelle [**LR69**]. What we call "thermodynamic limits" are called in the literature "Gibbs states" or "Gibbs measures." DLR measures are also called "Gibbsian random fields." We chose different terminology, to avoid confusion.

Ruelle gave a complete analysis of the one-dimensional lattice gas model in [**Rue68**]. Around the same time, Sinai constructed Markov partitions for Anosov diffeomorphisms [**Sin68a**], [**Sin68b**], and showed that the resulting symbolic models are simple generalizations of the lattice gas models of statistical physics. This discovery linked hyperbolic dynamics to mathematical statistical physics.

Sinai suggested to use the thermodynamic approach to study Anosov diffeomorphisms [**Sin72**]. He showed that some of the natural invariant measures in hyperbolic dynamics can be viewed as certain types of thermodynamic limits. Bowen [**Bow75**] and Ruelle [**Rue76**] explained how equilibrium measures appear naturally in the theory of Axiom A diffeomorphisms. Bowen showed that in this case, equilibrium measures are also what we call “Gibbs measures is the sense of Bowen” [**Bow75**]. Since then the thermodynamic formalism has been applied to a wide variety of dynamical problems.

Here is a partial list of excellent references to thermodynamic formalism, with indications of some of the applications they discuss: [**Bow75**] (Anosov diffeomorphisms), [**Rue78**] (equilibrium statistical physics), [**PP90**] (Anosov flows), [**Zin96**] (complex dynamics), [**Kel98**] (interval maps), [**MU03**] and [**PU10**] (geometric measure theory).

2. Topological Markov shifts (TMS)

2.1. One sided topological Markov shifts. Suppose $G = G(V, E)$ is a directed graph with a finite or countable collection of vertices V and edges $E \subset V \times V$. We always assume that every vertex v has at least one in-coming edge $a \to v$ and at least one out-going edge $v \to b$. The notation $a \to b$ means that $(a, b) \in E$, and the notation $a \xrightarrow{n} b$ means that there are vertices $\xi_1, \ldots, \xi_n$ s.t. $a \to \xi_1 \to \cdots \to \xi_{n-1} \to b$. In this case we say that *a connects to b in n steps*, and that $(a, \xi_1, \ldots, \xi_{n-1}, b)$ is *admissible*.

Definition 2.1. The *(one-sided) topological Markov shift (TMS)* associated to G is the set $\Sigma^+(G) := \{(x_0, x_1, \ldots) \in V^{\mathbb{N}_0} : x_i \to x_{i+1} \text{ for all } i\}$, together with the metric $d(\underline{x}, \underline{y}) := \exp[-\min\{n : x_n \neq y_n\}]$ and the action of the left shift map

$$\sigma(x_0, x_1, x_2, \ldots) = (x_1, x_2, \ldots).$$

V is called the *alphabet*, and elements of V are called *states*. The matrix $(t_{ab})_{V \times V}$, $t_{ab} = 1$ when $a \to b$ and $t_{ab} = 0$ when $a \not\to b$, is called the *transition matrix*.

$(\Sigma^+(G), d)$ is a complete and separable metric space. Its topology is generated by the *cylinders*

$$[a_0, \ldots, a_{n-1}] := \{\underline{x} \in \Sigma^+(G) : x_i = a_i \ (i = 0, \ldots, n-1)\}.$$

The cylinders are open and closed. $\Sigma^+(G)$ is compact iff V is finite. When V is infinite, $\Sigma^+(G)$ is not compact, and sometimes not even locally compact. It is easy to check that $\Sigma^+(G)$ is locally compact iff for every $a \in V$, $\#\{b \in V : a \to b\} < \infty$.

Compact TMS are also called *subshifts of finite type*. Non-compact TMS are also called *countable Markov shifts* (“countable” relates to V, not to $\Sigma^+(G)$).

Definition 2.2. A topological Markov shift $\Sigma^+(G)$ is called *topologically transitive*, if for all $a, b \in V$, $\exists n$ s.t. $a \xrightarrow{n} b$, and *topologically mixing* if for all $a, b \in V$, $\exists N = N(a, b)$ s.t. $a \xrightarrow{n} b$ for all $n \geq N$.

It is routine to show that $\Sigma^+(G)$ is *topologically transitive* iff σ is topologically transitive (i.e. for all open sets U, V there exists $n > 0$ s.t. $U \cap \sigma^{-n}(V) \neq \varnothing$). Similarly, $\Sigma^+(G)$ is *topologically mixing* iff σ is topologically mixing (i.e. for all open sets U, V there exists $N = N(U, V) > 0$ s.t. $U \cap \sigma^{-n}(V) \neq \varnothing$ for all $n > N$).

Remark 2.3. Let A denote the transition matrix of G. If $|V| < \infty$, then topological mixing is equivalent to the existence of N s.t. all the entries of A^N are positive. But when V is infinite, topological mixing is strictly weaker than this.

The following two theorems allow in many cases to reduce the discussion of general TMS to the topologically transitive or even topologically mixing case:

Theorem 2.4. *Suppose μ is a shift invariant ergodic measure on $\Sigma^+(G)$. Let $G_\mu = G(V_\mu, E_\mu)$ be the subgraph of G with vertices $V_\mu := \{a \in V : \mu[a] \neq 0\}$ and edges $E_\mu := \{(a,b) \in E : \mu[a,b] \neq 0\}$, then μ is supported inside $\Sigma^+(G_\mu)$, and $\Sigma^+(G_\mu)$ is topologically transitive.*

Theorem 2.5 (Spectral decomposition). *Let $\Sigma^+(G)$ denote a topologically transitive TMS. There exists a natural number p, called the* the period, *such that $\Sigma^+(G) = X_0 \uplus \cdots \uplus X_{p-1}$, $\sigma(X_i) = X_{i+1 (\mathrm{mod}\, p)}$, and s.t. $\sigma^p : X_i \to X_i$ is topologically conjugate to a topologically mixing TMS for all i.*

The period is given by $p := \gcd\{n : a \xrightarrow{n} a\}$ for some (all) states a.

2.2. Regularity of functions. The modulus of continuity of $\phi : \Sigma^+(G) \to \mathbb{R}$ is captured by the decay of the *variations* of ϕ:

$$\mathrm{var}_n\phi := \sup\{\phi(\underline{x}) - \phi(\underline{y}) : x_i = y_i \ (i = 0, \ldots, n-1)\}.$$

A function $\phi : \Sigma^+(G) \to \mathbb{R}$ is called

(1) *Markovian*, if $\mathrm{var}_2\phi = 0$. In this case $\phi(\underline{x}) = \phi(x_0, x_1)$.
(2) *Weakly Hölder continuous*, if $\exists A > 0, \theta \in (0,1)$ s.t. $\forall n \geq 2$, $\mathrm{var}_n\phi \leq A\theta^n$. If in addition $\mathrm{var}_1\phi < \infty$ then we say that ϕ is *locally Hölder*. If in addition ϕ is bounded, then we say that ϕ is Hölder. The number θ is called a *Hölder exponent* of ϕ.
(3) A function with *summable variations*, if $\sum_{n\geq 2} \mathrm{var}_n\phi < \infty$.

Every Markovian function is weakly Hölder continuous, and every weakly Hölder continuous function has summable variations. Neither of these conditions implies that ϕ is bounded, and it could even happen that $\mathrm{var}_1\phi = \infty$.[4]

The essence of the summable variations condition is the following estimate:

Lemma 2.6. *Let $\phi_n := \phi + \phi \circ \sigma + \cdots + \phi \circ \sigma^{n-1}$. For every $n \geq 1$ and every admissible word $\underline{a} = (a_0, \ldots, a_{n-1})$, if $\underline{x}, \underline{y} \in \sigma[a_{n-1}]$ and $x_0^{m-1} = y_0^{m-1}$, then $|\phi_n(\underline{a}\,\underline{x}) - \phi_n(\underline{a}\,\underline{y})| \leq \sum_{k\geq m+1} \mathrm{var}_k\phi$.*

In particular, if ϕ has summable variations, then $\sup_{n\geq 1}[\mathrm{var}_{n+k}\phi_n] \xrightarrow[k\to\infty]{} 0$.

2.3. Two-sided topological Markov shifts. Suppose $G = G(V,E)$ is a finite or countable directed graph, such that every vertex a has at least one outgoing edge and at least one in-coming edge.

Definition 2.7. The *two-sided topological Markov shift* associated to G is the set $\Sigma(G) := \{\underline{x} \in V^{\mathbb{Z}} : x_i \to x_{i+1}$ for all $i \in \mathbb{Z}\}$, together with the metric $d(\underline{x}, \underline{y}) := \exp[-\min\{|n| : x_n \neq y_n\}]$ and the action of the left-shift map $\sigma(\underline{x})_i = x_{i+1}$.

The conditions for the compactness, topological transitivity, and topological mixing of $\Sigma(G)$ are the same as for $\Sigma^+(G)$. Local compactness is different: a two-sided TMS is locally compact iff for every a, $\#\{(u,v) : u \to a \to v\} < \infty$. Cylinders are also slightly more complicated because of the need to keep track of the left-most coordinate of the constraint. We will use the notation

$${}_m[a_0, \ldots, a_{n-1}] := \{\underline{x} \in \Sigma(G) : x_{m+i} = a_i \ (i = 0, \ldots, n-1)\}.$$

---

[4] It is important to include unbounded functions in the discussion, because for non-compact TMS with infinite topological entropy, only unbounded functions can have finite pressure.

Every shift invariant Borel probability measure μ on $\Sigma(G)$ defines a shift invariant Borel probability measure μ^+ on $\Sigma^+(G)$ by $\mu^+([\underline{a}]) := \mu({}_0[\underline{a}])$ and the Carathéodory extension procedure. Conversely, every shift invariant Borel probability measure μ^+ on $\Sigma^+(G)$ determines a unique shift invariant measure μ on $\Sigma(G)$ through the equations $\mu({}_m[\underline{a}]) := \mu^+[\underline{a}]$. This is an instance of the "natural extension" procedure.

Every $\phi^+ : \Sigma^+(G) \to \mathbb{R}$ defines a function ϕ on $\Sigma(G)$ by $\phi(\underline{x}) := \phi^+(x_0, x_1, \ldots)$. The converse is not true, but the following is enough for most applications.

THEOREM 2.8. *Suppose $\phi : \Sigma(G) \to \mathbb{R}$ has summable variations, then there exists $\phi^+ : \Sigma^+(G) \to \mathbb{R}$ with summable variations and a bounded continuous function $h : \Sigma(G) \to \mathbb{R}$ s.t. for every $\underline{x} \in \Sigma(G)$, $\phi(\underline{x}) + h(\underline{x}) - h(\sigma\underline{x}) = \phi^+(x_0, x_1, \ldots)$.*

Most of the thermodynamic formalism is invariant under addition of coboundaries: what works for ϕ, works for $\phi + h - h \circ \sigma$. Therefore this theorem allows to reduce problems on two-sided shifts to problems for one-sided shifts.

Henceforth, unless stated otherwise, all topological Markov shifts are one-sided.

2.4. Notes and references. For the material in §2.1, see chapter 4 in [**Aar97**]. Theorem 2.8 was first proved for compact Markov shifts and Hölder continuous potentials by Sinai [**Sin72**], see also [**Bow75**]. For functions with summable variations, see [**CQ98**] (compact TMS) and [**Dao13**] (non-compact TMS).

3. Conformal measures and their ergodic properties

3.1. Conformal measures and DLR states. Let $X := \Sigma^+(G)$ denote a topological Markov shift with set of states V. Given a non-singular σ–finite Borel measure ν on X, let $\nu \circ \sigma$ denote the measure $(\nu \circ \sigma)(E) = \sum_{a \in V} \nu[\sigma(E \cap [a])]$. It is easy to verify that for all non-negative Borel functions, $\int f d\nu \circ \sigma = \sum_{a \in V} \int_{\sigma[a]} f(a\underline{x}) d\nu(\underline{x})$. Also, $\nu \ll \nu \circ \sigma$.

DEFINITION 3.1. The *Jacobian* of ν is $g_\nu := \frac{d\nu}{d\nu \circ \sigma}$.

DEFINITION 3.2. A non-singular Borel measure ν, which is finite and positive on cylinders, is called *conformal* for a function ϕ, if there is a constant λ s.t. $g_\nu = \lambda^{-1} \exp \phi$, $\nu \circ \sigma$–a.e.

There is a useful characterization of conformal measures in terms of Ruelle's operator $L_\phi f = \sum_{\sigma(\underline{y}) = \underline{x}} e^{\phi(\underline{y})} f(\underline{y})$:

THEOREM 3.3. *ν is conformal for ϕ iff ν is an eigenmeasure of L_ϕ: $\exists \lambda > 0$ s.t. $L_\phi^* \nu = \lambda \nu$, i.e. for all non-negative measurable functions f, $\int L_\phi f d\nu = \lambda \int f d\nu$.*

This follows immediately from the identity $\int f d\nu \circ \sigma = \sum_{a \in V} \int_{\sigma[a]} f(a\underline{x}) d\nu(\underline{x})$.

Every conformal measure ν s.t. $\nu(X) = 1$ satisfies the *DLR equations*

$$\nu(x_0, \ldots, x_{n-1} | x_n, x_{n+1}, x_{n+2}, \ldots) = \frac{\exp \phi_n(\underline{x})}{\sum_{\underline{y} \in \sigma^{-n}\{\sigma^n \underline{x}\}} \exp \phi_n(\underline{y})} \qquad \nu\text{–a.e.}$$

where $\phi_n = \phi + \phi \circ \sigma + \cdots + \phi \circ \sigma^{n-1}$. The proof is the same as in the case $X = \{0, 1\}^{\mathbb{N}}$, which was discussed in the first section.

It follows that any non-singular Borel probability measure is a DLR state for some "potential" ϕ (equal to $-\log g_\nu$). What is special about the measures studied in the thermodynamic formalism, is that their "potential" is not just any Borel function, but a function with good regularity properties.

3.2. Ergodic properties when the log Jacobian is regular. Recall the following general properties of non-singular measurable maps T on a σ–finite standard measure space $(\Omega, \mathscr{F}, \nu)$:

(1) *Conservativity:* All wandering sets have measure zero. A *wandering set* is a measurable set W s.t. $\{T^{-k}W\}_{k\geq 0}$ are pairwise disjoint.
(2) *Ergodicity:* Every T–invariant set E satisfies $\nu(E) = 0$ or $\nu(\Omega \setminus E) = 0$. An *invariant set* is a set E s.t. $T^{-1}E = E$.
(3) *Exactness:* Every tail set E satisfies $\nu(E) = 0$ or $\nu(\Omega \setminus E) = 0$. A *tail set* is a set which belongs to the *tail σ–algebra* $\bigcap_{n\geq 0} T^{-n}\mathscr{F}$. This implies mixing.

