EFFECTIVE INTRINSIC ERGODICITY FOR COUNTABLE STATE MARKOV SHIFTS

RENÉ RÜHR AND OMRI SARIG

Abstract. For strongly positively recurrent countable state Markov shifts, we bound the distance between an invariant measure and the measure of maximal entropy, in terms of their entropy difference. This extends an earlier result for subshifts of finite type, due to Kadyrov. We provide a similar bound for equilibrium measures of strongly positively recurrent potentials, in terms of the pressure difference. For measures with nearly maximal entropy, we have new, and sharp, bounds.

Dedicated to Benjy Weiss on the occasion of his eightieth birthday

1. Introduction and summary of main results

Topological dynamical systems with unique measures of maximal entropy are called *intrinsically ergodic* [Wei70]. This property is weaker than unique ergodicity, and this weakening is useful, because it allows for many more examples. Natural instrinsically ergodic but non-uniquely ergodic systems can be found in symbolic dynamics [Par64], [Gur70], [Bow75], [CT12], [Cli18], [Pav20]; one-dimensional dynamics [Hof79], [Buz97]; the theory of diffeomorphisms [AW70], [Bow75], [BCS19]; and in the theory of geodesic flows [Kni98], [BCFT18], [CKW21]. (This list of references is incomplete, the relevant literature is too plentiful to survey.)

Although weaker than unique ergodicity, intrinsic ergodicity is powerful enough to have many applications. These include classification problems in ergodic theory [AW70]; the foundations of statistical mechanics [Rue72], [Rue78], [Sin72], the analysis of periodic orbits [Bow72], [PP90]; and even number theory [ELMV12]. (Again, these are very partial lists.) Here we will focus on the connection between intrinsic ergodicity and equidistribution of measures with high entropy, in the special case of topological Markov shifts.

Consider for example a topologically transitive subshift of finite type $\sigma : \Sigma^+ \to \Sigma^+$. This system has a unique measure of maximal entropy μ_0 [Par64]. Subshifts of finite type are compact, and their entropy map $\mu \mapsto h_\mu(\sigma)$ is upper semi-continuous in the weak star topology. Together with intrinsic ergodicity, this easily implies that for every sequence of shift invariant probability measures μ_n,

$$\text{if } h_{\mu_n}(\sigma) \to h_{\mu_0}(\sigma), \text{ then } \mu_n \to \mu_0 \text{ weak star.} \tag{1.1}$$

Date: June 29, 2021.

2010 *Mathematics Subject Classification.* 37A35, 37D35 (primary), 37C30, 05C63 (secondary).

Key words and phrases. Countable Topological Markov Shifts, Intrinsic Ergodicity, Pressure, Entropy, Equilibrium Measure, Measure of Maximal Entropy, Spectral Gap, Strong Positive Recurrence, Thermodynamic Formalism.

O.S. was partially supported by ISF grant 1149/18.

Kadyrov gave a bound for the rate of convergence [Kad15]. He showed that there exist constants C_β such that for every shift invariant probability measure μ and for every Hölder continuous function $\psi : \Sigma^+ \to \mathbb{R}$ with Hölder exponent β,

$$\left| \int \psi d\mu - \int \psi d\mu_0 \right| \leq C_\beta \|\psi\|_\beta \sqrt{h_{\mu_0}(\sigma) - h_\mu(\sigma)}, \tag{1.2}$$

where $\|\psi\|_\beta$ is the β-Hölder norm of ψ (see §2).

Quantitative versions of (1.1) like (1.2) are called *effective intrinsic ergodicity estimates*. They first appeared in the doctoral thesis of F. Polo [Pol11] for the $\times 2$ map on $\mathbb{R}/\mathbb{Z}$ and for hyperbolic toral automorphisms, but with a cubic root instead of a square root. Polo credits M. Einsiedler for outlining the proof for the $\times 2$ map, and we will henceforth call (1.2) the *Einsiedler-Kadyrov-Polo (EKP) inequality.* For similar inequalities for other systems, see [Kad15], [Rüh16], [Kha17].

In this paper we extend the EKP inequality to topologically transitive countable state Markov shifts. The new features here are non-compactness, the possibility of escape of mass to infinity, and "phase transitions": non-analytic pressure functions.

Of course, the EKP inequality cannot be expected to hold for all countable Markov shifts. For (1.1) or (1.2) to make sense, we must at the very least assume that a measure of maximal entropy μ_0 exists, and that $h_{\mu_0}(\sigma) < \infty$.

A deeper observation, due to S. Ruette, is that even under these additional assumptions, (1.1) and (1.2) may fail. It follows from [Rue03] (see also [GS98], [GZ88]), that if (1.1) or (1.2) hold, then $\sigma : \Sigma^+ \to \Sigma^+$ must be *strongly positively recurrent (SPR)*, a condition whose definition is recalled in §2.

Thus the right context for studying effective intrinsic ergodicity for countable Markov shifts is the class of topologically transitive SPR shifts.

By the work of Gurevich [Gur70], topologically transitive SPR countable Markov shifts have unique measures of maximal entropy. G. Iommi, M. Todd and A. Velozo [ITV19, Theorem 8.12], [ITV20], have shown that all SPR topologically transitive countable state Markov shifts satisfy (1.1). Our main result is: *All SPR topologically mixing countable state Markov shifts satisfy the EKP inequality* (1.2).

Our proof is different from Kadyrov's, and it produces sharper bounds. We show in Theorem 6.1 that for every $\varepsilon > 0$ and β-Hölder continuous $\psi : \Sigma^+ \to \mathbb{R}$, there exists a $\delta > 0$ such that if $h_\mu(\sigma)$ is δ-close to $h_{\mu_0}(\sigma)$, then

$$\left| \int \psi d\mu_0 - \int \psi d\mu \right| \leq e^\varepsilon \sqrt{2} \sigma_{\mu_0}(\psi) \sqrt{h_{\mu_0}(\sigma) - h_\mu(\sigma)}. \tag{1.3}$$

Here $\sigma_{\mu_0}^2(\psi)$ is the *asymptotic variance* of ψ with respect to μ_0 (see §2). The square root and the constant $\sqrt{2}\sigma_{\mu_0}(\psi)$ are sharp (Theorem 6.1). The sharpness of the square root is an answer to a question of Kadyrov [Kad15, p. 240].

The measure of maximal entropy μ_0 is the equilibrium measure of the zero potential. Our results extend to equilibrium measures μ_ϕ of other potentials ϕ. Suppose Σ^+ has finite Gurevich entropy, $\sup \phi < \infty$ and ϕ is weakly Hölder continuous (see §2). In Theorem 7.1, we show that if ϕ is SPR in the sense of [Sar01a], then for every shift invariant measure μ and ψ β-Hölder continuous,

$$\left| \int \psi d\mu - \int \psi d\mu_\phi \right| \leq C_{\phi,\beta} \|\psi\|_\beta \sqrt{P_{\mu_\phi}(\phi) - P_\mu(\phi)}, \tag{1.4}$$

where $P_\nu(\sigma) := h_\nu(\sigma) + \int \phi d\nu$. A sharp version similar to (1.3) holds as well.

The SPR property is a necessary condition for (1.4): In Corollary 8.1, we show that (1.4) fails whenever ϕ is not SPR.

Let us compare Kadyrov's proof to our proof. Kadyrov's proof is better in two ways: It is much shorter, and it yields a finitary version of (1.2) with $\frac{1}{n}H_\mu(\alpha_0^{n-1})$ replacing $h_\mu(\sigma)$, see [Kad17]. The reader may wonder why we needed a different proof. One reason is that our proof gives sharper, optimal, bounds. But there is another reason, of a more technical nature, which we would like to explain.

Assume $\sigma : \Sigma^+ \to \Sigma^+$ is topologically mixing, and let $\widehat{T}$ denote the transfer operator[1] of the measure of maximal entropy. Let $(\mathcal{H}_\beta, \|\cdot\|_\beta)$ denote the space of β-Hölder continuous functions as defined in §2. In the case of finite alphabets, $\widehat{T}$ has spectral gap when acting on $(\mathcal{H}_\beta, \|\cdot\|_\beta)$, and since $\|\cdot\|_\beta \geq \|\cdot\|_\infty$, it follows that $\|\widehat{T}^n f - \int f d\mu_0\|_\infty \to 0$ exponentially fast. This exponential *uniform* convergence seems to us to be crucial for Kadyrov's proof, see [Kad15, pp. 244-245].

But in the case of an infinite alphabet, we cannot expect uniform exponential convergence like that for all Hölder continuous functions, even in the SPR case. Let $\mathscr{G}$ be the graph associated to the shift (see §2), and suppose every vertex in $\mathscr{G}$ has finite degree. If f is the indicator of a cylinder, then $\widehat{T}^n f$ vanishes outside a finite union of partition sets (which depends on n). So $\|\widehat{T}^n f - \int f d\mu_0\|_\infty \geq |\int f d\mu_0|$ for all n, and $\widehat{T}^n f - \int f d\mu_0 \not\to 0$ uniformly (exponentially or not).

This is the obstacle that forced us to seek a different proof.

There is an important class of countable Markov shifts which do have Banach spaces with spectral gap so that $\|\cdot\|_{\mathcal{L}} \geq \|\cdot\|_\infty$: The shifts with the *big images and pre-images (BIP) property* ([AD01], [Sar03], see also §2.5, Example 2). In the infinite alphabet BIP case, all measures of maximal entropy have infinite entropy, and there is no hope to get (1.2). But there may still be unbounded potentials with equilibrium measures with finite pressure. For the EKP inequality for such measures, in the form (1.4), see [Rüh21].

2. A review of the theory of topological Markov shifts

2.1. **Topological Markov shifts.** Let $\mathscr{G}$ denote a countable directed graph with set of vertices S and set of directed edges E. If there is an edge from a to b, we write $a \to b$. Set

$$\Sigma^+ = \Sigma^+(\mathscr{G}) := \{\underline{x} = (x_0, x_1, \ldots) : x_i \in S, x_i \to x_{i+1} \text{ for all } i\}.$$

For every $\underline{x} \neq \underline{y}$ in Σ^+, let $t(\underline{x}, \underline{y}) = \min\{i : x_i \neq y_i\}$. We equip Σ^+ with the metric

$$d(\underline{x}, \underline{y}) := \begin{cases} 0 & \underline{x} = \underline{y} \\ \exp[-t(\underline{x}, \underline{y})] & \text{otherwise.} \end{cases} \tag{2.1}$$

Definition 2.1. *The topological dynamical system $\sigma : \Sigma^+ \to \Sigma^+$ given by $\sigma(\underline{x})_i = x_{i+1}$ is called the* one-sided topological Markov shift (TMS) *associated to $\mathscr{G}$. The elements of S are called* states, *and σ is called the* left shift.

When $|S| = \aleph_0$, we will also call Σ^+ a *countable state Markov shift.*

The sets $[\underline{a}] = [a_0, \ldots, a_{n-1}] := \{\underline{x} \in \Sigma^+ : x_i = a_i \ (i = 0, \ldots, n-1)\}$ ($\underline{a} \in \bigcup_n S^n$) are called *cylinders* of length n. Cylinders of length one are also called *partition*

[1] In the notation of §4 in this paper, $\widehat{T}f = \lambda_0^{-1} M_{h_0}^{-1} L_0 M_{h_0}$.

sets. The cylinders form a basis for the topology, and they generate the Borel σ-algebra, which we denote by $\mathscr{B}$.

2.2. **Topological transitivity and topological mixing.** We write $a \xrightarrow{n} b$ when there is a non-empty cylinder of the form $[a, \xi_1, \ldots, \xi_{n-1}, b]$. In particular $a \xrightarrow{1} b \Leftrightarrow a \to b$. The following simple facts are well-known:

(1) $\sigma : \Sigma^+ \to \Sigma^+$ is topologically transitive if and only if $\forall a, b \in S\ \exists n$ such that $a \xrightarrow{n} b$. Equivalently, $\mathscr{G}$ is *strongly connected*: $\forall (a,b) \in S^2$, there is a path from a to b.
(2) If $\sigma : \Sigma^+ \to \Sigma^+$ is topologically transitive, then $p_a := \gcd\{n : a \xrightarrow{n} a\}$ $(a \in S)$ are all equal to the same value $p \geq 1$, and p is called the *period* of Σ^+.
(3) $\sigma : \Sigma^+ \to \Sigma^+$ is topologically mixing if and only if it is topologically transitive, and its period is equal to one.
(4) **The spectral decomposition:** Suppose Σ^+ is a topologically transitive TMS with period $p > 1$, then we can decompose $\Sigma^+ = \Sigma_0^+ \uplus \Sigma_1^+ \uplus \cdots \uplus \Sigma_{p-1}^+$ where $\sigma(\Sigma_i^+) = \Sigma_{i+1 \bmod p}^+$, and where $\sigma^p : \Sigma_i^+ \to \Sigma_i^+$ are all topologically conjugate to a topologically *mixing* countable Markov shift.

    Briefly, this is done as follows. There is an equivalence relation on the states of Σ^+ given by $a \sim b \Leftrightarrow a = b$ or $a \xrightarrow{n} b$ for some n divisible by p. There are p equivalence classes $S_0, \ldots, S_{p-1}$, and $\Sigma_i^+ = \{\underline{x} \in \Sigma^+ : x_0 \in S_i\}$. The map $\sigma^p : \Sigma_i^+ \to \Sigma_i^+$ is topologically conjugate to the TMS with set of states $\{[a_0, \ldots, a_{p-1}] : a_0 \in S_i\} \setminus \{\varnothing\}$ and edges $[\underline{a}] \to [\underline{b}]$ when $a_{p-1} \to b_0$, and this TMS is topologically mixing.

The spectral decomposition is a tool for reducing statements on topologically transitive TMS to the topologically mixing case. We will use this tool frequently.

2.3. **Weak Hölder continuity and summable variations.** The n^{th} *oscillation* (aka the n^{th} *variation*) of a function $\phi : \Sigma^+ \to \mathbb{R}$ is

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

A function $\phi : \Sigma^+ \to \mathbb{R}$ is called *θ-weakly Hölder continuous* if $\theta \in (0,1)$ and there exists $A > 0$ such that $\operatorname{osc}_n(\phi) \leq A\theta^n$ for all $n \geq 2$. This condition does not imply that ϕ is bounded. A bounded θ-weakly Hölder continuous is Hölder continuous with exponent $\beta := -\log\theta$ with respect to the metric (2.1). We define the space of such functions in (2.8).

Some of our results hold under the following weaker regularity assumption, called *summable variations*: $\sum_{n\geq 2} \operatorname{osc}_n(\phi) < \infty$.

2.4. **Pressure and equilibrium measures.** Suppose Σ^+ is topologically mixing, $\phi : \Sigma^+ \to \mathbb{R}$ has summable variations and let $\phi_n := \sum_{k=0}^{n-1} \phi \circ \sigma^k$. Given $a \in S$, let

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

The limit exists and is independent of a, see [Sar99].[2] It is always bigger than $-\infty$, but it could be equal to $+\infty$.

[2] This reference states the result under stronger regularity assumptions on ϕ, but the proofs there work verbatim for functions with summable variations.

Definition 2.2. *$P_G(\phi)$ is called the* Gurevich pressure *of ϕ. The* Gurevich entropy *of Σ^+ is $h := P_G(0) = \lim_{n\to\infty} \frac{1}{n} \log \#\{\underline{x} \in \Sigma^+ : x_0 = a\ , \sigma^n(\underline{x}) = \underline{x}\}$.*

Let $\mathscr{M}(\Sigma^+)$ denote the collection of all σ-invariant Borel probability measures on Σ^+, and let $h_\mu(\sigma)$ denote the metric entropy of $\mu \in \mathscr{M}(\Sigma^+)$. *Gurevich's variational principle* [Gur69] says that if Σ^+ is topologically mixing, then

$$h := P_G(0) = \sup\left\{h_\mu(\sigma) : \mu \in \mathscr{M}(\Sigma^+)\right\}.$$

Any measure which achieves the supremum is called a *measure of maximal entropy*. Such measures do not always exist, but if they do, and if Σ^+ is topologically transitive, then they are unique [Gur70]. We will describe the structure of the measure of maximal entropy in the following section.

The Gurevich pressure of a general ϕ with summable variations satisfies a similar variational principle, which we now explain.

Recall that ϕ is called *one-sided μ-integrable* it at least one of the integrals $\int_{[\phi>0]} \phi d\mu, \int_{[\phi<0]} \phi d\mu$ is finite. Let

$$\mathscr{M}_\phi(\Sigma^+) := \left\{ \mu \in \mathscr{M}(\Sigma^+) : \begin{array}{l} \phi \text{ is one-sided } \mu\text{-integrable, and} \\ (h_\mu(\sigma), \int \phi d\mu) \neq (+\infty, -\infty) \end{array} \right\}.$$

This is the collection of $\mu \in \mathscr{M}(\Sigma^+)$ for which the expression

$$P_\mu(\phi) := h_\mu(\sigma) + \int \phi d\mu,$$

is well-defined (we allow $P_\mu(\phi) = \pm\infty$ but forbid $P_\mu(\phi) = \infty - \infty$).

In this paper we will be mostly interested in the case when Σ^+ has finite Gurevich entropy and $\sup \phi < \infty$. In this case $P_\mu(\phi)$ is well-defined for all $\mu \in \mathscr{M}(\Sigma^+)$, and $\mathscr{M}_\phi(\Sigma^+) = \mathscr{M}(\Sigma^+)$.

The *variational principle* for TMS states that if Σ^+ is topologically mixing and if ϕ has summable variations, then

$$P_G(\phi) = \sup\left\{P_\mu(\phi) : \mu \in \mathscr{M}_\phi(\Sigma^+)\right\}. \tag{2.2}$$

See [Sar99] for the special case $\sup \phi < \infty$, and [IJT15] in general.

A measure $\mu \in \mathscr{M}_\phi(\Sigma^+)$ which achieves the supremum in (2.2) is called an *equilibrium measure for the potential ϕ*. The equilibrium measures for the constant potential are the measures of maximal entropy.

So far we have only discussed the topologically mixing case. In the topologically transitive case, with period p, we define

$$P_G(\phi) := (1/p) P_G(\phi_p|_{\Sigma_0^+}),$$

where $\phi_p := \sum_{i=0}^{p-1} \phi \circ \sigma^i$, and Σ_0^+ is some (any) of the components in the spectral decomposition of Σ^+.

With this definition, the variational principle holds, and m_0 is an equilibrium measure for $\phi_p|_{\Sigma_0^+}$ if and only if $m := \frac{1}{p}\sum_{i=0}^{p-1} m_0 \circ \sigma^{-i}$ is an equilibrium measure for ϕ. Furthermore, in this case $m_0 := m(\cdot|\Sigma_0^+)$, the conditional measure on Σ_0^+. For more details, see the end of the proof of Theorem 6.1.

2.5. **Existence and structure of equilibrium measures.** The topic is intimately related to the eigenvector problem for *Ruelle's operator*

$$(L_\phi f)(\underline{x}) := \sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})} f(\underline{y}). \tag{2.3}$$

We recall the connection ([Bow75, Sar01b, BS03]).

Fix a state $a \in S$ and $\underline{x} \in \bigcup_{n>0} \sigma^{-n}[a]$, let $\tau_a(\underline{x}) := 1_{[a]}(\underline{x}) \min\{n > 0 : x_n = a\}$. Given a function $\phi : \Sigma^+ \to \mathbb{R}$, let $Z_n^*(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[\tau_a=n]}(\underline{x})$. Recall that $\phi_n = \sum_{i=0}^{n-1} \phi \circ \sigma^i$, and $Z_n(\phi, a) := \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x})} 1_{[a]}(\underline{x})$.

Definition 2.3. *Suppose ϕ is a function with summable variations and finite Gurevich pressure on a topologically mixing TMS Σ^+. We say that ϕ is* positively recurrent*, if for some state a,*

$$\sum_{n=1}^{\infty} \lambda^{-n} Z_n(\phi, a) = \infty \text{ and } \sum_{n=1}^{\infty} n\lambda^{-n} Z_n^*(\phi, a) < \infty, \text{ where } \lambda := \exp P_G(\phi).$$

(This should not be confused with *strong* positive recurrence, a condition that is discussed in the next section.)

