ASYMPTOTIC BEHAVIOR OF THE PRESSURE FUNCTION FOR HÖLDER POTENTIALS

TAMARA KUCHERENKO AND ANTHONY QUAS

ABSTRACT. We study the behavior of the pressure function for Hölder continuous potentials on mixing subshifts of finite type. The classical theory of thermodynamic formalism shows that such pressure functions are convex, analytic and have slant asymptotes. We provide a sharp exponential lower bound on how fast the pressure function approaches its asymptotes. As a counterpart, we also show that there is no corresponding upper bound by exhibiting systems for which the convergence is arbitrarily slow. However, we prove that the exponential upper bound still holds for a generic Hölder potential. In addition, we determine that the pressure function satisfies a coarse uniform convexity property. Asymptotic bounds and quantitative convexity estimates are the first additional general properties of the pressure function obtained in the settings of Bowen and Ruelle since their groundbreaking work more than 40 years ago.

1. INTRODUCTION

To quote Ruelle, "the main object of the thermodynamic formalism is to study the differentiability and analyticity properties of the function P [topological pressure], and the structure of the equilibrium states and Gibbs states" [25, Introduction]. Arguably the most cited result in this context is that for mixing subshifts of finite type the pressure is real analytic on the space of Hölder continuous potentials and that for each such potential there is only one Gibbs state which is also the only equilibrium state. These statements served as catalysts for the growth of the ergodic theory of smooth hyperbolic systems starting with Anosov maps. In his breakthrough work [5] Bowen applied Ruelle's analytic tool of transfer operators to Anosov diffeomorphisms (in fact, more general Axiom A systems) using Markov partitions and symbolic coding introduced by Sinai. This allowed for the properties

Key words and phrases. thermodynamic formalism, topological pressure, variational principle, equilibrium states, Hölder potentials, subshifts of finite type.

T.K. is supported by grants from the Simons Foundation #430032.

A.Q. is supported by a grant from NSERC.

of the pressure and Gibbs states on shift spaces to be carried over to differentiable systems, resulting in a description of the behaviour of Lebesgue-almost every orbit.

Shortly after the introduction of the thermodynamic formalism, its relationship to dimension theory was discovered, where the concept of the topological pressure once again played a central role. A highly influential result due to Bowen [6] and Ruelle [26] is that the Hausdorff dimension of Julia sets for conformal maps can be computed as the root of the pressure function of a certain potential. It was used, in particular, to establish the analyticity of the Hausdorff dimension as a function of the parameter in the interior of the main cardioid of the Mandelbrot set. Since then dimensional estimates were obtained for numerous invariant sets and measures [19, 1, 20], the vast majority of which use a version of Bowen's pressure formula.

As part of dimension theory, multifractal analysis is concerned with the complexity of level sets of asymptotically defined quantities such as Birkhoff averages, Lyapunov exponents, and local entropies. Usually, the geometry of a level set is sufficiently complicated to necessitate tools such as Hausdorff dimension or topological entropy in order to describe its size and complexity. In most cases, the main technical device to identify the various multifractal spectra is the pressure function, see e.g. [2, 3, 9]. Through this approach the dimension of a level set is evaluated by the entropy of a suitable invariant measure which is produced as an equilibrium state for the appropriate potential. For instance, the pressure function of the geometric potential contains information about the spectrum of the maximum Lyapunov exponent for geodesic flows on compact manifolds [8].

Despite the fact that the pressure function has been used in applications more and more over time, the understanding of the behavior of the function itself has not gone beyond the general statements of analyticity, convexity and existence of asymptotes – properties already known to Bowen and Ruelle in the 1970's. Analyticity is the strongest possible regularity condition for real-valued functions. In the present work we examine the other two properties of the pressure in the classical settings of Hölder potentials and mixing subshifts of finite type. We are able to characterize the rate of convergence of the pressure function to its asymptote as well as strengthen the convexity statement.

Throughout the paper we assume that $\phi : X \to \mathbb{R}$ is a Hölder continuous potential associated with a mixing subshift of finite type (X, T).

The *topological pressure* of ϕ can be defined via the Variational Principle by

$$P_{\text{top}}(\phi) = \sup \left\{ h_T(\mu) + \int \phi \, d\mu \right\}$$

where the supremum is taken over the set of all T-invariant probability measures on X and $h_T(\mu)$ denotes the measure-theoretic entropy of the measure μ. The measures which realize the above supremum are called the *equilibrium states* of ϕ. The terminology comes from statistical physics: the quantity $E_\phi = -(h_T(\mu) + \int \phi \, d\mu)$ represents the free energy of the system in state μ and the equilibrium is given by the states which minimize the free energy. We refer the reader to the monographs [5, 25, 27] for a detailed exposition.

We study the *pressure function* of ϕ, $p_\phi(t) = P_{\text{top}}(t\phi)$, where t is a real-valued parameter. In statistical physics this function is regarded as a tool to observe an evolution of a system depending on a continuous external factor. One common interpretation of the parameter t is the inverse temperature of the system. Then the behavior of $p_\phi(t)$ when $t \to \infty$ is of significant interest, since it reveals certain changes within the system when the temperature is lowered to zero. It has been observed that on the microscopic level materials tend to be highly ordered at a low temperature, which mathematically means that corresponding equilibrium states should be supported on configurations of low complexity [23]. A system at absolute zero temperature exists in its ground state, hence the limit points of equilibrium states as temperature approaches zero are termed the ground states of the system. A long standing conjecture in ergodic theory (finally resolved in the affirmative by Contreras in 2016 [10]) states that for a generic Hölder potential on a subshift of finite type the ground state is unique and supported on a periodic orbit. The question we address here is how fast the energy level of a system can approach the energy of its ground state when the temperature is lowered to zero. This leads to the task of characterizing the asymptotic behavior of the pressure function.

For each t the potential $t\phi$ has a unique equilibrium state μ_t. The accumulation points of the family (μ_t) as $t \to \infty$ are the ground states of ϕ. If μ_t converges (in the weak*-topology) then the limit is called the zero-temperature measure. The matter of existence of such a measure received considerable attention in the literature. In 2001 Contreras, Lopes and Thieullen [11] established the existence of the zero-temperature limit for a generic set of Hölder potentials. Two years later, Bremont [7] proved that any locally constant potential admits a zero-temperature measure, which piqued the interest in the validity of the same statement for Hölder potentials. In 2010, Chazottes and Hochman [12] effectively ended the discussion by constructing an example of a Lipschitz continuous potential on a full shift such that the zero-temperature limit does not exist.

Although there might be multiple ground states for a Hölder potential ϕ, each ground state must maximize the integral of ϕ among all invariant probability measures (see e.g. [14]). Moreover, it has maximal entropy among the integral maximizing measures. We compare the free energy of the system in its equilibrium at temperature t, represented by $p_\phi(t)$, to the free energy of its ground state at the same temperature. Letting μ_∞ be one of the weak*-accumulation points of μ_t we see that the line with slope $\int \phi \, d\mu_\infty$ and vertical axis intercept $h_T(\mu_\infty)$ is the slant asymptote for the pressure function $p_\phi(t)$. It is clear from the variational principle that the pressure function lies above its asymptote. For any Hölder continuous ϕ that is not cohomologous to a constant, we establish a lower bound on the gap between the asymptote and the pressure function. We illustrate this statement in Figure 1.

Theorem 1. *Let X be a mixing subshift of finite type with positive entropy. Let ϕ be a Hölder continuous function that is not cohomologous to a constant. Then there exist C and t_0 such that $p_\phi(t) \geq \ell_\infty(t) + e^{-Ct}$ for all $t \geq t_0$, where $\ell_\infty(t)$ is the asymptote to $p_\phi(t)$ at infinity.*

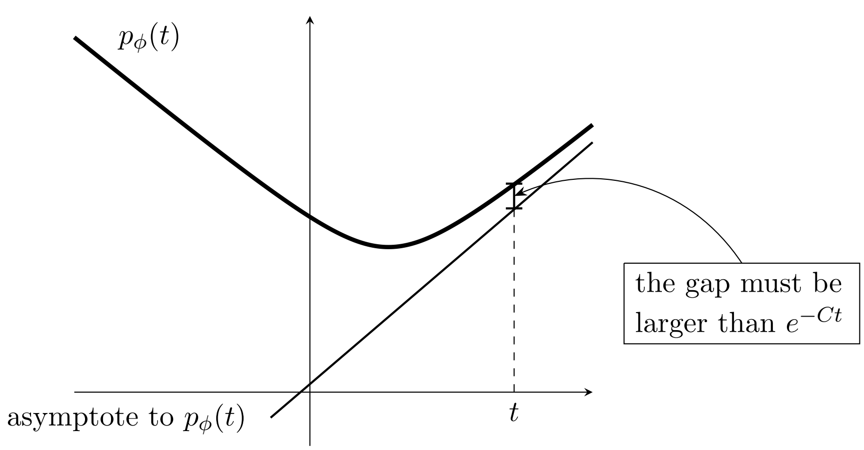

FIGURE 1. This figure illustrates Theorem 1.

The exponential lower bound on the gap is the best one can hope for. We see in Example 10 that for any locally constant potential on a full shift the rate with which the pressure function approaches its asymptote is exactly exponential. This might suggest that for a Hölder potential the upper bound on the rate should be of exponential type as well. However, this is far off the mark. As it turns out, there is no upper bound at all as shown by the following theorem.