For invertible probability preserving maps, we also have the following property:

(4) *Bernoulli property:* Measure theoretic isomorphism to a map of the form $\sigma : S^{\mathbb{Z}} \to S^{\mathbb{Z}}$ with $|S| \leq \aleph_0$ and the measure $m({}_m[a_m, \ldots, a_n]) = p_{a_m} \cdots p_{a_n}$, where $\vec{p}$ is a fixed probability vector.

Theorem 3.4. *Let X be a topologically transitive TMS and suppose ν is a conformal measure with Jacobian $g_\nu = \lambda^{-1}e^{\phi}$, where ϕ has summable variations. If X is compact (equivalently the number of states is finite), then*

(1) *ν is conservative and ergodic.*
(2) *There is a positive continuous function h s.t. $dm = hd\nu$ is an invariant probability measure.*
(3) *If X is topologically mixing, then ν and m are exact.*
(4) *If X is topologically mixing, the natural extension of (X, m, σ) is Bernoulli.*

When the number of states is infinite and X is not compact, the situation is more complicated. The simplest example of the new phenomena possible in this case is a transient countable Markov chain: The associated Markov measure is conformal with a Markovian potential, but the measure is not conservative.

Given a function $\phi : X \to \mathbb{R}$ and a state (vertex) a, let

$$Z_n(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[a]}(\underline{x}), \text{ where } \phi_n = \phi + \phi\circ\sigma + \cdots + \phi\circ\sigma^{n-1}.$$

Theorem 3.5. *Let X denote a topologically transitive TMS and suppose ν is a conformal measure with Jacobian $g_\nu = \lambda^{-1}e^{\phi}$, where ϕ has summable variations.*

(1) *If $\sum_{n=0}^{\infty} \lambda^{-n} Z_n(\phi, a) = \infty$ for some a, then $\sum_{n=0}^{\infty} \lambda^{-n} Z_n(\phi, a) = \infty$ for all a.*
(2) *ν is conservative iff $\sum_{n\geq 0} \lambda^{-n} Z_n(\phi, a) = \infty$ for some a. In this case ν is ergodic, and in the topologically mixing case exact.*
(3) *If $\sum_{n\geq 0} \lambda^{-n} Z_n(\phi, a) = \infty$ for some a, then there is a positive continuous function h s.t. $dm = hd\nu$ is shift invariant. It is possible that $\int hd\nu = \infty$.*
(4) *If $\sum_{n\geq 0} \lambda^{-n} Z_n(\phi, a) = \infty$, and the measure m in (3) is a probability measure, and X is topologically mixing, then the natural extension of $(X, \mathscr{B}, m, \sigma)$ is Bernoulli.*

3.3. Existence of conformal measures.

THEOREM 3.6. *Suppose X is a topologically transitive TMS, and $\phi : X \to \mathbb{R}$ is continuous. If X is compact (equivalently, the number of states is finite), then ϕ has a finite conformal measure.*

As noted by Ruelle, the theorem can be proved by considering the action of the continuous map $T(\nu) := (L_\phi^* \nu)/\int L_\phi 1 d\nu$ on the convex weak-star compact set $\mathscr{P}(X) := \{\text{shift invariant probability measures on } X\}$. The Schauder-Tychonoff theorem provides a fixed point $T(\nu) = \nu$, this fixed point is an eigenmeasure of L_ϕ, and eigenmeasures of L_ϕ are conformal measures by Theorem 3.3.

This argument fails when X is not compact, because in this case $\mathscr{P}(X)$ may lose its compactness. The following theorem summarizes our knowledge on the non-compact case:

THEOREM 3.7. *Suppose X is a topologically transitive TMS, and $\phi : X \to \mathbb{R}$ has summable variations, then ϕ has a conservative (possibly infinite) conformal measure ν s.t. $g_\nu = \lambda^{-1} e^\phi$ iff*

(1) $\limsup_{n\to\infty} \frac{1}{n} \ln Z_n(\phi, a) = \ln \lambda$ *for some (whence every) a, and*
(2) $\sum_{n\geq 0} \lambda^{-n} Z_n(\phi, a) = \infty$.

Infinite conformal measures do appear. To see how, suppose ν is a finite conformal measure with Jacobian $\lambda^{-1}\exp\phi$. For every positive $h : X \to \mathbb{R}$, $d\nu' := h d\nu$ has Jacobian $\lambda^{-1}\exp\phi'$ where $\phi' := \phi + \ln h - (\ln h)\circ\sigma$. There is no problem to cook a continuous h s.t. $\log h - \log h \circ \sigma$ has summable variations, but $\int h d\nu = \infty$. For such h, ϕ' has an infinite conformal measure $\nu' = h d\nu$.

3.4. Uniqueness of conformal measures. This depends on the regularity of ϕ, as can be seen from the following theorems.

THEOREM 3.8. *Suppose X is a topologically transitive topological Markov shift and $\phi : X \to \mathbb{R}$ has summable variations, then ϕ can have at most one conservative conformal measure.*

In the special case $X = \{0,1\}^{\mathbb{N}}$, more is known. A *g–function* is a continuous $g : X \to (0,1)$ s.t. $\sum_{\sigma \underline{y} = \underline{x}} g(\underline{y}) = 1$. It can be shown that a conformal measure with continuous Jacobian is shift invariant, iff its Jacobian is a g–function. Conformal measures for a g–function are also called *g–measures*.

THEOREM 3.9. *Suppose $X = \{0,1\}^{\mathbb{N}}$ and $g : X \to \mathbb{R}$ is a g–function.*

(1) *If $\sum(\text{var}_n \ln g)^2 < \infty$, then $\phi = \ln g$ has a unique conformal measure.*
(2) *For every $\epsilon > 0$ there is a g–function g satisfying $\sum(\text{var}_n \ln g)^{2+\epsilon} < \infty$, s.t. $\phi = \ln g$ has more than one conformal measure.*

3.5. Notes and references. Theorem 3.4 (1)–(3) is from [**ADU93**], and part (4) is from [**Sar11**], see also [**Dao13**], [**JÖP12**]. In the compact case, these results are older: ergodicity, existence of acip (and whence conservativity) can be traced to [**Rén57**], and the Bernoulli property of the natural extension of the acip (whence exactness) is due to Bowen [**Bow75a**], see also [**Gal73**].

The existence of conformal measures for compact topological Markov shifts is due to Ruelle [**Rue68**], [**Rue76**]. Theorem 3.7 on the non-compact case follows from results in [**Sar01b**] (the proof there is stated for locally Hölder potentials, but works verbatim under summable variations).

The uniqueness of conformal measures with potentials with summable variations is due to Bowen [**Bow75**] and Ruelle [**Rue76**] in the finite alphabet case. In the infinite alphabet case, it can be deduced from the fact if such measures are conservative, then they are ergodic [**ADU93**].

The concept of g–measures is due to Keane [**Kea72**], and its importance to thermodynamic formalism is explained in [**Led74**] and [**Wal75a**]. Examples of continuous g–functions with several g-measures are given in [**BK93**]. Part (1) of Theorem 3.9 is due to Johansson & Öberg [**JÖ03**], see [**JÖP12**] for generalizations. Part (2) is due to Berger, Hoffman, and Sidoravicius [**BHS03**].

4. Ruelle's operator, thermodynamic limits, and modes of recurrence

4.1. Thermodynamic limits and Ruelle's operator. When we discussed the lattice gas model $X = \{0,1\}^{\mathbb{N}}$ with the interaction U, we defined thermodynamic limits as a weak-star limit points of $(1/Z_n(\underline{x}))\sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})}\delta_{\underline{y}}$, where $\phi = -\beta U$, $\underline{x}$ is a boundary condition, and $Z_n(\underline{x}) := \sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})}$.

This definition extends without change to general compact TMS. But in the non-compact case, the number of states is infinite, and $Z_n(\underline{x})$ may diverge. Is it tempting to restrict the discussion to $\phi = -\beta U$ for which $Z_n(\underline{x}) < \infty$, but this is too strong for some purposes. It is better to use the following weaker restriction:

Definition 4.1. Suppose X is a TMS. A continuous function $\phi : X \to \mathbb{R}$ is called *admissible* if $\sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})} 1_{[a]}(\underline{y}) < \infty$ for all n, a, and $\underline{x}$. As always $\phi_n = \phi + \phi\circ\sigma + \cdots + \phi\circ\sigma^{n-1}$.

Definition 4.2. Suppose X is a TMS and $\phi := -\beta U$ is admissible. A *thermodynamic limit with potential U, inverse temperature β and boundary condition $\underline{x} \in X$* is any σ–finite ν which is finite on cylinders, such that for some $n_k \uparrow \infty$, for every finite union of partition sets F and for every cylinder $[\underline{a}]$

$$\frac{1}{Z_{n_k}(F,\underline{x})}\sum_{\sigma^{n_k}(\underline{y})=\underline{x}} e^{\phi_{n_k}(\underline{y})} 1_{F\cap[\underline{a}]}(\underline{y}) \xrightarrow[k\to\infty]{} \frac{\nu(F\cap[\underline{a}])}{\nu(F)},$$

where $Z_n(F,\underline{x}) = \sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})} 1_F(\underline{y})$.

We rewrite this in terms of *Ruelle's operator* $(L_\phi f)(\underline{x}) = \sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})} f(\underline{y})$. We are being intentionally vague as to the space on which this "operator" acts. As we shall see in later sections, it is useful to consider the action of L_ϕ on different spaces, depending on the case at hand.

A formal calculation shows that $(L_\phi^n f)(\underline{x}) = \sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})} f(\underline{y})$. Therefore, ν is a thermodynamic limit with boundary condition $\underline{x}$ iff there is a subsequence $n_k \to \infty$ s.t. for any cylinder $[\underline{a}]$ and every finite union of partition sets F, $(L_\phi^{n_k} 1_{[\underline{a}]})(\underline{x})/(L_\phi^{n_k} 1_F)(\underline{x}) \xrightarrow[k\to\infty]{} \nu[\underline{a}]/\nu(F)$. The analysis of thermodynamic limits reduces to the study of the asymptotic behavior of $L_\phi^n f$ as $n \to \infty$ for "sufficiently many" functions f.

4.2. Gurevich pressure. The first step in the analysis of $L_\phi^n f$ as $n \to \infty$ is to understand what happens to $\frac{1}{n}\ln L_\phi^n f$ as $n \to \infty$.

Suppose X is a topologically mixing TMS and $\phi : X \to \mathbb{R}$ has summable variations. Given a state a, let $Z_n(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[a]}(\underline{x})$.

THEOREM 4.3. *The limit* $P_G(\phi) := \lim_{n\to\infty} \frac{1}{n} \ln Z_n(\phi, a)$ *exists for all states* a*, and is independent of the choice of* a*. If* $\|L_\phi 1\|_\infty < \infty$*, then* $P_G(\phi) < \infty$*.*

THEOREM 4.4. *If* $P_G(\phi) < \infty$ *then* ϕ *is admissible, and for every* f *continuous, non-negative, not identically equal to zero, and supported inside a finite union of partition sets,* $\lim_{n\to\infty} \frac{1}{n} \ln(L_\phi^n f)(\underline{x}) = P_G(\phi)$ *for all* $\underline{x}$*.*

DEFINITION 4.5. $P_G(\phi)$ is called the *Gurevich pressure* of ϕ.

$P_G(\phi)$ plays a central role in the thermodynamic formalism. In section 5 we will see that $P_G(\phi) = \sup\{h_\mu(\sigma) + \int \phi d\mu\}$ (sup over all shift invariant measures), and in sections 6 and 7 we'll see that behavior of the functions $t \mapsto P_G(\phi + t\psi)$ characterizes the equilibrium measure of ϕ, and some of its statistical properties.

The following theorems provide additional information on $P_G(\phi)$:

THEOREM 4.6. *Let* X *be a topologically mixing TMS, and* ϕ, ψ *be functions with summable variations, then*

(1) $P_G(\phi + c) = P_G(\phi) + c$ *for every constant* c
(2) $\phi \le \psi \Rightarrow P_G(\phi) \le P_G(\psi)$
(3) $P_G(t\phi + (1-t)\psi) \le tP_G(\phi) + (1-t)P_G(\psi)$ *for all* $0 < t < 1$
(4) *If* $\phi - \psi = f - f \circ \sigma$*, then* $P_G(\phi) = P_G(\psi)$

THEOREM 4.7. *Let* $X = \Sigma^+(G)$ *be a topologically mixing TMS associated to a directed graph* G*. If* ϕ *has summable variations, then* $P_G(\phi) = \sup\{P_G(\phi|_{\Sigma^+(G')})\}$*, where the supremum ranges over all finite subgraphs* $G' \subset G$*.*

4.3. Generalized Ruelle's Perron-Frobenius Theorem. Suppose X is a topologically mixing TMS associated to a graph $G = G(V, E)$.

Before we tackle the asymptotic behavior of L_ϕ^n for general admissible functions with summable variations, we consider the special case when $L_\phi 1 = 1$ and ϕ is Markovian. This means that $\phi(\underline{x}) = f(x_0, x_1)$ for some function $f : E \to \mathbb{R}$.

Let G' denote the graph obtained from G by reversing the direction of its edges: $G' = G(V, E')$ where $E' = \{(b, a) : (a, b) \in E\}$. Since $L_\phi 1 = 1$,

$$P = (p_{ab})_{V\times V},\ p_{ab} := \begin{cases} e^{f(b,a)} & (a,b) \in E' \\ 0 & \text{otherwise} \end{cases}$$

is a stochastic matrix, compatible with G'. Reversing the edges guarantees the following identity: If $[\underline{b}] = [b_0, \ldots, b_{m-1}]$ and $x_0 = a$, then

$$L_\phi^{n+m} 1_{[\underline{b}]}(\underline{x}) = p_{ab_{m-1}}^{(n)} \cdot p_{b_{m-1}b_{m-2}} \cdots p_{b_1 b_0},$$

where $p_{ab_{m-1}}^{(n)}$ is the (a, b_{m-1})–entry of the matrix P^n:

$$p_{ab_{m-1}}^{(n)} = \sum_{\xi_1,\ldots,\xi_{n-1}\in V} p_{a\xi_1} p_{\xi_1\xi_2} \cdots p_{\xi_{n-2}\xi_{n-1}} p_{\xi_{n-1}b_{m-1}}.$$

Equivalently, $p_{ab_{m-1}}^{(n)} = \Pr(X_n = b_{m-1} | X_0 = a)$, where $\{X_n\}_{n\ge0}$ is the Markov chain on G' with transition probabilities $\Pr(a \to b) = p_{ab}$.