The *Generalized Ruelle's Perron-Frobenius theorem* [Sar01b][3] states that if Σ^+ is a topologically mixing TMS and ϕ has summable variations and finite Gurevich pressure, then ϕ is positively recurrent if and only if there is a positive continuous function $h_\phi : \Sigma^+ \to \mathbb{R}_+$ and a σ-finite measure ν_ϕ such that

$$\begin{aligned} &L_\phi h_\phi = e^{P_G(\phi)} h_\phi, \quad L_\phi^* \nu_\phi = e^{P_G(\phi)} \nu_\phi, \quad \int h_\phi d\nu_\phi = 1 \text{ and} \\ &e^{-nP_G(\phi)} (L_\phi^n 1_{[\underline{a}]})(\underline{x}) \xrightarrow[n\to\infty]{} h_\phi(\underline{x}) \nu_\phi[\underline{a}] \text{ pointwise for all cylinders } [\underline{a}]. \end{aligned} \tag{2.4}$$

In this case h_ϕ is continuous, bounded away from zero and infinity on partition sets, and ν_ϕ gives finite and positive measure to every non-empty cylinder. The measure m_ϕ defined as $dm_\phi = h_\phi d\nu_\phi$ turns out to be a σ-invariant probability measure, and is called the *Ruelle-Perron-Frobenius (RPF) measure.*

Theorem 2.1 ([BS03], [CS09])**.** *Let ϕ be a potential with summable variations and finite Gurevich pressure on a topologically mixing TMS, and suppose* $\sup \phi < \infty$*.*

(1) If ϕ admits an equilibrium measure m, then this measure is unique, ϕ is positively recurrent, and m equals the RPF measure of ϕ.
(2) Conversely, if ϕ is positively recurrent and the RPF measure m_ϕ has finite entropy, then m_ϕ is the unique equilibrium measure of ϕ.

Corollary 2.1. *Let ϕ be a potential with summable variations and finite Gurevich pressure on a topologically mixing TMS with finite Gurevich entropy, and suppose* $\sup \phi < \infty$*. Then ϕ has an equilibrium measure if and only if ϕ is positively recurrent. In this case, the equilibrium measure is unique.*

An important consequence of the description of the equilibrium measure as the RPF measure is the following identity for conditional probabilities [Led74, Wal75]:

$$\mathbb{E}_{m_\phi}(f|\sigma^{-n}\mathscr{B}) = \lambda^{-n}(h_\phi^{-1} L_\phi^n(h_\phi f)) \circ \sigma^n \quad m_\phi\text{-a.e.} \tag{2.5}$$

[3]See the footnote on page 4.

Example 1 (SFT). In the finite alphabet case (when Σ^+ is a topologically mixing subshift of finite type), every ϕ with summable variations is positively recurrent and the RPF is always an equilibrium measure. This is a consequence of *Ruelle's Perron-Frobenius theorem*, see [Bow75].

Example 2 (BIP). A topologically mixing Markov shift is said to have the *big images and pre-images (BIP) property* if there is a finite set of states $b_1, \ldots, b_N$ such that for every state a there are some edges $a \to b_i, b_j \to a$. On a TMS with the BIP property, every ϕ with summable variations such that $\mathrm{osc}_1(\phi) < \infty$ and $P_G(\phi) < \infty$ is positively recurrent [Sar03], [MU01]. The corresponding RPF measure is an equilibrium measure if and only if it belongs to $\mathscr{M}_\phi(\Sigma^+)$.

Sometimes this is not the case. Fix a probability vector $\vec{p} = (p_i)_{i \in \mathbb{N} \cup \{0\}}$ with infinite entropy, and suppose $\Sigma^+ = \mathbb{N}^{\mathbb{N} \cup \{0\}}$ and $\phi(\underline{x}) = \log p_{x_0}$. In this case the RPF measure is the Bernoulli measure μ with probability vector $\vec{p}$. But for this measure $h_\mu(\sigma) + \int \phi dm$ is not well-defined, because $h_\mu(\sigma) = +\infty$ and $\int \phi d\mu = -\infty$.

Example 3 (Measures of maximal entropy). Returning to the countable alphabet case, let's consider the special case of measures of maximal entropy, $\phi \equiv 0$. Let $h := P_G(0)$ be the Gurevich entropy, and set $\lambda := \exp(h)$. It is not difficult to check using (2.4) that in this case $h_\phi(\underline{x})$ depends only on x_0, so there is a positive vector $(\ell_a)_{a \in S}$ such that $h_\phi(\underline{x}) = \ell_{x_0}$. Since $L_\phi h_\phi = \lambda h_\phi$, $\sum_{a: a \to b} \ell_a = \lambda \ell_b$.

Similarly, if $(r_a)_{a \in S}$ is the vector $r_a := \nu_\phi[a]$, then the equation $L_\phi^* \nu_\phi = \lambda \nu_\phi$ implies that $r_a = \nu_\phi(1_{[a]}) = \lambda^{-1} \nu_\phi(L_\phi 1_{[a]}) = \lambda^{-1} \nu_\phi(\sigma[a])$, whence $\sum_{b: a \to b} r_b = \lambda r_a$. So $\vec{r}, \vec{\ell}$ are positive right and left eigenvectors of the transition matrix of Σ^+. In addition, $\sum \ell_a r_a = \int h_\phi d\nu_\phi = 1$.

Substituting $f = 1_{[x_0, \ldots, x_{n-1}]}$ in (2.5), we find that

$$m_\phi(x_0, \ldots, x_{n-1} | x_n, \ldots) = \frac{\ell_{x_0}}{\lambda^n \ell_{x_n}} \quad m_\phi\text{-a.e.}$$

The right-hand side is independent of x_k with $k > n$. This and the shift invariance of m_ϕ implies that

$$\frac{m_\phi[x_i, x_{i+1}, \ldots, x_{n-1}]}{m_\phi[x_{i+1}, \ldots, x_{n-1}]} = m_\phi(x_i | x_{i+1}, \ldots, x_{n-1}) = m_\phi(x_i | x_{i+1}, \ldots) = \frac{\ell_{x_i}}{\lambda \ell_{x_{i+1}}}. \tag{2.6}$$

Let $p_{a,b} := \frac{\ell_a}{\lambda \ell_b}$ and $p_a := \ell_a r_a$. Taking the product of (2.6) over $0 \le i \le n-2$, we recover the identification of the measure of maximal entropy as *Parry's measure* [Par64], [Gur70]:

$$m_\phi[x_0, \ldots, x_{n-1}] = p_{x_0} p_{x_0, x_1} \cdots p_{x_{n-2}, x_{n-1}}. \tag{2.7}$$

In particular, the measure of maximal entropy is a Markov measure.

2.6. **Strong positive recurrence.** This is a strengthening of the positive recurrence condition, which implies that L_ϕ acts quasi-compactly on a "sufficiently rich" Banach space of functions (see the next section).

We begin with the special case when Σ^+ is topologically mixing, and $\phi \equiv 0$. Fix a state a, and let $f_n(a)$ denote the number of first return loops with length n, at a:

$$f_n(a) := \#\{\underline{x} \in \Sigma^+ : \sigma^n(\underline{x}) = \underline{x},\ x_0 = a,\ x_1, \ldots, x_{n-1} \neq a\}.$$

Let $F_a(z) := \sum_{n \ge 1} f_n(a) z^n$ be the associated generating function.

Definition 2.4. *A topologically mixing TMS Σ^+ with finite Gurevich entropy is called* strongly positively recurrent (SPR)*, if for some state a, $F_a(R_a) > 1$ where $R_a :=$ radius of convergence of $F_a(z)$.*

It follows from [VJ62] that (a) SPR implies that $\phi \equiv 0$ is positively recurrent; (b) the SPR property is independent of a: if $F_a(R_a) > 1$ for some a, then $F_a(R_a) > 1$ for all a; and (c) SPR implies that Parry's measure is an exponentially mixing Markov chain.

The SPR condition was extended to general potentials with summable variations in [Sar01a] (see [GS98] for the special case when $\mathrm{osc}_2(\phi) = 0$, i.e. ϕ is *Markovian*).

Suppose Σ^+ is topologically mixing, and a is a state. The *induced map on* $[a]$ is the map $\sigma^{\tau_a} : [a]' \to [a]'$, where $[a]' := \{\underline{x} \in [a] : x_i = a \text{ infinitely often}\}$, and

$$\tau_a(\underline{x}) := 1_{[a]}(\underline{x}) \min\{n > 0 : x_n = a\}.$$

This map has a coding as a full shift: Let $\overline{S} := \{[a, \xi_1, \ldots, \xi_n, a] : \xi_i \neq a\} \setminus \{\varnothing\}$, $\overline{\Sigma}^+ := \overline{S}^{\mathbb{N}\cup\{0\}}$ and define $\pi : \overline{\Sigma}^+ \to [a]'$ by

$$\pi([a, \underline{\xi}_1, a], [a, \underline{\xi}_2, a], \ldots) := (a, \underline{\xi}_1, a, \underline{\xi}_2, a, \ldots).$$

Then $\pi^{-1} \circ \sigma^{\tau_a} \circ \pi$ is the left shift on $\overline{\Sigma}^+$. Given a function $\phi : \Sigma^+ \to \mathbb{R}$, we let

$$\overline{\phi} := \Big(\sum_{k=0}^{\tau_a - 1} \phi \circ \sigma^k\Big) \circ \pi.$$

If ϕ is weakly Hölder continuous on Σ^+, then $\overline{\phi}$ is weakly Hölder continuous on $\overline{\Sigma}^+$.

Definition 2.5. *Suppose ϕ is a weakly Hölder continuous potential with finite Gurevich pressure on a topologically mixing TMS, and let a be a state. The* discriminant *of ϕ at state a is the (possibly infinite) expression*

$$\Delta_a[\phi] := \sup\{P_G(\overline{\phi + p}) : p \in \mathbb{R} \text{ such that } P_G(\overline{\phi + p}) < \infty\}.$$

We say that ϕ is strongly positively recurrent (SPR) *if $\Delta_a[\phi] > 0$ for some a.*

The definition extends to ϕ with summable variations, but some care is needed since $\overline{\phi}$ may not have summable variations in this case, see [Sar01a].

Lemma 2.1. *If Σ^+ is topologically mixing, ϕ has summable variations, and $P_G(\phi) < \infty$, then ϕ is strongly positively recurrent if and only if for some state a,*

$$P_G^*(\phi, a) := \limsup_{n\to\infty} \frac{1}{n} \log Z_n^*(\phi, a) < \lim_{n\to\infty} \frac{1}{n} \log Z_n(\phi, a) = P_G(\phi).$$

If this happens for some state, then it happens for all states.

Proof. This is a consequence of the discriminant theorem of [Sar01a]. □

Example. Consider the special case $\phi \equiv 0$. In this case $\overline{\phi + p} = \overline{p} = p \cdot \tau_a \circ \pi$. So for every $\overline{a} := [a, \underline{\xi}, a] \in \overline{S}$,

$$Z_n(\overline{\phi + p}, \overline{a}) = \sum_{[a,\underline{\xi}_1,a],\ldots,[a,\underline{\xi}_{n-1},a] \in \overline{S}} e^{p(|a\underline{\xi}| + |a\underline{\xi}_1| + \cdots + |a\underline{\xi}_{n-1}|)} = e^{p|a\underline{\xi}|}\Bigg(\sum_{[a\underline{\eta}a] \in \overline{S}} e^{p|a\underline{\eta}|}\Bigg)^{n-1},$$

and $P_G(\overline{\phi + p}) = \log \sum_{[a\underline{\eta}a] \in \overline{S}} e^{p|a\underline{\eta}|}$. The terms with $|a\underline{\eta}| = k$ represent first return time loops with length k at the vertex a. So $P_G(\overline{\phi + p}) = \log \sum_{k=1}^{\infty} f_k(a) e^{kp}$. It follows that $\Delta_a[0] = \log \sum_{n=1}^{\infty} f_n(a) R_a^n$, where R_a is the radius of convergence of $F_a(z) = \sum_{n\geq 1} f_n(a) z^n$. So $\phi \equiv 0$ is SPR if and only if $F_a(R_a) > 1$.

It is shown in [Sar01a] that if Σ^+ is topologically mixing, ϕ has summable variations, and $P_G(\phi) < \infty$, then: (a) SPR implies positive recurrence; (b) The SPR property is independent of the choice of a: If $\Delta_a[\phi] > 0$ for some a, then $\Delta_a[\phi] > 0$ for all a. Additionally, in [CS09] it is shown that: (c) If ϕ is weakly Hölder continuous, then the RPF measure of ϕ has exponential decay of correlations for Hölder continuous test functions.

So far we have only discussed the topologically mixing case. If Σ^+ is topologically transitive with period $p > 1$, and ϕ is a potential with summable variations and finite Gurevich pressure, then we say that ϕ is *strongly positively recurrent* if and only if $\phi_p|_{\Sigma_0^+}$ is strongly positively recurrent, where $\phi_p := \sum_{i=0}^{p-1} \phi \circ \sigma^i$ and Σ_0^+ is some (any) component of the spectral decomposition.

2.7. **SPR and spectral gap.** Let Σ^+ be a topologically mixing TMS. A θ-weakly Hölder continuous function ϕ is said to have the *spectral gap property* if there exists a Banach space $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ of functions $f : \Sigma^+ \to \mathbb{C}$ with the following properties:

(a) $(L_\phi f)(\underline{x}) = \sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})} f(\underline{y})$ converges absolutely whenever $f \in \mathcal{L}$, and $\mathcal{L}$ contains all the indicators of cylinder sets.
(b) $f \in \mathcal{L} \Rightarrow |f| \in \mathcal{L}$ and $\big\||f|\big\|_{\mathcal{L}} \leq \|f\|_{\mathcal{L}}$.
(c) $\mathcal{L}$-convergence implies uniform convergence on cylinders.
(d) $L_\phi(\mathcal{L}) \subset \mathcal{L}$ and $L_\phi : \mathcal{L} \to \mathcal{L}$ is bounded.
(e) $L_\phi = \lambda P + N$ where $\lambda = \exp P_G(\phi)$, P, N are bounded linear operators on $\mathcal{L}$ such that $PN = NP = 0$, $P^2 = P$, $\dim \mathrm{Image}(P) = 1$, and the spectral radius of N is strictly less than λ.
(f) For every bounded θ-weakly Hölder continuous function $\psi : \Sigma^+ \to \mathbb{R}$, $z \in \mathbb{C}$, $L_{\phi+z\psi}$ is bounded on $\mathcal{L}$, and $z \mapsto L_{\phi+z\psi}$ is analytic on some complex neighborhood of zero. (See [Kat95, Chapter VII §1] for the definition of analyticity for families of operators depending on a complex parameter.)

Property (e) says that the spectrum of $L_\phi : \mathcal{L} \to \mathcal{L}$ consists of a simple eigenvalue λ and a compact subset of $\{z : |z| < \lambda\}$. P is the eigenprojection of λ. The "gap" is the difference between $|\lambda|$ and the spectral radius of N.

The paper [CS09] proves that *a weakly Hölder continuous function ϕ with finite Gurevich pressure on a topologically mixing TMS has the spectral gap property, if and only if ϕ is strongly positively recurrent.*

The space $\mathcal{L}$ constructed there has the following additional property [CS09, p. 650]. Let $\mathcal{H}_\beta$ be the space of β-Hölder continuous functions,

$$\mathcal{H}_\beta := \left\{ \psi : \Sigma^+ \to \mathbb{R} : \|\psi\|_\beta := \|\psi\|_\infty + \sup_{\underline{x}\neq\underline{y}} \frac{|\psi(\underline{x}) - \psi(\underline{y})|}{e^{-\beta t(\underline{x},\underline{y})}} < \infty \right\}. \tag{2.8}$$

(g) If $\theta = e^{-\beta}$, then for all $f \in \mathcal{H}_\beta$ and $g \in \mathcal{L}$, $fg \in \mathcal{L}$ and $\|fg\|_{\mathcal{L}} \leq \|f\|_\beta \|g\|_{\mathcal{L}}$.

We caution the reader that the SPR property does *not* imply the spectral gap property for topologically transitive TMS with period $p > 1$. We may still find a Banach space with (a)–(d) on which L_ϕ acts quasi-compactly, but there will be p points in the spectrum with modulus $\exp P_G(\phi)$, and not just one as in (e).

3. The pressure function

Throughout this section, we fix a topologically mixing one-sided countable Markov shift $\sigma : \Sigma^+ \to \Sigma^+$ with finite positive Gurevich entropy h, and potentials ϕ, ψ with summable variations such that $\sup \phi < \infty$ and $\|\psi\|_\infty < \infty$. It follows that

$$\mathscr{M}_{\phi+t\psi}(\Sigma^+) = \mathscr{M}(\Sigma^+) \text{ for all } t.$$

Recall the definition of the Gurevich pressure from §2.4.

Definition 3.1. *The* pressure function of ϕ in direction ψ *is the function* $\mathfrak{p}_{\phi,\psi} : \mathbb{R} \to \mathbb{R}$ *given by* $\mathfrak{p}_{\phi,\psi}(t) := P_G(\phi + t\psi)$.

By the variational principle (2.2),

$$\mathfrak{p}_{\phi,\psi}(t) = \sup\{P_\mu(\phi) + t \int \psi d\mu : \mu \in \mathscr{M}(\Sigma^+)\},$$

where $P_\mu(\phi) := h_\mu(\sigma) + \int \phi d\mu$.

Theorem 3.1. *Let Σ^+ be a topologically mixing TMS with finite Gurevich entropy. If ϕ, ψ have summable variations, $\sup \phi < \infty$ and $\|\psi\|_\infty < \infty$, then*

(1) $\mathfrak{p}_{\phi,\psi}(t)$ is finite, convex, and continuous on $\mathbb{R}$.
(2) $\mathfrak{p}_{\phi,\psi}(t)$ has well-defined finite one-sided derivatives

$$(D^\pm \mathfrak{p}_{\phi,\psi})(t) := \lim_{h\to\pm 0} \frac{1}{h}[\mathfrak{p}_{\phi,\psi}(t+h) - \mathfrak{p}_{\phi,\psi}(t)].$$

(3) $\lim_{t\to\infty} (D^\pm \mathfrak{p}_{\phi,\psi})(t)$ exist, are equal, and finite. We call their common value $\mathfrak{p}'_{\phi,\psi}(+\infty)$. Similarly, $\lim_{t\to-\infty} (D^\pm \mathfrak{p}_{\phi,\psi})(t)$ are equal and finite. We call their common value $\mathfrak{p}'_{\phi,\psi}(-\infty)$.
(4) $\mathfrak{p}'_{\phi,\psi}(+\infty) = \sup\{\int \psi d\mu : \mu \in \mathscr{M}(\Sigma^+)\}$, $\mathfrak{p}'_{\phi,\psi}(-\infty) = \inf\{\int \psi d\mu : \mu \in \mathscr{M}(\Sigma^+)\}$.
(5) If ψ is not cohomologous to a constant by a continuous transfer function, then $\mathfrak{p}'_{\phi,\psi}(-\infty) < \mathfrak{p}'_{\phi,\psi}(+\infty)$.