Theorem 2. *Let X be a mixing subshift of finite type with positive entropy and let $f : \mathbb{R} \to \mathbb{R}$ be any convex function with an asymptote at infinity $\ell_\infty(t) = at + b$, where $0 \le b < h_{\text{top}}(X)$. Then there exists a Hölder potential $\phi : X \to \mathbb{R}$ such that $p_\phi(t)$ is asymptotic to $\ell_\infty(t)$ as $t \to \infty$ and $p_\phi(t) > f(t)$ for all sufficiently large t.*

Note that in order for the line $at + b$ to be an asymptote to the pressure function of some potential, b must be the entropy of its ground state. Therefore, by the variational principle, b cannot exceed the topological entropy of X (or be negative). If $b = h_{\text{top}}(X)$ then this ground state is necessarily the unique measure of maximal entropy of X, in which case the pressure function coincides with its asymptote. Hence, $0 \le b < h_{\text{top}}(X)$ is the weakest condition under which the statement holds.

We see from Theorem 2 that the pressure function can decrease to its asymptote arbitrarily slowly, while Theorem 1 tells us that it cannot decrease faster than exponentialy. This raises natural questions: (i) what is the typical asymptotic behavior for Hölder potentials? (ii) do properties of the associated ground state have any impact on the rate of convergence? We provide a resolution to both. Question (ii) has a negative answer. In Section 6 we construct two Hölder potentials on the full shift on two symbols which have the same zero temperature limit – a point-mass measure. However, for one of them the pressure function approaches its asymptote exponentially fast, while for the other one the convergence is no faster than $\log\log t/\log t$. To answer Question (i), we prove that generically Hölder potentials admit an exponential upper bound on the gap between the pressure function and its asymptote.

Theorem 3. *Let X be a mixing shift of finite type and let $\mathcal{H}$ denote a fixed Hölder class on X. Then there is a dense open subset $\mathcal{U}$ of potentials in $\mathcal{H}$ such that for all $\phi \in \mathcal{U}$, there exist $C > 0$ and t_0 such that $p_\phi(t) \le \ell_\infty(t) + e^{-Ct}$ for all $t \ge t_0$.*

This leads us to Question (iii): could a typical asymptotic behavior for Hölder potentials be given simply by $p_\phi(t) - \ell_\infty(t) \approx e^{-Ct}$ for some constant C? We conjecture that the answer is yes, however the proof of such a statement appears to require an approach different from the one used in this work.

Next, we turn our attention to the convexity of the pressure function. Although strictly convex analytic functions could be "almost flat" on some intervals, we show that this is not possible for the pressure function of a Hölder potential. For a fixed $t \in \mathbb{R}$ we consider a symmetric interval $(t-h, t+h)$, where $h > 0$. Since the pressure function is strictly convex, the midpoint of the secant line of the graph of the pressure function corresponding to points $t-h$ and $t+h$ is above the value of the pressure at t (see Figure 2). We show that the difference cannot be smaller than $c_1 e^{-c_2|t|/h}$ for some fixed positive constants c_1 and c_2 which do not depend on the point t. We interpret this as a quantitative lower bound on the curvature, where the curvature bounds improve (a lot) when one considers coarse intervals.

Theorem 4. *Let X be any mixing subshift of finite type with positive entropy and let ϕ be a Hölder continuous function that is not cohomologous to a constant. Then there exist $c_1 > 0$ and $c_2 > 0$ such that for any $t \in \mathbb{R}$ and any $h \in \mathbb{R}^+$*

$$\frac{p_\phi(t+h) + p_\phi(t-h)}{2} - p_\phi(t) > c_1 e^{-c_2|t|/h}. \tag{1}$$

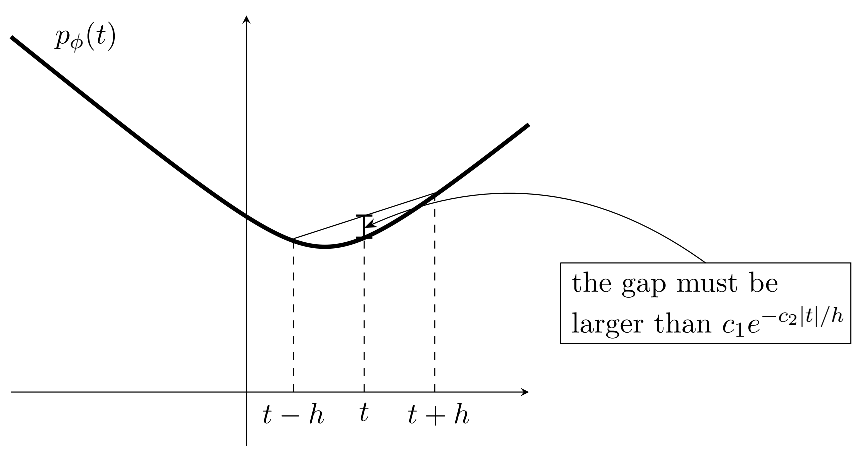

Figure 2. This figure illustrates Theorem 4.

One motivation for the present study comes from our previous work [15] where we consider the pressure function of a *continuous* potential on a full shift over a finite alphabet. It is known that such a pressure function is Lipschitz, convex, and has an asymptote at infinity. It turns out that in the case of continuous potentials no additional properties are present. In [15] we explicitly construct a continuous potential on a full shift whose pressure function coincides with *any* prescribed convex Lipschitz asymptotically linear function starting at a given positive value of the parameter. Immediately the question arose whether an analogously strong statement holds for the pressure function of a Hölder potential, where "Lipschitz" is replaced by "analytic". It follows from the assertions we made above that the answer is no. While this paper was in preparation we learned that the above question was also addressed in [17], where a negative answer was obtained by establishing an inequality involving powers of the second, third, and fourth derivatives of the pressure function.

We finish this section with a brief outline of the paper. After preliminary material is given in Section 2, the next two sections are devoted to the proofs of Theorem 1 and Theorem 4. It is convenient to write $g_t(s)$ for the gap between $p_\phi(t)$ and its tangent line at t: $g_t(s) = p_\phi(s) - \Big(h_T(\mu_t) + s\int\phi\,d\mu_t\Big)$ where μ_t is the (unique) equilibrium state for the potential $t\phi$. We extend this notation to $g_\infty(s)$ for the gap between p_ϕ and its slant asymptote $\ell_\infty(s) = h_T(\mu_\infty) + s\int\phi\,d\mu_\infty$. Recall that even though μ_∞ may not be unique, $\int\phi\,d\mu_\infty$ and $h_T(\mu_\infty)$ are the same for all accumulation points, μ_∞, of the corresponding family of equilibrium states. In this notation, Theorem 1 may be re-expressed as

$$g_\infty(t) \geq e^{-Ct}$$

for all large t; and Theorem 4 may be re-expressed as

$$g_t(t-h) + g_t(t+h) \geq c_1 e^{-c_2|t|/h}$$

for all $t \in \mathbb{R}$ and all $h > 0$.

To prove the two theorems we estimate the gap functions g_∞ and g_t from below by building a new invariant measure μ' starting from μ_∞ (for Theorem 1) or μ_t (for Theorem 4) using "coupling and splicing" techniques described in [21]. A general construction of this kind is carried out in Section 3. The objective is to increase the entropy of μ' compared to μ_∞(or μ_t) while controlling the decrease in the value of the integral of ϕ. In Section 4 we verify that the measure μ' can be constructed in such a way that the gain in entropy exceeds the drop in the integral ensuring that the quantity $h_T(\mu') + t\int\phi\,d\mu'$ is above the asymptote (or the tangent line) to the pressure function by the required amount.

In Section 5 we prove Theorem 2. We show that one can find a potential ϕ whose pressure function approaches its asymptote as slowly as desired. The idea of the proof is to define ϕ in terms of the distance to a carefully chosen subshift of X. We mimic the Rothstein shift [22], whose complexity is slightly below exponential.

Lastly, in Section 6 we establish the generic exponential upper bound for Hölder potentials (Theorem 3). To prove this result we use the fact due to Contreras [10] that the set of potentials for which the zero temperature measure supported on a periodic orbit contains an open and dense set. However, this set must be necessarily modified for our purpose since there are potentials for which the zero temperature measure is supported on a periodic orbit, but for which the convergence of the pressure function to the asymptote is much slower than exponential, see Example 14.

2. Notation and conventions

Our analysis takes place in the setting of two-sided shifts of finite type, which we always assume to have forbidden blocks of length 2 only. We denote the alphabet by $\mathcal{A}$ and the shift map by T. We use the metric $d(x,y) = 2^{-n(x,y)}$ where $n(x,y) = \inf\{|n|\colon x_n \neq y_n\}$.

For a word $w = w_0 \cdots w_{n-1} \in \mathcal{A}^n$, the cylinder generated by w is denoted $[w] = \{x \in \mathcal{A}^{\mathbb{Z}} : x_i = w_i \text{ for i} = 0, \ldots, n-1\}$. We write $|w|$ for its length, n (and also refer to n as the length of the cylinder set $[w]$). Given any two words w and v we write wv for their concatenation. For a subshift $X \subset \mathcal{A}^{\mathbb{Z}}$ we denote by $\mathcal{L}_n(X)$ the set of all admissible words in X of length n and write $\mathcal{L}(X) = \bigcup_{n=1}^{\infty} \mathcal{L}_n(X)$ for the language of X. A subshift of finite type X is *mixing* if there exists an integer L such that for any two words $u, v \in \mathcal{L}(X)$ and any $n \geq L$ there is a word $w \in \mathcal{L}_n(X)$ such that $uwv \in \mathcal{L}(X)$. In this case we call the smallest such L the *mixing length* of X.