Mixing countable state Markov chains fall into three classes: positive recurrent chains, null recurrent chains, and transient chains. To describe these cases, let ${}_a p_{aa}^{(n)} := \sum_{\xi_1,\ldots,\xi_n \in V\setminus\{a\}} p_{a\xi_1} p_{\xi_1\xi_2}\cdots p_{\xi_{n-2}\xi_{n-1}} p_{\xi_{n-1}a}$ (the probability that if the chain starts at a, then it will return to a *for the first time* at time n).

Positive recurrent behavior: $p_{ab}^{(n)} \xrightarrow[n\to\infty]{} p_b$, where $\underline{\pi} = (p_b)_{b\in V}$ is a stationary probability vector ($\underline{\pi}P = \underline{\pi}$). If we start a positive recurrent Markov chain at some state a, then it will return to a infinitely many times with full probability, and the average time till the first return to a is finite. The frequency of visits to a is positive. A Markov chain is positive recurrent iff $\sum\limits_{n=0}^{\infty} p_{aa}^{(n)} = \infty$ and $\sum\limits_{n=1}^{\infty} n \cdot {}_a p_{aa}^{(n)} < \infty$.

Null recurrent behavior: $p_{ab}^{(n)} \xrightarrow[n\to\infty]{} 0$ but $\sum_{n>0} p_{ab}^{(n)} = \infty$ and there is a sequence $a_n \uparrow \infty$, $a_n = o(n)$ s.t. for all states b, $\frac{1}{a_n}\sum_{k=0}^{n-1} p_{ab}^{(k)} \xrightarrow[n\to\infty]{} p_b$ where $\underline{\pi} = (p_b)_{b\in V}$ is a stationary positive vector s.t. $\sum \pi_b = \infty$. If we start a null recurrent Markov chain at a state a, then it will return to a infinitely many times with full probability, but the average time till the first return to a is infinite. In this case the asymptotic frequency of visits to a is zero. Null recurrence happens iff $\sum\limits_{n=0}^{\infty} p_{aa}^{(n)} = \infty$ and $\sum\limits_{n=1}^{\infty} n \cdot {}_a p_{aa}^{(n)} = \infty$.

Transient behavior: $\sum_{n>0} p_{ab}^{(n)} < \infty$. A transient Markov chain started at a state a has positive probability never to return to a. Transience is characterized by the condition $\sum\limits_{n=0}^{\infty} p_{aa}^{(n)} < \infty$.

All finite state Markov chains are positive recurrent. For examples of null recurrence and transience, consider the random walk on $\mathbb{Z}^d$. This is the Markov chain with set of states $\mathbb{Z}^d$, allowed transitions $\vec{v} \to \vec{v} \pm \vec{e}_i$, $\vec{e}_i = (\delta_{ik})_{k=1}^{d}$, and transition probabilities $1/2d$. The random walk on $\mathbb{Z}^d$ is null recurrent when $d = 1, 2$, and transient when $d \geq 3$.

The discussion above provides a full description of the asymptotic behavior of $L_\phi^k 1_{[\underline{b}]}$ as $k \to \infty$ in the case when ϕ is Markovian, and $L_\phi 1 = 1$. The theory extends to general potentials with summable variations:

Definition 4.8. Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is a function with summable variation and finite Gurevich pressure $P_G(\phi)$. For every state a, let $\varphi_a(\underline{x}) := 1_{[a]}(\underline{x}) \inf\{n \geq 1 : x_n = a\}$ and set

$$Z_n(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[a]}(\underline{x}) \text{ and } Z_n^*(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[\varphi_a = n]}(\underline{x}).$$

Let $\lambda := \exp P_G(\phi)$, and fix some state a. We say that

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

It can be shown that these definitions do not depend on the choice a.

$Z_n(\phi, a)$ and $Z_n^*(\phi, a)$ generalize $p_{aa}^{(n)}$ and ${}_a p_{aa}^{(n)}$: Summing over n–periodic $\underline{x} \in [a]$ is the same as summing over all paths $a \to x_1 \to \cdots \to x_{n-1} \to a$, and summing over n–periodic $\underline{x} \in [\varphi_a = n]$ is the same as summing over all paths $a \to \xi_1 \to \cdots \to \xi_{n-1} \to a$ such that $\xi_i \neq a$ for all i.

With this dictionary in mind, definition 4.8 is a translation of the characterization of the modes of recurrence of a Markov chain in terms of $p_{aa}^{(n)}$ and ${}_ap_{aa}^{(n)}$.

Theorem 4.9 (Generalized Ruelle's Perron-Frobenius Theorem). *Let X be a topologically mixing topological Markov shift, and let $\phi : X \to \mathbb{R}$ be a function with summable variations and finite Gurevich pressure.*

(1) ϕ is positive recurrent *iff there are $\lambda > 0$, a positive continuous function h, and a conservative measure ν which is finite and positive 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 $[\underline{a}]$,*

$$\lambda^{-n} L_\phi^n 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}] \text{ uniformly on compacts.}$$

(2) ϕ is null recurrent *iff there are $\lambda > 0$, a positive continuous function h, and a conservative measure ν which is finite and positive 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. There is a sequence $a_n \uparrow \infty$, $a_n = o(n)$ s.t. for all cylinders $[\underline{a}]$,*

$$\frac{1}{a_n}\sum_{k=0}^{n-1} \lambda^{-k} L_\phi^k 1_{[\underline{a}]} \xrightarrow[n\to\infty]{} h\nu[\underline{a}] \text{ uniformly on compacts.}$$

(3) ϕ is transient *iff there is no conservative measure ν, which is finite and positive on cylinders, such that $L_\phi^*\nu = \lambda\nu$ for some $\lambda > 0$. In this case $\sum e^{-nP_G(\phi)} L_\phi^n 1_{[\underline{a}]} < \infty$ pointwise for every cylinder $[\underline{a}]$.*

If X has finitely many states, every ϕ with summable variations is positive recurrent.

Theorem 4.10. *Under the assumptions of the previous theorem, if ϕ is positive recurrent, then there is a unique thermodynamic limit up to normalization, equal to ν. If in addition $\nu(X) < \infty$, then $\nu/\nu(X)$ is a DLR state.*

Proof. That ν is the thermodynamic limit follows from the expression of this property in terms of the Ruelle operator.

Now suppose $\nu(X) = 1$. To see that ν is a DLR state for ϕ, we show that the Jacobian of ϕ equals $\lambda^{-1}\exp\phi$. For every non-negative measurable $f : X \to \mathbb{R}$,

$$\int f\lambda^{-1}e^{\phi} d\nu\circ\sigma = \sum_{a\in V}\int_{\sigma[a]} f(a\underline{x})\lambda^{-1}e^{\phi(a\underline{x})}d\nu(\underline{x})$$
$$= \int \lambda^{-1}\sum_{a\in V} 1_{\sigma[a]}(\underline{x})e^{\phi(a\underline{x})}f(a\underline{x})d\nu(\underline{x}) = \int \lambda^{-1}(L_\phi f)d\nu = \int f d\nu,$$

because $L_\phi^*\nu = \lambda\nu$. Thus $\lambda^{-1}e^{\phi} = \frac{d\nu}{d\nu\circ\sigma}$ a.e., and ν is conformal for ϕ. As we saw in section 3, conformal probability measures are DLR states. □

Definition 4.11. Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is a positive recurrent function with finite Gurevich pressure and summable variations. The measure $dm = h d\nu$, where $L_\phi h = \lambda h$, $L_\phi^*\nu = \lambda\nu$, and $\int h d\nu = 1$ is called the *Ruelle-Perron-Frobenius (RPF) measure* of ϕ.

This is a shift invariant measure, because for every measurable non-negative function f, $L_\phi[(f\circ\sigma)h] = \lambda f h$, so $\int (f\circ\sigma)\, h d\nu = \int \lambda^{-1}L_\phi[(f\circ\sigma)h]d\nu = \int f\, h d\nu$. We will see in the next section that $h d\nu$ is the solution to the variational problem, whenever such a solution exists.

4.4. Notes and references. The Gurevich pressure first appeared in [**Gur69**], in the special case $\phi \equiv 0$. Gurevich extended the theory to general Markovian potentials in [**Gur78**]. In these works, the formula in Theorem 4.7 was used as the definition, and what we presented as a definition was proved as a theorem. The extension of the theory to non-Markovian potentials was done in [**Sar99**]. There, Theorems 4.3 and 4.7 are proved for locally Hölder potentials, but the same proof works for potentials with summable variations.

The Generalized Ruelle's Perron-Frobenius Theorem (Theorem 4.9) has a more complicated history.

When ϕ is Markovian, L_ϕ preserves the space $W = \text{span}\{1_{[a]} : a \in V\}$, and $L_\phi : W \to W$ is encoded by a non-negative $|V| \times |V|$ matrix. If $|V| < \infty$, then Theorem 4.9 reduces to the classical Perron–Frobenius Theorem. If V is countable, then Theorem 4.9 follows from Vere-Jones's generalization of the Perron-Frobenius theorem to countable positive matrices [**VJ67**], [**VJ68**], see also [**Gur78**], [**GS98**]. A central idea in Vere-Jones's papers is to model the analysis on the theory of countable state Markov chains, see e.g. [**Chu60**].

When ϕ is not Markovian and X is a compact TMS, Theorem 4.9 is due to Ruelle in [**Rue68**], [**Rue76**] under the assumption that ϕ is Hölder. Bowen extended the theorem to functions with summable variations, and gave it the name "Ruelle's Perron-Frobenius Theorem" [**Bow75**]. Walters extended the theorem to even larger classes of regularity [**Wal01**], and Pollicott extended it to complex valued potentials [**Pol86**]. All these results are for compact TMS, where only positive recurrent behavior is possible.

The generalization of Ruelle's Perron-Frobenius Theorem to non-compact topological Markov shifts was done in [**Sar01b**]. The proof is written there for weakly Hölder functions, but works verbatim under the weaker summable variations condition. For even weaker regularity, see [**Dao13**].

5. Pressure, equilibrium measures, and Gibbs measures in the sense of Bowen

5.1. The variational problem. Let T be a continuous map on a complete metric separable space (Y, d), and suppose $\phi : Y \to \mathbb{R}$ is Borel measurable.

Definition 5.1. The *variational pressure* of ϕ is $\sup\{h_\mu(T) + \int \phi d\mu\}$, where the supremum ranges over all T–invariant Borel probability measures for which $\int \phi d\mu$ makes sense and $h_\mu(T) + \int \phi d\mu \neq \infty - \infty$. The measures which attain the supremum are called *equilibrium measures* (for ϕ).

The *variational problem* is to find, for a given ϕ, the variational pressure of ϕ, and to determine its equilibrium measures. We will focus on TMS.

Remark 5.2. The word "pressure" is a relic of the first papers on the subject by Ruelle, which treated lattice gas models. Actually, the role of the "pressure" in thermodynamic formalism is much closer to the role of (minus) the *free energy* in thermodynamics, see Theorems 6.5, 7.4–7.6 and the discussion preceding them.

5.2. The variational pressure and the Gurevich pressure. How to calculate the variational pressure? Recall that we defined the Gurevich pressure of a function ϕ with summable variations to be

$$P_G(\phi) = \lim_{n\to\infty} \frac{1}{n} \ln Z_n(\phi, a), \text{ where } Z_n(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[a]}(\underline{x}).$$

THEOREM 5.3 (Variational Principle). *If X is a topologically mixing TMS and $\phi : X \to \mathbb{R}$ has summable variations, then*

$$P_G(\phi) = \sup\{h_\mu(\sigma) + \int \phi d\mu\}$$

where the supremum ranges over all shift invariant Borel probability measures μ s.t. ϕ is μ–integrable, and $(h_\mu(\sigma), \int \phi d\mu) \neq (\infty, -\infty)$.

In particular, if $P_G(\phi) < \infty$ and ϕ is positive recurrent, then $\sup\{h_\mu(\sigma) + \int \phi d\mu\}$ is an eigenvalue of Ruelle's operator.

5.3. Equilibrium measures. Our next task is to find the *equilibrium measures* of ϕ: the measures which bring $\sup\{h_\mu(\sigma) + \int \phi d\mu\}$ to a maximum.

Recall the definition of the *Ruelle Perron Frobenius (RPF) measure*: $dm = hd\nu$ where $L_\phi h = \lambda h$, $L^*_\phi \nu = \lambda \nu$, $\int hd\nu = 1$. This measure exists whenever ϕ is a positive recurrent function with summable variations and finite Gurevich pressure.

THEOREM 5.4. *Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is positive recurrent, with summable variations, and finite Gurevich pressure. If the RPF measure of ϕ has finite entropy, then it is an equilibrium measure for ϕ. This is always the case when X has finitely many states.*

For an example of an RPF measure with infinite entropy, take $X = \mathbb{N}^{\mathbb{N}}$ and $\phi(\underline{x}) = \ln p_{x_0}$, where $\underline{p} = (p_k)_{k\in\mathbb{N}}$ is a probability vector with infinite entropy. In this case $P_G(\phi) = 0$, $h \equiv 1$, ν =Bernoulli measure with probability vector $\underline{p}$, so the RPF measure is ν, and $h_\nu(\sigma) = -\sum p_i \ln p_i = \infty$, $\int \phi d\nu = \sum p_i \ln p_i = -\infty$. We see that $h_\nu(\sigma) + \int \phi d\nu = \infty - \infty$, and is meaningless.