For subshifts of finite type (or countable Markov shifts with the BIP property), these results are proved in [BL98, §3], [Mor07]. In this case more is true: $\mathfrak{p}_{\phi,\psi}(t)$ is real-analytic on $\mathbb{R}$, and the one-sided derivatives can be replaced by ordinary derivatives. For general countable Markov shifts we do not necessarily have differentiability on $\mathbb{R}$, but the theorem is saved by standard convexity arguments:

Proof. Fix some $\nu \in \mathscr{M}(\Sigma^+)$ carried by a periodic orbit, then $\mathfrak{p}_{\phi,\psi}(t) \geq P_\nu(\phi + t\psi) > -\infty$, proving that $\mathfrak{p}_{\phi,\psi}(t) \neq -\infty$ on $\mathbb{R}$. Next, $\mathfrak{p}_{\phi,\psi}(t) < \infty$, because for every $\mu \in \mathscr{M}(\Sigma^+)$, $P_\mu(\phi + t\psi) = h_\mu(\sigma) + \int \phi d\mu + t \int \psi d\mu \leq h + \sup \phi + |t|\|\psi\|_\infty < \infty$. Passing to the supremum over $\mu \in \mathscr{M}(\Sigma^+)$ gives $\mathfrak{p}_{\phi,\psi}(t) \leq h + \sup \phi + |t|\|\psi\|_\infty < \infty$. So $\mathfrak{p}_{\phi,\psi}$ is finite on $\mathbb{R}$.

To see convexity, we use the identity $P_G(\phi + t\psi) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\phi + t\psi, a)$. By Hölder's inequality, if $t = st_1 + (1-s)t_2$ with $s \in [0,1]$, then

$$Z_n(\phi + t\psi, a) \leq Z_n(\phi + t_1\psi, a)^s Z_n(\phi + t_2\psi, a)^{1-s}.$$

It follows that $t \mapsto \frac{1}{n} \log Z_n(\phi + t\psi, a)$ is convex on $\mathbb{R}$. Pointwise limits of convex functions are convex, therefore $\mathfrak{p}_{\phi,\psi}(t)$ is convex on $\mathbb{R}$. We proved (1), (2).

To see (3) we use convexity to note that $t \mapsto (D^{\pm}\mathfrak{p}_{\phi,\psi})(t)$ is increasing (in the broad sense), and for all $t_1 < t_2$

$$(D^{-}\mathfrak{p}_{\phi,\psi})(t_1) \leq (D^{+}\mathfrak{p}_{\phi,\psi})(t_1) \leq (D^{-}\mathfrak{p}_{\psi,\phi})(t_2) \leq (D^{+}\mathfrak{p}_{\psi,\phi})(t_2).$$

This implies the existence and equality of the limits which define $\mathfrak{p}'_{\phi,\psi}(\pm\infty)$. To see that these quantities are finite, we note that $P_G(\phi) - |t|\|\psi\|_\infty \leq \mathfrak{p}_{\phi,\psi}(t) \leq P_G(\phi) + |t|\|\psi\|_\infty$, so $\mathfrak{p}_{\phi,\psi}(t)$ is a convex function with asymptotes of finite slope.

We prove (4). Let $p(t) := \mathfrak{p}_{\phi,\psi}(t)$, $p'_{\pm} := D^{\pm}p$, $p'(\pm\infty) := \mathfrak{p}'_{\phi,\psi}(\pm\infty)$. By convexity, for every $\mu \in \mathscr{M}(\Sigma^+)$ and $t > 0$,

$$p'(\infty) \geq p'_{+}(t) \geq \frac{p(t) - p(0)}{t} \geq \frac{(P_\mu(\phi) + t\int \psi d\mu) - p(0)}{t} \xrightarrow[t\to\infty]{} \int \psi d\mu.$$

Thus $p'(\infty) \geq \sup\{\int \psi d\mu : \mu \in \mathscr{M}(\Sigma^+)\}$.

Next, we fix $\varepsilon > 0$ arbitrarily small and $t > 0$ so large that $p'(+\infty) \leq p'_{-}(t) + \varepsilon$. This is possible by (3). For every $0 < \delta < \varepsilon$, we choose a σ-invariant $m := m_{t,\delta}$ such that $p(t+\delta) \leq P_m(\phi) + (t+\delta)\int \psi dm + \delta^2$. By the variational principle, $p(t) \geq P_m(\phi) + t\int \psi dm$, so

$$\begin{aligned} p'(\infty) &\leq p'_{-}(t) + \varepsilon \leq \frac{p(t+\delta) - p(t)}{\delta} + \varepsilon \\ &\leq \frac{[P_m(\phi) + (t+\delta)\int \psi dm + h^2] - [P_m(\phi) + t\int \psi dm]}{\delta} + \varepsilon \\ &= \int \psi dm + \delta + \varepsilon \leq \sup\{\int \psi d\mu\} + 2\varepsilon. \end{aligned}$$

The sup runs over all σ-invariant probabilities. Passing to the limit $\varepsilon \to 0$, we obtain $p'(+\infty) \leq \sup\{\int \psi d\mu : \mu \in \mathscr{M}(\Sigma^+)\}$. This shows (4) for $p'(+\infty)$. The claim for $p'(-\infty)$ follows immediately using the identity $\mathfrak{p}_{\phi,-\psi}(t) = \mathfrak{p}_{\phi,\psi}(-t)$.

To see (5), assume the contrary: $p'(-\infty) = p'(\infty) = c$. Then (4) tells us that $\int \psi d\mu = c$ for all $\mu \in \mathscr{M}(\Sigma^+)$. Applying this to invariant measures carried by a periodic orbit we find that $\sigma^n(\underline{x}) = \underline{x} \Rightarrow \sum_{k=0}^{n-1}\psi(\sigma^k\underline{x}) = cn$. By the Livshits theorem, this implies that ψ is cohomologous to c via a continuous transfer function. (The proof for subshifts of finite type given in [Bow75, Theorem 1.28] works verbatim in the countable alphabet case.) □

Remark. The assumption that the Gurevich entropy h is finite is used in two places: In (1) we use it to show that $P_G(\phi + t\psi) < \infty$ for all t, and in (4) we use it implicitly to make sure that $P_m(\phi)$ and $P_\mu(\phi)$ are well-defined. If we relax the assumption that $h < \infty$ to the assumption that $P_G(\phi) < \infty$, then the theorem remains true, except that the suprema in (4) should be taken over $\mathscr{M}_\phi(\Sigma^+)$, instead of $\mathscr{M}(\Sigma^+)$.

Theorem 3.2. *Suppose Σ^+ is a topologically mixing countable Markov shift with finite Gurevich entropy. Let ϕ be an SPR θ-weakly Hölder continuous potential such that $\sup\phi < \infty$. Then there exist $M := M_\theta(\phi) > 1$ and $\varepsilon := \varepsilon_\theta(\phi) > 0$ such that for every $\beta > |\log\theta|$ and ψ such that $\|\psi\|_\beta \leq 1$, the following holds:*

(1) $\mathfrak{p}_{\phi,\psi}(t)$ is real-analytic on $(-\varepsilon, \varepsilon)$.
(2) If $|t| \leq \varepsilon$ then there exists a unique equilibrium measure $m_t := \mu_{\phi+t\psi}$ for $\phi + t\psi$, i.e. $\mathfrak{p}_{\phi,\psi}(t) = P_{m_t}(\phi + t\psi)$, and $\mathfrak{p}_{\phi,\psi}(t) > P_\mu(\phi + t\psi)$ for $\mu \neq m_t$.
(3) If $|t| < \varepsilon$, then $\mathfrak{p}'_{\phi,\psi}(t) = \mathbb{E}_{m_t}(\psi) := \int \psi dm_t$.

(4) *If* $|t| < \varepsilon$, *then* $\mathfrak{p}''_{\phi,\psi}(t) = \sigma^2_{m_t}(\psi)$ *where* $\sigma^2_{m_t}(\psi) := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{m_t}[(\psi_n - \mathbb{E}_{m_t}(\psi_n))^2]$ *and* $\psi_n := \sum_{k=0}^{n-1} \psi \circ \sigma^k$.
(5) *If* $|t| < \varepsilon$, *then* $|\mathfrak{p}'_{\phi,\psi}(t)|, |\mathfrak{p}''_{\phi,\psi}(t)|, |\mathfrak{p}'''_{\phi,\psi}(t)| \leq M$.
(6) *If* $\varphi \in \mathcal{H}_\beta$ *such that* $\|\varphi\|_\beta \leq 1$, *then* $(t,s) \mapsto P_G(\phi + t\psi + s\varphi)$ *has continuous partial derivatives of all orders on* $(-\varepsilon, \varepsilon)^2$.

The proof of the theorem is long, and is deferred to §4.

Remark 1. Theorem 3.2 is false without the SPR assumption, because of "phase transitions", see [Sar01a], [Sar06]. Some version of the theorem is also true for TMS with infinite Gurevich entropy, but in this case the measures m_t should be taken to be the RPF measures of $\phi + t\psi$ and not their equilibrium measures.

Remark 2. A weaker version of the theorem with ε, M depending also on ψ is known. For the finite alphabet case, see [PP90, Chapter 4], [GH88], [Rue78]. For infinite alphabets, see [CS09]. We need ε, M to be independent of ψ to describe the regime when the "sharp" EKP inequality (1.3) holds, see the end of Section 6.

We make a few more comments on the quantity $\sigma^2_{m_t}(\psi)$ in part (4) of the theorem. In general, the *variance* of $\psi \in L^2(m)$ is $\mathrm{Var}_m(\psi) := \int [(\psi - \int \psi dm)^2] dm$. If m is σ-invariant, then $\psi \in L^2(m) \Rightarrow \psi_n := \sum_{k=0}^{n-1} \psi \circ \sigma^k \in L^2(m)$ for all n, and the *asymptotic variance* of ψ is the following limit whenever it exists:

$$\sigma^2_m(\psi) := \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}_m(\psi_n).$$

Theorem 3.2(4) says that the limit exists under the assumptions of that theorem, and relates it to the second derivative of the pressure function.

The asymptotic variance is important because it is the variance of the Gaussian distributional limit of the m-distributions of $(\psi_n - n\int \psi dm)/\sqrt{n}$, whenever m is the equilibrium measure of an SPR weakly Hölder continuous potential ϕ, and $\psi \in \mathcal{H}_\beta$. See [PP90, Chapter 4] and [CS09].

The following theorem gives some of the properties of the asymptotic variance.

Theorem 3.3. *Let* Σ^+ *be a topologically mixing countable Markov shift with finite Gurevich entropy. Let* ϕ *be a* θ*-weakly Hölder continuous SPR potential such that* $\sup \phi < \infty$. *Let* m *be the unique equilibrium measure of* ϕ.

(1) *For every* $\psi \in \mathcal{H}_\beta$, $\sigma_m(\psi) = 0$ *if and only if* $\psi = r - r\circ\sigma + c$ *where* $c \in \mathbb{R}$, *and* r *is continuous (but perhaps not bounded).*
(2) *If* $\psi, \varphi \in \mathcal{H}_\beta$ *and* $\psi - \varphi = r - r\circ\sigma + c$ *with* r *continuous and* $c \in \mathbb{R}$, *then* $\sigma_m(\psi) = \sigma_m(\varphi)$.
(3) *Let* $M := M_\theta(\phi)$ *be as in Theorem* 3.2, *then* $\sigma_m(\psi) \leq M\|\psi\|_\beta$ *whenever* $e^{-\beta} \leq \theta$ *and* $\psi \in \mathcal{H}_\beta$.

The theorem is well-known for subshifts of finite type [PP90, Chapter 4], [GH88], and follows from results in [CS09] in the infinite alphabet case. See Section 4.

4. The Proofs of Theorems 3.2 and 3.3

The material in this section is not used in other parts of the paper and can be skipped at first reading.

As in [Rue78], [PP90], [GH88], the proof uses the *transfer operator method*: First we represent $\mathfrak{p}_{\phi,\psi}(t)$ as the logarithm of the leading eigenvalue $\lambda(t)$ of Ruelle's operator $L_{\phi+t\psi}$ (see (2.3)); Then we will analyze the dependence of $\lambda(t)$ on t, using perturbation theory. The perturbative analysis hinges on the following fact from §2.7: If ϕ is SPR, then L_ϕ acts with spectral gap on some "nice" Banach space.

Standing assumptions for this section: Throughout this section, we suppose that Σ^+ is a topologically mixing TMS with positive finite Gurevich entropy, ϕ is an SPR θ-weakly Hölder continuous potential such that $\sup\phi < \infty$, and $\psi \in \mathcal{H}_\beta$ where $e^{-\beta} \le \theta$ and $\|\psi\|_\beta \le 1$. Since $\beta \mapsto \|\psi\|_\beta$ is monotonically increasing, if Theorem 3.2 holds for the β such that $e^{-\beta} = \theta$, then it holds with the same ε, M for all β such that $e^{-\beta} \le \theta$. Henceforth we assume $e^{-\beta} = \theta$.

Lemma 4.1. *For every state a, $\underline{x} \in \Sigma^+$, $t \in \mathbb{R}$ and $n \ge 1$, $(L^n_{\phi+t\psi}1_{[a]})(\underline{x}) < \infty$ and $\mathfrak{p}_{\phi,\psi}(t) = P_G(\phi + t\psi) = \lim\limits_{n\to\infty} \frac{1}{n}\log(L^n_{\phi+t\psi}1_{[a]})(\underline{x})$.*

Proof. Fix t, and let $\varphi := \phi + t\psi$. Set $B := \sum_{n\ge 2} \mathrm{osc}_n(\varphi)$. Then for every cylinder $[a, \xi_1, \dots, \xi_{n-1}, b]$ of length $n+1$, and for every $\underline{x}, \underline{y} \in [a, \xi_1, \dots, \xi_{n-1}, b]$,

$$|\varphi_n(\underline{x}) - \varphi_n(\underline{y})| \le B.$$

Let $\varphi_n^\pm[a, \xi_1, \dots, \xi_{n-1}, a]$ denote the supremum $(+)$ or infimum $(-)$ of φ_n on $[a, \xi_1, \dots, \xi_{n-1}, a]$. These differ from each other by at most B. It follows that

$$e^{-B} \sum_{[a,\xi_1,\dots,\xi_{n-1},a]\neq\varnothing} e^{\varphi_n^+[a,\xi_1,\dots,\xi_{n-1},a]} \le Z_n(\varphi, a) \le e^B \sum_{[a,\xi_1,\dots,\xi_{n-1},a]\neq\varnothing} e^{\varphi_n^-[a,\xi_1,\dots,\xi_{n-1},a]}$$

Similarly, one shows that if $[a, \underline{\xi}, a]$ and $[a, \underline{\eta}, a]$ are non-empty cylinders of lengths $m+1$ and $n+1$, then $|\varphi^+_{m+n}[a, \underline{\xi}, a, \underline{\eta}, a] - \varphi^-_m[a, \underline{\xi}, a] - \varphi^-_n[a, \underline{\eta}, a]| \le 2B$.

These estimates can be used to show that $Z_n(\varphi, a)^k \le \exp(2kB) Z_{kn}(\varphi, a)$. By the standing assumptions, $\lim \frac{1}{\ell}\log Z_\ell(\varphi, a) = P_G(\varphi) \le P_G(0) + \sup\varphi = h + \sup\varphi < \infty$, and therefore $Z_{kn}(\varphi, a) < \infty$ for every n, for all k large enough. So

$$Z_n(\varphi, a) < \infty \text{ for all } n. \tag{4.1}$$

If $\underline{x} \in [a]$, $(L^n_\varphi 1_{[a]})(\underline{x}) = \sum\limits_{\sigma^n(\underline{y})=\underline{x}} e^{\varphi_n(\underline{y})} 1_{[a]}(\underline{y}) = \sum\limits_{[a,\xi_1,\dots,\xi_{n-1},a]\neq\varnothing} e^{\varphi_n(a,\underline{\xi},a,x_1,x_2,\dots)}$.

The exponent is sandwiched between $\varphi_n^\pm[a, \underline{\xi}, a]$, so by the previous paragraph,

$$e^{-B} Z_n(\varphi, a) \le (L^n_\varphi 1_{[a]})(\underline{x}) \le e^B Z_n(\varphi, a) \tag{4.2}$$

for all n. It follows that $L^n_\varphi 1_{[a]}$ is finite on $[a]$ for all n, and $\frac{1}{n}\log(L^n_\varphi 1_{[a]})(\underline{x}) \xrightarrow[n\to\infty]{} P_G(\varphi)$ for all $\underline{x} \in [a]$. This proves the lemma in the special case when $\underline{x} \in [a]$.

Suppose $\underline{x} \in [b]$ where $b \neq a$. By topological mixing, there is a finite admissible path $b\underline{\xi}a = (b, \xi_1, \dots, \xi_{p-1}, a)$. If $\underline{y}$ is a σ^n-pre-image of $\underline{x}$ in $[a]$, then $(b, \underline{\xi}, \underline{y})$ is a σ^{n+p}-preimage of $\underline{x}$ in $[b]$. Therefore

$$e^{\varphi_p^-[b\underline{\xi}a]} L^n_\varphi 1_{[a]}(\underline{x}) = e^{\varphi_p^-[b\underline{\xi}a]} \sum_{\sigma^n(\underline{y})=\underline{x}} e^{\varphi_n(\underline{y})} 1_{[a]}(\underline{y}) \le e^B (L^{n+p}_\varphi 1_{[b]})(\underline{x}).$$

Since $\underline{x} \in [b]$, $(L^{n+p}_\varphi 1_{[b]})(\underline{x}) \le \mathrm{const}\, Z_{n+p}(\varphi, b)$ on $[b]$, by (4.2). It follows that

$$L^n_\varphi 1_{[a]} \le \mathrm{const}\, Z_{n+p}(\varphi, b) < \infty \text{ on } [b],$$

and $\limsup \frac{1}{n}\log L^n_\phi 1_{[a]} \le P_G(\varphi)$ on $[b]$.

Topological mixing also gives us a finite admissible path $(a, \eta_1, \dots, \eta_{q-1}, b)$. Let $\underline{z} := (a, \underline{\eta}, \underline{x})$. If $\sigma^n(\underline{y}) = \underline{z}$, then $\sigma^{n+q}(\underline{y}) = \underline{x}$. So

$$e^{-\varphi_q(\underline{z})} (L^{n+q}_\varphi 1_{[a]})(\underline{x}) \ge (L^n_\varphi 1_{[a]})(\underline{z}) \ge e^{-B} Z_n(\varphi, a),$$

and $\liminf \frac{1}{n}\log L_\phi^n 1_{[a]} \geq P_G(\varphi)$ on $[b]$. $\square$

The following lemma generalizes [Sar01a, Theorem 3] by making ε independent of ψ (the importance of this is explained at the end of §6). The weak Hölder continuity and the SPR property of ϕ are essential.

Lemma 4.2. *There exists $\varepsilon := \varepsilon_\theta(\phi) > 0$ such that for every $\psi \in \mathcal{H}_\beta$ such that $\|\psi\|_\beta \leq 1$, $\mathfrak{p}_{\phi,\psi}(t)$ is real-analytic on $(-\varepsilon, \varepsilon)$.*

Proof. By our assumptions ϕ is θ-weakly Hölder continuous, and strongly positively recurrent. As explained in § 2.7, this implies that L_ϕ acts with spectral gap on a Banach space $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ over $\mathbb{C}$, with properties (a)–(g) as listed there. The space $\mathcal{L}$ depends on ϕ and on θ.[4]

Property (e) says that the spectrum of $L_\phi : \mathcal{L} \to \mathcal{L}$ consists of a simple eigenvalue $\lambda := \exp P_G(\phi)$, and a compact set $K_0 \subset \{z \in \mathbb{C} : |z| < \lambda\}$. It is well-known that this spectral picture persists for small perturbations of L_ϕ.

Specifically, let γ' be a smooth parametrization of a circle in $\mathbb{C}$, with center zero, radius $R' < \lambda$, and such that γ' contains K_0 in its interior. Let γ be a smoothly parameterized closed circle with center λ, and radius R so small that γ is completely outside γ'. By the theory of analytic perturbations of linear operators on Banach spaces, there exists $\delta_0 := \delta_0(\phi, \theta) > 0$ as follows. If L is a bounded linear operator such that $\|L - L_\phi\| < \delta_0$, then

(A) the spectrum of L consists of a simple eigenvalue $\lambda(L)$ inside γ, and a compact subset K_L inside γ'.
(B) $P = P(L) := \frac{1}{2\pi i}\oint_\gamma (\xi I - L)^{-1} d\xi$ is a well-defined operator such that $P^2 = P$, $PL = LP = \lambda(L)P$, $\dim \mathrm{Image}(P) = 1$, and
$$\mathrm{spectrum}(LP) = \{\lambda(L)\}\ ,\ \mathrm{spectrum}(L(I - P)) = K_L.$$

See [Kat95, Theorem IV-3.16].