We write $\mathcal{P}$ for the (generating) partition consisting of all cylinder sets of length 1 and write $\mathcal{P}_n$ for the partition consisting of all cylinder sets of length n. If μ is a T-invariant measure we denote its entropy by $h_T(\mu) := \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}_n)$, where as usual $H_\mu(\mathcal{Q})$ denotes the entropy of the countable partition $\mathcal{Q}$ with respect to measure μ (and we use natural logarithms in the definition).

A function ϕ is Hölder continuous if there exist $c \geq 0$ and $0 < \alpha < 1$ such that $|\phi(x) - \phi(y)| \leq c\alpha^{n(x,y)}$ where $n(x,y)$ is as above. For a shift of finite type X and a Hölder continuous potential ϕ, our principal object of study is the function $p_\phi(\cdot)$ given by $p_\phi(t) = P_{\text{top}}(t\phi)$, where $P_{\text{top}}(\psi)$ denotes the topological pressure, equal by the Variational Principle to $\sup_\mu \left(h_T(\mu) + \int \psi \, d\mu\right)$, where the supremum is taken over the collection of T-invariant probability measures on X, $\mathcal{M}(X)$.

We now recall basic properties of the function p_ϕ which are easily deduced from the Variational Principle. One may check that p_ϕ is convex. Monotonicity of the map $\psi \mapsto P_{\text{top}}(\psi)$ together with the equality $P_{\text{top}}(\psi + c) = P_{\text{top}}(\psi) + c$ implies that p_ϕ is a Lipschitz function. When t is large the term $\int \phi\, d\mu$ predominates, so p_ϕ has a slant asymptote: there is an affine function $\ell_\infty(s)$ such that $p_\phi(s) \geq \ell_\infty(s)$ with $p_\phi(s) - \ell_\infty(s) \to 0$ as $s \to \infty$. In fact, the slant asymptote is given by $\ell_\infty(s) = h_T(\mu_\infty) + s \int \phi\, d\mu_\infty$ where μ_∞ is any measure with maximal entropy among the collection of *maximizing measures*, that is those invariant measures achieving the supremum $\sup_{\nu \in \mathcal{M}(X)}\{\int \phi\, d\nu\}$ (see e.g. [14]).

Another property comes from the description of equilibrium states as tangent functionals to the pressure given by Walters [28]. If μ is any equilibrium state for $t\phi$, then the affine function $\ell_\mu(s) = h_\mu(T) + s \int \phi\, d\mu$ is a sub-gradient of p_ϕ at t: $p_\phi(s) \geq \ell_\mu(s)$ for all s; and $\ell_\mu(t) = p_\phi(t)$. Conversely, for any sub-gradient $\ell(s)$ of p_ϕ at t, there is an equilibrium state μ such that $\ell(s) = \ell_\mu(s)$ for all s. Since $h_T(\mu)$ is a non-negative quantity, bounded above by $\log|\mathcal{A}|$, we see that all subgradients of p_ϕ intercept the vertical axis in a bounded sub-interval of $[0, \infty)$. One can show that boundedness of the vertical axis intercepts of the supporting lines implies both the Lipschitz condition and the existence of a slant asymptote.

Although there are no other general properties of the pressure function for continuous potentials, one can say much more about the pressure under the restriction that X is a subshift of finite type and potential $\phi : X \to \mathbb{R}$ is Hölder. It was shown by Ruelle in [24] (see also [5]) that in this case ϕ has a unique equilibrium state μ_ϕ which satisfies *Gibbs property*, namely there is a constant $C_\phi > 0$ such that for all $n \in \mathbb{N}$, $w \in \mathcal{L}_n(X)$ and $x \in [w]$ we have

$$\frac{1}{C_\phi} \leq \frac{\mu_\phi([w])}{\exp\left(S_n\phi(x) - nP_{\text{top}}(\phi)\right)} \leq C_\phi, \tag{2}$$

where, as usual, $S_n\phi(x)$ denotes the Birkhoff sum $\sum_{i=0}^{n-1} \phi(T^i x)$. Furthermore, the function $p_\phi(t)$ is analytic and strictly convex [25]. Clearly in this case p_ϕ has a unique sub-gradient (in fact a tangent line) at each t, which we denote by $\ell_t(\cdot)$, so that $\ell_t(s) = h(\mu_t) + s \int \phi\, d\mu_t$ where μ_t is the unique equilibrium state for $t\phi$.

3. Construction of invariant measures

In this section we use what have been termed "coupling and splicing" techniques [21] to build a family of invariant measures μ' on a mixing subshift of finite type (X, T) by modifying realizations of an initial measure μ. A word w in the language of X is fixed and the measures μ' are, roughly speaking, obtained by starting from a realization x of

μ and randomly replacing words of length $|w|$ in x by copies of w with some frequency. To ensure that the new point belongs to X, blocks of the mixing length L prior to and following the inserted w's have to be modified also.

We start with some auxiliary measures that will be used in the construction. We build a family of ergodic measures on $\{0,1\}^{\mathbb{Z}}$ with two parameters η and M such that gaps between 1's are independent and geometrically distributed with parameter η, taking values in $\{n\colon n \geq M\}$. To distinguish from the subshift X we denote the shift map on $\{0,1\}^{\mathbb{Z}}$ by σ.

Lemma 5. *There exists a family of ergodic invariant measures $\nu_{\eta,M}$ on $\{0,1\}^{\mathbb{Z}}$ where η runs over $(0,1)$ and M runs over $\mathbb{N}$ with the following properties:*

- *(Spacing of 1's) For $\nu_{\eta,M}$-a.e. $y \in \{0,1\}^{\mathbb{Z}}$, if $i < j$ and $y_i = y_j = 1$, then $j \geq i + M$;*
- *(Frequency of 1's) $\nu_{\eta,M}([1]) = \Big(M + e^{-\eta}/(1-e^{-\eta})\Big)^{-1}$, or approximately $\eta/(1+M\eta)$ for small values of η;*
- *(Entropy) $h(\nu_{\eta,M}) = -\eta\log\eta + O(\eta M)$ for small values of η.*

Proof. Set $S = \{M, M+1, M+2, \dots\}^{\mathbb{Z}}$ with the shift map σ. For any parameter $\eta > 0$, equip S with the Bernoulli measure χ where the symbol $M+k$ occurs with probability $(1-e^{-\eta})e^{-k\eta}$ for $k = 0,1,2,\dots$. By a standard construction we take the suspension of S by the height function $h(x) = x_0$ to obtain the space $\bar{S} = \{(x,n)\colon x \in S; 0 \leq n < x_0\}$, equipped with the standard suspension map

$$\bar{\sigma}(x,n) = \begin{cases} (x, n+1) & \text{if } n < x_0 - 1; \\ (\sigma(x), 0) & \text{if } n = x_0 - 1. \end{cases}$$

The measure χ on S lifts to the probability measure $\bar{\chi}$ on $\bar{S}$ defined by

$$\begin{aligned} \bar{\chi}(A \times \{j\}) &= \chi(A)/\Big(\sum_{n \geq M} n(1-e^{-\eta})e^{-\eta(n-M)}\Big) \\ &= \chi(A)/\Big(M + (1-e^{-\eta})\sum_{n=0}^{\infty} ne^{-\eta n}\Big) \qquad (3) \\ &= \chi(A)/\Big(M + e^{-\eta}/(1-e^{-\eta})\Big), \end{aligned}$$

for any set $A \subset \{x \in S\colon x_0 > j\}$. The denominator in this expression is simply a normalization factor, which is just the integral of the height function. Since χ is an ergodic σ-invariant measure, it follows that $\bar{\chi}$ is ergodic and $\bar{\sigma}$-invariant.

We then build a factor map from $\bar{S}$ to $Y = \{0,1\}^{\mathbb{Z}}$, defined by

$$\Phi(x,n)_j = \begin{cases} 1 & \text{if } \bar{\sigma}^j(x,n) \in \{(y,0)\colon y \in S\}; \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The push-forward of $\bar{\chi}$ under Φ will be denoted by $\nu_{\eta,M}$, or simply by ν. The statement on the spacing of 1's is now immediate from the construction of Y. Taking $A = S$ and $j = 0$ in (3) gives the statement on the frequency of 1's.

Since Φ is one-to-one, the entropies $h_{\bar{S}}(\bar{\chi})$ and $h_\sigma(\nu)$ are equal. A calculation shows that

$$\begin{aligned} h_\sigma(\chi) &= -\sum_{n=0}^{\infty}(1-e^{-\eta})e^{-\eta n}\log\left((1-e^{-\eta})e^{-\eta n}\right) \\ &= -\log(1-e^{-\eta}) + \eta(1-e^{-\eta})\sum_{n=0}^{\infty} n e^{-\eta n} \\ &= -\log(1-e^{-\eta}) + \eta e^{-\eta}/(1-e^{-\eta}) \\ &= -\log\eta + O(1). \end{aligned}$$

Hence by Abramov's formula,

$$\begin{aligned} h_\sigma(\nu) = h_{\bar{\sigma}}(\bar{\chi}) &= \frac{-\log(1-e^{-\eta}) + \eta e^{-\eta}/(1-e^{-\eta})}{M + e^{-\eta}/(1-e^{-\eta})} \\ &= -\eta\log\eta + O(\eta M). \end{aligned}$$

□

Now let X be a shift of finite type (defined by forbidden blocks of length 2) and μ be an ergodic invariant measure on X. Let $w = w_0 \dots w_{m-1}$ be a word in the language of X with the property that there does not exist $j < \frac{2m}{3}$ such that $w_0 \dots w_{n-1-j} = w_j \dots w_{n-1}$ (so that no prefix of w of length at least $\frac{m}{3}$ recurs within w), then we say that w satisfies the *no long overlaps condition*. We now use the measures $\nu_{\eta,M}$ constructed above to build a modified measure $\mu'_{\eta,w}$ on X inserting additional w's as described at the beginning of the section.