THEOREM 5.5. *Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ has summable variations, $\sup \phi < \infty$, and $P_G(\phi) < \infty$, then*

1. *ϕ has at most one equilibrium measure.*
2. *This equilibrium measure, if it exists, is 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.*

THEOREM 5.6. *If X is a topologically mixing TMS and $\phi : X \to \mathbb{R}$ has summable variations, $\sup \phi < \infty$ and $P_G(\phi) < \infty$, then every equilibrium measure m of ϕ is exact, $h_m(\sigma) = -\int \log \frac{dm}{dm\circ\sigma} dm$, and the natural extension of m is Bernoulli.*

5.4. Gibbs measures in the sense of Bowen. Recall that a shift invariant probability measure m is called a *Gibbs measure (in the sense of Bowen)* if there are constants $M > 1$ and $P \in \mathbb{R}$ s.t. for every cylinder

$$M^{-1} \leq \frac{m[a_0, \ldots, a_{n-1}]}{\exp(\sum_{k=0}^{n-1} \phi(\sigma^k \underline{x}) - nP)} \leq M \text{ for all } \underline{x} \in [a_0, \ldots, a_{n-1}]. \tag{5.1}$$

THEOREM 5.7. *Suppose X is a topologically mixing TMS. If X is compact (i.e. the number of states is finite), then every $\phi : X \to \mathbb{R}$ has a unique Gibbs measure in the sense of Bowen. This measure is also the equilibrium measure of ϕ, and is equal to the RPF measure of ϕ. The P in (5.1) equals $P_G(\phi)$.*

When the number of states is infinite, the situation is more complicated: Gibbs measures can only exist on some TMS, but not on others. The combinatorial property which distinguishes the TMS which carry Gibbs measures is the following:

DEFINITION 5.8. A topological Markov shift is said to have the *big images and pre-images (BIP) property*, if there is a finite collection of states $b_1, \ldots, b_N$ s.t. for every state a there are i, j s.t. $a \to b_i$ and $b_j \to a$.

Every compact topological Markov shift has the BIP property (take $\{b_1, \ldots, b_N\}$ to be the full collection of states), but many non-compact TMS do not.

THEOREM 5.9. *Let X be a topologically mixing TMS. A function $\phi : X \to \mathbb{R}$ with summable variations possesses a Gibbs measure in the sense of Bowen iff the following three conditions hold:*

(1) $P_G(\phi) < \infty$
(2) $\mathrm{var}_1\phi < \infty$
(3) *X has the big images and pre-images property.*

In this case, ϕ is positive recurrent, the Gibbs measure m equals the RPF measure of ϕ, the P in (5.1) equals $P_G(\phi)$, and the natural extension of m is Bernoulli.

5.5. Notes and references. The variational principle for non-compact TMS (theorem 5.3) was proved in [**Sar99**] under the additional assumption that $\sup \phi < \infty$. This assumption was removed in [**IJT13**]. Theorems 5.4, 5.5 on equilibrium measures are from [**BS03**]. The following important cases were done before:

(1) Compact TMS, $\phi \equiv 0$ (Parry [**Par64**]). This was the first calculation of the measure of maximal entropy for a dynamical system.
(2) Compact TMS, ϕ Hölder continuous (Ruelle [**Rue67**], [**Rue76**] and Bowen [**Bow75b**]). These were the first papers to define topological pressure and equilibrium measures for dynamical systems.
(3) Non-compact TMS, ϕ Markovian (Gurevich [**Gur69**], [**Gur70**], [**Gur78**], [**Gur84**]). These were the first papers to treat the non-compact case.

See also [**MU01**], [**FFY02**], [**Yur03a**], [**Zar85**] and references therein.

The formula for the entropy of an equilibrium measure in Theorem 5.6 can be found in [**Led74**] for compact Markov shifts and in [**BS03**] for non-compact TMS (see also the appendix to [**CS09**]). The rest of the theorem follows from the results for conformal measures in §3. See the end of that section for further references.

Bowen introduced his notion of a Gibbs measure in [**Bow75**], and showed that such measures always exist for compact TMS. The ergodic and stochastic properties of Gibbs measures for non-compact TMS shifts were studied by Aaronson and Denker in [**AD01**], and Aaronson, Denker & Urbanski in [**ADU93**]. The BIP condition and Theorem 5.9 are from [**Sar03**]. The direction "BIP⇒existence" in Theorem 5.9 follows from earlier work of Mauldin & Urbanski [**MU01**]. For generalizations of Bowen's definition, see [**Yur00**], [**Yur03b**] and references therein.

Various authors considered the variational problem for general dynamical systems. Walters defined the *topological* pressure $P_{top}(\cdot)$ for continuous maps on general compact metric spaces, and showed that $P_{top}(\phi) = \sup\{h_\mu(T) + \int \phi d\mu\}$ for general continuous functions ϕ [**Wal75b**]. Bowen gave an alternative formula for $P_{top}(\phi)$ in terms of the metric structure of Y [**Bow75**]. For the special case $\phi \equiv 0$, see the earlier works [**AKM65**], [**Goo69**], [**Goo71**]. The general non-compact case is still not understood, except when ϕ has a "nice" continuous extension to some "nice" compactification of X, see [**Wal78**], [**Zar85**], [**GS98**]. In the absence of such conditions, various possible definitions of $P_{top}(\phi)$ have been suggested [**Bow73**], [**PP84**], which provide upper bounds for $\sup\{h_\mu(T)+\int\phi d\mu\}$, see also [**Tho11**]. For definitions of pressure in complex dynamics, see [**PRLS04**].

We turn to equilibrium measures. Upper-semi-continuous functions attain their maximum over compact sets. So, if a continuous map T on a *compact* metric space X has the property that the entropy map $\mu\mapsto h_\mu(T)$ is upper-semi-continuous with respect to the weak-star topology, then every continuous potential on X has an equilibrium measure. Sufficient conditions for the upper-semi-continuity of the entropy map were given by Misiurewicz [**Mis76**] and Newhouse [**New89**], see §8.1.

6. Strong positive recurrence and spectral gap

6.1. Spectral Gap Property. Ruelle's operator $(L_\phi f)(\underline{x})=\sum\limits_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})}f(\underline{y})$ has played a central role in our discussion of the variational problem and the thermodynamic limit. In this section we discuss a technical property of L_ϕ which provides detailed information on L_ϕ^n as $n\to\infty$.

Let X be a topologically mixing TMS, and let $\phi:X\to\mathbb{R}$ be a *weakly Hölder* continuous function with finite Gurevich pressure (recall that weak Hölder continuity means that for some $0<\theta<1$, $\mathrm{var}_n\phi\le A\theta^n$ for all $n\ge 2$). Let

$$\mathrm{dom}(L_\phi):=\{f:X\to\mathbb{R}:\sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})}f(\underline{y})\text{ converges absolutely for all }\underline{x}\}.$$

Definition 6.1. We say that ϕ has the *spectral gap property (SGP)*, if there is a Banach space $\mathscr{L}$ of continuous functions on X s.t.:

(1) $\mathscr{L}$ is "rich":
   (a) $\mathscr{L}\subset\mathrm{dom}(L_\phi)$ and $\mathscr{L}\supset\{1_{[\underline{a}]}:[\underline{a}]\neq\varnothing\}$
   (b) $f\in\mathscr{L}\Longrightarrow|f|\in\mathscr{L}$ and $\big\||f|\big\|_{\mathscr{L}}\le\|f\|_{\mathscr{L}}$
   (c) $\mathscr{L}$–convergence implies uniform convergence on cylinders.

(2) $L_\phi:\mathscr{L}\to\mathscr{L}$ has spectral gap:
   (a) $L_\phi(\mathscr{L})\subset\mathscr{L}$ and $L_\phi:\mathscr{L}\to\mathscr{L}$ is bounded
   (b) $L_\phi=\lambda P+N$ where $\lambda=\exp P_G(\phi)$, $PN=NP=0$, $P^2=P$, $\dim\mathrm{Im}P=1$, and the spectral radius of N is less than λ.

(3) If $g:X\to\mathbb{R}$ is weakly Hölder continuous and bounded, then $L_{\phi+zg}:\mathscr{L}\to\mathscr{L}$ is bounded, and $z\mapsto L_{\phi+zg}$ is holomorphic on some complex neighborhood U of zero: For all $z_0\in U$, $\lim\limits_{h\to 0}\frac{1}{h}(L_{\phi+(z_0+h)g}-L_{\phi+z_0 g})$ exists in the operator norm.

Property (2)(b) is an algebraic way of saying that $\lambda=e^{P_G(\phi)}$ is a simple eigenvalue of $L_\phi:\mathscr{L}\to\mathscr{L}$, and that the remainder of the spectrum lies in $\{z:|z|\le\rho\}$ where ρ (the spectral radius of N) is strictly smaller than λ. The relations $NP=PN=0$ and $P^2=P$ imply that P is the eigenprojection of λ, and $\lambda^{-n}L_\phi^n=\lambda^{-n}(\lambda P+N)^n=P+\lambda^{-n}N^n$. Since $\rho<\lambda$,

$$\|\lambda^{-n}L_\phi^n-P\|=\lambda^{-n}\|N^n\|\xrightarrow[n\to\infty]{}0\text{ exponentially fast.}$$

Thus (2)(b) implies that $\lambda^{-n}L_\phi^n$ converges exponentially fast in norm to P.

Property (3) is saying that $L_{\phi+zg}$ is an "analytic perturbation" of L_ϕ. Perturbation theory for linear operators says that isolated simple eigenvalues (such as λ) survive analytic perturbations, and vary analytically. Property (3) allows us to apply this theory to $P_G(\phi + zg) = \ln \lambda(L_{\phi+zg})$, for $|z|$ small.

6.2. Implications of the spectral gap property. Throughout this section we make the following assumptions: X is a topologically mixing TMS, ϕ is weakly Hölder continuous, $P_G(\phi) < \infty$, and ϕ has the spectral gap property. Let $L_\phi = \lambda P + N$ be the decomposition given by the SGP.

THEOREM 6.2 (Stable Positive Recurrence). *ϕ is positive recurrent, and P takes the form $Pf = h \int f d\nu$, where $L_\phi h = \lambda h$, $L_\phi^* \nu = \lambda \nu$, and $\int h d\nu = 1$. Moreover, for every weakly Hölder continuous bounded function $g : X \to \mathbb{R}$, there is an $\epsilon > 0$ s.t. $\phi + tg$ is positive recurrent for all $|t| < \epsilon$.*

THEOREM 6.3 (Exponential decay of correlations). *Let m denote the RPF measure of ϕ. There exists $0 < \kappa < 1$ such that for every $f : X \to \mathbb{R}$ bounded Hölder continuous and $g \in L^\infty(m)$ there is a constant $C(f, g)$ s.t.*

$$\left| \int f(g \circ \sigma^n) dm - \int f dm \int g dm \right| \leq C(f, g)\kappa^n \text{ for all } n \geq 1.$$

THEOREM 6.4 (Central Limit Theorem). *Let m denote the RPF measure of ϕ, and suppose $g : X \to \mathbb{R}$ is a bounded Hölder continuous function. If $\int g dm = 0$ and g cannot be put in the form $\varphi - \varphi \circ \sigma$ with φ continuous, then there is a positive constant $\sigma_\phi(g)$ s.t. for every $t \in \mathbb{R}$*

$$m\left\{ \underline{x} \in X : \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} g(\sigma^k(\underline{x})) < t \right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi\sigma_\phi^2(g)}} \int_{-\infty}^{t} e^{-s^2/2\sigma_\phi^2(g)} ds.$$

The identity $P_G(\phi) = \sup\{h_\mu(\sigma) + \int \phi d\mu\}$ represents $P_G(\phi)$ as a dynamical analogue of (minus) the Helmholtz free energy F. The next theorem is a translation of a well-known property of the free energy into the language of dynamics.

First some background from thermodynamics. The functional dependence of F on its natural parameters completely characterizes the thermodynamic state in equilibrium: All thermodynamic quantities can be written as first order partial derivatives of F with respect to a suitable parameter.

If first-order partial derivatives of F are thermodynamic quantities, then second-order partial derivatives of F equal the rate of change of thermodynamic quantities when an external parameter is changed. Such quantities are called "linear response functions." The "linear response theorem" relates the linear response functions to the fluctuations of the corresponding microscopically defined quantities.

The following theorem shows that something similar happens in the world of dynamical systems:

THEOREM 6.5 (Derivatives of Pressure). *Let m denote the RPF measure of ϕ, and suppose $g : X \to \mathbb{R}$ is a bounded Hölder continuous function, then $t \mapsto P_G(\phi + tg)$ is analytic on a neighborhood of zero, and*

$$\left.\frac{d}{dt}\right|_{t=0} P_G(\phi + tg) = \int g dm \ , \ \left.\frac{d^2}{dt^2}\right|_{t=0} P_G(\phi + tg) = \sigma_\phi^2(g),$$

where $\sigma_\phi(g)$ is given by the previous theorem.

Thus the functional $\phi \mapsto P_G(\phi)$ completely characterizes the RPF measure m through its directional derivatives, and it relates the second order directional derivatives of $P_G(\cdot)$ to the fluctuations of the Birkhoff sums with respect to this measure. We see that $P_G(\cdot)$ behaves like a dynamical "free energy."

6.3. Strong positive recurrence. We give a necessary and sufficient condition for the spectral gap property, and then we discuss the prevalence of this condition.

The condition relies on the *induction procedure*, which we now explain. Suppose X is a topologically mixing TMS, and a is some state. The *induced system* on $[a]$ is $\sigma_a : X_a \to X_a$, where $X_a := \{\underline{x} \in X : x_0 = a, x_i = a \text{ infinitely often}\}$ and $\sigma_a(\underline{x}) := \sigma^{\varphi_a(\underline{x})}(\underline{x})$, where $\varphi_a(\underline{x}) := \min\{n \geq 1 : x_n = a\}$.

The resulting transformation can be given the structure of a TMS as follows: Let $\overline{S} := \{[a, \xi_1, \ldots, \xi_{n-1}, a] : n \geq 1, \xi_1, \ldots, \xi_{n-1} \neq a\} \setminus \{\varnothing\}$ and let $\overline{\sigma} : \overline{X} \to \overline{X}$ denote the left shift on $\overline{X} = (\overline{S})^{\mathbb{N}}$, then $\overline{\sigma} : \overline{X} \to \overline{X}$ is topologically conjugate to $\sigma_a : X_a \to X_a$. The conjugacy $\overline{\pi} : \overline{X} \to X_a$ is given by

$$\overline{\pi}([a, \underline{\xi}^0, a], [a, \underline{\xi}^1, a], [a, \underline{\xi}^2, a], \cdots) := (a, \underline{\xi}^1, a, \underline{\xi}^2, a, \underline{\xi}^3, a, \ldots).$$

Functions $\phi : X \to \mathbb{R}$ can also be induced: The *induced potential* $\overline{\phi} : \overline{X} \to \mathbb{R}$ is

$$\overline{\phi} := \left(\sum_{i=0}^{\varphi_a - 1} \phi \circ \sigma^i \right) \circ \overline{\pi}.$$

If ϕ is weakly Hölder continuous, then $\overline{\phi}$ is locally Hölder continuous (Lemma 2.6). Notice that if we induce ϕ on $[a]$, then for all $p \in \mathbb{R}$, $\overline{\phi + p} = \overline{\phi} + p\varphi_a \geq \overline{\phi} + p$.