Define the operators $(L_z f)(\underline{x}) := \sum_{\sigma(\underline{y})=\underline{x}} e^{\phi(\underline{y})+z\psi(\underline{y})} f(\underline{y})$ and $M_\psi f = \psi f$. By property (g), M_ψ is a bounded linear operator on $\mathcal{L}$, and $\|M_\psi\| \leq \|\psi\|_\beta \leq 1$. Thus
$$L_z = L_\phi M_{\exp z\psi} = \sum_{n=0}^{\infty} \frac{z^n}{n!} L_\phi M_\psi^n, \tag{4.3}$$
and this series converges absolutely on $\mathcal{L}$ in the operator norm, because $\|L_\phi M_\psi^n\| \leq \|L_\phi\|\|\psi\|_\beta^n \leq \|L_\phi\|$. It follows that L_z is a bounded linear operator on $\mathcal{L}$, and
$$\|L_z - L_\phi\| \leq |z| \sum_{n=1}^{\infty} \frac{|z|^{n-1}}{n!}\|L_\phi\| \leq 2\|L_\phi\| \cdot |z|, \text{ for all } |z| \leq 1.$$
Let $\varepsilon^{(1)} := \varepsilon^{(1)}(\theta, \phi) := \frac{\delta_0}{3\|L_\phi\|}$. If $|z| < \varepsilon^{(1)}$, then $\|L_z - L_\phi\| < \delta_0$, whence
$$P_z := P(L_z)$$
is well-defined. Notice that $\varepsilon^{(1)}$ is independent of ψ.

Next we claim that there exist $\varepsilon^{(2)} := \varepsilon^{(2)}(\theta, \phi)$ and $K := K(\theta, \phi)$ such that for all $|z| < \varepsilon^{(2)}$ and $\xi \in \gamma$, $\xi I - L_z$ is invertible, and $\|(\xi I - L_z)^{-1}\| \leq K$.

To see this, recall that by choice of γ, the spectrum of L_ϕ does not intersect γ, and therefore $\xi I - L_\phi$ has bounded inverse for all $\xi \in \gamma$. The resolvent set of

[4]The dependence on θ is genuine: For some ϕ, e.g. $\phi \equiv 0$, there is no "canonical" θ.

L_ϕ is open, therefore the norm of this inverse is locally bounded on γ, whence by compactness, less than some global constant $K' = K'(\theta, \phi)$ everywhere on γ. The formal calculation

$$(\xi I - L_z)^{-1} = (\xi I - L_\phi - (L_z - L_\phi))^{-1} = (\xi I - L_\phi)^{-1} \sum_{n=0}^{\infty} [(L_z - L_\phi)(\xi I - L_\phi)^{-1}]^n$$

shows that if $K' \| L_z - L_\phi \| \leq \frac{1}{2}$, then $\xi I - L_z$ is invertible, and the norm of the inverse is bounded by $2K'$.

Let $\varepsilon^{(2)} := \varepsilon^{(2)}(\theta, \phi) := \min\{\varepsilon^{(1)}, \frac{1}{4\|L_\phi\| K'}\}$, $K := K(\theta, \phi) := 2K'$. These constants are independent of ψ, and if $|z| < \varepsilon^{(2)}$, then $\|(\xi I - L_z)^{-1}\| \leq K$.

We now investigate the properties of $P_z := P(L_z)$.

P_0 is the eigenprojection of $\lambda_0 := \lambda$. By the spectral gap property for ϕ and [CS09, Lemma 8.1],

$$P_0 f = h_0 \int f d\nu_0 \tag{4.4}$$

where h_0 is a positive function, uniformly bounded away from zero and infinity on partition sets, and ν_0 is a positive measure which is finite and positive on cylinders. In fact, by (4.4) $h_0 d\nu_0$ is the RPF measure, whence the equilibrium measure, of ϕ.

Claim: *$z \mapsto P_z$ is analytic on $\{z \in \mathbb{C} : |z| < \varepsilon^{(2)}\}$.*

Proof of the claim: If $|z|, |w| < \varepsilon^{(2)}$, then P_z, P_w are well-defined, and by (B), $\frac{P_z - P_w}{z-w} = \frac{1}{2\pi i} \oint_\gamma \frac{(\xi I - L_z)^{-1} - (\xi I - L_w)^{-1}}{z-w} d\xi$. The integrand satisfies the identity

$$\frac{(\xi I - L_z)^{-1} - (\xi I - L_w)^{-1}}{z - w} = -(\xi I - L_z)^{-1} \left(\frac{L_z - L_w}{z - w} \right) (\xi I - L_w)^{-1}.$$

To see this, start with the left-hand side, pull $(\xi I - L_z)^{-1}$ to the left, and pull $(\xi I - L_w)^{-1}$ to the right.

It is not difficult to verify, using (4.3), that for every $|z| < \varepsilon^{(2)}$,

$$-(\xi I - L_z)^{-1} \left(\frac{L_z - L_w}{z - w} \right) (\xi I - L_w)^{-1} \xrightarrow[w \to z]{} (\xi I - L_z)^{-1} L_z' (\xi I - L_z)^{-1},$$

where $L_z' := \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} L_\phi M_\psi^n$.

If $|z|, |w| < \varepsilon^{(2)}$ and $\xi \in \gamma$, then $\|(\xi I - L_z)^{-1}\|, \|(\xi I - L_w)^{-1}\| < K$, and

$$\begin{aligned} \left\| \frac{L_z - L_w}{z - w} \right\| &\leq \sum_{n=1}^{\infty} \frac{|z^n - w^n|}{n! |z - w|} \|L_\phi\| \|\psi\|_\beta^n \quad \text{, by (4.3),} \\ &\leq \|L_\phi\| \sum_{n=1}^{\infty} \frac{|z|^{n-1} + |z|^{n-2}|w| + \cdots + |w|^{n-1}}{n!} < 3\|L_\phi\|. \end{aligned}$$

Therefore $\| -(\xi I - L_z)^{-1} \left(\frac{L_z - L_w}{z - w} \right) (\xi I - L_w)^{-1} \| \leq 3K^2 \|L_\phi\|$.

Now fix some arbitrary bounded linear functional φ on the Banach space of bounded linear operators on $\mathcal{L}$. By the previous discussion and the bounded convergence theorem,

$$\lim_{w \to z} \varphi \left(\frac{P_z - P_w}{z - w} \right) = \frac{1}{2\pi i} \oint_\gamma \varphi \left[(\xi I - L_z)^{-1} L_z' (\xi I - L_z)^{-1} \right] d\xi.$$

Thus P_z is differentiable in the weak sense at z. By a well-known consequence of the Banach-Steinhaus theorem, this implies differentiability in the strong sense, and indeed holomorphy on $\{z \in \mathbb{C} : |z| < \varepsilon^{(2)}\}$. The claim is proved.

We note for future reference a by-product of the previous proof: If $|z|, |w| < \varepsilon^{(2)}$, then

$$\left\| \frac{P_z - P_w}{z-w} \right\| \leq \left\| \oint_\gamma \frac{(\xi I - L_z)^{-1} - (\xi I - L_w)^{-1}}{z-w} d\xi \right\| \leq \frac{3K^2 \|L_\phi\|}{2\pi} \cdot \text{length}(\gamma).$$

So $\|P_z - P_w\| \leq C|z-w|$ with $C := C(\theta, \phi)$ independent of ψ.

Our next task is to prove the analyticity of $\lambda_z := \lambda(L_z)$ on some neighborhood of zero. Pick a bounded linear functional φ on $\mathcal{L}$ such that $\varphi(h_0) \neq 0$. Since $\|P_z - P_w\| \leq C|z-w|$, there exists $0 < \varepsilon^{(3)}(\theta, \phi) < \varepsilon^{(2)}$ independent of ψ such that $\varphi(P_z h_0) \neq 0$ for all $|z| < \varepsilon^{(3)}$. From (B) and linearity of φ, we get that

$$\lambda(L_z) = \frac{\varphi(L_z P_z h_0)}{\varphi(P_z h_0)},$$

a ratio of two holomorphic non-vanishing functions on $\{z : |z| < \varepsilon^{(3)}\}$. So λ_z is holomorphic there on $\{z : |z| < \varepsilon^{(3)}\}$.

We are now ready to prove the lemma. By (4.4), $P_0 1_{[a]} = \nu_0[a] h_0 \neq 0$, whence $\|P_0 1_{[a]}\| \neq 0$. Using $\|P_z - P_w\| \leq C|z-w|$, we find that for some $0 < \varepsilon^{(4)}(\theta, \phi) < \varepsilon^{(3)}_\beta$ independent of ψ, $\|P_z 1_{[a]}\| \neq 0$ for all $|z| < \varepsilon^{(4)}$.

Fix $|z| < \varepsilon^{(4)}$ and choose some $\underline{x} \in \Sigma^+$ such that $(P_z 1_{[a]})(\underline{x}) \neq 0$. Let $N_z := L_z(I - P_z)$. Since $P_z^2 = P_z$ and $P_z L_z = \lambda(L_z) P_z$, we have $P_z N_z = N_z P_z = 0$. By (A) and (B), the spectrum of N_z is inside γ' and λ_z is outside γ', therefore $\lambda_z^{-n} N_z^n \to 0$ in norm. So

$$\lambda_z^{-n} L_z^n = \lambda_z^{-n} (\lambda_z P_z + N_z)^n = \lambda_z^{-n} (\lambda_z^n P_z + N_z^n) \to P_z \text{ in } \mathcal{L}.$$

By property (c) of $\mathcal{L}$, $\lambda_z^{-n} (L_z^n 1_{[a]})(\underline{x}) \to (P_z 1_{[a]})(\underline{x}) \neq 0$ for all $\underline{x}$.

We now specialize to the case of *real* t, $|t| < \varepsilon^{(4)}$. For such t, $(L_t^n 1_{[a]})(\underline{x}) = (L_{\phi+t\psi}^n 1_{[a]})(\underline{x})$ is real-valued. Necessarily, λ_t and $(P_t 1_{[a]})(\underline{x})$ are also real-valued.[5] Passing to natural logarithms, we find using Lemma 4.1 that

$$\log \lambda_t = \lim_{n\to\infty} \frac{1}{n} \log (L_{\phi+t\psi}^n 1_{[a]})(\underline{x}) = \mathfrak{p}_{\phi,\psi}(t) \text{ whenever } |t| < \varepsilon^{(4)}.$$

By (A), (B) and the choice of $\varepsilon^{(3)}$, $z \mapsto \lambda_z$ maps $\{z : |z| < \varepsilon^{(3)}\}$ holomorphically into the interior of γ. Zero is outside γ. If $\mathrm{Log}(z)$ is a branch of the complex logarithm which is holomorphic on the interior of γ, then $\mathrm{Log}\, \lambda_z$ is holomorphic on $\{z \in \mathbb{C} : |z| < \varepsilon^{(3)}\}$, and agrees with $\mathfrak{p}_{\phi,\psi}(t)$ on $(-\varepsilon^{(4)}, \varepsilon^{(4)})$.

It follows that $\mathfrak{p}_{\phi,\psi}(t)$ is real-analytic on $(-\varepsilon_\theta(\phi), \varepsilon_\theta(\phi))$, where $\varepsilon_\theta(\phi) := \varepsilon^{(4)}$. □

Lemma 4.3. *There exists $\varepsilon_\theta(\phi) > 0$ such that for every $\psi \in \mathcal{H}_\beta$ for which $\|\psi\|_\beta \leq 1$, $\phi + t\psi$ has the spectral gap property for each $|t| < \varepsilon_\theta(\phi)$.*

Proof. We continue with the notation of the proof of the previous lemma. Take $\varepsilon := \varepsilon^{(4)}(\theta, \phi)$ and t real such that $|t| < \varepsilon$. We claim that if $\|\psi\|_\beta \leq 1$, then $\phi + t\psi$ satisfies the spectral gap property with the Banach space $\mathcal{L}$ from the previous proof.

[5]If $z^n r_n \to w \neq 0$ and r_n are real, then z, w must also be real, otherwise $\frac{z^n r_n}{|z^n r_n|} \not\to \frac{w}{|w|}$.

Property (a) for $\phi + t\psi$ follows from property (a) for ϕ, because ψ is bounded. Properties (b), (c) and (g) are obvious, because they do not involve t. Property (d) is because thanks to (g), the series (4.3) converges in norm.

Property (e) is because for all $|t| < \varepsilon^{(1)}$, $\|L_t - L_\phi\| < \delta_0$, and therefore there is a projection P_t onto a one-dimensional space such that $P_t L_t = L_t P_t$, $L_t P_t = \lambda_t P_t$, and so that the spectrum of $N_t := L_t(I - P_t)$ is contained in a disc with center at the origin and radius strictly less than $|\lambda_t|$. In addition, as we saw at the end of the previous proof, if $|t| < \varepsilon^{(4)}$, then $\lambda_t = \exp \mathfrak{p}_{\phi,\psi}(t) = \exp P_G(\phi + t\psi)$.

Property (f) is because the following series converges in norm for every $\psi' \in \mathcal{H}_\beta$: $L_{\phi+t\psi+z\psi'} = L_{\phi+t\psi} M_{\exp(z\psi')} = \sum_{n=0}^{\infty} \frac{z^n}{n!} L_t M^n_{z\psi'}$, see (4.3). □

Lemma 4.4. *There exist $\varepsilon_\theta(\phi), M_\theta(\phi) > 0$ so that for every $\psi \in \mathcal{H}_\beta$ such that $\|\psi\|_\beta \leq 1$ and for every $|t| < \varepsilon_\theta(\phi)$, $|\mathfrak{p}'_{\phi,\psi}(t)|, |\mathfrak{p}''_{\phi,\psi}(t)|, |\mathfrak{p}'''_{\phi,\psi}(t)| \leq M_\theta(\phi)$.*

Proof. Let $\varepsilon^{(4)}$ and δ_0 be as in the proof of Lemma 4.2. Then $\mathfrak{p}_{\phi,\psi}(t)$ extends to a holomorphic function $f(z) = \operatorname{Log} \lambda_z$ on a neighborhood of $\{z : |z| < \varepsilon^{(4)}\}$, where $\lambda_z := \lambda(P_z)$ is an eigenvalue of operator L_z such that $\|L_z - L_\phi\| < \delta_0$, and $\operatorname{Log}(z)$ is a suitable branch of the complex logarithm.

By the choice of δ_0, λ_z is in the interior of some fixed circle γ and outside some fixed circle γ' surrounding zero, with γ, γ' independent of ψ. So $|f(z)|$ is uniformly bounded on $\{z : |z| < \varepsilon^{(4)}\}$ by some constant which is independent of ψ.

Take $\varepsilon := \frac{1}{2}\varepsilon^{(4)}$. The lemma now follows from Cauchy's integral formula for the derivatives of $f(z)$ ("Cauchy's bounds"). □

Lemma 4.5. *There exists $\varepsilon_\theta(\phi) > 0$ such that for all $\psi, \varphi \in \mathcal{H}_\beta$ for which $\|\psi\|_\beta, \|\varphi\|_\beta \leq 1$, the function $p(s,t) := P_G(\phi + t\psi + s\varphi)$ has continuous partial derivatives of all orders on $(-\varepsilon_\theta(\phi), \varepsilon_\theta(\phi))^2$.*

Proof. The proof is similar to the proof of Lemma 4.2, so we only sketch it. Let $\mathcal{L}$ be the Banach space in that proof, and define the operator

$$L_{z,w} f := L_{\phi+z\psi+w\varphi} f \qquad (z, w \in \mathbb{C}).$$

Then $L_{z,w} = L_\phi M_{\exp z\psi} M_{\exp w\varphi} = \sum_{n,m=0}^{\infty} \frac{z^n w^m}{n!m!} L_\phi M^n_\psi M^m_\varphi$, and the series converges in norm for all $z, w \in \mathbb{C}$, because $\|L_\phi M^n_\psi M^m_\varphi\| \leq \|L_\phi\| \|\psi\|^n_\beta \|\varphi\|^m_\beta \leq \|L_\phi\|$ by property (g) of $\mathcal{L}$. So $L_{z,w}$ are well-defined bounded linear operators on $\mathcal{L}$.

The series representation for $L_{z,w}$ implies that for all $|z|, |w| \leq 1$,$|z_0|, |w_0| \leq 1$,

$$\|L_{z,w} - L_{z_0,w_0}\| \leq \|L_\phi\| \sum_{n,m=0}^{\infty} \frac{|z^n w^m - z_0^n w_0^m|}{n!m!}$$

$$\leq \|L_\phi\| \sum_{\substack{n,m=0 \\ (n,m)\neq(0,0)}}^{\infty} \frac{|z|^n |w^m - w_0^m| + |w_0|^m |z^n - z_0^n|}{n!m!} \leq e^2 \|L_\phi\| (|z - z_0| + |w - w_0|).$$

In particular, if δ_0 is as in the proof of Lemma 4.2, then there exists $\kappa^{(1)} > 0$ independent of ψ, φ such that $\|L_{z,w} - L_\phi\| < \delta_0$ for all $|z|, |w| \leq \kappa^{(1)}$.

By the definition of δ_0, for such z, w, the spectrum of $L_{z,w}$ does not intersect γ, therefore $(\xi I - L_{z,w})^{-1}$ is well-defined and bounded for all $\xi \in \gamma$. Arguing as in the proof of Lemma 4.2, we find $0 < \kappa^{(2)} < \kappa^{(1)}$ and $K > 0$ independent of ψ, φ such that for all $|z|, |w| \leq \kappa^{(2)}$ and for all ξ on γ,

$$\|(\xi I - L_{z,w})^{-1}\| < K.$$

As in the proof of that lemma, this can be used to show that $(z,w)\mapsto P_{z,w}$ is analytic separately in each of its variables on $\{(z,w)\in\mathbb{C}^2 : |z|,|w|<\kappa_\beta^{(2)}\}$.

It follows that for every bounded linear functional F on the Banach space of bounded linear operators on $\mathcal{L}$, $(z,w)\mapsto F(P_{z,w})$ is holomorphic in z and in w, whence by Hartogs' theorem, in both variables on $\{(z,w) : |z|,|w|<\kappa^{(2)}\}$. In particular, $F(P_{z,w})$ has continuous partial derivatives of all orders there.

Since this holds for all bounded linear functionals F, $P_{z,w}$ has continuous partial derivatives of all orders on $\{(z,w) : |z|,|w|<\kappa^{(2)}\}$. We now continue exactly as in the proof of Lemma 4.2 to construct $0<\kappa^{(4)}<\kappa^{(3)}<\kappa^{(2)}$ independent of ψ,φ and a branch of the complex logarithm $\mathrm{Log}(z)$ such that

(1) $\mathrm{Log}\,\lambda(P_{z,w})$ has partial derivatives of all orders on $\{(z,w) : |z|,|w|<\kappa^{(3)}\}$
(2) $\mathrm{Log}\,\lambda(P_{t,s})=P_G(\phi+t\psi+s\varphi)$ for all t,s real such that $|t|,|s|<\kappa^{(4)}$.

It follows that $(t,s)\mapsto P_G(\phi+t\psi+s\varphi)$ has continuous partial derivatives of all orders on $\{(t,s)\in\mathbb{R}^2 : |t|,|s|<\kappa_\beta^{(4)}\}$. □

Proof of Theorem 3.2. Let $\varepsilon:=\varepsilon(\theta,\phi)$ be the minimum of the epsilons in the previous lemmas, and let $M:=M_\theta(\phi)$ be as in Lemma 4.4.

Part (1): This is Lemma 4.2.

Part (2): By Lemma 4.3, $\phi+t\psi$ has the spectral gap property. As explained in sections 2.7 and 2.5, this implies that ϕ is strongly positively recurrent, whence positively recurrent. Let m_t denote the RPF measure of $\phi+t\psi$ (see § 2.5).