Let $M = m+2L+1$, where L is the mixing length of X. Then let $\nu_{\eta,M}$ and (Y,σ) be as previously constructed. We build a new measure $\mu'_{\eta,w}$ as follows. For each $a \in \mathcal{A}$ denote by $u'(a)$ and $u''(a)$ the lexicographically minimal words of length L such that $au'(a)w_0$ and $w_{m-1}u''(a)a$ belong to $\mathcal{L}(X)$. Then for $a,b \in \mathcal{A}$ let $v(a,b) = u'(a)wu''(b)$. Define a map $\Phi\colon X \times Y \to X$ where $\Phi(x,y)_j$ is given by

$$\begin{cases} v(x_{k-1},x_{k+L+m})_{j-k} & \text{if } y_k = 1 \text{ for some } k \in \{j-m-2L-1,\dots,j\}; \\ x_j & \text{otherwise.} \end{cases}$$

Informally, $\Phi(x,y)$ is the point x' obtained by simultaneously replacing, for each k such that $y_k=1$, the word $x_{k+L}\dots x_{k+L+m-1}$ by w, and choosing $x'_k\dots x'_{k+L-1}$ and $x'_{k+L+m}\dots x'_{k+m+2L-1}$ to be the minimal words so that the resulting word belongs to X. We then obtain a measure $\mu'_{\eta,w}$ on X defined by

$$\mu'_{\eta,w}=\Phi_*(\mu\times\nu_{\eta,M}). \tag{5}$$

Lemma 6. *Let (X,T) be a mixing subshift of finite type with mixing length L, μ be an ergodic invariant measure on X, and w be a word of length at least $3L$ with no long overlaps. Suppose $\eta>0$ is such that $\delta>e^{-1}\mu([w])$, where $\delta=\nu_{\eta,2L+|w|+1}([1])$ is as in Lemma 5. Then the invariant measure $\mu'_{\eta,w}$ on X, constructed in (5) above, satisfies*

$$h_T(\mu'_{\eta,w})\ge h_T(\mu)+h_\sigma(\nu_{\eta,2L+|w|+1})-\delta H_\mu(\mathcal{P}_{2L+|w|})-6\delta\log 2. \tag{6}$$

Proof. We write $\bar X$ for the space $X\times Y\times X$ and $\bar\Phi\colon X\times Y\to\bar X$ for the map $\bar\Phi(x,y)=(x,y,\Phi(x,y))$, where Φ is as defined in (4). Let $\nu=\nu_{\eta,2L+|w|+1}$, $\mu'=\mu'_{\eta,w}$, and $\bar\mu$ be the measure $\bar\Phi_*(\mu\times\nu)$ on $\bar X$. The product of the three shift maps (one in each coordinate) is denoted by $\bar T$, i.e. $\bar T=T\times\sigma\times T$. We introduce three partitions of $\bar X$:

$$\begin{aligned}\mathcal{P}_X&=\Big\{\{(x,y,z):x_0=a\}:a\in\mathcal{A}\Big\};\\ \mathcal{P}_Y&=\Big\{\{(x,y,z):y_0=\epsilon\}:\epsilon\in\{0,1\}\Big\};\text{ and}\\ \mathcal{P}_Z&=\Big\{\{(x,y,z):z_0=a\}:a\in\mathcal{A}\Big\}.\end{aligned}$$

We have the following equalities:

$$\begin{aligned}h_T(\mu')&=h_{\bar T}(\bar\mu,\mathcal{P}_Z)\\ h_{\bar T}(\bar\mu)&=h_{T\times\sigma}(\mu\times\nu)=h_T(\mu)+h_\sigma(\nu)\\ h_{\bar T}(\bar\mu)&=h_{\bar T}(\bar\mu,\mathcal{P}_Z)+h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)+h_{\bar T}(\bar\mu,\mathcal{P}_X|\mathcal{F}_Y\vee\mathcal{F}_Z),\end{aligned}$$

where $\mathcal{F}_Z=\bigvee_{j=-\infty}^{\infty}\bar T^{-j}P_Z$ with a similar definition for $\mathcal{F}_Y$; and where the second equality follows since $\bar\Phi$ is an isomorphism from $X\times Y$ equipped with the measure $\mu\times\nu$ to $\bar X$ equipped with the measure $\bar\mu$.

Combining the equalities gives

$$h_T(\mu')=h_T(\mu)+h_\sigma(\nu)-h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)-h_{\bar T}(\bar\mu,\mathcal{P}_X|\mathcal{F}_Y\vee\mathcal{F}_Z) \tag{7}$$

Hence obtaining a lower bound for $h_T(\mu')$ requires us to obtain upper bounds for $h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)$ and $h_{\bar T}(\bar\mu,\mathcal{P}_X|\mathcal{F}_Y\vee\mathcal{F}_Z)$.

Let $\delta=\nu([1])$, so that by Lemma 5, using $M=2L+|w|+1$,

$$\delta=\frac{1}{2L+|w|+1/(1-e^{-\eta})}=\eta+O(|w|\eta^2).$$

We claim that we can obtain the bound

$$h_{\bar{T}}(\bar{\mu}, \mathcal{P}_X | \mathcal{F}_Y \vee \mathcal{F}_Z) \le \delta H_\mu(\mathcal{P}_{2L+|w|}).$$

This is very intuitive in terms of information: given the y and z strings, x matches z outside blocks of length $2L + |w|$ starting at each k where $y_k = 1$. Since y is independent of x, the average amount of information in a single reconstruction is $H_\mu(\mathcal{P}_{2L+|w|})$. One then expects that $\delta H_\mu(\mathcal{P}_{2L+|w|})$ is an upper bound for $h_{\bar{T}}(\bar{\mu}, \mathcal{P}_X | \mathcal{F}_Y \vee \mathcal{F}_Z)$ since if the blocks are sequentially reconstructed, knowledge of previous reconstructions may give you some information about the current reconstruction. For a formal proof, consider the induced map on $E = \{(x, y, z) \in \bar{X} \colon y_0 = 1\}$ and note that $\bar{\mu}(E) = \delta$. Let $r_E(x, y, z) = \min\{n > 0 \colon y_n = 1\}$. We introduce countable partitions of E: $\tilde{\mathcal{P}}_X$ and $\tilde{\mathcal{P}}_Z$ where the elements of $\tilde{\mathcal{P}}_X$ are of the form

$$\left\{E \cap B_{i_0} \cap \bar{T}^{-1} B_{i_1} \cap \ldots \cap \bar{T}^{-(n-1)} B_{i_{n-1}} \cap \{\bar{x} \colon r_E(\bar{x}) = n\}\right\},$$

where n runs over $\mathbb{N}$ and $B_{i_0}, \ldots, B_{i_{n-1}}$ are elements of $\mathcal{P}_X$. We define $\tilde{\mathcal{P}}_Z$ analogously. Thus $\tilde{\mathcal{P}}_X$ and $\tilde{\mathcal{P}}_Z$ partition $\bar{X}$ according to the return time to E and the symbols in the x- and z-coordinates until that return respectively.

By Abramov's formula, we have

$$h_{\bar{T}}(\bar{\mu}, \mathcal{P}_X | \mathcal{F}_Y \vee \mathcal{F}_Z) = \delta h_{\bar{T}_E}(\bar{\mu}_E, \tilde{\mathcal{P}}_X | \tilde{\mathcal{F}}_Y \vee \tilde{\mathcal{F}}_Z). \tag{8}$$

Since $E \in \mathcal{F}_Y \vee \mathcal{F}_Z$, r_E is $(\mathcal{F}_Y \vee \mathcal{F}_Z)$-measurable.