Definition 6.6. Suppose X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is weakly Hölder continuous with finite Gurevich pressure. The *a–discriminant* of ϕ is $\Delta_a[\phi] := \sup\{P_G(\overline{\phi + p}) : p \in \mathbb{R} \text{ s.t. } P_G(\overline{\phi + p}) < \infty\}$.

Theorem 6.7 (Discriminant Theorem). *Let X be a topologically mixing TMS, and suppose $\phi : X \to \mathbb{R}$ is a weakly Hölder continuous function s.t. $P_G(\phi) < \infty$. For every state a,*

1. $\Delta_a[\phi] > 0$ *iff ϕ is positive recurrent with the spectral gap property,*
2. $\Delta_a[\phi] = 0$ *iff ϕ is null recurrent or ϕ is positive recurrent without the spectral gap property,*
3. $\Delta_a[\phi] < 0$ *iff ϕ is transient.*

In case (1) we call ϕ strongly positive recurrent*.*

Suppose ϕ is strongly positive recurrent. We will describe a Banach space $\mathscr{L}$ where L_ϕ acts with spectral gap. We need the following observation:

Lemma 6.8. *If ϕ is weakly Hölder continuous, positive recurrent, and with finite Gurevich pressure, then $P_G(\phi) = 0 \Leftrightarrow P_G(\overline{\phi}) = 0$.*

Let ϕ be a weakly Hölder continuous potential with finite Gurevich pressure. Assume without loss of generality that $P_G(\phi) = 0$ (otherwise work with $\phi - P_G(\phi)$). By weak Hölder continuity, there is $0 < \theta < 1$ (which we now fix once and for all) s.t. $\text{var}_n\phi \leq \text{const.}\theta^n$ for all $n \geq 2$. By strong positive recurrence, there is a state a s.t. $\Delta_a[\phi] > 0$, therefore there exists $\epsilon > 0$ small s.t. $0 < P_G(\overline{\phi + 2\epsilon}) < \infty$. By the lemma, $P_G(\phi) = 0 \Rightarrow P_G(\overline{\phi}) = 0$. Since $\epsilon \mapsto P_G(\overline{\phi + \epsilon})$ is convex, it is continuous, and one can choose ϵ small enough so that $0 < \theta e^p < 1$ for $p := P_G(\overline{\phi + \epsilon})$.

Let $\psi := \phi + \epsilon - p1_{[a]}$, then $P_G(\overline{\psi}) = P_G(\overline{\phi+\epsilon}) - p = 0$. Also, $P_G(\overline{\psi+\epsilon}) < \infty$ and $P_G(\overline{\psi+\epsilon}) \geq P_G(\overline{\psi}) + \epsilon > 0$. Consequently, $\Delta_a[\psi] > 0$, whence ψ is (strongly) positive recurrent. By the lemma, $P_G(\psi) = 0$, and by the generalized Ruelle's Perron-Frobenius theorem, there is a continuous function h_0 s.t. $L_\psi h_0 = h_0$.

Using the weak Hölder continuity of ψ and the convergence $L^n_\psi 1_{[a]} \xrightarrow[n\to\infty]{} \nu[a]h_0$, it is easy to check that $\mathrm{var}_1(\ln h_0) < \infty$. This allows us to define

$$h_0[b] := \sup_{[b]} h_0$$

for all states b. We will use these numbers to modulate the size and smoothness of the elements of our Banach space on the partition sets $[b]$.

Given $\underline{x}, \underline{y} \in X$, let $t(\underline{x}, \underline{y}) := \min\{n \geq 0 : x_n \neq y_n\}$, and $s_a(\underline{x}, \underline{y}) := \#\{0 \leq i \leq t(x,y) - 1 : x_i = y_i = a\}$. Define for a function $f : X \to \mathbb{C}$,

$$\|f\|_{\mathscr{L}} := \sup_b \frac{1}{h_0[b]} \left[\sup_{[b]} |f| + \sup \left\{ |f(\underline{x}) - f(\underline{y})| / \theta^{s_a(\underline{x},\underline{y})} : \underline{x}, \underline{y} \in [b], \underline{x} \neq \underline{y} \right\} \right].$$

THEOREM 6.9. *L_ϕ acts with the spectral gap property on $(\mathscr{L}, \|\cdot\|_{\mathscr{L}})$, where $\mathscr{L} := \{f : X \to \mathbb{C} : \|f\|_{\mathscr{L}} < \infty\}$.*

We finish this section with two useful facts on the discriminant. The first is a consequence of the obvious fact that if $\psi = 1_{[a]}$, then the induced version of ψ on $[a]$ is $\overline{\psi} \equiv 1$. Using this observation it is easy to see that

$$\Delta_a[\phi + t1_{[a]}] = \Delta_a[\phi] + t. \tag{6.1}$$

In particular, if $\Delta_a[\phi] < \infty$, then the one parameter family of potentials $\phi + t1_{[a]}$ exhibits for some parameters transience, and for other parameters recurrence.

The second fact we would like to mention is a useful estimate for $\Delta_a[\phi]$. Recall that $Z^*_n(\phi, a) = \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[\varphi_a=n]}(\underline{x})$.

THEOREM 6.10. *Suppose X is a topologically mixing TMS and ϕ has summable variations and finite Gurevich pressure. For every a, either $\Delta_a[\phi] = \infty$ or*

$$\left| \Delta_a[\phi] - \ln \sum_{n=1}^\infty R^n Z^*_n(\phi, a) \right| \leq \sum_{n=2}^\infty \mathrm{var}_n \phi,$$

where R is the radius of convergence of the series $\sum t^n Z^*_n(\phi, a)$.

In particular, if $\sum_{n=1}^\infty t^n Z^*_n(\phi, a)$ diverges at its radius of convergence, then ϕ has the spectral gap property. Since it is rather common for a positive power series to diverge at its radius of convergence, it seems likely that the spectral gap property happens "often." In the next section we will investigate this further.

6.4. How common is spectral gap? We mention some particular cases when the spectral gap property holds, and then discuss the general case.

THEOREM 6.11. *Suppose X is a topologically mixing TMS with finitely many states, then every weakly Hölder continuous potential has the spectral gap property. In this case one can use the Banach space*

$$\mathscr{L} = \left\{ f : X \to \mathbb{C} : \|f\| := \max |f| + \sup \left\{ |f(\underline{x}) - f(\underline{y})| / \theta^{t(\underline{x},\underline{y})} \right\} < \infty \right\},$$

where θ is a Hölder exponent of ϕ.

Recall that a topologically mixing topological Markov shift has the *big images and pre-images (BIP) property*, if there is a finite collection of states $b_1, \ldots, b_N$ s.t. for every state a there are i, j s.t. $b_i \to a$ and $a \to b_j$.

THEOREM 6.12. *Suppose X is a topologically mixing TMS with the BIP property, then every weakly Hölder continuous potential ϕ with finite Gurevich pressure and such that $\mathrm{var}_1\phi < \infty$ has the spectral gap property. In this case one can use the Banach space*

$$\mathscr{L} = \{f : X \to \mathbb{C} : \|f\| := \sup|f| + \sup_{B\in\beta}\sup_{\underline{x},\underline{y}\in B} |f(\underline{x}) - f(\underline{y})|/\theta^{t(\underline{x},\underline{y})} < \infty\},$$

where θ is the Hölder exponent of ϕ and β is the smallest partition whose σ–algebra contains $\{\sigma[a] : a \text{ is a state}\}$.

Next we characterize the topological Markov shifts for which *every* weakly Hölder continuous potential with finite Gurevich pressure has the spectral gap property (including potentials for which $\mathrm{var}_1\phi = \infty$).

DEFINITION 6.13. Let $G = G(V, E)$ be a directed graph. A subset $F \subset V$ is called a *uniform Rome* if every path of length N in G contains at least one vertex in F ("all roads lead to Rome, in less than N steps").

THEOREM 6.14. *A topologically mixing TMS X has the property that all weakly Hölder continuous potentials on X have the spectral gap property, iff its associated graph has a finite uniform Rome.*

Notice that while every finite graph has a finite uniform Rome (equal to the full set of vertices), this property is rare for infinite graphs. Graphs with "infinite rays" $a_1 \to a_2 \to a_3 \to \cdots$ with a_i distinct do *not* have finite uniform Romes. Thus Theorem 6.14 says that while for all compact TMS all "reasonable" potentials have spectral gap, in the non-compact case, barring very pathological combinatorial scenarios, there will exist some potentials *without* the spectral gap property.

Next we try to determine the topological "size" of the set of strongly positive recurrent potentials within

$$\Phi := \{\phi : X \to \mathbb{R} : \phi \text{ is weakly Hölder continuous, and } P_G(\phi) < \infty\}.$$

Several topologies come to mind. To define them efficiently, we fix $\omega = (\omega_n)_{n\geq 1}$, where $0 \leq \omega_n \leq \infty$, and let $\|f\|_\omega := \sup|f| + \sum_{n=2}^{\infty}\omega_n \mathrm{var}_n(f)$, where $0\cdot\infty := 0$.

DEFINITION 6.15. The *ω–topology* on Φ is the topology generated by the *basic ω–neighborhoods* $V(\phi, \epsilon) := \{\phi' \in \Phi : \|\phi - \phi'\|_\omega < \epsilon\}$.

THEOREM 6.16. *The set of $\phi \in \Phi$ with the spectral gap property is open and dense in Φ with respect to every ω–topology.*

In particular, it is open in the sup–norm topology $\omega = (0, 0, 0, \ldots)$, and dense in the θ–Hölder topology $\omega = (0, \theta^{-1}, \theta^{-2}, \theta^{-3}, \ldots)$, $0 < \theta < 1$.

We see that "most" potentials has spectral gap. But, because of Theorem 6.14, except in very strange cases (finite uniform Rome), some potentials do not. We will see in the next section that breakdown of the spectral gap property opens the way to critical phenomena similar to what one observes in a thermodynamic system undergoing a phase transition. Viewed from this perspective, Theorems 6.14 and 6.16 are in the spirit of the physical intuition that thermodynamic systems exhibit critical behavior for *some* configuration of external parameters, but that the set of the "critical parameters" where this happens is small.

6.5. Notes and references. For references on spectral gaps and their dynamical implications, see [**PP90**], [**Aar97**], [**Bal00**] or [**HH01**]. Proofs and references to the results in §6.2 in the specific case of TMS can be found in [**CS09**]. For a discussion of the connection between the central limit theorem and the derivatives of the pressure, see [**Rue78**]. For other stochastic implications of spectral gap such as almost sure invariance principles and local limit theorems see [**RE83**], [**DP84**], [**GH88**], [**AD01**], [**Gou10a**], [**Gou10b**].

Gurevich and Savchenko characterized the stability of positive recurrence for Markovian potentials in [**GS98**]. The discriminant was introduced in [**Sar01a**], as a tool for characterizing stable positive recurrence for potentials with summable variations. The discriminant theorem (Theorem 6.7) is taken from [**Sar01a**], except for the equivalence $\Delta > 0 \Leftrightarrow$Spectral Gap Property, which was shown in [**CS09**]. Lemma 6.8 and Theorem 6.10 are from [**Sar01a**].

The construction of Banach spaces with spectral gap for averaging operators similar to L_ϕ has a long history, starting with the paper of Doeblin & Fortet [**DF37**]. Theorem 6.11 is due to Ruelle [**Rue67**], [**Rue76**], see also [**PP90**]. Theorem 6.12 is due to Aaronson and Denker [**AD01**]. The Banach space which demonstrates the spectral gap property for general strongly positive recurrent potentials was constructed in [**CS09**], and was motivated by Young [**You98**].

The uniform Rome condition and Theorem 6.14 are due to Cyr [**Cyr11**]. Theorem 6.16 on the genericity of the spectral gap property is shown in [**CS09**].

7. Absence of spectral gap and critical phenomena

7.1. Changes in mode of recurrence. Throughout this section we assume that X is a topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is a positive recurrent weakly Hölder continuous function with finite Gurevich pressure but *without* the spectral gap property.

By the discriminant theorem, such potentials have zero discriminant, and by (6.1) $\Delta_a[\phi + t1_{[a]}] = t$. It follows that $\phi + t1_{[a]}$ is transient for all $t < 0$, and positive recurrent (with the spectral gap property) for $t > 0$. Thus the one-parameter family $\phi + t1_{[a]}$ exhibits a change in the mode of recurrence.

There are examples of TMS X and potentials ϕ s.t. the one-parameter family $\{\beta\phi\}_{\beta>0}$ changes its mode of recurrence infinitely many times, or stay "stuck" in the "critical" phase $\Delta_a = 0$ on a full interval of parameters. I am not aware of any restrictions on the possible behavior in the general case.

Changes in mode of recurrence can often result in non-analyticity for the pressure function. The following theorem gives the mechanism. Recall the definition and notation for the induced potential from the previous section.

Theorem 7.1. *Suppose X is a topologically mixing TMS, ϕ has summable variations and finite Gurevich pressure, and fix some state a.*

1. *The equation $P_G(\overline{\phi + p}) = 0$ has a unique solution $p = p(\phi)$ if $\Delta_a[\phi] \geq 0$ and no solution if $\Delta_a[\phi] < 0$.*
2. *If $\Delta_a[\phi] \geq 0$, then $P_G(\phi) = -p(\phi)$.*

(3) *If* $\Delta_a[\phi] < 0$*, then* $P_G(\phi) = \limsup_{n\to\infty} \frac{1}{n}\log Z_n^*(\phi, a)$.

Thus when a one-parameter family ϕ_t changes its mode of recurrence from transient to recurrent, the formula for $P_G(\phi_t)$ changes from $-p(\phi)$ to $\limsup_{n\to\infty} \frac{1}{n}\log Z_n^*(\phi_t, a)$. This can result in lack of analyticity for $t \mapsto P_G(\phi_t)$.