Σ^+ is topologically mixing with finite Gurevich entropy h. So $h_{m_t}(\sigma)\le h<\infty$, and by Theorem 2.1, m_t is the unique equilibrium measure of $\phi+t\psi$.

Part (3): The proof of this and the following part is essentially in [GH88] and [PP90, Chapter 4], but we give it for completeness.

Let λ_z, L_z, P_z be as in the proof of Lemma 4.2. By part (2), if t is real and $|t|<\varepsilon$, then $\phi+t\psi$ is positively recurrent, whence by the generalized Ruelle's Perron-Frobenius theorem, there is a positive continuous function h_t and a measure ν_t finite and positive on cylinders such that

$$L_t h_t=\lambda_t,\, L_t^*\nu_t=\lambda_t\nu_t,\, \int h_t d\nu_t=1.$$

By [CS09, Lem 8.1], $h_t\in\mathcal{L}$, and $P_t f=h_t\int f d\nu_t$ for all $f\in\mathcal{L}$.

We saw in the proof of Lemma 4.2 that $t\mapsto\lambda_t,P_t,L_t$ are differentiable, and $\lambda_t=\exp\mathfrak{p}_{\phi,\psi}(t)$. By (4.3), $L_t' f:=\frac{d}{dt}L_t f=L_t(\psi f)=L_t M_\psi f$. Differentiating both sides of the identity $L_tP_t=\lambda_tP_t$ gives

$$L_tM_\psi P_t+L_tP_t'=\mathfrak{p}_{\phi,\psi}'(t)\lambda_tP_t+\lambda_tP_t'.$$

We multiply by P_t and cancel the equal terms $P_tL_tP_t'=\lambda_tP_tP_t'$, with the result

$$\lambda_tP_tM_\psi P_t=\mathfrak{p}_{\phi,\psi}'(t)\lambda_tP_t.$$

We now apply the operators on the two sides of the equation to h_t (which belongs to $\mathcal{L}$), and obtain $\lambda_tP_t(\psi h_t)=\mathfrak{p}_{\phi,\psi}'(t)\lambda_t h_t$. Substituting $t=0$ and noting that $\lambda_0=\exp P_G(\phi)\neq 0$, we obtain $\mathfrak{p}_{\phi,\psi}'(0)=\int\psi h_0 d\nu_0$. Since $h_0\nu_0$ is the equilibrium measure of ϕ, we are done.

Part (4): It is enough to consider the special case when $P_G(\phi) = 0$ and $\int \psi dm_0 = 0$, since the general case can be reduced to this one by subtracting constants from ϕ and ψ. In this case $\lambda_0 = 1$ and $\lambda_0' = \frac{d}{dt}\big|_{t=0}\lambda_t = \int \psi dm_0 = 0$.

Fix n, differentiate $L_t^n P_t = \lambda_t^n P_t$ twice, then apply P_t from the left, and substitute $t = 0$, dropping all the terms which contain λ_0'. Collecting terms we obtain

$$n\mathfrak{p}_\psi''(0)P_0 = P_0 M_{\psi_n}^2 P_0 + 2P_0 M_{\psi_n} P_0'.$$

Applying this to h_0/n, writing $H_0' = P_0' h_0$ and using $P_0 f = h_0 \nu_0(f)$, we get

$$\mathfrak{p}_\psi''(0) = \frac{1}{n} m_0\left((\psi_n)^2\right) + 2\nu_0\left(\frac{\psi_n}{n} H_0'\right). \tag{4.5}$$

To complete the proof, it remains to show that the second summand tends to zero.

The integrand $\frac{\psi_n}{n} H_0'$ tends to zero ν_0-a.e.: First, ψ is bounded and m_0 is ergodic, so $\frac{\psi_n}{n} \to \int \psi dm_0 = 0$ m_0-almost everywhere. Second, $h_0 > 0$, so $\nu_0 = \frac{1}{h_0} m_0 \ll m_0$, and $\frac{\psi_n}{n} \to 0$ ν_0-almost everywhere.

The integrand $\frac{\psi_n}{n} H_0'$ is dominated by an $L^1(\nu_0)$–function: $h_0 \in \mathcal{L}$, so $H_0' \in \mathcal{L}$, whence by property (b), $|H_0'| \in \mathcal{L}$. So $P_0 H_0' \in \mathcal{L}$. We saw above that $P_0 f = h_0 \nu_0(f)$ for all $f \in \mathcal{L}$. In particular, $\nu_0(|H_0'|) < \infty$. So $|\frac{\psi_n}{n} H_0'| \le \|\psi\|_\infty |H_0'|$ and $H_0' \in L^1(\nu_0)$. By the dominated convergence theorem, $\nu_0\left(\frac{\psi_n}{n} H_0'\right) \to 0$.

Parts (5) and (6): This is the content of Lemmas 4.4 and 4.5. □

Proof of Theorem 3.3.

Part (2): Let $\psi_n := \psi + \psi \circ \sigma + \cdots + \psi \circ \sigma^{n-1}$. Recall the definitions

$$\mathfrak{p}_{\phi,\psi}(t) = P_G(\phi + t\psi) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\phi + t\psi, a),$$
$$\text{where } Z_n(\phi + t\psi, a) = \sum_{\sigma^n(\underline{x})=\underline{x}} e^{\phi_n(\underline{x}) + t\psi_n(\underline{x})} 1_{[a]}(\underline{x}).$$

If $\psi = \varphi + r - r \circ \sigma + c$, then $\psi_n(\underline{x}) = \varphi_n(\underline{x}) + cn$ for all x such that $\sigma^n(\underline{x}) = \underline{x}$. It follows that $\mathfrak{p}_{\phi,\psi}(t) = \mathfrak{p}_{\phi,\varphi}(t) + ct$, whence $\mathfrak{p}_{\phi,\psi}'' = \mathfrak{p}_{\phi,\varphi}''$. By Theorem 3.2, $\sigma_m^2(\psi) = \mathfrak{p}_{\phi,\psi}''(0) = \mathfrak{p}_{\phi,\varphi}''(0) = \sigma_m^2(\varphi)$.

Part (1): Suppose $\psi = r - r \circ \sigma + c$ with r continuous and c a constant. Then $\sigma_m(\psi) = \sigma_m(0) = 0$, by part 2.

Now suppose $\sigma_m(\psi) = 0$. Without loss of generality, $\int \psi dm = 0$ (otherwise work with $\psi - \int \psi dm$). By Lemma 4.3, ϕ has the spectral gap property, and we are in the situation discussed in [CS09, Appendix B]. The expression σ^2 defined there agrees with $\sigma_m^2(\psi)$ by [CS09, Equation (8.4)] and part (4) of Theorem 3.2.[6] Using the argument in [CS09, pp. 664-665], we find that ψ must be a coboundary with a continuous transfer function.

Part (3): Let M be the constant from Theorem 3.2(5). If $\psi \equiv 0$, then $\sigma_m(\psi) = 0$ and there is nothing to prove. Suppose $\psi \not\equiv 0$. It is easy to check that $\sigma_m(t\psi) = |t|\sigma_m(\psi)$. So $\sigma_m(\psi) = \|\psi\|_\beta \sigma_m(\varphi)$, where $\varphi := \frac{\psi}{\|\psi\|_\beta}$. Since $\|\varphi\|_\beta = 1$, $\sigma_m^2(\varphi) = \mathfrak{p}_{\phi,\varphi}''(0) \le M$. So $\sigma_m(\psi) \le \|\psi\|_\beta \sigma_m(\varphi) \le M^{1/2}\|\psi\|_\beta \le M\|\psi\|_\beta$, where the last inequality is because $M > 1$. □

[6] Note that the λ_t in [CS09] is what we call in this paper λ_{it}.

5. The restricted pressure function

Suppose $\sigma : \Sigma^+ \to \Sigma^+$ is a topologically mixing countable Markov shift with finite Gurevich entropy h, and ϕ is a weakly Hölder continuous function such that $\sup \phi < \infty$. Fix $\psi \in \mathcal{H}_\beta$ which is not cohomologous to a constant via a continuous transfer function. By Theorem 3.1 $\mathfrak{p}'_{\phi,\psi}(-\infty) < \mathfrak{p}'_{\phi,\psi}(+\infty)$, and we can make the following definition:

Definition 5.1. *The* restricted pressure function of ϕ, constrained on ψ, *is* $\mathfrak{q}_{\phi,\psi} : (\mathfrak{p}'_{\psi,\phi}(-\infty), \mathfrak{p}'_{\psi,\phi}(+\infty)) \to \mathbb{R}$, *given by*

$$\mathfrak{q}_{\phi,\psi}(a) := \sup\{h_\mu(\sigma) + \int \phi d\mu : \mu \in \mathscr{M}(\Sigma^+), \int \psi d\mu = a\}.$$

Lemma 5.1. *Under the assumptions above, $\mathfrak{q}_{\phi,\psi}$ is a well-defined finite valued concave function, which is uniformly bounded from above.*

Proof. Since $h < \infty$ and $\sup \phi < \infty$, $\mathscr{M}_\phi(\Sigma^+) = \mathscr{M}(\Sigma^+)$ and $h_\mu(\sigma) + \int \phi d\mu$ is well-defined for every $\mu \in \mathscr{M}(\Sigma^+)$. By Theorem 3.1, $a \in (\mathfrak{p}'_{\phi,\psi}(-\infty), \mathfrak{p}'_{\phi,\psi}(+\infty))$ if and only if $\inf\{\int \psi d\mu\} < a < \sup\{\int \psi d\mu\}$, therefore there exist two invariant measures μ_1, μ_2 such that $\int \psi d\mu_1 < a < \int \psi d\mu_2$. Then $\int \psi d\mu = a$ for some convex combination μ of μ_1, μ_2, and the set $\{\mu : \int \psi d\mu = a\}$ is non-empty. So the supremum in the definition of $\mathfrak{q}_{\phi,\psi}(a)$ is over a non-empty set, and $\mathfrak{q}_{\phi,\psi}(a)$ is well-defined.

Let's choose the measures μ_1, μ_2 more carefully. A standard ergodic decomposition argument shows that we may choose μ_1, μ_2 to be ergodic. On countable Markov shifts, μ_i generic orbits are limits of periodic orbits, therefore we may choose μ_1, μ_2 to be ergodic measures sitting on periodic orbits. For such measures $\int \phi d\mu_i > -\infty$, whence $\int \phi d\mu > -\infty$ for every convex combination of μ_1, μ_2. Looking at the argument in the previous section, we find that $\mathfrak{q}_{\phi,\psi}(a) > -\infty$.

Next, by the variational principle, $\mathfrak{q}_{\phi,\psi}(a) \leq P_G(\phi) < \infty$. So $\mathfrak{q}_{\phi,\psi}$ is finite on its domain, and bounded from above.

Concavity is because for all a_1, a_2 in the domain, for all $0 \leq t \leq 1$, and for all $\varepsilon > 0$, if μ_1, μ_2 are invariant measures such that

$$P_{\mu_i}(\phi) \geq \mathfrak{q}_{\phi,\psi}(a_i) - \varepsilon, \text{ and } \int \psi d\mu_i = a_i,$$

then $\mu := t\mu_1 + (1-t)\mu_2$ is an invariant measure such that $\int \psi d\mu = ta_1 + (1-t)a_2$ and hence by the affine properties of the Kolmogorov-Sinai entropy, $P_\mu(\phi) = tP_{\mu_1}(\phi) + (1-t)P_{\mu_2}(\phi)$. So $\mathfrak{q}_{\phi,\psi}(ta_1 + (1-t)a_2) \geq P_\mu(\phi) \geq t\mathfrak{q}_{\phi,\psi}(a_1) + (1-t)\mathfrak{q}_{\phi,\psi}(a_2) - \varepsilon$. Since ε was arbitrary, the concavity of $\mathfrak{q}_{\phi,\psi}$ follows. □

The restricted pressure function is understood very well for subshifts of finite type [BL98, §3], and for countable Markov shifts with the BIP property [Mor07]. In these cases, and if ψ is not cohomologous to a constant, then this function is smooth and strictly concave.

This is not true in general in the infinite alphabet case, because of the phenomenon of "phase transitions" [Sar01a],[Sar06]. However, as the following theorem shows, in the SPR case there is an explicit subinterval of the domain where $\mathfrak{q}_{\phi,\psi}$ is smooth and uniformly concave. Crucially to the applications we have in mind, this interval can be chosen in a way which depends on ψ only through $\mathbb{E}_m(\psi)$ and $\sigma^2_m(\psi)$, where m is the equilibrium measure of ϕ.

Lemma 5.2. *Let Σ^+ be a topologically mixing countable Markov shift with finite Gurevich entropy. Let ϕ be a θ-weakly Hölder continuous SPR potential such that $\sup\phi < \infty$. Let m denote the unique equilibrium measure of ϕ, and suppose $e^{-\beta} \leq \theta$. Then $\exists \delta_\theta(\phi), H_\theta(\phi) > 0$ as follows. For all $\psi \in \mathcal{H}_\beta$ such that $\|\psi\|_\beta \leq 1$, and $\sigma_m(\psi) \neq 0$:*

(1) $I_\psi := \{t \in \mathbb{R} : |t - \int \psi dm| < \delta_\theta(\phi)\sigma_m^4(\psi)\}$ is contained in the domain of $\mathfrak{q}_{\phi,\psi}$.
(2) $\mathfrak{q}_{\phi,\psi}$ is uniformly bounded and differentiable infinitely many times on I_ψ.
(3) If $a \in I_\psi$, then $-2\sigma_m^{-2}(\psi) \leq \mathfrak{q}''_{\phi,\psi}(a) \leq -\frac{1}{2}\sigma_m^{-2}(\psi)$, $\quad |\mathfrak{q}'''_{\phi,\psi}(a)| \leq H_\theta(\phi)\sigma_m^{-6}(\psi)$. In particular, $\mathfrak{q}_{\phi,\psi}$ is strictly concave on I_ψ.
(4) If $a_0 := \int \psi dm$, then $\mathfrak{q}_{\phi,\psi}(a_0) = P_G(\phi)$, $\quad \mathfrak{q}'_{\phi,\psi}(a_0) = 0$, $\quad \mathfrak{q}''_{\phi,\psi}(a_0) = -\frac{1}{\sigma_m^2(\psi)}$.

Proof. Fix $0 < \beta \leq |\log\theta|$ and let $\varepsilon := \varepsilon_\theta(\phi)$, $M := M_\theta(\phi)$ as in Theorem 3.2. Without loss of generality, $0 < \varepsilon < 0.01$ and $M > 100$. Next, fix $\psi \in \mathcal{H}_\beta$ with norm $\|\psi\|_\beta \leq 1$ and such that $\sigma_m(\psi) \neq 0$. Let

$$\sigma := \sigma_m(\psi),\ a_0 := \int \psi dm,\ p(t) := \mathfrak{p}_{\phi,\psi}(t),\ q(a) := \mathfrak{q}_{\phi,\psi}(a).$$

Claim. *If $|a - a_0| < \frac{\varepsilon\sigma^4}{2M^2}$, then $\exists! t \in \mathbb{R}$ such that $p'(t) = a$. In addition, $\phi + t\psi$ has a unique equilibrium measure m_t, $q(a) = P_{m_t}(\phi)$, and*

$$|t| \leq \frac{\varepsilon\sigma^2}{M^2}\ ,\ (1-\varepsilon)\sigma^2 \leq p''(t) \leq (1+\varepsilon)\sigma^2. \tag{5.1}$$

Proof of the claim. By Theorem 3.2, p' is real-analytic on $(-\varepsilon, \varepsilon)$, $p'(0) = a_0$, $p''(0) = \sigma^2$, and $|p'''| \leq M$ on $(-\varepsilon, \varepsilon)$. Taylor's approximation for p' gives

$$p'(\tfrac{\varepsilon\sigma^2}{M^2}) \geq p'(0) + p''(0)\tfrac{\varepsilon\sigma^2}{M^2} - \tfrac{1}{2!}M\left(\tfrac{\varepsilon\sigma^2}{M^2}\right)^2 > a_0 + \frac{\varepsilon\sigma^4}{2M^2} \geq a.$$

Similarly, $p'(-\frac{\varepsilon\sigma^2}{M^2}) \leq p'(0) - p''(0)\frac{\varepsilon\sigma^2}{M^2} + \frac{1}{2!}M\left(\frac{\varepsilon\sigma^2}{M^2}\right)^2 < a_0 - \frac{\varepsilon\sigma^4}{2M^2} \leq a$. By the intermediate value theorem, $\exists t \in (-\frac{\varepsilon\sigma^2}{M^2}, \frac{\varepsilon\sigma^2}{M^2})$ such that $p'(t) = a$.

To see that this t is unique, we recall that p is convex, therefore p' is monotonically increasing in the broad sense, therefore by the previous inequalities, all solutions to $p'(t) = a$ must belong to $(-\frac{\varepsilon\sigma^2}{M^2}, \frac{\varepsilon\sigma^2}{M^2})$. Inside this interval, there can be at most one solution, because if there were $t_1 \neq t_2$ such that $p'(t_i) = a$, then p'' would have vanished somewhere in $(-\frac{\varepsilon\sigma^2}{M^2}, \frac{\varepsilon\sigma^2}{M^2})$, whereas

$$|p''(t) - \sigma^2| = |p''(t) - p''(0)| \leq M|t| \leq \varepsilon\sigma^2 \quad (\because |p'''| \leq M, |t| \leq \tfrac{\varepsilon\sigma^2}{M^2}),$$

so $p''(t) > 0$ on this interval. Indeed, $(1-\varepsilon)\sigma^2 \leq p''(t) \leq (1+\varepsilon)\sigma^2$ there.

Let t be the unique solution to $p'(t) = a$. We saw that $|t| \leq \frac{\varepsilon\sigma^2}{M^2}$. By Theorem 3.3, $\sigma \leq M$, so $|t| \leq \varepsilon$. By Theorem 3.2 and by the choice of ε, $\phi + t\psi$ has a unique equilibrium measure m_t. By the choice of t also $a = p'(t) = \int \psi dm_t$.

It follows that

$$P_{m_t}(\phi) = P_{m_t}(\phi + t\psi) - \int t\psi dm_t = p(t) - \int t\psi dm_t = p(t) - tp'(t) = p(t) - ta.$$

For all other $\mu \in \mathscr{M}(\Sigma^+)$ such that $\int \psi d\mu = a$, we have

$$P_\mu(\phi) = P_\mu(\phi + t\psi) - \int t\psi d\mu \leq p(t) - \int t\psi d\mu = p(t) - ta = P_{m_t}(\phi).$$

So $P_{m_t}(\phi) = \sup\{P_\mu(\phi) : \mu \in \mathscr{M}(\Sigma^+), \int \psi d\mu = a\} = q(a)$, proving the claim.

Let $\delta := \delta_\theta(\phi) := \frac{\varepsilon}{2M^2}$, and $I_\psi := (a_0 - \delta\sigma^4, a_0 + \delta\sigma^4)$. The claim shows that every $a \in I_\psi$ equals $p'(t)$ for some $|t| \leq \frac{\varepsilon\sigma^2}{2M^2}$ and $q(a) = P_{m_t}(\phi) = p(t) - ta$ is uniformly bounded. It follows that $I_\psi \subset (p'(-\infty), p'(+\infty))$, the domain of q, and $|q|$ is uniformly bounded on I_ψ. This proves part (1) of the theorem.

When we proved the claim, we mentioned in passing the following fact:

$$q(a) = p(t) - tp'(t) \text{ for the unique } t \text{ such that } p'(t) = a. \tag{5.2}$$

In other words, $q : I_\psi \to \mathbb{R}$ is minus the Legendre transform of the restriction of p to $(p')^{-1}(I_\psi)$. Indeed, we have the identity $q = (p - ap') \circ (p')^{-1}$.

By Theorem 3.2, p' is C^∞, and by Theorem 3.3 $p'' \neq 0$ (because $p''(t) = 0 \Rightarrow \sigma_{m_t}(\psi) = 0 \Rightarrow \psi$ is cohomologous to a constant $\Rightarrow \sigma_m(\phi) = 0$ in contradiction to our assumptions.