As a partition of E, $(\mathcal{P}_X)_{2L+|w|} \vee \tilde{\mathcal{P}}_Z = \tilde{\mathcal{P}}_X \vee \tilde{\mathcal{P}}_Z$ since z_k agrees with x_k everywhere except on the $2L + |w|$ symbols following an occurrence of 1 in the y coordinate. That is $\mathcal{F}_Y \vee \mathcal{F}_Z \vee (\mathcal{P}_X)_{2L+|w|}$ is a refinement of $\tilde{\mathcal{P}}_X$. It follows that

$$\begin{aligned} h_{\bar{T}_E}(\bar{\mu}_E, \tilde{\mathcal{P}}_X | \tilde{\mathcal{F}}_Y \vee \tilde{\mathcal{F}}_Z) &\le h_{\bar{T}_E}(\bar{\mu}_E, (\mathcal{P}_X)_{2L+|w|}) \\ &\le H_{\bar{\mu}_E}((\mathcal{P}_X)_{2L+|w|}) \\ &= H_\mu(\mathcal{P}_{2L+|w|}), \end{aligned}$$

where for the last equality we used the independence of $\mathcal{F}_X$ and $\mathcal{F}_Y$. Combining this with (8), we see

$$h_{\bar{T}}(\bar{\mu}, \mathcal{P}_X | \mathcal{F}_Y \vee \mathcal{F}_Z) \le \delta H_\mu(\mathcal{P}_{2L+|w|}). \tag{9}$$

We next estimate $h_{\bar{T}}(\bar{\mu}, \mathcal{P}_Y | \mathcal{F}_Z)$. We use the hypothesis that $\delta > e^{-1}\mu([w])$. Given this, we need an estimate for $\mu'([w])$. There are three ways that the word w may appear in $z = \Phi(x, y)$: firstly if $y_{k-L} = 1$, then $z_k \ldots z_{k+|w|-1} = w$; secondly if $x_k \ldots x_{k+|w|-1} = w$ then $z_k \ldots z_{k+|w|-1}$ may also be w (unless the map Φ overwrites some of that part of x); and thirdly if a w is "inadvertently" created involving some parts of the original sequence x and some symbols that are modified by the map Φ. Since the instances of w are at least $\frac{2}{3}|w|$ apart by the no long overlaps condition, we claim there can be at most two of this third type of w for each instance of the first type. To see this, notice that the next possible w occurs at least $\frac{2m}{3}$ positions to the right of $k+L$; and the next w after that is at least $\frac{4m}{3}$ positions to the right of $k+L$. However since $L<\frac{m}{3}$, this second w lies outside the coordinate range $k,\dots,k+2L+m-1$ that is modified as a result of y_k being equal to 1. This shows that for each w inserted as a result of y_k being equal to 1, there is at most one "inadvertent" w formed to the right of the inserted w. A similar argument shows that there is at most one inadvertent w to the left of the inserted w. We see that $\mu'([w])\le\mu([w])+3\delta\le 6\delta$.

Given this, we estimate $h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)$ using the induced system of $\bar T$ with returns to $G:=\{(x,y,z): z_0\ldots z_{|w|-1}=w\}$. Note that $\bar\mu(G)=\mu'([w])$. We then let $\mathcal{Q}_Y$ be the partition of G according to the first return time map $r_G(\bar x)=\min\{n>0:\bar T^n(\bar x)\in G\}$, and certain symbols in the y sequence. Specifically $\mathcal{Q}_Y$ is the countable partition of G with elements of the form

$$G\cap\sigma^{-L}B_{i_0}\cap\sigma^{-L+1}B_{i_1}\cap\ldots\cap\sigma^{n-1-L}B_{i_{n-1}}\cap\{\bar x\in G\colon r_G(\bar x)=n\},$$

where n runs over the positive integers and the B's are elements of $\mathcal{P}_Y$. That is G is partitioned according to the return time and the y-symbols between time $-L$ and r_G-L-1. By the construction of $\bar\Phi$, if $\bar x=(x,y,z)\in G$, y_{-L} is either 1 or 0 (according to whether the copy of w was deliberately inserted or not) and $y_{-L+1},\dots,y_{r_G(\bar x)-L-1}$ are all 0. By Abramov's theorem,

$$h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)=\bar\mu(G)h_{\bar T_G}(\bar\mu_G,\mathcal{Q}_Y|\mathcal{F}_Z).$$

By the above description, since G is $\mathcal{F}_Z$-measurable, $H_{\bar\mu_G}(\mathcal{Q}_Y|\mathcal{F}_Z)\le\log 2$, so that

$$h_{\bar T}(\bar\mu,\mathcal{P}_Y|\mathcal{F}_Z)\le 6\delta\log 2. \tag{10}$$

Substituting (9) and (10) in (7) gives (6) as required. □

Lemma 7. *Let X be a mixing subshift of finite type with mixing length L and μ be an ergodic measure on X. Suppose $\phi: X\to\mathbb{R}$ is Hölder, satisfying $|\phi(x)-\phi(y)|\le c\alpha^{n(x,y)}$. Then*

$$\int\phi\,d\mu'_{\eta,w}\ge\int\phi\,d\mu+\delta\Big(S_{|w|}\phi(w)-|w|\int\phi\,d\mu\Big)-2\delta c(L+\tfrac{1}{1-\alpha}), \tag{11}$$

where $\mu'_{\eta,w}$ is as above, $\delta=\nu_{\eta,2L+|w|+1}([1])$ as in the previous lemmas and $S_{|w|}\phi(w)$ denotes $\inf_{x\in[w]}S_{|w|}\phi(x)$.

Proof. We let $\bar{X}$ and $\bar{\mu}$ be as in the proof of Lemma 6. As before, let $E = \{(x,y,z) \in \bar{X} \colon y_0 = 1\}$ and $r_E(x,y,z)$ denote the return time to E. By definition, $\bar{\mu}(E) = \delta$. We let $\bar{\mu}_E$ be the normalized induced measure on the set E. Notice that we have

$$\begin{aligned} \int \phi\, d\mu &= \delta \int_E S_{r_E(\bar{x})}\phi \circ \pi_1(\bar{x})\, d\bar{\mu}_E(\bar{x}); \quad \text{and} \\ \int \phi\, d\mu'_{\eta,w} &= \delta \int_E S_{r_E(\bar{x})}\phi \circ \pi_3(\bar{x})\, d\bar{\mu}_E(\bar{x}), \end{aligned} \tag{12}$$

where $\pi_1(x,y,z) = x$ and $\pi_3(x,y,z) = z$.

Let $m = |w|$. For $0 \le j < L$ and $L + m \le j < 2L + m$, we have

$$\phi(T^j\pi_3(\bar{x})) - \phi(T^j\pi_1(\bar{x})) \ge -c. \tag{13}$$

If $\bar{x} = (x,y,z) \in E$, then $x_L \dots x_{L+m-1} = w$, so that we have

$$\int_E \sum_{j=L}^{L+m-1} \phi(T^j\pi_3(\bar{x}))\, d\bar{\mu}_E \ge S_m\phi(w).$$

We also have

$$\begin{aligned} &\int_E \sum_{j=L}^{L+m-1} \phi(T^j\pi_1(\bar{x}))\, d\bar{\mu}_E = \int_E S_m\phi(T^L(\pi_1(\bar{x})))\, d\bar{\mu}_E \\ &= \frac{1}{\bar{\mu}(E)} \int_{\{\bar{x}\colon y_0=1\}} S_m\phi(T^L x)\, d\bar{\mu}(x,y,z) \\ &= \int S_m\phi(T^L x)\, d\mu(x) \\ &= m \int \phi(x)\, d\mu(x), \end{aligned}$$

where in the third line, we used the independence of $\pi_1(\bar{x})$ and $\pi_2(\bar{x})$. Combining the two previous facts, we see

$$\int_E \big(S_m\phi(T^L\pi_3(\bar{x})) - S_m\phi(T^L\pi_1(\bar{x}))\big)\, d\bar{\mu} \ge S_m\phi(w) - m\int \phi\, d\mu. \tag{14}$$

Finally, if $\bar{x} = (x,y,z) \in E$ and $2L + m \le j < r_E(\bar{x})$, then

$$\phi(T^j z) - \phi(T^j x) \ge -c\alpha^{\min(j-(2L+m-1), r_E(\bar{x})-j)}.$$

Summing the geometric series over this range of j's yields

$$\sum_{j=2L+m}^{r_E(\bar{x})-1} \Big(\phi(T^j\pi_3(\bar{x})) - \phi(T^j\pi_1(\bar{x}))\Big) \ge \frac{-2c}{1-\alpha}. \tag{15}$$

Combining equations (13), (14) and (15), we obtain

$$\int_E \big(S_{r_E(\bar{x})}\phi \circ \pi_3(\bar{x}) - S_{r_E(\bar{x})}\phi \circ \pi_1(\bar{x})\big)\, d\bar{\mu}_E$$
$$\geq S_m\phi(w) - m\int \phi\, d\mu - 2Lc - \frac{2c}{1-\alpha},$$

so that the claimed result follows from (12). □

Lemma 8. *Let ϕ be a Hölder potential on a mixing shift of finite type X and let μ be the corresponding Gibbs measure with constant C_ϕ. For the partition $\mathcal{P}$ and all $n \in \mathbb{N}$ we have*

$$|H_\mu(\mathcal{P}_n) - nh_T(\mu)| \leq \log C_\phi.$$

Proof. We have

$$H(\mathcal{P}_n) = -\int \log \mu(\mathcal{P}_n(x))\, d\mu(x),$$

where $\mathcal{P}_n(x)$ is the element of $\mathcal{P}_n$ containing x. By (2),

$$|\log \mu(\mathcal{P}_n(x)) - S_n\phi(x) + nP(\phi)| \leq \log C_\phi,$$

so that

$$\left| -\int \log \mu(\mathcal{P}_n(x))\, d\mu(x) - \int \Big(nP_{\text{top}}(\phi) - S_n\phi(x)\Big)\, d\mu \right| \leq \log C_\phi,$$

which yields

$$\left| H_\mu(\mathcal{P}_n) - n\left(P_{\text{top}}(\phi) - \int \phi\, d\mu\right)\right| \leq \log C_\phi.$$

Since μ is an equilibrium state for the potential ϕ, we have the equality $P_{\text{top}}(\phi) = h_T(\mu) + \int \phi\, d\mu$, so the above gives $|H_\mu(\mathcal{P}_n) - nh_T(\mu)| \leq \log C_\phi$ as claimed. □

If ϕ is a Hölder continuous potential on a mixing shift of finite type X, and μ is the equilibrium state for the potential ϕ, it is well known that the Gibbs inequality (2) is satisfied (see for example [5]) for a constant C that may depend on ϕ. In this work, we need to control the dependence of C on ϕ. For this reason, we call the constant C_ϕ. The following lemma controls the behaviour of $C_{t\phi}$ as t runs over the reals.