7.2. Statistical implications of non-analyticity of the pressure. Recall that the Gurevich pressure is an analogue of the free energy, and the free energy F is a "thermodynamic potential": Its first partial derivatives are thermodynamic quantities. If F is not differentiable in some direction, then some thermodynamic quantity (a partial derivative) is discontinuous. This is called a "first order phase transition." If the free energy is differentiable, but one of its second order partial derivatives blows up, then some linear response function explodes. This is called a "second order phase transition." In both cases, the precise type of singularity carries information on the behavior at the phase transition.

It is natural to look for a similar theory in the world of dynamical systems. First some preparations from probability theory. Let Y_n denote a sequence of real random variables, possibly on different probability spaces $(\Omega_n, \mathscr{F}_n, \mathrm{Pr}_n)$.

Definition 7.2. Y_n/n *converges exponentially in distribution to* $y_0 \in \mathbb{R}$, if for every $\epsilon > 0$ there is some $I(\epsilon) > 0$ s.t. $\mathrm{Pr}_n[|Y_n/n - y_0| > \epsilon] \le e^{-I(\epsilon)n}$ for all n large enough. In this case we write $Y_n/n \xrightarrow[n\to\infty]{\exp} y_0$.

Theorem 7.3. *Suppose* $\mathbb{E}(e^{tY_n}) < \infty$ *for all* $n \in \mathbb{N}, t \in \mathbb{R}$*, and assume that the limit* $F(t) := \lim_{n\to\infty} \frac{1}{n}\ln \mathbb{E}(e^{tY_n})$ *exists and is finite for all* $t \in \mathbb{R}$*, then* $Y_n/n \xrightarrow[n\to\infty]{exp} y_0$ *iff* $F(t)$ *is differentiable at zero and* $F'(0) = y_0$.

Here is an application to our context. Suppose X is topologically mixing, and $\phi : X \to \mathbb{R}$ has summable variations and finite Gurevich pressure. Fix a state a and consider the sequence of measures defining the thermodynamic limit with boundary condition $\underline{x}$, conditioned on $[a]$:

$$\mathrm{Pr}_n = \frac{\sum_{\sigma^n(\underline{y})=\underline{x}} e^{\phi_n(\underline{y})} 1_{[a]}(\underline{y})\delta_{\underline{y}}}{(L_\phi^n 1_{[a]})(\underline{x})} \tag{7.1}$$

Next we fix a *bounded* $\psi : X \to \mathbb{R}$ with summable variations, and consider the distribution of $Y_n := \psi_n = \sum_{k=0}^{n-1} \psi\circ\sigma^k$ with respect to Pr_n. Observe that

$$\mathbb{E}(e^{tY_n}) = (L^n_{\phi+t\psi}1_{[a]})(\underline{x})/(L^n_\phi 1_{[a]})(\underline{x}),$$

which is finite for all t because ψ is bounded and ϕ is admissible. By lemma 4.4,

$$F(t) := \lim_{n\to\infty} \frac{1}{n}\ln \mathbb{E}(e^{tY_n}) = P_G(\phi + t\psi) - P_G(\phi).$$

We obtain the following interpretation of differentiability for the pressure function:

Theorem 7.4. *If* $t \mapsto P_G(\phi + t\psi)$ *is differentiable at zero, then* $\frac{1}{n}Y_n$ *converges exponentially in distribution to a constant* c*, and if* $t \mapsto P_G(\phi + t\psi)$ *is not differentiable at zero, then it doesn't. The value of* c *is* $\frac{d}{dt}\big|_{t=0} P_G(\phi + t\psi)$.

Next we consider a case when $t \mapsto P_G(\phi + t\psi)$ is differentiable at zero, but not twice differentiable there.

In the physics literature, it is often assumed that the singularity is a power law singularity, and we will discuss the implications of this assumption. We specify our assumptions more precisely. A positive Borel function $L : (a, \infty) \to (0, \infty)$ is called *slowly varying (s.v.)* (at infinity), if $L(st)/L(t) \xrightarrow[t\to\infty]{} 1$ for all $s > 0$. A sequence $\ell(n)$ is *slowly varying*, if $L(x) := \ell(\lfloor x \rfloor)$ is s.v. at infinity. Typical examples include $L(t) = \text{const.}$, $L(t) = \ln t$, $L(t) = 1/\ln t$ etc.

We will consider the situation when $P_G(\phi + t\psi) = a + ct + t^\alpha L(1/t)$ with $L(t)$ s.v. at infinity. We'll take $\alpha > 1$ to guarantee differentiability at zero, and $\alpha < 2$ to guarantee that the second derivative has a singularity at 0^+.

Our results are simplest to state in the special case when $\psi := 1_{[a]}$ and $P_G(0) < \infty$, although these assumptions can be significantly weakened. See the end of the section for references to more general results.

Let $X_a := \{\underline{x} \in [a] : x_n = a \text{ infinitely often}\}$, and define for every $\underline{x} \in X_a$, $\varphi_a(\underline{x}) := 1_{[a]}(\underline{x}) \inf\{n \geq 1 : x_n = a\}$ and $\sigma_a(\underline{x}) = \sigma^{\varphi_a(\underline{a})}(\underline{x})$. By the Kac formula, $\int \varphi_a d\mu = 1$ for every ergodic shift invariant probability measure μ s.t. $\mu[a] \neq 0$.

THEOREM 7.5. *Suppose X is topologically mixing TMS s.t. $P_G(0) < \infty$ and let ϕ be a locally Hölder continuous function s.t. $\sup \phi < \infty$ and with an equilibrium measure μ_ϕ. The following are equivalent for every state a and $1 < \alpha < 2$:*

(1) $P_G(\phi + t1_{[a]}) = P_G(\phi) + ct + t^\alpha L(1/t)$ *as $t \to 0^+$, with L s.v. at infinity.*

(2) $\frac{1}{B_n}\left(\sum_{k=0}^{n-1} \varphi_a \circ \sigma_a^k - cn\right) \xrightarrow[n\to\infty]{} cG_\alpha$ *where $B_n = n^{1/\alpha}\ell(n)$ and ℓ is s.v.,* $c = \frac{1}{\mu_\phi[a]}\int_{[a]} \varphi_a d\mu_\phi$, *and G_α is the stable law s.t.* $\int e^{-s\xi} G_\alpha(d\xi) = e^{s^\alpha}$.

(3) $\mu_\phi[\varphi_a(\underline{x}) > t] = [1 + o(1)]\frac{\mu_\phi[a]^{-\alpha}}{|\Gamma(1-\alpha)|} t^{-\alpha} L(t)$ *as $t \to \infty$.*

Thus the theorem expresses a power law singularity for $P_G(\phi + t1_{[a]})$ in terms of abnormal fluctuations and heavy tails for the gaps between appearances of a.

Next we consider the effect of singular expansion on the decay of correlations. The *covariance* of two functions $f, g \in L^2(\mu)$ is $\text{Cov}_\mu(f, g) := \int fgd\mu - \int fd\mu \int gd\mu$.

THEOREM 7.6. *Under the assumptions of Theorem 7.5, if $P_G(\phi + t1_{[a]}) = P_G(\phi) + ct + t^\alpha L(1/t)$ as $t \to 0^+$, where $1 < \alpha < 2$ and L is s.v. at infinity, then*

$$\text{Cov}_{\mu_\phi}(f_1, f_2 \circ \sigma^n) = \frac{1 + o(1)}{\mu_\phi[a]^\alpha \Gamma(2 - \alpha)} \frac{L(n)}{n^{\alpha-1}} \int f_1 d\mu_\phi \int f_2 d\mu_\phi$$

for all f_1, f_2 weakly Hölder continuous s.t. $\text{var}_1 f_i < \infty$, f_i are supported inside a finite union of cylinders, and $\int f_i d\mu_\phi > 0$.

The theorem follows from a general asymptotic formula for the decay of correlations, discussed in the following section.

7.3. Asymptotic behavior of L_ϕ^n in the absence of spectral gap. Suppose X is a general topologically mixing TMS, and $\phi : X \to \mathbb{R}$ is a positive recurrent weakly Hölder continuous with exponent θ and finite Gurevich pressure $P_G(\phi) = \ln \lambda$. Let $\mu_\phi := hd\nu$ denote the RPF measure of ϕ. Fix a state a and define a Banach algebra of functions on $[a]$ by

$$\mathscr{L}_a := \left\{ f : [a] \to \mathbb{C} : \|f\|_{\mathscr{L}_a} := \sup |f| + \sup_{\underline{x}, \underline{y} \in [a]} \frac{|f(\underline{x}) - f(\underline{y})|}{\theta^{s_a(\underline{x}, \underline{y})}} < \infty \right\},$$

where $s_a(\underline{x},\underline{y}) := \#\{0 \leq i \leq t(\underline{x},\underline{y}) - 1 : x_i = y_i = a\}$, $t(\underline{x},\underline{y}) := \min\{i : x_i \neq y_i\}$. Every $f \in \mathscr{L}_a$ extends to a function $f1_{[a]}$ on X.

THEOREM 7.7. *If $\mu_\phi[\varphi > n] = O(n^{-\beta})$ with $\beta > 1$, then for every $f \in \mathscr{L}_a$ there are functions $\epsilon_n \in \mathscr{L}_a$ s.t.*

$$\lambda^{-n} L_\phi^n(f1_{[a]}) = h\int_{[a]} f d\nu + h\sum_{k=n+1}^{\infty} \mu_\phi[\varphi > k] \int_{[a]} f d\nu + \epsilon_n \text{ on } [a],$$

where $\|\epsilon_n\|_{\mathscr{L}_a} = O(n^{-\beta})$ when $\beta > 2$, $\|\epsilon_n\|_{\mathscr{L}_a} = O(n^{-2}\ln n)$ when $\beta = 2$, and $\|\epsilon_n\|_{\mathscr{L}_a} = O(n^{-2(\beta-1)})$ when $1 < \beta < 2$.

The following theorem follows immediately, once we recall that $\ln h$ is locally Hölder continuous and therefore $h, h^{-1} \in \mathscr{L}_a$.

THEOREM 7.8. *If $\mu_\phi[\varphi_a > n] = O(1/n^\beta)$ for some $\beta > 1$, then there exist $\theta' \in (0,1)$ such that for all $f, g : X \to \mathbb{R}$ bounded, Hölder continuous with exponent θ', and supported inside $[a]$ there is a constant $C = C(f,g)$ s.t.*

$$\left| \mathrm{Cov}_{\mu_\phi}(f, g \circ \sigma^n) - \left(\sum_{k=n+1}^{\infty} \mu_\phi[\varphi_a > k] \right) \int f d\mu_\phi \int g d\mu_\phi \right| \leq C F_\beta(n),$$

where $F_\beta(n) = n^{-\beta}$ if $\beta > 2$, $F_\beta(n) = n^{-2}\ln n$ if $\beta = 2$, and $F_\beta(n) = n^{-2(\beta-1)}$ if $1 < \beta < 2$. In particular, if $C_1 n^{-\beta} \leq \mu_\phi[\varphi_a > n] \leq C_2 n^{-\beta}$, then

$$\mathrm{Cov}_{\mu_\phi}(f, g \circ \sigma^n) \sim \left(\sum_{k=n+1}^{\infty} \mu_\phi[\varphi_a > k] \right) \int f d\mu_\phi \int g d\mu_\phi \text{ as } n \to \infty.$$

7.4. Notes and references. Theorem 7.1 is from [**Sar01a**]. There, one can also find examples of one-parameter families $\{\beta\phi\}_{\beta>0}$ with complicated changes in mode of recurrence. For other examples of dynamical systems undergoing a "phase transition", see [**FF70**], [**Hof77**], [**Wan89**], [**PS92**], [**Lop93**], [**BK98**], [**Sar00**], [**Hu08**], [**PZ06**] (this is a partial list). For a description of the critical phenomena encountered in thermodynamic systems, see e.g. [**Sta71**].

For a discussion of exponential convergence in distribution and a proof of Theorem 7.3, see [**Ell06**]. For the theory of slowly varying functions see [**BGT89**].

Theorem 7.5 is implicit in [**Sar06**] (see the discussion on pages 635–636). That paper also contains versions of Theorems 7.5 and 7.6 for a more general class of ψ and X. Theorems 7.7 and 7.8 are due to Gouëzel [**Gou04**], and improve earlier results in [**Sar02**]. For other results on subexponential decay of correlations which apply to topological Markov shifts, see [**You99**] and [**Pol00**].

We focused in this section of stochastic implications of singular pressure functions. For consequences in multifractal analysis and geometric measure theory, see [**PW99**], [**Iom05**], [**Iom10**], [**BI11**] and references therein.

8. Application to surface diffeomorphisms

8.1. Symbolic dynamics for surface diffeomorphisms. Suppose $f : M \to M$ is a $C^{1+\epsilon}$ diffeomorphism on a compact smooth orientable surface. Assume that f possesses invariant measures with positive entropy.

THEOREM 8.1. *For every $\delta > 0$, no matter how small, there exists a locally compact two-sided TMS $\sigma : \Sigma \to \Sigma$ with alphabet S, and there exists a Hölder continuous map $\pi : \Sigma \to M$ with the following properties:*

(1) $\pi \circ \sigma = f \circ \pi$.
(2) $\pi(\Sigma)$ *has full measure for every ergodic f–invariant μ s.t. $h_\mu(f) > \delta$.*
(3) $\pi : \Sigma^\# \to M$ *is finite-to-one (but perhaps not bounded-to-one), where*
$$\Sigma^\# = \{\underline{x} : \{x_i\}_{i<0}, \{x_i\}_{i>0} \text{ have constant susbsequences}\}.$$
(4) *Moreover, $\exists C_{ab}$ $(a, b \in S)$ s.t. $\#\{\underline{x} \in \Sigma^\# : \pi(\underline{x}) = p\} < C_{ab}$ for every $p \in M$ s.t. $p = \pi(\underline{x})$ where $x_i = a$ for infinitely many $i < 0$ and $x_i = b$ for infinitely many $i > 0$.*
(5) *For every ergodic f–invariant measure μ on M s.t. $h_\mu(f) > \delta$, there exists a shift invariant ergodic measure $\widehat{\mu}$ on Σ s.t. $\mu = \widehat{\mu} \circ \pi^{-1}$ and $h_{\widehat{\mu}}(\sigma) = h_\mu(f)$.*

The theorem allows to reduce some questions on the thermodynamic formalism for $f : M \to M$ to questions on the thermodynamic formalism for $\sigma : \Sigma \to \Sigma$. We explore some of the applications in the following sections.