It follows that $(p')^{-1}$ is C^∞ on $\mathbb{R}$, whence q is C^∞ on I_ψ. Since we have already seen that q is uniformly bounded on I_ψ, this proves part (2).

By (5.2), $q(p'(t)) = p(t) - tp'(t)$. Repeated differentiation with respect to t gives

$$q'(p'(t)) = -t,\ q''(p'(t)) = -1/p''(t),\ q'''(p'(t)) = p'''(t)/p''(t)^3. \tag{5.3}$$

So $q''(a) = q''(p'(t)) = -\frac{1}{p''(t)} \in (-2\sigma^{-2}, -\frac{1}{2}\sigma^{-2})$ by (5.1), and $|q'''(a)| = \left|\frac{p'''(t)}{p''(t)^3}\right| \leq 8M\sigma^{-6}$. Part (3) follows with $H_\beta(\phi) := 8M$. Part (4) also follows from (5.3), because $p(0) = P_G(\phi)$, $p'(0) = a_0$ and $p''(0) = \sigma^2$. □

Corollary 5.1. *Under the assumptions of Lemma 5.2, there are $\delta_\theta(\phi), H_\theta(\phi) > 0$ such that if $0 < \delta \leq \delta_\theta(\phi)$, then the following holds for every $\psi \in \mathcal{H}_\beta$ such that $\|\psi\|_\beta \leq 1$ and $\sigma_m(\psi) \neq 0$. Let $a_0 := \int \psi dm$, $\sigma := \sigma_m(\psi)$.*

(1) If $|a - a_0| \leq \frac{\delta}{H_\theta(\phi)}\sigma^4$, then

$$e^{-\delta}\frac{1}{2\sigma^2}(a - a_0)^2 \leq \mathfrak{q}_{\phi,\psi}(a_0) - \mathfrak{q}_{\phi,\psi}(a) \leq e^{\delta}\frac{1}{2\sigma^2}(a - a_0)^2.$$

(2) If $|a - a_0| > \frac{\delta_\theta(\phi)}{H_\theta(\phi)}\sigma^4$, then $\frac{\delta_\theta(\phi)\sigma^2}{8H_\theta(\phi)}|a - a_0| \leq (\mathfrak{q}_{\phi,\psi}(a_0) - \mathfrak{q}_{\phi,\psi}(a))$.

Proof. Let $\sigma := \sigma_m(\psi), a_0 := \int \psi dm$, $q := \mathfrak{q}_{\phi,\psi}$, $p := \mathfrak{p}_{\phi,\psi}$. Let $H := H_\theta(\phi)$, $\delta_\theta(\phi)$ be as in the previous lemma. Without loss of generality $\delta_\theta(\phi) < \frac{1}{3}$ and $H > 1$ (actually, the proof of Lemma 5.2 gives $\delta_\theta(\phi) = \frac{\varepsilon}{2M^2} < 10^{-6}$ and $H = 8M > 800$). Let's write $A \pm B = C$ if $A \in [C - |B|, C + |B|]$.

Suppose $|a - a_0| \leq \frac{\delta\sigma^4}{H}$ where $0 < \delta \leq \delta_\theta(\phi)$. Then $a \in I_\psi$ and we can use the properties listed in Lemma 5.2. Taylor's expansion gives

$$q(a) = q(a_0) + q'(a_0)(a - a_0) + \frac{1}{2}q''(a_0)(a - a_0)^2 + \frac{1}{6}q'''(\eta)(a - a_0)^3$$

for some η such that $|\eta - a_0| \leq \frac{\delta\sigma^4}{H}$. By Lemma 5.2,

$$\begin{aligned}
q(a) &= q(a_0) - \frac{1}{2\sigma^2}(a - a_0)^2 \pm \frac{1}{6}\frac{H}{\sigma^6}|a - a_0|^3 \\
&= q(a_0) - \frac{1}{2\sigma^2}(a - a_0)^2\left(1 \pm \frac{1}{3}\frac{H}{\sigma^4}|a - a_0|\right) = q(a_0) - \frac{1}{2\sigma^2}(a - a_0)^2(1 \pm \tfrac{\delta}{3}) \\
&= q(a_0) - e^{\pm\delta} \cdot \frac{1}{2\sigma^2}(a - a_0)^2, \text{ because } 0 < \delta < \delta_\beta(\phi) < \tfrac{1}{3}, \text{ so } 1 \pm \tfrac{\delta}{3} \in (e^{-\delta}, e^{\delta}).
\end{aligned}$$

Rearranging terms, we obtain the first part of the corollary.

The second part of the corollary requires more care, since it deals with parts of the domain where we do not know that $q(\cdot)$ is differentiable.

Suppose $a - a_0 \geq \frac{\delta\sigma^4}{H}$ with $\delta = \delta_\theta(\phi)$, and let $a_1 := a_0 + \frac{\delta\sigma^4}{2H}$, then $a - a_1 \geq \frac{1}{2}(a - a_0)$. Since $\delta := \delta_\theta(\phi)$, $q(\cdot)$ is C^∞ on a neighborhood of $[a_0, a_1]$, and $q'' \leq -\frac{1}{2\sigma^2}$ on $[a_0, a_1]$. So by the mean value theorem for q',

$$q'(a_1) = q'(a_0) + q''(\xi)(a_1 - a_0), \text{ for some } \xi \in [a_0, a_1]$$
$$\leq -\frac{1}{2\sigma^2}(a_1 - a_0) = -\frac{\delta\sigma^2}{4H}.$$

Although we cannot assume that q is differentiable on $[a_0, a]$, we do know that it is concave there. This is sufficient to deduce that $\frac{q(a)-q(a_1)}{a-a_1} \leq (D^+q)(a_1) = q'(a_1) \leq -\frac{\delta\sigma^2}{4H}$. Rearranging terms, and recalling that $(a - a_1) \geq \frac{1}{2}(a - a_0)$, we find that

$$q(a) \leq q(a_1) - \frac{\delta\sigma^2}{4H}(a - a_1) \leq q(a_0) - \frac{\delta\sigma^2}{8H}(a - a_0),$$

where the last inequality uses the inequality $q(a_1) \leq q(a_0)$, a consequence of part (1). Rearranging terms, we obtain part (2) in the case when $a > a_0 + \frac{\delta\sigma^4}{H}$. The case $a < a_0 - \frac{\delta\sigma^4}{H}$ is obtained from the symmetry $\psi \leftrightarrow -\psi$. □

6. The EKP Inequality for measures with large pressure

Suppose Σ^+ is a topologically transitive countable Markov shift with finite Gurevich entropy. Let ϕ be a θ-weakly Hölder continuous SPR potential such that $\sup \phi < \infty$. Let m be the unique equilibrium measure of ϕ. Suppose $e^{-\beta} \leq \theta$, and recall that $(\mathcal{H}_\beta, \|\cdot\|_\beta)$ denotes the space of β-Hölder continuous functions of Σ^+, see (2.8).

Theorem 6.1. *There exist $\varepsilon^*_\theta(\phi), C^*_\theta(\phi) > 0$ such that for every $0 \not\equiv \psi \in \mathcal{H}_\beta$, $0 < \varepsilon < \varepsilon^*_\theta(\phi)$, and $\mu \in \mathscr{M}(\Sigma^+)$, if $P_\mu(\phi) \geq P_G(\phi) - C^*_\theta(\phi)\varepsilon^2 \frac{\sigma_m^6(\psi)}{\|\psi\|_\beta^6}$ then*

$$\left| \int \psi d\mu - \int \psi dm \right| \leq \sqrt{2} e^\varepsilon \sigma_m(\psi) \sqrt{P_G(\phi) - P_\mu(\phi)}.$$

The bound is sharp in the following sense: For any $\psi \in \mathcal{H}_\beta$ such that $\sigma_m(\psi) > 0$ there exists a sequence of ergodic measures $\nu_n \in \mathscr{M}(\Sigma^+)$ such that $P_{\nu_n}(\phi) \to P_G(\phi)$, and

$$\frac{|\int \psi d\nu_n - \int \psi dm|}{\sqrt{P_G(\phi) - P_{\nu_n}(\phi)}} \xrightarrow[n\to\infty]{} \sqrt{2}\sigma_m(\psi).$$

Remark. In the special case $\phi \equiv 0$, m is the measure of maximal entropy, $P_G(\phi)$ is the entropy of m, and the condition that ϕ is SPR is the same as the condition that Σ^+ is SPR.

Proof. Suppose first that $\sigma : \Sigma^+ \to \Sigma^+$ is topologically *mixing*.

If $\sigma_m(\psi) = 0$, then $P_\mu(\phi) \geq P_G(\phi) - C^*_\theta(\phi)\varepsilon^2 \frac{\sigma_m^6(\psi)}{\|\psi\|_\beta^6}$ implies that $P_\mu(\phi) = P_G(\phi)$, whence by the uniqueness of the equilibrium measure $\mu = m$ and the inequality is trivial. So it is enough to consider the case $\sigma_m(\psi) \neq 0$. It is easy to verify that $\sigma_m(t\psi) = |t|\sigma_m(\psi)$. This allows us to work with normalized functions $\psi/\|\psi\|_\beta$. Henceforth we assume that $\|\psi\|_\beta = 1$ and $\sigma := \sigma_m(\psi) \neq 0$. Let $a_0 := \int \psi dm$. Let $H_\theta(\phi)$ and $\delta_\theta(\phi)$ denote the constants from Corollary 5.1. Let $H := H_\theta(\phi)$, $C^* :=$

$\frac{1}{9H^2}$. Fix some $0 < \delta < \delta_\theta(\phi)$, and suppose $\mu \in \mathscr{M}(\Sigma^+)$ satisfies $P_\mu(\phi) \geq P_G(\phi) - C^*\delta^2\sigma^6$. Let $a := \int \psi d\mu$. By the definition of the restricted pressure,

$$\mathfrak{q}_{\phi,\psi}(a) \geq P_\mu(\phi).$$

We claim that $|a - a_0| \leq \frac{\delta_\theta(\phi)\sigma^4}{H}$. Otherwise, by the assumption on μ,

$$\begin{aligned}
C^*\delta_\theta^2(\phi)\sigma^6 &\geq P_G(\phi) - P_\mu(\phi) \geq P_G(\phi) - \mathfrak{q}_{\phi,\psi}(a), \text{ because } \mathfrak{q}_{\phi,\psi}(a) \geq P_\mu(\phi)\\
&= \mathfrak{q}_{\phi,\psi}(a_0) - \mathfrak{q}_{\phi,\psi}(a), \text{ because } \mathfrak{q}_{\phi,\psi}(a_0) = P_G(\phi) \text{ by Lemma 5.2}\\
&\geq \frac{\delta_\theta(\phi)\sigma^2}{8H}|a - a_0| \text{ by the 2nd part of Corollary 5.1}\\
&\geq \frac{\delta_\theta(\phi)\sigma^2}{8H} \cdot \frac{\delta_\theta(\phi)\sigma^4}{H} = \frac{\delta_\theta^2(\phi)\sigma^6}{8H^2}, \text{ by the assumption } |a - a_0| \geq \frac{\delta_\theta(\phi)\sigma^4}{H}.
\end{aligned}$$

But this contradicts the definition of C^*.

So $|a - a_0| \leq \frac{\delta_\theta(\phi)\sigma^4}{H}$, and the first part of Corollary 5.1 gives us

$$(a - a_0)^2 \leq 2\sigma^2 e^\delta(\mathfrak{q}_{\phi,\psi}(a_0) - \mathfrak{q}_{\phi,\psi}(a)).$$

Taking the square root, and recalling that $a = \int \psi d\mu$, $a_0 = \int \psi dm$, $\mathfrak{q}_{\phi,\psi}(a_0) = P_G(\phi)$, and $\mathfrak{q}_{\phi,\psi}(a) \geq P_\mu(\phi)$, we obtain

$$\left|\int \psi d\mu - \int \psi dm\right| \leq e^\delta\sqrt{2}\sigma\sqrt{P_G(\phi) - \mathfrak{q}_{\phi,\psi}(a)} \leq e^\delta\sqrt{2}\sigma\sqrt{P_G(\phi) - P_\mu(\phi)}.$$

This proves the first part of the theorem with $\varepsilon_\theta^*(\phi) := \delta_\theta(\phi)$.

To see the second part, take $a_n \to a_0$. When we proved Lemma 5.2, we saw that if a_n is sufficiently close to a_0, then $\exists! t_n$ such that $\mathfrak{p}'_{\phi,\psi}(t_n) = a_n$, the equilibrium measure $\nu_n := m_{t_n}$ of $\phi + t_n\psi$ exists, and $\mathfrak{q}_{\phi,\psi}(a_n) = P_{\nu_n}(\phi)$.

Repeating the previous argument with ν_n replacing μ, but now with the full force of Corollary 5.1(1), we find that $\sqrt{2}\sigma e^{-\delta_n} \leq \frac{|\int \psi d\nu_n - \int \psi dm|}{\sqrt{P_G(\phi) - P_{\nu_n}(\phi)}} \leq \sqrt{2}\sigma e^{\delta_n}$, where $\delta_n := \frac{H}{\sigma^4}|a_n - a_0| \to 0$.

This completes the proof in the topologically mixing case. We will now briefly outline the proof in the general topologically transitive case. Suppose $\sigma : \Sigma^+ \to \Sigma^+$ is topologically transitive, with period p, and let

$$\Sigma^+ = \Sigma_0^+ \uplus \cdots \uplus \Sigma_{p-1}^+$$

be the spectral decomposition from §2.2. The assumption that ϕ is SPR on Σ^+ means, by definition, that $\phi_p := \sum_{i=0}^{p-1} \phi \circ \sigma^i$ is SPR with respect to the topologically mixing $\sigma^p : \Sigma_i^+ \to \Sigma_i^+$.

$\Sigma_i^+ = \sigma^{i-j}(\Sigma_j^+)$, therefore for every σ-invariant measure μ, $\mu(\Sigma_i^+)$ are all equal (to $1/p$), and $\mu = \frac{1}{p}\sum_{i=0}^{p-1} \mu_i$, where $\mu_i := \mu(\cdot|\Sigma_i^+)$ (the conditional measure on Σ_i^+). In addition:

(1) $h_{\mu_i}(\sigma^p) = p h_\mu(\sigma)$: First, $\sigma^{j+p-i} : (\Sigma_i^+, \mu_i) \to (\Sigma_j^+, \mu_j)$ is a factor map for all i, j so $h_{\mu_i}(\sigma^p)$ are all equal. Second, $\frac{1}{p}\sum_{i=0}^{p-1} h_{\mu_i}(\sigma^p) = h_\mu(\sigma^p) = p h_\mu(\sigma)$.
(2) $\int \phi_p d\mu_i = p \int \phi d\mu$.
(3) By (1) and (2), $P_{\mu_i}(\phi_p) = p P_\mu(\phi)$.
(4) If m is an equilibrium measure of ϕ, then m_i is an equilibrium measure of ϕ_p.
(5) By definition, $P_G(\phi_p|_{\Sigma_i^+}, \sigma^p) = p P_G(\phi|_\Sigma, \sigma)$.

(6) $\sigma_{m_0}(\psi_p) = \sqrt{p}\sigma_m(\psi)$: By (2), $\mathbb{E}_{m_i}(\psi_{np}) = np\mathbb{E}_m(\psi)$, therefore

$$\sigma_m(\psi)^2 = \lim_{n\to\infty} \frac{1}{np} \operatorname{Var}_m(\psi_{np}) = \lim_{n\to\infty} \frac{1}{np^2} \sum_{i=0}^{p-1} \operatorname{Var}_{m_i}(\psi_{np})$$
$$= \lim_{n\to\infty} \frac{1}{np^2} \sum_{i=0}^{p-1} \operatorname{Var}_{m_0}\Big(\sum_{j=0}^{n-1} (\psi_p) \circ \sigma^{jp} \circ \sigma^i\Big) = \frac{1}{p^2} \sum_{i=0}^{p-1} \sigma_{m_0}(\psi_p \circ \sigma^i)^2 = \frac{1}{p}\sigma_{m_0}(\psi_p)^2,$$

where the last equality is because $\psi_p \circ \sigma^i$ is σ^p-cohomologous to ψ_p.

It is now an easy exercise to deduce the theorem for $\sigma : \Sigma^+ \to \Sigma^+, \psi, m, \mu$ from the theorem for the topologically mixing $\sigma^p : \Sigma_0^+ \to \Sigma_0^+$, with ψ_p, m_0, μ_0. □

The previous result gives the "optimal" form of the EKP inequality for measures with high entropy. Note that the bound $const.\|\psi\|_\beta \sqrt{h - h_\mu(\sigma)}$ in the original EKP inequality is replaced by

$$e^\varepsilon \sqrt{2}\sigma_{\mu_0}(\psi)\sqrt{h - h_\mu(\sigma)},$$

where $\sigma^2_{\mu_0}(\psi)$ is the asymptotic variance of ψ with respect to the measure of maximal entropy μ_0. This is better than (1.2), because $\sigma_{\mu_0}(\psi) \le M\|\psi\|_\beta$ (Theorem 3.3), and because of the following fact.

Lemma 6.1. *Suppose Σ^+ is topologically transitive, with positive Gurevich entropy. For every $\beta > 0$, there is a sequence of $\psi_n \in \mathcal{H}_\beta$ such that $\sigma_{\mu_0}(\psi_n) \ne 0$, $\int \psi_n d\mu = 0$, and* $\dfrac{\sigma_{\mu_0}(\psi_n)}{\|\psi_n\|_\beta} \xrightarrow[n\to\infty]{} 0.$

Proof. There are two periodic points $\underline{x}, \underline{y}$ with the same (perhaps non-minimal) period p, and with disjoint orbits: $\sigma^m(\underline{x}) \ne \sigma^n(\underline{y})$ for all m, n. Otherwise there is only one periodic orbit, and the Gurevich entropy equals zero.

Construct $\psi \in \mathcal{H}_\beta$ such that $\sum_{j=0}^{p-1} \psi(\sigma^j(\underline{x})) \ne \sum_{j=0}^{p-1} \psi(\sigma^j(\underline{y}))$ and $\int \psi d\mu_0 = 0$. Then ψ cannot be cohomologous to a constant (otherwise it would give all periodic points of fixed period the same weight). By Theorem 3.3, $\sigma_{\mu_0}(\psi) \ne 0$. In addition, since ψ is clearly non-constant, $\psi \ne \psi \circ \sigma$, whence $\|\psi - \psi \circ \sigma\|_\beta \ne 0$.

Now take $\psi_n := \frac{1}{n}\psi + (\psi - \psi \circ \sigma)$. On the one hand $\sigma_{\mu_0}(\psi_n) = \sigma_{\mu_0}(\psi/n) = \sigma_{\mu_0}(\psi)/n$, a sequence of positive numbers which converges to zero. On the other hand $\|\psi_n\|_\beta \ge \|\psi - \psi \circ \sigma\|_\beta - \|\psi/n\|_\beta \to \|\psi - \psi \circ \sigma\|_\beta \ne 0$. □

We can now explain why we needed the constants $\varepsilon_\theta(\phi), M_\theta(\phi)$ in Theorem 3.2 to be independent of ψ. We do this in the case of main interest $\phi \equiv 0$, when μ_0 is the measure of maximal entropy. In this case Theorem 6.1 gives a nearly optimal EKP inequality in the regime

$$h_\mu(\sigma) \ge h - \varepsilon^2 C^*_\theta(0)(\sigma_{\mu_0}(\psi)/\|\psi\|_\beta)^6.$$

In the absence of the uniformity in ψ in Theorem 3.2 the best we could have hoped for was to prove this bound in the regime $h_\mu(\sigma) \ge h - \varepsilon^2 C^*_\theta(\psi)$, but without further information on the structure of $C^*_\theta(\psi)$.

7. The EKP inequality for arbitrary measures

Our next result (which reduces in the case of subshifts of finite type and $\phi \equiv 0$ to a result of S. Kadyrov [Kad15]) is an inequality for all σ-invariant measures, also those with low entropy or pressure.