Lemma 9. *Let ϕ be a Hölder potential on a mixing shift of finite type X. Then there exist a and b such that $C_{t\phi} \leq e^{a+b|t|}$.*

Proof. The lemma can be established by carefully following the arguments in Ruelle's book [25, Chapter 5]. By an argument in [13], we can write the Hölder continuous function ϕ as a infinite sum of locally constant functions, $\phi = \sum_n \phi_n$ where $\phi_n(x)$ depends only on $x_0, \ldots, x_{n-1}$

and $\|\phi_n\|_\infty$ decays exponentially. This, in turn, allows us to build an *interaction* in the language of statistical physics (see [25, Section 3.2]), so that the equilibrium state for $t\phi$ is also the Gibbs measure for the corresponding interaction $(t\Phi_{[j,k]})_{[j,k]\subset\mathbb{Z}}$. Based on this interaction, Ruelle [25, Section 5.12] writes down an explicit leading eigenfunction h_t (denoted by $\psi_>$ by Ruelle) for the Perron-Frobenius operator $\mathcal{L}_t$ defined by $\mathcal{L}_t f(x) = \sum_{y\in T^{-1}x} e^{t\phi(y)} f(y)$. One can check from the expression for h_t that it satisfies a bound of the form $\|\log h_t\|_\alpha \le a + b|t|$. If one sets $\psi_t = t\phi + \log h_t - \log h_t \circ T - \log \lambda_t$, then ψ_t is normalized in the sense that

$$\sum_{y\in T^{-1}x} e^{\psi_t(y)} = 1$$

for all $x \in X$. This implies that $\mathcal{L}^*_{\psi_t}\mu_t = \mu_t$ and $\mathcal{L}_{\psi_t}\mathbf{1} = \mathbf{1}$. The equilibrium measure μ_t then satisfies

$$\mu_t([u]) = \int \exp(S_n\psi_t(ux))\, d\mu_t(x),$$

where ux denotes the concatenation of the word u with the infinite string x and $S_n\psi_t(ux)$ is declared to be $-\infty$ if the transition from the last symbol of u to the first symbol of x is not allowed in X. From this, it follows that $C_t \le \exp(\sum_{n=0}^\infty \operatorname{var}_n(\psi_t))$, where

$$\operatorname{var}_n(\psi_t) = \sup\{|\psi_t(x) - \psi_t(y)| : x_0 \dots x_{n-1} = y_0 \dots y_{n-1}\}.$$

The above estimates establish a bound of the form $C_t \le \exp(a|t| + b)$ as required. □

4. Proofs of Main theorems

We now turn to the proofs of the main theorems, which mostly consist of estimating quantities of the form $h_T(\mu') + s\int \phi\, d\mu'$ from below where μ' is one of the measures built in the previous section (and μ is μ_∞ in the case of Theorem 1 or μ_t in the case of Theorem 4). This then gives a lower bound for $p_\phi(s)$. To bound $h_T(\mu') + s\int \phi\, d\mu'$ from below, we rely on bounds from the previous section, showing that $h_T(\mu')$ exceeds $h_T(\mu)$ plus a term of order $-\eta\log\eta$ and that $s\int\phi\, d\mu'$ is at least $s\int \phi\, d\mu$ minus a term of order $s\eta$). If η is taken to be less than e^{-Cs} for a suitable C, the gain dominates the loss by an amount of order η.

We restate Theorem 1 for convenience.

Theorem. *Let (X,T) be a mixing subshift of finite type with positive entropy. Let ϕ be a Hölder potential that is not cohomologous to a constant. Then there exist C and t_0 such that $p_\phi(t) \ge \ell_\infty(t) + e^{-Ct}$ for all $t \ge t_0$, where $\ell_\infty(t)$ is the asymptote to p_ϕ at infinity.*

Proof. By [4], there exists a Hölder continuous function ψ that is cohomologous to ϕ such that $A(\phi) \leq \psi(x) \leq B(\phi)$ for all $x \in X$, where $A(\phi) = \min_{\nu \in \mathcal{M}_T(X)} \int \phi \, d\nu$ and $B(\phi) = \max_{\nu \in \mathcal{M}_T(X)} \int \phi \, d\nu$. The assumption that ϕ is not cohomologous to a constant implies that $A(\phi) \neq B(\phi)$. Since the pressure functions of ϕ and ψ coincide, we derive the required estimate for $p_\psi(t)$.

Let μ_∞ be a measure achieving the slant asymptote ℓ_∞, so that the support of μ_∞ is contained in the proper subset of X:

$$S_{\max} := \{x \in X : \psi(x) = B(\phi)\}.$$

As before, denote by L the mixing length of X. Let $[w]$ be a cylinder set lying in the complement of $\operatorname{supp}(\mu_\infty)$ where w is a word of length at least $3L$ with no long overlaps. Such a word always exists, see e.g. [16, Theorem 8.3.9]. Let $M = |w| + 2L + 1$. We then equip $Y = \{0,1\}^{\mathbb{Z}}$ with the measure $\nu = \nu_{\eta,M}$ constructed in the previous section and use ν to build a measure $\mu'_{\eta,w}$ as in (5) (where μ is taken to be the measure μ_∞).

We see from Lemma 7 that

$$\int \psi \, d\mu'_{\eta,w} \geq \int \psi \, d\mu_\infty - 2\delta c \left(L + \frac{1}{1-\alpha} \right) - \delta c |w|,$$

where c and α are constants in the Hölder condition for ψ. Since $m = |w|$ is fixed, and $\delta = \left(2L + m + (1 - e^{-\eta})^{-1}\right)^{-1} = O(\eta)$, we see that there exists c_1 such that

$$\int \psi \, d\mu'_{\eta,w} \geq \int \psi \, d\mu_\infty - c_1 \eta.$$

Since $\mu_\infty([w]) = 0$ the estimate derived in Lemma 6 works for any $\eta \in (0,1)$. Combining it with the entropy formula for measure $\nu_{\eta,M}$ from Lemma 5 we conclude that there is a second constant c_2 such that

$$h_T(\mu'_{\eta,w}) \geq h_T(\mu_\infty) - \eta \log \eta - c_2 \eta.$$

It follows that

$$\begin{aligned} P_{\text{top}}(t\psi) &\geq h_T(\mu'_{\eta,w}) + t \int \psi \, d\mu'_{\eta,w} \\ &\geq \left(h_T(\mu_\infty) + t \int \phi \, d\mu_\infty \right) - \eta \log \eta - (c_2 + c_1 |t|) \eta. \end{aligned}$$

Since this bound holds for all small values of η, we substitute $\eta = \exp(-1 - c_2 - c_1|t|)$ and deduce that $p_\phi(t) = p_\psi(t) \geq \ell_\infty(t) + e^{-(c_2+1)} e^{-c_1|t|}$ as required for large t. □

For locally constant functions, the true gap between $p_\phi(t)$ and $\ell_\infty(t)$ is asymptotically exponential, matching the form of the lower bound in the previous theorem:

Example 10. *Let (X,T) be the full two-sided shift on the alphabet $\{1,...,k\}$ and $\phi : X \to \mathbb{R}$ be a potential which is constant on cylinders of length 1, i.e. $\phi(x) = c_{x_0}$ where $c_1,\dots,c_k$ are fixed real numbers. Then $p_\phi(t) = \log(e^{c_1 t} + \dots + e^{c_k t})$.*

In the case $c_i > \max_{j\neq i} c_j$, $p_\phi(t) = c_i t + O(e^{-\Delta t})$ where $\Delta = c_i - \max_{j\neq i} c_j$.

We now restate Theorem 4.

Theorem. *Let (X,T) be a mixing subshift of finite type with positive entropy. Let ϕ be a Hölder potential that is not cohomologous to a constant. Then there exist c_1 and c_2 such that for any $t \in \mathbb{R}$, and any $h \in \mathbb{R}^+$,*

$$p_\phi(t+h) + p_\phi(t-h) - 2p_\phi(t) > c_1 e^{-c_2|t|/h}.$$

Proof of Theorem 4. Let $t \in \mathbb{R}$ be given and write $\ell_t(s) = h(\mu_t) + s\int \phi\, d\mu_t$, where μ_t denotes the (unique) equilibrium state for the Hölder continuous potential $t\phi$. Using the facts that $\ell_t(t) = p_\phi(t)$ and that ℓ_t is linear, we see that the desired inequality is equivalent to

$$(p_\phi(t+h) - \ell_t(t+h)) + (p_\phi(t-h) - \ell_t(t-h)) \geq c_1 e^{-c_2 t/h},$$

which is equivalent to showing the existence of c_1 and c_2 such that

$$\max\big(g_t(t+h), g_t(t-h)\big) \geq c_1 e^{-c_2 t/h}, \tag{16}$$

where $g_t(s)$ is the 'gap' function $p_\phi(s) - \ell_t(s)$. Note that $g_t(s)$ is strictly positive for all $s \neq t$, since ℓ_t is a tangent line to p_ϕ and p_ϕ is strictly convex.