8.2. Growth of periodic points. Let $f : M \to M$ denote a $C^{1+\epsilon}$–surface diffeomorphism on a compact surface M. A *periodic point* is a point $p \in M$ s.t. $f^n(p) = p$. Let $\mathrm{Per}_n(f) := \#\{p \in M : f^n(p) = p\}$.

Recall that the *topological entropy* of a surface diffeomorphism equals $h = \sup h_\mu(f)$, where the supremum is taken over all f–invariant probability measures. By Kushnireko's Theorem, the supremum is finite. Measures which attain the maximum are called *measures of maximal entropy.*

THEOREM 8.2. *Suppose f has positive topological entropy h. If f possesses a measure of maximal entropy, then there are $p \in \mathbb{N}$ and $C > 0$ s.t. $\mathrm{Per}_n(f) \geq Ce^{hn}$ for all $n > 0$ divisible by p.*

REMARK 8.3. The condition that f possesses a measure of maximal entropy is always fulfilled when f is C^∞, because of the following theorem of S. Newhouse. Let $\mathscr{M}(f)$ denote the collection of f–invariant Borel probability measures μ, equipped with the weak-star topology, then $\mathscr{M}(f)$ is compact and

THEOREM 8.4. *If f is C^∞, then $\mu_n \xrightarrow[n\to\infty]{w^*} \mu \Rightarrow h_\mu(f) \geq \limsup_{n\to\infty} h_{\mu_n}(f)$.*

PROOF OF THEOREM 8.2. Fix $0 < \delta < h$, and let Σ denote the two–sided topological Markov shift given by Theorem 8.1. Since every f–invariant probability measure on M with entropy larger than δ lifts to a shift invariant probability measure on Σ with the same entropy,
$$\sup\{h_{\widehat{\mu}}(\sigma) : \widehat{\mu} \text{ is shift a invariant probability measure on } \Sigma\} \geq h.$$
Every shift invariant measure $\widehat{\mu}$ projects to an f–invariant measure $\mu := \widehat{\mu} \circ \pi^{-1}$. The factor map $\pi : (\Sigma, \widehat{\mu}) \to (M, \mu)$ is finite-to-one on $\Sigma^\#$. Since $\Sigma^\#$ has full measure (Poincaré Recurrence Theorem), and finite-to-one factors preserve entropy, $h_{\widehat{\mu}}(\sigma) = h_\mu(f) \leq h$. It follows that
$$\sup\{h_{\widehat{\mu}}(\sigma) : \widehat{\mu} \text{ is shift a invariant probability measure on } \Sigma\} \leq h.$$
Thus $\sup\{h_{\widehat{\mu}}(\sigma) : \widehat{\mu}$ is shift a invariant probability measure on $\Sigma\} = h$.

This argument also shows that $f : M \to M$ has a measure of maximal entropy iff $\sigma : \Sigma \to \Sigma$ has a measure of maximal entropy, and the entropy is the same.

Now let Σ^+ denote the *one-sided* TMS corresponding to Σ:

$$\Sigma^+ = \{(x_0, x_1, \ldots) : \underline{x} \in \Sigma\}.$$

Abusing notation, we denote the left shift map on Σ^+ by σ. Every shift-invariant measure on Σ defines a shift invariant measure on Σ^+ with the same entropy, and every shift invariant measure on Σ^+ arises this way (because of the natural extension construction). It follows that $\sigma : \Sigma^+ \to \Sigma^+$ also possesses a measure of maximal entropy, with entropy h.

If $\sigma : \Sigma^+ \to \Sigma^+$ possesses a measure of maximal entropy, then $\sigma : \Sigma^+ \to \Sigma^+$ possesses an *ergodic* measure of maximal entropy. This is because the entropy map is affine, so a.e. ergodic component of a measure of maximal entropy is an ergodic measure of maximal entropy.

Let m_0 denote an ergodic measure of maximal entropy for $\sigma : \Sigma^+ \to \Sigma^+$. By the discussion in §2.1, m_0 is carried by a *topologically transitive* topological Markov shift $\Sigma_0^+ \subset \Sigma^+$. By Theorem 2.5, there is a positive integer p and a decomposition $\Sigma_0^+ = \biguplus_{i=0}^{p-1} X_i$ s.t. $\sigma(X_i) = X_{(i+1) \bmod p}$ and s.t. $\sigma^p : X_i \to X_i$ is topologically conjugate to a *topologically mixing* topological Markov shift. This shift must also possess a measure of maximal entropy, with entropy ph. Denote this topological Markov shift by $T : X \to X$.

Applying Theorems 5.3 and 5.5 to T, we find that the zero potential on X is positive recurrent, and its Gurevich pressure equals ph. By the generalized Ruelle's Perron-Frobenius Theorem $e^{-nph} L_0^n 1_{[a]} \xrightarrow[n \to \infty]{}$ positive constant, where $(L_0 f)(\underline{x}) = \sum_{T\underline{y}=\underline{x}} f(\underline{y})$.

Notice that if $\underline{x} \in X$ starts with a, then $L_0^n 1_{[a]}(\underline{x})$ equals the number of admissible words $(a, \xi_1, \ldots, \xi_{n-1}, a)$. Equivalently,

$$(L_0^n 1_{[a]})(\underline{x}) = \#\{\underline{z} \in X : z_0 = a, T^n \underline{z} = \underline{z}\} = \#\{\underline{y} \in \Sigma_0^+ : y_0 = a, \sigma^{np}(\underline{y}) = \underline{y}\}.$$

Thus for some positive C and all n large, $\#\{\underline{y} \in \Sigma_0^+ : y_0 = a, \sigma^{np}(\underline{y}) = \underline{y}\} \geq Ce^{nph}$, whence also $\#\{\underline{y} \in \Sigma : y_0 = a, \sigma^{np}(\underline{y}) = \underline{y}\} \geq Ce^{nph}$ for the two-sided shift Σ.

Every periodic sequence in Σ projects to a periodic point of f, and the map $\underline{y} \mapsto \pi(\underline{y})$ is at worst C_{aa}–to-one on the collection of sequences which contains the symbol a infinitely many times in the past and in the future. It follows that $\mathrm{Per}_{np}(f) \geq (C/C_{aa})e^{nph}$ for all n large. □

8.3. Ergodic properties of equilibrium measures. Let $f : M \to M$ denote a $C^{1+\epsilon}$ surface diffeomorphism, and suppose $\phi : M \to \mathbb{R}$ is continuous. An *equilibrium measure* for ϕ is a measure that maximizes $h_\mu(f) + \int \phi d\mu$. Such measures always exist when f is C^∞, because of Newhouse's Theorem 8.4.

In this section we will determine the ergodic theoretic structure of ergodic equilibrium measures of Hölder continuous potentials, subject to the assumption that their entropy is positive.

THEOREM 8.5. *If $\phi : M \to \mathbb{R}$ is Hölder continuous, then every ergodic equilibrium measure of ϕ which has positive entropy is measure theoretically isomorphic to the product of a Bernoulli scheme and a finite rotation.*

(A *finite rotation* is a map of the form $x \mapsto x + 1 \mod p$ on $\{0, 1, 2, \ldots, p-1\}$.)

Sketch of proof. Suppose μ is an ergodic equilibrium measure for ϕ, and μ has positive entropy. Fix $0 < \delta < h_\mu(f)$, and construct the two-sided topological Markov shift Σ corresponding to δ. Let $\widehat{\mu}$ denote an ergodic lift of μ to Σ: $\mu = \widehat{\mu} \circ \pi^{-1}$. Arguing as in the previous section, one can see that $\widehat{\mu}$ is an ergodic equilibrium measure for the lifted potential $\widehat{\phi} := \phi \circ \pi : \Sigma \to \mathbb{R}$.

Since $\phi : M \to \mathbb{R}$ and $\pi : \Sigma \to \mathbb{R}$ are Hölder, $\widehat{\phi} : \Sigma \to \mathbb{R}$ is Hölder. By theorem 2.8 there is a bounded continuous function h s.t. $\widehat{\phi}(\underline{x}) + h(\underline{x}) - h(\sigma\underline{x}) = \psi(x_0, x_1, \ldots)$, where ψ is a function with summable variations on Σ^+, the one-sided version of Σ (ψ can actually be chosen to be weakly Hölder and bounded). Since h is bounded, $\widehat{\mu}$ is also the equilibrium measure of $\widehat{\phi} + h - h \circ \sigma$. So $\widehat{\mu}^+$, the measure $\widehat{\mu}$ induces on Σ^+, is the equilibrium measure of ψ.

Since $\widehat{\mu}$ is ergodic, $\widehat{\mu}^+$ is ergodic, and therefore $\widehat{\mu}^+$ is carried by a topologically transitive topological Markov shift Σ_0^+.

Assume for simplicity that Σ_0^+ is actually topologically mixing. Since $\widehat{\mu}^+$ is the equilibrium measure of ψ and ψ is Hölder, the natural extension of $\widehat{\mu}^+$ is Bernoulli. So $\widehat{\mu}$ is Bernoulli. A theorem of Ornstein says that factors of Bernoulli schemes are Bernoulli, so $\mu = \widehat{\mu} \circ \pi^{-1}$ has the Bernoulli property, and we proved the theorem with $p = 1$.

In general, Σ_0^+ is not topologically mixing. In such cases one appeals to the spectral decomposition to show that Σ_0^+ splits into a cycle of p mutually disjoint pieces so that the restriction of σ^p to each piece is topologically mixing. Then one argues as above to show that σ^p is Bernoulli on each piece. Once this is shown the theorem follows in a standard way. □

8.4. Notes and references. Theorem 8.1 is proved in [**Sar13**]. Earlier examples of symbolic codings of diffeomorphisms by TMS include the Smale Horseshoe [**Sma65**], hyperbolic automorphisms of $\mathbb{T}^2$ [**AW67**], [**AW70**], Anosov diffeomorphisms [**Sin68a**], [**Sin68b**], and Axiom A diffeomorphisms [**Bow70**]. Sinai's paper [**Sin72**], Bowen's monograph [**Bow75**], and Ruelle's book [**Rue78**] were particularly influential in positioning symbolic dynamics and topological Markov shifts as central tools for studying smooth dynamical systems.

Theorem 8.2 is from [**Sar13**]. Earlier, Katok showed $\limsup \frac{1}{n} \ln \mathrm{Per}_n(f) \geq h$ [**Kat80**]. Katok's bound does not require the existence of a measure of maximal entropy. In the case of uniformly hyperbolic diffeomorphisms, it is also true that $\mathrm{Per}_n(f) \leq C' e^{nh}$ and much more can be said on the periodic points of f, see [**Bow71**], [**PP90**]. But for general diffeomorphisms, $\mathrm{Per}_n(f)$ could grow super-exponentially, see [**Kal00**].

Theorem 8.4 is due to Newhouse [**New89**].

Theorem 8.5 is from [**Sar11**]. In the case of Anosov and Axiom A diffeomorphisms, the result is due to Ratner [**Rat74**] and Bowen [**Bow75a**]. Pesin proved the Bernoulli property for smooth invariant measures with positive entropy for general smooth surface diffeomorphisms [**Pes77**], and Ledrappier did this for SRB measures [**Led84**].

The thermodynamic formalism of countable Markov shifts has been used in a similar way to prove similar results for other dynamical systems, such as interval maps, multi-dimensional beta-transformations, and piecewise affine homeomorphisms, see [**Buz97**], [**Buz05**], [**Buz09**], [**Buz10**] and references therein.

References

[Aar97] Jon Aaronson. *An introduction to infinite ergodic theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.

[AD01] Jon Aaronson and Manfred Denker. Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps. *Stoch. Dyn.*, 1(2):193–237, 2001.

[ADU93] Jon Aaronson, Manfred Denker, and Mariusz Urbański. Ergodic theory for Markov fibred systems and parabolic rational maps. *Trans. Amer. Math. Soc.*, 337(2):495–548, 1993.

[AKM65] R. L. Adler, A. G. Konheim, and M. H. McAndrew. Topological entropy. *Trans. Amer. Math. Soc.*, 114:309–319, 1965.

[AW67] R. L. Adler and B. Weiss. Entropy, a complete metric invariant for automorphisms of the torus. *Proc. Nat. Acad. Sci. U.S.A.*, 57:1573–1576, 1967.

[AW70] Roy L. Adler and Benjamin Weiss. *Similarity of automorphisms of the torus*. Memoirs of the American Mathematical Society, No. 98. American Mathematical Society, Providence, R.I., 1970.

[Bal00] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co. Inc., River Edge, NJ, 2000.

[BGT89] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[BHS03] Noam Berger, Christofer Hoffman, and Vladas Sidoravicius. Nonuniqueness for specifications in $\ell^{2+\epsilon}$. *arXiv:math/0312344*, 2003.

[BI11] Luis Barreira and Godofredo Iommi. Multifractal analysis and phase transitions for hyperbolic and parabolic horseshoes. *Israel J. Math.*, 181:347–379, 2011.

[BK93] Maury Bramson and Steven Kalikow. Nonuniqueness in g-functions. *Israel J. Math.*, 84(1-2):153–160, 1993.

[BK98] Henk Bruin and Gerhard Keller. Equilibrium states for S-unimodal maps. *Ergodic Theory Dynam. Systems*, 18(4):765–789, 1998.

[Bow70] Rufus Bowen. Markov partitions and minimal sets for Axiom A diffeomorphisms. *Amer. J. Math.*, 92:907–918, 1970.

[Bow71] Rufus Bowen. Periodic points and measures for Axiom A diffeomorphisms. *Trans. Amer. Math. Soc.*, 154:377–397, 1971.

[Bow73] Rufus Bowen. Topological entropy for noncompact sets. *Trans. Amer. Math. Soc.*, 184:125–136, 1973.

[Bow75] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Lecture Notes in Mathematics, Vol. 470. Springer-Verlag, Berlin, 1975.

[Bow75a] Rufus Bowen. Bernoulli equilibrium states for Axiom A diffeomorphisms. *Math. Systems Theory*, 8(4):289–294, 1974/75.

[Bow75b] Rufus Bowen. Some systems with unique equilibrium states. *Math. Systems Theory*, 8(3):193–202, 1974/75.

[BS03] Jérôme Buzzi and Omri Sarig. Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps. *Ergodic Theory Dynam. Systems*, 23(5):1383–1400, 2003.