Suppose Σ^+ is a topologically mixing countable Markov shift with finite and positive Gurevich entropy. Let ϕ be a θ-weakly Hölder continuous SPR potential such that $\sup \phi < \infty$. Let m denote the unique equilibrium measure of ϕ. Fix β such that $e^{-\beta} \leq \theta$, and let $(\mathcal{H}_\beta, \|\cdot\|_\beta)$ denote the space of β-Hölder continuous functions on Σ^+, see (2.8).

Lemma 7.1. *There exist constants $K_\beta, Q(\phi) > 0$ such that for every σ-invariant probability μ there exists some function $A \in \mathcal{H}_\beta$ such that $\int A dm = \int A d\mu = 0$, $\|A\|_\beta \leq K_\beta$, and $\sigma_m(A) > Q(\phi)$.*

Proof. As in Lemma 6.1, we can find $p \geq 1$ and four periodic points $\underline{x}, \underline{y}, \underline{z}, \underline{w} \in \Sigma^+$ with period p, such that

$$x_0 = y_0 = z_0 = w_0$$

and so that the orbits of $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{w}$ are disjoint. Let $\underline{x}^{\mathrm{p}} := (x_0, \ldots, x_{p-1}, x_0)$, $\underline{y}^{\mathrm{p}} := (y_0, \ldots, y_{p-1}, y_0)$, $\underline{z}^{\mathrm{p}} := (z_0, \ldots, z_{p-1}, z_0)$, $\underline{w}^{\mathrm{p}} := (w_0, \ldots, w_{p-1}, w_0)$. Define

$$a(\cdot) := 1_{[\underline{x}^{\mathrm{p}}]}(\cdot) - m([\underline{x}^{\mathrm{p}}]),\ b(\cdot) := 1_{[\underline{z}^{\mathrm{p}}]}(\cdot) - m([\underline{z}^{\mathrm{p}}]).$$

Recall the notation $\varphi_n = \sum_{k=0}^{n-1} \varphi \circ \sigma^k$. Since $\{\sigma^k \underline{x}\}$, $\{\sigma^k \underline{y}\}$, $\{\sigma^k \underline{z}\}$, $\{\sigma^k \underline{w}\}$ are disjoint, the orbit of $\underline{y}$ does not enter $[\underline{x}^{\mathrm{p}}]$, and the orbit of $\underline{w}$ does not enter $[\underline{z}^{\mathrm{p}}]$. In particular, $a_p(\underline{x}) \neq a_p(\underline{y})$ and $b_p(\underline{z}) \neq b_p(\underline{w})$.

This implies that a, b are not cohomologous to constants (otherwise $a_p(\underline{x}) = a_p(\underline{y}) = \mathrm{const}\, p$ for any pair of p-periodic orbits). So $\sigma_m^2(a), \sigma_m^2(b) \neq 0$.

By construction, $\int a dm = \int b dm = 0$. If $\int a d\mu = 0$ take $A := a$. If $\int b d\mu = 0$, take $A := b$. Notice that A is independent of μ, therefore $\|A\|_\beta$, $\sigma_m^2(A)$ are independent of μ, and the lemma follows.

In the remaining case, $\int a d\mu \neq 0$ and $\int b d\mu \neq 0$, and we define

$$A := \frac{a \int b d\mu - b \int a d\mu}{\sqrt{(\int a d\mu)^2 + (\int b d\mu)^2}}.$$

Clearly $\int A d\mu = \int A dm = 0$, and

$$\|A\|_\beta \leq \frac{|\int a d\mu| + |\int b d\mu|}{\sqrt{(\int a d\mu)^2 + (\int b d\mu)^2}} \max\{\|a\|_\beta, \|b\|_\beta\} \leq 2 \max\{\|a\|_\beta, \|b\|_\beta\} =: K_\beta.$$

K_β is independent of μ, and only depends on β and Σ^+. To complete the proof, it remains to bound $\sigma_m^2(A)$ from below by a constant which is independent of μ.

We need for this purpose the function $Q(s,t) := \sigma_m^2(sa + tb)$. By Theorem 3.2,

$$\sigma_m^2(sa+tb) = \frac{d^2}{d\tau^2}\bigg|_{\tau=0} \mathfrak{p}_{\phi, sa+tb}(\tau) \equiv \frac{d^2}{d\tau^2}\bigg|_{\tau=0} P_G(\phi + \tau sa + \tau tb)$$

$$\stackrel{!}{=} s^2 \frac{\partial^2 P}{\partial u^2}(0,0) + 2st \frac{\partial^2 P}{\partial u \partial v}(0,0) + t^2 \frac{\partial^2 P}{\partial v^2}(0,0), \text{ where } P(u,v) := P_G(\phi + ua + vb).$$

To justify $\stackrel{!}{=}$, we use Theorem 3.2 (6). We see that $Q(s,t)$ is a quadratic form.

By the definition of the asymptotic variance, $Q(s,t) \geq 0$ for all (s,t). We claim that Q is positive definite.

Suppose $Q(s,t) = 0$, then $\sigma_m^2(sa+tb) = 0$, whence $sa+tb$ is cohomologous to a constant. In this case, by Livshits theorem, $sa+tb$ gives the same weight to all periodic orbits with the same period. In particular

$$sa_p(\underline{x}) + tb_p(\underline{x}) = sa_p(\underline{y}) + tb_p(\underline{y})$$
$$sa_p(\underline{z}) + tb_p(\underline{z}) = sa_p(\underline{w}) + tb_p(\underline{w}).$$

Recall that the orbits of $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{w}$ are disjoint, so the orbits of $\underline{y}$, $\underline{z}$, $\underline{w}$ do not enter $[\underline{x}^{\mathrm{p}}]$ whence $a_p(\underline{y}) = a_p(\underline{z}) = a_p(\underline{w}) = -pm([\underline{x}^{\mathrm{p}}])$, and the orbits of $\underline{x}$, $\underline{y}$, $\underline{w}$ do not enter $[\underline{z}^{\mathrm{p}}]$, so $b_p(\underline{x}) = b_p(\underline{y}) = b_p(\underline{w}) = -pm([\underline{z}^{\mathrm{p}}])$. On the other hand $a_p(\underline{x}) = n_{\underline{x}} - pm([\underline{x}^{\mathrm{p}}]), b_p(\underline{z}) = n_{\underline{z}} - pm([\underline{z}^{\mathrm{p}}])$ with $n_{\underline{x}}, n_{\underline{z}}$ positive integers. Substituting this above, we obtain $n_{\underline{x}}s = 0, n_{\underline{z}}t = 0$, whence $s = t = 0$. So $Q(s,t) = 0 \Rightarrow (s,t) = (0,0)$, and Q is positive definite.

Since $Q(s,t)$ is a positive definite quadratic form, there exists $Q_0 > 0$ such that $Q(s,t) \geq Q_0^2(s^2+t^2)$ for all $(s,t) \in \mathbb{R}^2$. In particular,

$$\sigma_m^2(A) = Q\left(\frac{\int b d\mu}{\sqrt{(\int a d\mu)^2 + (\int b d\mu)^2}}, \frac{-\int a d\mu}{\sqrt{(\int a d\mu)^2 + (\int b d\mu)^2}}\right) \geq Q_0^2.$$

Notice that Q_0 depends only on a, b and ϕ, and is therefore independent of μ. We let $Q(\phi) := Q_0$. □

Lemma 7.2. *Given $\beta > 0$ $\exists K'_\beta(\phi) > 0$ as follows: For every $\psi \in \mathcal{H}_\beta$ and $\mu \in \mathscr{M}(\Sigma^+)$, $\exists \varphi \in \mathcal{H}_\beta$ such that $\int \varphi dm = \int \psi dm$, $\int \varphi d\mu = \int \psi d\mu$, $\|\varphi\|_\beta \leq K'_\beta(\phi)\|\psi\|_\beta$ and*

$$\frac{\|\varphi\|_\beta}{\sigma_m(\varphi)} \leq K'_\beta(\phi). \tag{7.1}$$

Proof. Let $Q := Q(\phi)$, K_β and $A(\cdot)$ be as in the previous lemma.

If $\sigma_m(\psi) \geq \frac{1}{3}Q\|\psi\|_\beta$, we take $\varphi := \psi$, and note that

$$\frac{\|\varphi\|_\beta}{\sigma_m(\varphi)} \leq \frac{3}{Q}.$$

If $\sigma_m(\psi) < \frac{1}{3}Q\|\psi\|_\beta$, we take $\varphi := \psi + \|\psi\|_\beta A$. Then $\int \varphi dm = \int \psi dm$, $\int \varphi d\mu = \int \psi d\mu$, and $\|\varphi\|_\beta \leq (1 + \|A\|_\beta)\|\psi\|_\beta \leq (1+K_\beta)\|\psi\|_\beta$. In addition,

$$\begin{aligned}
\sigma_m^2(\varphi) &= \lim_{n\to\infty} \frac{1}{n}\operatorname{Var}(\varphi_n), \text{ where } \varphi_n := \sum_{k=0}^{n-1} \varphi \circ \sigma^k, \ \operatorname{Var}(B) := \int (B - \textstyle\int B dm)^2 dm \\
&= \lim_{n\to\infty} \frac{1}{n}\operatorname{Var}(\psi_n) + \lim_{n\to\infty} \frac{1}{n}\operatorname{Var}(\|\psi\|_\beta A_n) + 2\lim_{n\to\infty} \frac{1}{n}\operatorname{Cov}(\psi_n, \|\psi\|_\beta A_n) \\
&\qquad \text{where } \operatorname{Cov}(B,C) := \int (B - \textstyle\int B dm)(C - \textstyle\int C dm) dm \\
&\geq 0 + \|\psi\|_\beta^2 \sigma_m^2(A) - 2\|\psi\|_\beta \lim_{n\to\infty} \frac{1}{n}\sqrt{\operatorname{Var}(\psi_n)\operatorname{Var}(A_n)} \quad \text{(Cauchy-Schwarz)} \\
&= \|\psi\|_\beta^2 \sigma_m^2(A) - 2\|\psi\|_\beta \sigma_m(\psi)\sigma_m(A) \\
&\geq \frac{1}{3}Q^2\|\psi\|_\beta^2 \quad (\because \sigma_m(A) \geq Q, \ \sigma_m(\psi) < \tfrac{1}{3}Q\|\psi\|_\beta),
\end{aligned}$$

so $\sigma_m(\varphi) \geq \frac{1}{\sqrt{3}}Q\|\psi\|_\beta$.

We saw above that $\|\varphi\|_\beta \leq (1+K_\beta)\|\psi\|_\beta$. It follows that $\frac{\|\varphi\|_\beta}{\sigma_m(\varphi)} \leq \sqrt{3}\frac{1+K_\beta}{Q}$. The lemma follows with $K'_\beta(\phi) := \max\{\frac{3}{Q(\phi)}, \sqrt{3}\big(\frac{1+K_\beta}{Q(\phi)}\big), K_\beta+1\}$. □

Theorem 7.1. *Suppose Σ^+ is a topologically transitive countable Markov shift with finite Gurevich entropy. Let ϕ be a θ-weakly Hölder continuous SPR potential such that $\sup\phi < \infty$, and let m be the unique equilibrium measure of ϕ. If $e^{-\beta} \leq \theta$, then $\exists C_{\theta,\beta}(\phi) > 0$ such that for every $\psi \in \mathcal{H}_\beta$, and every $\mu \in \mathscr{M}(\Sigma^+)$,*

$$\left|\int \psi d\mu - \int \psi dm\right| \leq C_{\theta,\beta}(\phi)\|\psi\|_\beta\sqrt{P_G(\phi) - P_\mu(\phi)}.$$

Remark. In the special case $\phi \equiv 0$, m is the measure of maximal entropy, $P_G(\phi)$ is the entropy of m, and the inequality becomes (1.2).

Proof. As in the proof of Theorem 6.1, it is sufficient to prove the theorem in the topologically mixing case. Fix a σ-invariant measure μ and some $\psi \in \mathcal{H}_\beta$. If $\int \psi d\mu = \int \psi dm = 0$ then there is nothing to prove so suppose one of these integrals is non-zero.

Let $K := K'_\beta(\phi)$ be a constant independent of μ, ψ as in the previous lemma, and choose $\varphi \in \mathcal{H}_\beta$ such that $\|\varphi\|_\beta \leq K\|\psi\|_\beta$, $\int \psi d\mu = \int \varphi d\mu$, $\int \psi dm = \int \varphi dm$, and

$$\frac{\|\varphi\|_\beta}{\sigma_m(\varphi)} \leq K.$$

Notice that $\|\varphi\|_\beta \neq 0$, because at least one of the integrals $\int \varphi d\mu, \int \varphi dm$ is non-zero. So we can normalize $\overline{\varphi} := \dfrac{\varphi}{\|\varphi\|_\beta}$.

Let $\delta := \delta_\theta(\phi)$, $H := H_\theta(\phi)$ be the constants in Corollary 5.1, and let $M := M_\theta(\phi)$ be the constant from Theorem 3.3. Without loss of generality, $e^\delta < \sqrt{2}$. Set $a := \int \overline{\varphi} d\mu$, $a_0 = \int \overline{\varphi} dm$, $\sigma := \sigma_m(\overline{\varphi})$. If $|a - a_0| \leq \frac{\delta\sigma^4}{H}$, then

$$\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right| \leq e^\delta\sqrt{2}\sigma\sqrt{\mathfrak{q}_{\phi,\overline{\varphi}}(a_0) - \mathfrak{q}_{\phi,\overline{\varphi}}(a)} \overset{!}{\leq} 2M\sqrt{P_G(\phi) - P_\mu(\phi)},$$

where $\overset{!}{\leq}$ is because $\mathfrak{q}_{\phi,\overline{\varphi}}(a_0) = P_G(\phi)$, $\mathfrak{q}_{\phi,\overline{\varphi}}(a) \geq P_\mu(\phi)$, and $\sigma_m(\overline{\varphi}) \leq M\|\overline{\varphi}\|_\beta = M$.

Similarly, if $|a - a_0| > \frac{\delta\sigma^4}{H}$, then by the 2nd part of Corollary 5.1

$$\frac{1}{2}\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right| \leq \frac{1}{2}\frac{8H}{\delta\sigma^2}(P_G(\phi) - P_\mu(\phi)) = \frac{4H}{\delta\sigma_m^2(\overline{\varphi})}(P_G(\phi) - P_\mu(\phi))$$
$$= \frac{4H\|\varphi\|_\beta^2}{\delta\sigma_m^2(\varphi)}(P_G(\phi) - P_\mu(\phi)) \leq \frac{4HK^2}{\delta}(P_G(\phi) - P_\mu(\phi)).$$

Since $\|\overline{\varphi}\|_\beta = 1$, $\frac{1}{2}|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm| \leq 1$, and therefore

$$\frac{1}{2}\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right| \leq \sqrt{\frac{1}{2}\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right|} \leq \sqrt{\frac{4HK^2}{\delta}(P_G(\phi) - P_\mu(\phi))},$$

whence $\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right| \leq (4K\sqrt{H/\delta})\sqrt{P_G(\phi) - P_\mu(\phi)}$.

Let $C'_{\theta,\beta} := C'_{\theta,\beta}(\phi) := \max\{2M, 4K\sqrt{H/\delta}\}$. This depends only on β, θ and ϕ, and the following inequality holds no matter the value of $|a - a_0|$:

$$\left|\int \overline{\varphi} d\mu - \int \overline{\varphi} dm\right| \leq C'_{\theta,\beta}\sqrt{P_G(\phi) - P_\mu(\phi)}. \tag{7.2}$$

By the choice of $\overline{\varphi}$, the left hand side of (7.2) equals $\left|\int \psi d\mu - \int \psi dm\right| / \|\varphi\|_\beta$, so $\left|\int \psi d\mu - \int \psi dm\right| \leq C'_{\theta,\beta}\|\varphi\|_\beta \sqrt{P_G(\phi) - P_\mu(\phi)}$. Since $\|\varphi\|_\beta \leq K\|\psi\|_\beta$, the theorem follows with $C_{\theta,\beta}(\phi) := C'_\beta K$. □

8. SPR is a necessary condition for the EKP inequality

Let $\sigma : \Sigma^+ \to \Sigma^+$ be a topologically transitive TMS associated to a countable directed graph $\mathscr{G}$, and suppose $\phi : \Sigma^+ \to \mathbb{R}$ is a function with summable variations and finite Gurevich pressure. In this section we prove that if μ_ϕ satisfies the EKP inequality (1.4) (or even just the weaker property (1.1)), then ϕ must be strongly positively recurrent. Ruette proved this for $\phi \equiv 0$ [Rue03], and we will extend her proof to non-constant potentials.

Theorem 8.1. *Suppose Σ^+ is topologically mixing. If ϕ is not SPR, then $\mathscr{G}$ contains a subgraph $\mathscr{G}'$ such that $\Sigma^+(\mathscr{G}') \subsetneq \Sigma^+(\mathscr{G})$, $\sigma : \Sigma^+(\mathscr{G}') \to \Sigma^+(\mathscr{G}')$ is topologically mixing, and $P_G(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi|_{\Sigma^+})$.*

Proof. The case $\phi \equiv 0$ is done in [Rue03].

There is no loss of generality in assuming that $\mathscr{G}$ has at most one edge $a \to b$ for every ordered pair $(a,b) \in S \times S$, and that every vertex a has an outgoing edge $a \to b$ in E. Otherwise, pass to the graph $\mathscr{G}^*$ with set of vertices $S^* := \{a \in S : [a] \neq \varnothing\}$ and set of edges $\{(a,b) \in E : [a,b] \neq \varnothing\}$, then $\Sigma^+(\mathscr{G}^*) = \Sigma^+(\mathscr{G})$ and $\mathscr{G}^*$ has the required properties.

It follows that if $\mathscr{G}'$ is a proper subgraph of $\mathscr{G}$, then $\Sigma^+(\mathscr{G}') \subsetneq \Sigma^+(\mathscr{G})$.

If ϕ is not SPR, then $\mathscr{G}$ must have infinitely many vertices. Otherwise Σ is a subshift of finite type, and $\phi + t1_{[a]}$ has an equilibrium measure for all $t \in \mathbb{R}$ by [Bow75]. But in the non-SPR case, $\Delta_a[\phi] \leq 0$, and then $\Delta_a[\phi + t1_{[a]}] = \Delta_a[\phi] + t < 0$ for t negative. By the discriminant theorem of [Sar01a], $\phi + t1_{[a]}$ is not positively recurrent for $t < 0$. This is a contradiction, because potentials with equilibrium measures are always positively recurrent [BS03].

By our standing assumptions, $\Sigma^+(\mathscr{G})$ is topologically mixing. Therefore there exist two closed loops starting and ending at the same vertex with co-prime lengths. The union of these loops defines a finite subgraph $\mathscr{G}_{core}$ of $\mathscr{G}$, and every strongly connected graph which contains $\mathscr{G}_{core}$ defines a topologically mixing TMS.

We will obtain the graph $\mathscr{G}'$ from $\mathscr{G}$, by removing some edge $a \to b$ outside $\mathscr{G}_{\text{core}}$. We use the following edge removal procedure from [Rue03].

Let S_{core} denote the set of vertices of $\mathscr{G}_{\text{core}}$. Let $\mathcal{E}$ denote the collection of finite paths γ with the following properties:

(1) the first vertex and the last vertex of γ belong to S_{core};
(2) all other vertices are outside S_{core};
(3) for every vertex $x_i \in S \setminus S_{\text{core}}$ of $\gamma = (x_0, \ldots, x_{n-1})$ either $(x_0, \ldots, x_i)$ or $(x_i, \ldots, x_{n-1})$ is a $\mathscr{G}$–geodesic.

Since σ is topologically transitive, $\mathscr{G}$ is strongly connected. Every vertex in $S \setminus S_{\text{core}}$ can be connected by forward and backward $\mathscr{G}$–geodesics to S_{core}. Their union is a path in $\mathcal{E}$. So every vertex in S^c_{core} belongs to some path in $\mathcal{E}$, and $\mathcal{E}$ is infinite.