By the strict convexity of p_ϕ, and using the fact that $p'_\phi(t) = \int \phi\, d\mu_t$, we see that $\int \phi\, d\mu_t$ is a strictly increasing function taking values in the range $(A(\phi), B(\phi))$ for all t, and tending to $A(\phi)$ as $t \to -\infty$ and $B(\phi)$ as $t \to \infty$. Let $\gamma(\phi) = B(\phi) - A(\phi)$.

As in the proof of Theorem 1 there exists a Hölder continuous function f such that $\psi = \phi + f\circ T - f$ takes values in the range $[A(\phi), B(\phi)]$ where $A(\phi) = \min_{\mu\in\mathcal{M}_T(X)} \int \phi\, d\mu$ and $B(\phi) = \max_{\mu\in\mathcal{M}_T(X)} \int \phi\, d\mu$. It follows that there exist (disjoint) subshifts X_α and X_β on which ψ takes the constant values $A(\phi)$ and $B(\phi)$ respectively. Since the pressure functions of ϕ and ψ coincide, we work with $p_\psi(t)$.

A word w of length m is called *heavy* if $S_m\psi(x) > m(B(\phi) - \frac{1}{4}\gamma(\phi))$ for all $x \in [w]$ and similarly a word w is *light* if $S_m\psi(x) < m(A(\phi) + \frac{1}{4}\gamma(\phi))$ for all $x \in [w]$. All sufficiently long words in the languages of

X_α or X_β are light and heavy respectively by the Hölder continuity of ψ. In particular, one may find a pair of heavy words u_h and v_h of the same length such that arbitrary concatenations of u_h and v_h are legal in X and so that u_h does not appear as a sub-word of the infinite concatenation of v_h's and v_h does not appear as a sub-word of the infinite concatenation of u_h's. Likewise there exist light u_l and v_l with the analogous properties. By Lemma 9, there exists $C_1 > 0$ such that for all words w and all t, $\mu_t[w]$ agrees with $e^{t(S_{|w|}\psi(x)-p_\psi(t))}$ up to a multiplicative factor in the range $e^{-C_1(1+|t|)}$ to $e^{C_1(1+|t|)}$ where x is an arbitrary point of $[w]$.

We first deal with the case where $\int \psi\, d\mu_t \le \frac{1}{2}(A(\phi)+B(\phi))$. We let w_k be a heavy block of the form $u_h^k v_h^k$ and build measures μ'_{η,w_k} by editing realizations of $\mu := \mu_t$ as in Section 3. We then tune k and η and use the results of the previous section to give a lower bound for $g_t(t+h)$. By the properties of u_h and v_h, any two occurrences of w_k are separated by a gap of at least $|w_k| - \max(|u_h|,|v_h|)$.

As before, let $m = |w_k|$ and $\delta = 1/(2L+m+(1-e^{-\eta})^{-1})$ be the frequency of 1's in realizations of $\nu_{\eta,2L+m+1}$. Recall that $\delta = \eta + O(\eta^2|w_k|)$. Provided that $k \ge 2$ and $2k|u_h| \ge 3L$, the word w_k has no long overlaps. Applying Lemma 6, with $\mu = \mu_t$, provided η is large enough that $\delta > e^{-1}\mu([w_k])$, we have

$$h_T(\mu'_{\eta,w_k}) \ge h_T(\mu_t) + \delta\big(-\log\delta - 6\log 2 - H_{\mu_t}(\mathcal{P}_{2L+m})\big).$$

By Lemmas 8 and 9, $H_{\mu_t}(\mathcal{P}_{2L+m}) \le (2L+m)h_T(\mu_t) + a + b|t|$ for constants a and b that don't depend on t. Substituting into the previous equation, and using the facts that $2Lh_T(\mu_t)$ is uniformly bounded as t varies and that $\delta = \eta + O(\eta^2 m)$, we obtain a bound of the form

$$h_T(\mu'_{\eta,w_k}) \ge h_T(\mu_t) + \eta\big(-\log\eta - m\,h_T(\mu_t) - a - b|t|\big),$$

where a and b are independent of t.

We also require a bound on $(t+h)\int\psi\,d\mu'$. By Lemma 7, we have

(17)
$$\begin{aligned}(t+h)\int\psi\,d\mu' \ge{}& (t+h)\int\psi\,d\mu_t\\ &+ \delta\cdot(t+h)\left(S_m\psi(w) - m\int\psi\,d\mu_t\right) - c(|t|+h)\delta,\end{aligned}$$

where $S_m\psi(w) = \min_{x\in[w]} S_m\psi(x)$.

Combining the above two inequalities, we obtain

$$\begin{aligned}
&h_T(\mu') + (t+h)\int \psi\, d\mu' \\
&\geq \ell_t(t+h) - \eta\log\eta \\
&\quad + \eta\cdot\left(S_m(t\psi)[w] - m\left(h(\mu_t) + \int t\psi\, d\mu_t\right)\right) \\
&\quad + \eta h\cdot\left(S_m(\psi)[w] - m\int \psi\, d\mu_t\right) - \eta\cdot(c|t| + ch + d) \\
&\geq \ell_t(t+h) - \eta\log\eta + \eta\log\mu_t([w]) + \tfrac{1}{4}\eta hm\gamma(\psi) \\
&\quad - \eta(c|t| + ch + d),
\end{aligned} \tag{18}$$

where we used the Gibbs condition and Lemma 9 in the last line; and the constants c and d vary from line to line (and are not the same as the c above) as more error terms are combined, but where none of these constants depend on t.

One can check using elementary calculus that provided $a < 1$, the function $g\colon \eta \mapsto -\eta\log\eta + a\eta$ takes its maximum value as η ranges over $[0,1]$ at $\eta^* = e^{a-1}$ and $g(\eta^*) = \eta^*$.

We then choose m, the length of the word $|w|$, so that $\frac{1}{4}mh\gamma(\phi) > c|t| + ch + d$. Note that since w is of the form $u^k v^k$, the length of w can be specified up to an additive constant no more than $|u| + |v|$. Once this choice is made, by (18), we have for any $\eta \in [0,1]$, by taking μ' to be the measure obtained from w and η as in (5), that

$$h_T(\mu') + (t+h)\int \psi\, d\mu' \geq \ell_t(t+h) - \eta\log\eta + a\eta,$$

where $a \geq \log\mu_t([w])$. This ensures that taking η to be the argument maximizing the above bound, namely $\eta = e^{a-1}$, then $\eta > e^{-1}\mu_t([w])$, which was the condition that was imposed above. This gives

$$P_{\text{top}}((t+h)\psi) \geq h_T(\mu') + (t+h)\int \psi\, d\mu' \geq \ell_t(t+h) + e^{-1}\mu_t([w]).$$

By the choice of $|w|$ above, and using (2) with the bounds in Lemma 9, we obtain the required lower bound for $g_t(t+h)$.

Turning to the case where $\int \psi\, d\mu \geq \frac{1}{2}(A(\phi) + B(\phi))$, we argue similarly to obtain a lower bound for $g_t(t-h)$. More specifically, we form a block $w = u_l^k v_l^k$ that is a concatenation of light words and we build the measure μ' as before by (5).

The derivation of

$$h_T(\mu') \geq h(\mu_t) - \eta\log\eta - \eta\big(|w|h(\mu_t) + a + b|t|\big)$$

proceeds exactly as before (Lemma 6). The analogue of (17) is

$$(t-h)\int \psi\, d\mu' \geq (t-h)\int \phi\, d\mu_t$$
$$+\eta\cdot(t-h)\Big(S_m\phi(w) - m\int \phi\, d\mu_t\Big) - c(|t|+|h|)\eta.$$

Combining these inequalities similarly to the above computation gives for any $\eta \in (0,1)$,

$$g_t(t-h) \geq -\eta\log\eta + \eta\log\mu_0([w]) + h\eta m^{\frac{\gamma}{4}} - \eta\big(c(|t|+h)+d\big),$$

where $m = |w| = k(|u_l|+|v_l|)$. As before, we choose k to ensure that $h\eta m^{\frac{\gamma}{4}} - \eta\big(c(|t|+h)+d\big) > 0$. Then the η giving the largest lower bound for $g_t(t-h)$ satisfies $\eta > e^{-1}\mu_t([w])$, and as before we obtain $g_t(t-h) \geq e^{-1}\mu_t([w])$. By the choice of k, we deduce a bound of the form in the statement of the theorem. □

5. Arbitrarily slow convergence

So far we have proven that the pressure function cannot approach its asymptote "too fast" by providing an exponential lower bound on the gap. In this section we show the non-existence of a corresponding upper bound. For a given mixing subshift of finite type we construct potentials for which the convergence of the pressure to the asymptote is arbitrarily slow, thus proving Theorem 2.

First we construct a subshift Y of a given entropy where a fixed word u appears with constant frequency. For that we make use of β-shifts. For $\beta > 1$ the β-shift is defined as the smallest two-sided subshift of $\{0, ..., \lceil\beta\rceil - 1\}^{\mathbb{Z}}$ which contains all sequences of the coefficients in β-expansions of real numbers in $[0,1)$, see e.g. [18]. It is well known that the entropy of the β-shift is $\log\beta$ and there is a unique measure of maximal entropy which is fully supported.

Lemma 11. *Let (X,T) be a mixing subshift of finite type with positive entropy, let $0 \leq b < h_{\text{top}}(X)$ and let $u \in \mathcal{L}(X)$. There exists an $\ell > 0$ and a subshift Y with the following properties:*

(1) *The alphabet of Y is a subset those words in $\mathcal{L}_\ell(X)$ of the form uv such that $vu \in \mathcal{L}_\ell(X)$;*
(2) *$h_{top}(Y) = b\ell$.*

The first condition ensures that the symbols appearing in the shift base of Y may be concatenated to give a point of X (with u's occurring every ℓ steps); the second condition will ensure that the subshift of X formed from these concatenations has topological entropy b.