[Buz97] Jérôme Buzzi. Intrinsic ergodicity of smooth interval maps. *Israel J. Math.*, 100:125–161, 1997.

[Buz05] Jérôme Buzzi. Subshifts of quasi-finite type. *Invent. Math.*, 159(2):369–406, 2005.

[Buz09] Jérôme Buzzi. Maximal entropy measures for piecewise affine surface homeomorphisms. *Ergodic Theory Dynam. Systems*, 29(6):1723–1763, 2009.

[Buz10] Jérôme Buzzi. Puzzles of quasi-finite type, zeta functions and symbolic dynamics for multi-dimensional maps. *Ann. Inst. Fourier (Grenoble)*, 60(3):801–852, 2010.

[Chu60] Kai Lai Chung. *Markov chains with stationary transition probabilities*. Die Grundlehren der mathematischen Wissenschaften, Bd. 104. Springer-Verlag, Berlin, 1960.

[CQ98] Zaqueu Coelho and Anthony N. Quas. Criteria for $\overline{d}$-continuity. *Trans. Amer. Math. Soc.*, 350(8):3257–3268, 1998.

[CS09] Van Cyr and Omri Sarig. Spectral gap and transience for Ruelle operators on countable Markov shifts. *Comm. Math. Phys.*, 292(3):637–666, 2009.

[Cyr11] Van Cyr. Countable Markov shifts with transient potentials. *Proc. Lond. Math. Soc. (3)*, 103(6):923–949, 2011.

[Dao13] Yair Daon. Bernoullicity of equilibrium measures on countable Markov shifts. *Discrete Contin. Dyn. Syst.*, 33(9):4003–4015, 2013.

[DF37] Wolfgang Doeblin and Robert Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148, 1937.

[Dob68a] R. L. Dobrušin. Description of a random field by means of conditional probabilities and conditions for its regularity. *Teor. Verojatnost. i Primenen*, 13:201–229, 1968.

[Dob68b] R. L. Dobrušin. Gibbsian random fields for lattice systems with pairwise interactions. *Funkcional. Anal. i Priložen.*, 2(4):31–43, 1968.

[Dob68c] R. L. Dobrušin. The problem of uniqueness of a Gibbsian random field and the problem of phase transitions. *Funkcional. Anal. i Priložen.*, 2(4):44–57, 1968.

[DP84] Manfred Denker and Walter Philipp. Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems*, 4(4):541–552, 1984.

[Ell06] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original.

[FF70] M.E. Fisher and B.U. Felderhof. Phase transition in one-dimensional cluster–interaction fluids:ia. thermodynamics, ib. critical behavior.ii,simple logarithmic model. *Ann. Phys.*, 58:177–280, 1970.

[FFY02] Doris Fiebig, Ulf-Rainer Fiebig, and Michiko Yuri. Pressure and equilibrium states for countable state Markov shifts. *Israel J. Math.*, 131:221–257, 2002.

[Gal73] G. Gallavotti. Ising model and Bernoulli schemes in one dimension. *Comm. Math. Phys.*, 32:183–190, 1973.

[GH88] Y. Guivarc'h and J. Hardy. Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(1):73–98, 1988.

[Goo69] L. Wayne Goodwyn. Topological entropy bounds measure-theoretic entropy. *Proc. Amer. Math. Soc.*, 23:679–688, 1969.

[Goo71] T. N. T. Goodman. Relating topological entropy and measure entropy. *Bull. London Math. Soc.*, 3:176–180, 1971.

[Gou04] Sébastien Gouëzel. Sharp polynomial estimates for the decay of correlations. *Israel J. Math.*, 139:29–65, 2004.

[Gou10a] Sébastien Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.*, 38(4):1639–1671, 2010.

[Gou10b] Sébastien Gouëzel. Characterization of weak convergence of Birkhoff sums for Gibbs-Markov maps. *Israel J. Math.*, 180:1–41, 2010.

[GS98] B. M. Gurevich and S. V. Savchenko. Thermodynamic formalism for symbolic Markov chains with a countable number of states. *Uspekhi Mat. Nauk*, 53(2(320)):3–106, 1998.

[Gur69] B. M. Gurevič. Topological entropy of a countable Markov chain. *Dokl. Akad. Nauk SSSR*, 187:715–718, 1969.

[Gur70] B. M. Gurevič. Shift entropy and Markov measures in the space of paths of a countable graph. *Dokl. Akad. Nauk SSSR*, 192:963–965, 1970.

[Gur78] B. M. Gurevič. A variational principle for one-dimensional latticed Gibbs random fields with a countable number of states. *Dokl. Akad. Nauk SSSR*, 241(4):749–752, 1978.

[Gur84] B. M. Gurevich. A variational characterization of one-dimensional countable state Gibbs random fields. *Z. Wahrsch. Verw. Gebiete*, 68(2):205–242, 1984.

[HH01] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.

[Hof77] Franz Hofbauer. Examples for the nonuniqueness of the equilibrium state. *Trans. Amer. Math. Soc.*, 228(223–241.), 1977.

[Hu08] Huyi Hu. Equilibriums of some non-Hölder potentials. *Trans. Amer. Math. Soc.*, 360(4):2153–2190, 2008.

[IJT13] Godofredo Iommi, Thomas Jordan, and Mike Todd. Recurrence and transience for suspension flows. *Israel J. of Math (to appear)*, 2013.

[Iom05] Godofredo Iommi. Multifractal analysis for countable Markov shifts. *Ergodic Theory Dynam. Systems*, 25(6):1881–1907, 2005.

[Iom10] Godofredo Iommi. Multifractal analysis of the Lyapunov exponent for the backward continued fraction map. *Ergodic Theory Dynam. Systems*, 30(1):211–232, 2010.

[JÖ03] Anders Johansson and Anders Öberg. Square summability of variations of g-functions and uniqueness of g-measures. *Math. Res. Lett.*, 10(5-6):587–601, 2003.

[JÖP12] Anders Johansson, Anders Öberg, and Mark Pollicott. Unique Bernoulli g-measures. *J. Eur. Math. Soc. (JEMS)*, 14(5):1599–1615, 2012.

[Kal00] Vadim Yu. Kaloshin. Generic diffeomorphisms with superexponential growth of number of periodic orbits. *Comm. Math. Phys.*, 211(1):253–271, 2000.

[Kat80] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, (51):137–173, 1980.

[Kea72] Michael Keane. Strongly mixing g-measures. *Invent. Math.*, 16:309–324, 1972.

[Kel98] Gerhard Keller. *Equilibrium states in ergodic theory*, volume 42 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

[Led74] François Ledrappier. Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 30:185–202, 1974.

[Led84] F. Ledrappier. Propriétés ergodiques des mesures de Sinaï. *Inst. Hautes Études Sci. Publ. Math.*, (59):163–188, 1984.

[Lop93] A. O. Lopes. The zeta function, nondifferentiability of pressure, and the critical exponent of transition. *Adv. Math.*, 101(2):133–165, 1993.

[LR69] O. E. Lanford, III and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.*, 13:194–215, 1969.

[Mis76] Michał Misiurewicz. Topological conditional entropy. *Studia Math.*, 55(2):175–200, 1976.

[MU01] R. Daniel Mauldin and Mariusz Urbański. Gibbs states on the symbolic space over an infinite alphabet. *Israel J. Math.*, 125:93–130, 2001.

[MU03] R. Daniel Mauldin and Mariusz Urbański. *Graph directed Markov systems*, volume 148 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets.

[New89] Sheldon E. Newhouse. Continuity properties of entropy. *Ann. of Math. (2)*, 129(2):215–235, 1989.

[Par64] William Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.*, 112:55–66, 1964.

[Pes77] Ja. B. Pesin. Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat. Nauk*, 32(4 (196)):55–112, 287, 1977.

[Pol86] Mark Pollicott. Meromorphic extensions of generalised zeta functions. *Invent. Math.*, 85(1):147–164, 1986.

[Pol00] Mark Pollicott. Rates of mixing for potentials of summable variation. *Trans. Amer. Math. Soc.*, 352(2):843–853, 2000.

[PP84] Ya. B. Pesin and B. S. Pitskel′. Topological pressure and the variational principle for noncompact sets. *Funktsional. Anal. i Prilozhen.*, 18(4):50–63, 96, 1984.

[PP90] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.

[PRLS04] Feliks Przytycki, Juan Rivera-Letelier, and Stanislav Smirnov. Equality of pressures for rational functions. *Ergodic Theory Dynam. Systems*, 24(3):891–914, 2004.

[PS92] Thomas Prellberg and Joseph Slawny. Maps of intervals with indifferent fixed points: thermodynamic formalism and phase transitions. *J. Statist. Phys.*, 66(1-2):503–514, 1992.

[PU10] Feliks Przytycki and Mariusz Urbański. *Conformal fractals: ergodic theory methods*, volume 371 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2010.

[PW99] Mark Pollicott and Howard Weiss. Multifractal analysis of Lyapunov exponent for continued fraction and Manneville-Pomeau transformations and applications to Diophantine approximation. *Comm. Math. Phys.*, 207(1):145–171, 1999.

[PZ06] Yakov Pesin and Ke Zhang. Phase transitions for uniformly expanding maps. *J. Stat. Phys.*, 122(6):1095–1110, 2006.

[Rat74] M. Ratner. Anosov flows with Gibbs measures are also Bernoullian. *Israel J. Math.*, 17:380–391, 1974.

[RE83] J. Rousseau-Egele. Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux. *Ann. Probab.*, 11(3):772–788, 1983.

[Rén57] A. Rényi. Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hungar*, 8:477–493, 1957.

[Rue67] D. Ruelle. A variational formulation of equilibrium statistical mechanics and the Gibbs phase rule. *Comm. Math. Phys.*, 5:324–329, 1967.

[Rue68] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Comm. Math. Phys.*, 9:267–278, 1968.

[Rue76] David Ruelle. A measure associated with axiom-A attractors. *Amer. J. Math.*, 98(3):619–654, 1976.

[Rue78] David Ruelle. *Thermodynamic formalism*, volume 5 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1978. The mathematical structures of classical equilibrium statistical mechanics, With a foreword by Giovanni Gallavotti and Gian-Carlo Rota.

[Sar99] Omri M. Sarig. Thermodynamic formalism for countable Markov shifts. *Ergodic Theory Dynam. Systems*, 19(6):1565–1593, 1999.

[Sar00] Omri M. Sarig. On an example with a non-analytic topological pressure. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(4):311–315, 2000.

[Sar01a] Omri M. Sarig. Phase transitions for countable Markov shifts. *Comm. Math. Phys.*, 217(3):555–577, 2001.

[Sar01b] Omri M. Sarig. Thermodynamic formalism for null recurrent potentials. *Israel J. Math.*, 121:285–311, 2001.

[Sar02] Omri Sarig. Subexponential decay of correlations. *Invent. Math.*, 150(3):629–653, 2002.

[Sar03] Omri Sarig. Existence of Gibbs measures for countable Markov shifts. *Proc. Amer. Math. Soc.*, 131(6):1751–1758 (electronic), 2003.

[Sar06] Omri Sarig. Continuous phase transitions for dynamical systems. *Comm. Math. Phys.*, 267(3):631–667, 2006.

[Sar11] Omri M. Sarig. Bernoulli equilibrium states for surface diffeomorphisms. *J. Mod. Dyn.*, 5(3):593–608, 2011.

[Sar13] Omri M. Sarig. Symbolic dynamics for surface diffeomorphisms with positive entropy. *J. Amer. Math. Soc.*, 26(2):341–426, 2013.

[Sin68a] Ja. G. Sinaĭ. Construction of Markov partitionings. *Funkcional. Anal. i Priložen.*, 2(3):70–80 (Loose errata), 1968.

[Sin68b] Ja. G. Sinaĭ. Markov partitions and U-diffeomorphisms. *Funkcional. Anal. i Priložen*, 2(1):64–89, 1968.

[Sin72] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[Sma65] Stephen Smale. Diffeomorphisms with many periodic points. In *Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse)*, pages 63–80. Princeton Univ. Press, Princeton, N.J., 1965.

[Sta71] E. Eugene Stanley. *Introduction to phase transitions and critical phenomena*. The International series of monographs on physics. Oxford Univ. Press, New York and Oxford, 1971.

[Tho11] Daniel J. Thompson. A thermodynamic definition of topological pressure for non-compact sets. *Ergodic Theory Dynam. Systems*, 31(2):527–547, 2011.

[VJ67] D. Vere-Jones. Ergodic properties of nonnegative matrices. I. *Pacific J. Math.*, 22:361–386, 1967.

[VJ68] D. Vere-Jones. Ergodic properties of nonnegative matrices. II. *Pacific J. Math.*, 26:601–620, 1968.

[Wal75a] Peter Walters. Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.*, 214:375–387, 1975.

[Wal75b] Peter Walters. A variational principle for the pressure of continuous transformations. *Amer. J. Math.*, 97(4):937–971, 1975.

[Wal78] Peter Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153, 1978.

[Wal01] Peter Walters. Convergence of the Ruelle operator for a function satisfying Bowen's condition. *Trans. Amer. Math. Soc.*, 353(1):327–347 (electronic), 2001.

[Wan89] X.-J. Wang. Abnormal fluctuations and thermodynamic phase transitions in dynamical systems. *Phys. Rev. A*, 39:3214–3217, 1989.

[You98] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.

[You99] Lai-Sang Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

[Yur00] Michiko Yuri. Weak Gibbs measures for certain non-hyperbolic systems. *Ergodic Theory Dynam. Systems*, 20(5):1495–1518, 2000.

[Yur03a] Michiko Yuri. Thermodynamic formalism for countable to one Markov systems. *Trans. Amer. Math. Soc.*, 355(7):2949–2971 (electronic), 2003.

[Yur03b] Michiko Yuri. Weak Gibbs measures for intermittent systems and weakly Gibbsian states in statistical mechanics. *Comm. Math. Phys.*, 241(2-3):453–466, 2003.

[Zar85] A. S. Zargaryan. A variational principle for a symbolic Markov chain with a countable number of states. In *Probability theory, theory of random processes, and functional analysis (Russian) (Moscow, 1984)*, pages 24–26. Moskov. Gos. Univ., Moscow, 1985.

[Zin96] Michel Zinsmeister. *Formalisme thermodynamique et systèmes dynamiques holomorphes*, volume 4 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 1996.

Faculty of Mathematics and Computer Science, Weizmann Institute of Science, 234 Herzl Street, Rehovot 7610001, Israel

E-mail address: omsarig@gmail.com