All paths in $\mathcal{E}$ begin at S_{core}. Since S_{core} is finite and $\mathcal{E}$ is infinite, there exist two different paths $\gamma_0, \gamma_1 \in \mathcal{E}$ which begin at the same vertex in S_{core} and end at the same vertex in S_{core}. Let γ denote the maximal common prefix of γ_0, γ_1, and let a denote the last vertex in γ.

By construction, a has at least two different outgoing edges $a \to b_i$ $(i = 0, 1)$ such that $b_i \notin S_{\rm core}$. Divide the outgoing edges from a to S^c_{core} into two arbitrary subsets, E_0 and E_1, such that $(a, b_i) \in E_i$.

Let $\mathscr{G}_i$ be the graph obtained by removing the edges E_{1-i} from $\mathscr{G}$ and restricting to the strongly connected component containing a. It is not difficult to see that $\mathscr{G}_i$ contain $\mathscr{G}_{\rm core}$, so $\sigma : \Sigma^+(\mathscr{G}_i) \to \Sigma^+(\mathscr{G}_i)$ is topologically mixing for $i = 1, 2$. By the assumption on $\mathscr{G}$, $\Sigma^+(\mathscr{G}_i) \subsetneq \Sigma^+(\mathscr{G})$.

If $x_0 = a, x_1, \ldots, x_{n-1}, x_n = a$ is a first return loop to a, then $x_i \neq a$ for $1 \leq i \leq n-1$, therefore edges from E_i can only appear as the first edge (x_0, x_1).

If $a \in S_{core}$, we let $E_{\rm core}$ be the edges $(a, b) \in \mathscr{G}$ such that $a, b \in S_{core}$ (in particular, $E_{\rm core} = \varnothing$ whenever $a \notin S_{core}$). $Z^*_n(\phi, a)$ splits into the sums,

$$Z^*_n(\phi, a) = Z^*_n(E_0) + Z^*_n(E_1) + Z^*_n(E_{\rm core}) \tag{8.1}$$

where

$$Z^*_n(E_i) := \sum_{\substack{\sigma^n(\underline{x})=\underline{x} \\ (x_0,x_1)\in E_i}} e^{\phi_n(\underline{x})} 1_{[\tau_a=n]}(\underline{x})$$

and

$$Z^*_n(E_{\rm core}) := \sum_{\substack{\sigma^n(\underline{x})=\underline{x} \\ (x_0,x_1)\in E_{\rm core}}} e^{\phi_n(\underline{x})} 1_{[\tau_a=n]}(\underline{x}).$$

We claim that

$$Z^*_n(\phi|_{\Sigma^+(\mathscr{G}_i)}, a) = Z^*_n(E_i) + Z^*_n(E_{\rm core}). \tag{8.2}$$

The non-trivial inequality is $\geq$. Clearly, the first return loops at a which begin with an edge in E_i or $E_{\rm core}$ do not contain other edges in E_{1-i}, because a cannot appear in the middle of a first return loop at a. Therefore all such loops must be contained in the irreducible component of a in $\mathscr{G} \setminus E_{1-i}$, whence in $\mathscr{G}_i$. So the loops participating in the sums in the right-hand-side must also appear in the sum on the left-hand-side. As the sums on the right are over disjoint sets, $(\geq)$ follows.

Recall that $P^*_G(\phi, a) = \limsup \frac{1}{n} \log Z^*_n(\phi, a)$. By (8.1) and (8.2),

$$\begin{aligned} P^*_G(\phi, a) &\leq \limsup \frac{1}{n} \log\big(2 \max_i (Z^*_n(\phi|_{\Sigma^+(\mathscr{G}_i)}, a)\big) \\ &= \limsup \frac{1}{n} \log\big(\max_i (Z^*_n(\phi|_{\Sigma^+(\mathscr{G}_i)}, a)\big) \leq \max_i P^*_G(\phi|_{\Sigma^+(\mathscr{G}_i)}, a) \leq P^*_G(\phi, a). \end{aligned}$$

So $P^*_G(\phi, a) = P^*_G(\phi|_{\mathscr{G}_i}, a)$ for at least one of the two indices $i = 0, 1$. Call this index i_0. Then $P_G(\phi) \overset{(1)}{=} P^*_G(\phi, a) = P^*_G(\phi|_{\Sigma^+(\mathscr{G}_{i_0})}, a) \overset{(2)}{\leq} P_G(\phi|_{\Sigma^+(\mathscr{G}_{i_0})}) \overset{(3)}{\leq} P_G(\phi)$: $\overset{(1)}{=}$ follows from Lemma 2.1 and the assumption that ϕ is not SPR; and $\overset{(2)}{\leq}$ and $\overset{(3)}{\leq}$ are due to the trivial inequalities $Z^*_n(\phi|_{\Sigma^+(\mathscr{G}_{i_0})}, a) \leq Z_n(\phi|_{\Sigma^+(\mathscr{G}_{i_0})}, a) \leq Z_n(\phi, a)$. So $P_G(\phi) = P_G(\phi|_{\Sigma^+(\mathscr{G}_{i_0})})$ as required. □

Corollary 8.1. *Suppose $\sigma : \Sigma^+ \to \Sigma^+$ is a topologically transitive TMS, and $\phi : \Sigma^+ \to \mathbb{R}$ is a potential with summable variations, finite Gurevich pressure, and an equilibrium measure μ_ϕ. If the EKP inequality* (1.4) *holds, then ϕ must be SPR.*

Proof. It is enough to consider the topologically mixing case, because if Σ^+ is topologically transitive with period p, and $\Sigma^+ = \biguplus_{i=0}^{p-1} \Sigma^+_i$ is the spectral decomposition from §2.2, then σ satisfies the EKP inequality if and only if the topologically *mixing* $\sigma^p : \Sigma^+_0 \to \Sigma^+_0$ satisfies the EKP inequality. See the proof of Theorem 6.1.

Let $\mathscr{G}$ denote the graph associated to Σ^+. Assume without loss of generality that $\mathscr{G}$ has at most one edge $a \to b$ for every ordered pair of vertices (a, b).

By assumption, ϕ has an equilibrium measure, whence by [BS03], ϕ is positively recurrent. Suppose by way of contradiction that ϕ is not SPR. By the previous theorem, there is a proper subgraph $\mathscr{G}' \subset \mathscr{G}$ such that $\sigma : \Sigma^+(\mathscr{G}') \to \Sigma^+(\mathscr{G}')$ is topologically mixing, and

$$P_G(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi).$$

By the variational principle, there are invariant probability measures μ_n on $\Sigma^+(\mathscr{G}')$ such that $P_{\mu_n}(\phi|_{\Sigma^+(\mathscr{G}')}) \xrightarrow[n\to\infty]{} P_G(\phi|_{\Sigma^+(\mathscr{G}')})$. In particular $\sqrt{P_G(\phi) - P_{\mu_n}(\phi)} \xrightarrow[n\to\infty]{} 0$, so by the EKP inequality or by (1.1) *on the original Markov shift* Σ^+,

$$|\mu_n([\underline{a}]) - \mu_\phi[\underline{a}]| \xrightarrow[n\to\infty]{} 0$$

for all non-empty cylinders $[\underline{a}]$ in Σ^+.

But this is false for any cylinder of the form $[a, b]$ where $a \to b$ is an edge which appears in $\mathscr{G}$ but not in $\mathscr{G}'$. For this cylinder $\mu_n[a, b] = 0$ for all n, because μ_n are supported in $\Sigma^+(\mathscr{G}')$. But $\mu_\phi[a, b] \neq 0$ because equilibrium measures of potentials with summable variations on topologically mixing TMS always have full support, see [BS03]. □

The following generalizes a result of Salama [Sal88, Theorem 2.3], whose proof has been corrected in [Rue03, Theorem 2.7] and in [Fie96]. The case of Markovian potentials (ϕ such that $\mathrm{osc}_2(\phi) = 0$) was done in [GS98, Theorem 3.15].

Corollary 8.2. *Let $\sigma : \Sigma^+ \to \Sigma^+$ be a topologically mixing TMS with finite Gurevich entropy, and associated to a countable directed graph $\mathscr{G}$. Suppose $\phi : \Sigma^+ \to \mathbb{R}$ is bounded from above, has summable variations, and has finite Gurevich pressure. Then ϕ is SPR if and only if $P_G(\phi|_{\Sigma^+(\mathscr{G}')}) < P_G(\phi|_{\Sigma^+(\mathscr{G})})$ for every proper subgraph $\mathscr{G}'$ of $\mathscr{G}$ for which $\sigma : \Sigma^+(\mathscr{G}') \to \Sigma^+(\mathscr{G}')$ is topologically mixing.*

Remark. The assumption that the Gurevich entropy is finite is only needed for the *if* direction.

Proof. Theorem 8.1 implies the $\Leftarrow$ implication by contraposition: If ϕ were not SPR, then Theorem 8.1 would have provided a proper subgraph $\mathscr{G}'$ such that $\sigma : \Sigma^+(\mathscr{G}') \to \Sigma^+(\mathscr{G}')$ is topologically mixing, and so that $P_G(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi)$.

We prove the other direction by contradiction. Assume that ϕ is SPR, but there is a proper subgraph $\mathscr{G}'$ such that $\sigma : \Sigma^+(\mathscr{G}') \to \Sigma^+(\mathscr{G}')$ is topologically mixing and $P_G(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi)$. Fix some vertex a of $\mathscr{G}'$. By Lemma 2.1, $P^*_G(\phi, a) < P_G(\phi)$, whence

$$P^*_G(\phi|_{\Sigma^+(\mathscr{G}')}, a) \leq P^*_G(\phi, a) < P_G(\phi) = P_G(\phi|_{\Sigma^+(\mathscr{G}')}),$$

whence $\phi|_{\Sigma^+(\mathscr{G}')}$ is SPR. In particular, $\phi|_{\Sigma^+(\mathscr{G}')}$ is positively recurrent. Let m denote the RPF measure of $\phi|_{\Sigma^+(\mathscr{G}')}$. Since Σ^+ has finite Gurevich entropy, m has finite entropy. By Theorem 2.1, m is an equilibrium measure of $\phi|_{\Sigma^+(\mathscr{G}')}$.

Now let m' denote the shift invariant probability measure on $\Sigma^+(\mathscr{G})$, given by $m'(E) := m(E \cap \Sigma^+(\mathscr{G}'))$. Clearly

$$P_{m'}(\phi) = P_m(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi|_{\Sigma^+(\mathscr{G}')}) = P_G(\phi|_{\Sigma^+(\mathscr{G})}).$$

So m' defines an equilibrium measure on $\Sigma^+(\mathscr{G})$. By Theorem 2.1, this is the unique equilibrium measure. But now we have a contradiction, because $\operatorname{supp} m' = \Sigma^+(\mathscr{G}')$, whereas the equilibrium measure of a potential with summable variations on a topologically transitive TMS is always globally supported, see §2.5. □

9. Two sided topological Markov shifts

Suppose $\mathscr{G}$ is a countable directed graph. The *two-sided* topological Markov shift associated to $\mathscr{G}$ is the dynamical system with the space

$$\Sigma = \Sigma(\mathscr{G}) := \{\underline{x} = (\ldots, x_{-1}, x_0, x_1, \ldots) : x_i \in S, x_i \to x_{i+1} \text{ for all } i\},$$

the action $\sigma(\underline{x})_i = x_{i+1}$, and the metric $d(\underline{x}, \underline{y}) = \exp(-\min\{|i| : x_i \neq y_i\})$.

The definitions we gave in §2 for one-sided shifts of the Gurevich pressure, Gurevich entropy, the spectral decomposition and SPR topological Markov shifts extend verbatim to the two-sided case, after replacing all the Σ^+ by Σ.

The definition of the SPR property for potentials has a similar extension to the two sided case, except that now to define the discriminant, we need to work with the induced shift on $\overline{\Sigma} := \overline{S}^{\mathbb{Z}}$, instead of on $\overline{\Sigma} := \overline{S}^{\mathbb{N}\cup\{0\}}$.

With these definitions in place, Theorems 6.1, 7.1 on the EKP inequality for SPR potentials extend to the two-sided setup without much difficulty.

We explain why. Any weakly Hölder continuous $\phi : \Sigma \to \mathbb{R}$ is cohomologous via a bounded continuous transfer function to a "one-sided" function of the form $\phi^+ \circ \pi$, where ϕ^+ is a weakly Hölder continuous function on Σ^+ [Dao13], [Sin72], [Bow75].

The pressure function is defined in terms of sums over periodic orbits. Such sums do not change if we change ϕ by a coboundary. Therefore, it is easy to see that ϕ is SPR if and only if ϕ^+ is SPR, and $P_G(\phi) = P_G(\phi^+)$.

Consider the projection $\pi : \Sigma \to \Sigma^+$, $\pi(\underline{x}) = (x_0, x_1, \ldots)$. If μ is a shift invariant measure on Σ, then $\mu^+ := \mu \circ \pi^{-1}$ is a shift invariant probability measure on Σ^+, and it is easy to see that $h_{\mu^+}(\sigma) = h_\mu(\sigma)$. In addition, coboundaries with bounded continuous transfer functions are absolutely integrable with zero integral for all shift invariant probability measures, so $\int \phi d\mu = \int \phi^+ d\mu^+$. So $\mu \in \mathscr{M}_\phi(\Sigma)$ if and only if $\mu^+ \in \mathscr{M}_{\phi^+}(\Sigma^+)$, and in this case $P_\mu(\phi) = P_{\mu^+}(\phi^+)$.

In particular, μ maximizes $P_\mu(\phi)$ if and only if μ^+ maximizes $P_{\mu^+}(\phi^+)$, and therefore μ_ϕ is the equilibrium measure of ϕ on Σ, if and only if $(\mu_\phi)^+$ is the equilibrium measure of ϕ^+ on Σ^+. We are therefore at liberty to write $\mu_\phi^+ = \mu_{\phi^+}$.

It is now a simple matter to see that the EKP inequalities (1.2) and (1.4) for Σ^+ and ϕ^+, imply the EKP inequalities (1.2) and (1.4) for Σ and ϕ, *provided the test function ψ belongs to* $\mathcal{H}_\beta^+(\Sigma) := \{\psi^+ \circ \pi \circ \sigma^n : \psi \in \mathcal{H}_\beta(\Sigma^+), n \in \mathbb{Z}\}$. Since $\mathcal{H}_\beta^+(\Sigma)$ is dense in the space of β-Hölder continuous functions on Σ, (1.2) and (1.4) follow for all β-Hölder continuous functions on Σ.

References

[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.

[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.

[BCFT18] Keith Burns, Vaughn Climenhaga, Todd Fisher, and Daniel J. J. Thompson. Unique equilibrium states for geodesic flows in nonpositive curvature. *Geom. Funct. Anal.*, 28(5):1209–1259, 2018.

[BCS19] Jérôme Buzzi, Sylvain Crovisier, and Omri Sarig. Measures of maximal entropy for surface diffeomorphisms, 2019.

[BL98] Martine Babillot and François Ledrappier. Lalley's theorem on periodic orbits of hyperbolic flows. *Ergodic Theory Dynam. Systems*, 18(1):17–39, 1998.

[Bow72] Rufus Bowen. The equidistribution of closed geodesics. *Amer. J. Math.*, 94:413–423, 1972.

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

[Bow75] 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.

[CKW21] Vaughn Climenhaga, Gerhard Knieper, and Khadim War. Uniqueness of the measure of maximal entropy for geodesic flows on certain manifolds without conjugate points. *Adv. Math.*, 376:107452, 44, 2021.

[Cli18] Vaughn Climenhaga. Specification and towers in shift spaces. *Comm. Math. Phys.*, 364(2):441–504, 2018.

[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.

[CT12] Vaughn Climenhaga and Daniel J. Thompson. Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors. *Israel J. Math.*, 192(2):785–817, 2012.

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

[ELMV12] Manfred Einsiedler, Elon Lindenstrauss, Philippe Michel, and Akshay Venkatesh. The distribution of closed geodesics on the modular surface, and Duke's theorem. *Enseign. Math. (2)*, 58(3-4):249–313, 2012.

[Fie96] Ulf-Rainer Fiebig. *Symbolic dynamics and locally compact Markov shifts*. PhD thesis, U. Heidelberg, 1996.

[GH88] Yves Guivarc'h and Jean 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.

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

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

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

[GZ88] Boris M. Gurevich and A. S. Zargaryan. Conditions for the existence of a maximal measure for a countable symbolic Markov chain. *Vestnik Moskov. Univ. Ser. I Mat. Mekh.*, (5):14–18, 103, 1988.

[Hof79] Franz Hofbauer. On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. *Israel J. Math.*, 34(3):213–237 (1980), 1979.

[IJT15] Godofredo Iommi, Thomas Jordan, and Mike Todd. Recurrence and transience for suspension flows. *Israel Journal of Mathematics*, 209(2):547–592, 2015.

[ITV19] Godofredo Iommi, Mike Todd, and Anibal Velozo. Escape of entropy for countable markov shifts, 2019.

[ITV20] Godofredo Iommi, Mike Todd, and Aníbal Velozo. Upper semi-continuity of entropy in non-compact settings. *Math. Res. Lett.*, 27(4):1055–1078, 2020.

[Kad15] Shirali Kadyrov. Effective uniqueness of Parry measure and exceptional sets in ergodic theory. *Monatsh. Math.*, 178(2):237–249, 2015.

[Kad17] Shirali Kadyrov. Effective equidistribution of periodic orbits for subshifts of finite type. In *Colloquium Mathematicum*, volume 149, pages 93–101. Polska Akademia Nauk, 2017.

[Kat95] Tosio Kato. *Perturbation theory for linear operators.* Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[Kha17] Ilya Khayutin. Large deviations and effective equidistribution. *International Mathematics Research Notices*, 2017(10):3050–3106, 2017.

[Kni98] Gerhard Knieper. The uniqueness of the measure of maximal entropy for geodesic flows on rank 1 manifolds. *Ann. of Math. (2)*, 148(1):291–314, 1998.

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

[Mor07] Ian D. Morris. Entropy for zero-temperature limits of Gibbs-equilibrium states for countable-alphabet subshifts of finite type. *J. Stat. Phys.*, 126(2):315–324, 2007.

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

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

[Pav20] Ronnie Pavlov. On entropy and intrinsic ergodicity of coded subshifts. *Proc. Amer. Math. Soc.*, 148(11):4717–4731, 2020.

[Pol11] Fabrizio Polo. *Equidistribution in chaotic dynamical systems.* ProQuest LLC, Ann Arbor, MI, 2011. Thesis (Ph.D.)–The Ohio State University.

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

[Rue72] David Ruelle. Statistical mechanics on a compact set with Z^ν action satisfying expansiveness and specification. *Bull. Amer. Math. Soc.*, 78:988–991, 1972.

[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.

[Rue03] Sylvie Ruette. On the Vere-Jones classification and existence of maximal measures for countable topological Markov chains. *Pacific J. Math.*, 209(2):366–380, 2003.

[Rüh16] René Rühr. Effectivity of uniqueness of the maximal entropy measure on p-adic homogeneous spaces. *Ergodic Theory Dynam. Systems*, 36(6):1972–1988, 2016.

[Rüh21] René Rühr. Pressure inequalities for gibbs measures of countable markov shifts. *Dynamical Systems*, pages 1–8, 2021.

[Sal88] Ibrahim Salama. Topological entropy and recurrence of countable chains. *Pacific journal of mathematics*, 134(2):325–341, 1988.

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

[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.

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

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

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

[VJ62] David Vere-Jones. Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford Ser. (2)*, 13:7–28, 1962.

[Wal75] Peter Walters. Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.*, 214:375–387, 1975.

[Wei70] Benjamin Weiss. Intrinsically ergodic systems. *Bull. Amer. Math. Soc.*, 76:1266–1269, 1970.

Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, POB 26, Rehovot, Israel

Email address: omsarig@gmail.com reneruehr@gmail.com