Proof. Since $b < h_{\text{top}}(X)$, and $\#\mathcal{L}_k(X) \geq e^{k\,h_{\text{top}}(X)}$ for all k, there exists an n such that $\#\mathcal{L}_n(X) > e^{b(n+|u|+2L)}$ where L is the mixing length of X. Let $\ell = n + |u| + 2L$ and set $N = \lceil e^{\ell b} \rceil$ (so that $N \leq \#\mathcal{L}_n(X)$). Enumerate a subcollection of N elements of $\mathcal{L}_n(X)$ as $v_0, \dots, v_{N-1}$. By the mixing condition, there exist words $p_1, \dots, p_N$ and $q_1, \dots, q_N$ in $\mathcal{L}_L(X)$ such that $w_i := up_iv_iq_i \in \mathcal{L}_\ell(X)$ and $q_iu \in \mathcal{L}_{L+|u|}(X)$ for each i. Let $\mathcal{W} = \{w_0, \dots, w_{N-1}\}$. By construction, these words are distinct. Let B denote the standard β-shift on the alphabet $\{0, \dots, N-1\}$ where $\beta = e^{b\ell}$. As was mentioned above that $h_{\text{top}}(B) = \log \beta = b\ell$. We then let Y be the image of B under the bijective one-block map $\theta\colon B \to \mathcal{W}^{\mathbb{Z}}$ defined by $\theta(b)_j = w_{b_j}$, so that Y satisfies the conditions in the statement of the lemma. □

In the next lemma we describe an auxiliary shift $Z_\infty \subset \{0,1\}^{\mathbb{Z}}$ required for our construction. A shift of this type was used by Rothstein in [22] to produce an example of a loosely Bernoulli process which is not very weak Bernoulli. Hence, we will refer to Z_∞ as the Rothstein shift.

Lemma 12. *Given an increasing sequence of positive integers $(n_j)_{j=1}^{\infty}$ there is a sequence of nested subshifts Z_j of $\big(\{0,1\}^{\mathbb{Z}}, \sigma\big)$ and a shift $Z_\infty = \bigcap Z_j$ satisfying*

- $h_{\text{top}}(Z_j) = \dfrac{\log 2}{2^j}$ *(so that $h_{\text{top}}(Z_\infty) = 0$);*
- *For $j \in \mathbb{N}$ and $z \in Z_j$ we have $d(z, Z_\infty) < 2^{-n_{j+1}}$.*

Proof. For a set of words W on alphabet $\{0,1\}$ denote by W^n the set of all possible concatenations of n elements of W. We define a sequence (W_j) of sets of words inductively. Let $(n_j)_{j=1}^{\infty}$ be an increasing sequence of positive integers. We start with $W_0 = \{0,1\}$, which are two words of length 1. We set

$$W_{j+1} = \left\{ww : w \in W_j^{n_{j+1}}\right\} \text{ for } j = 0, 1, 2...$$

Denote by l_j the length of words in W_j. Then $l_0 = 1$ and one can easily check that $l_j = 2^j n_1 ... n_j$. Also, since W_0 consists of two words, we see that $|W_1|$ has 2^{n_1} words, and the number of words in W_j is

$$|W_j| = |W_{j-1}|^{n_j} = |W_{j-2}|^{n_{j-1}n_j} = ... = 2^{n_1 n_2 ... n_j}.$$

Let Z_j be the subshift which consists of all possible concatenations of words from W_j. Since there are $|W_j|^{l_j}$ words of length nl_j in Z_j, we compute

(19)
$$h_{\text{top}}(Z_j) = \lim_{n\to\infty} \frac{\log(l_j 2^{n(n_1 ... n_j)})}{nl_j} = \frac{n_1 ... n_j \log 2}{l_j} = \frac{n_1 ... n_j \log 2}{2^j n_1 ... n_j} = \frac{\log 2}{2^j}.$$

The Rothstein shift Z_∞ is defined as the set of all elements $x \in \{0,1\}^{\mathbb{Z}}$ such that every finite word in x is a subword of $w \in W_j$ for some j. It follows from (19) that $h_{\text{top}}(Z_\infty) = 0$.

We need an estimate on the distance between points in Z_j and the shift Z. Suppose $z \in Z_j$ for some $j \in \mathbb{N}$. We claim that for $n = \frac{1}{2}(l_{j+1} - l_j)$ the word $z_{-n} \dots z_{n-1}$ is a subword of a word in W_{j+2}. Indeed, $z_{-n} \dots z_{n-1}$ has length $(2n_{j+1}-1)l_j$ and is in the language of Z_j. Therefore, there are words $w_i \in W_j$ with $i = -n_{j+1}, ..., n_{j+1} - 1$ such that $z_{-n}...z_{n-1}$ is a subword of their concatenation $w_{-n_{j+1}} \cdots w_{n_{j+1}-1}$. Denote $u = w_{-n_{j+1}} \cdots w_{-1}$ and $v = w_0 \cdots w_{n_{j+1}-1}$, so that $z_{-n}...z_{n-1}$ is a subword of uv. Then by definition of W_{j+1}, $uu, vv \in W_{j+1}$ and hence $uuvv = (uu)(vv)$ is a subword of a word in W_{j+2} as a concatenation of two words in W_{j+1}. It follows that $z_{-n} \dots z_{n-1}$ is a subword of a word in W_{j+2} and hence a subword of a word in W_k for all $k \geq j+2$. It follows that $z_{-n} \dots z_{n-1} \in \mathcal{L}(Z)$. We deduce

$$d(z, Z) \leq 2^{-\frac{l_{j+1}-l_j}{2}}, \text{ whenever } z \in Z_j. \tag{20}$$

Recall that $l_j = 2^j n_1..n_j$, so that $\frac{1}{2}(l_{j+1} - l_j) = 2^{j-1}n_1...n_j(2n_{j+1} - 1) \geq n_{j+1}$. It follows from above that for $z \in Z_j$ we have $d(z, Z) \leq 2^{-n_{j+1}}$, as required. □

We are now ready to provide the proof of Theorem 2. We remind the reader of the statement.

Theorem. *Suppose (X, T) is any mixing subshift of finite type with positive entropy and $f : \mathbb{R} \to \mathbb{R}$ is any convex function asymptotic to a line $\ell_\infty(t) := at + b$, where $0 \leq b < h_{\text{top}}(X)$. Then there exists a Hölder potential $\phi : X \to \mathbb{R}$ such that*

(1) $p_\phi(t)$ is asymptotic to the line $\ell_\infty(t) = at + b$ as $t \to \infty$;
(2) $p_\phi(t) > f(t)$ for all sufficiently large t.

Proof. We observe that for any function ϕ, $p_{a+\phi}(t) = at + p_\phi(t)$. Hence it suffices to show that if f is any convex function asymptotic to a constant b such that $0 \leq b < h_{\text{top}}(X)$ then there exists a Hölder potential ϕ such that p_ϕ is asymptotic to b and $p_\phi(t) > f(t)$ for all sufficiently large b.

Since $h_{\text{top}}(X) > 0$, we may pick two distinct words u, v in $\mathcal{L}(X)$ of the same length with the same first and last symbols. Let ℓ and Y be as constructed in Lemma 11 based on the word u, so that Y is a subshift of topological entropy $b\ell$ whose alphabet consists of a subset of $\mathcal{L}_\ell(X)$ with each word in the alphabet starting with u.

We next construct a suitable Rothstein shift. Since we assume that $f(t)$ is asymptotic to the constant b, we may pick an increasing sequence of real numbers (t_j) such that for $t \geq t_j$ we have $f(t) < b + \ell^{-1}2^{-(j+2)}$. Set $n_j = \lceil \log(\ell t_j) \rceil + j$. By Lemma 12 there is a nested family of subshifts (Z_j) on the alphabet $\{0,1\}$ such that $h_{\mathrm{top}}(Z_j) = 2^{-j}\log 2$ so that $Z_\infty := \bigcap Z_j$ has $h_{\mathrm{top}}(Z_\infty) = 0$. Further $d(z, Z_\infty) \leq 2^{-n_{j+1}}$ for all $z \in Z_j$. Denote by λ_j the measure of maximal entropy for Z_j and by λ_∞ any invariant measure supported on Z_∞ (necessarily of entropy zero).

We now let $\bar{Y}_j = Y \times Z_j \times \{0, \ldots, \ell - 1\}$ and consider the constant height ℓ suspension over the product $Y \times Z_j$. That is we consider the map

$$\bar{T}(y, z, i) = \begin{cases} (Ty, Tz, 0) & \text{if } i = \ell - 1; \\ (y, z, i+1) & \text{otherwise.} \end{cases} \tag{21}$$

Clearly $T \times T$ acting on $Y \times Z_j$ has topological entropy $b\ell + 2^{-j}\log 2$ so that $h_{\mathrm{top}}(\bar{Y}_j) = b + \log 2/(2^j \ell)$ and $h_{\mathrm{top}}(\bar{Y}_\infty) = b$. Let $\bar{\nu}_j$ be the invariant measure $\zeta \times \lambda_j \times c$ where ζ is the measure of maximal entropy on Y and c is the normalized counting measure on $\{0, \ldots, \ell - 1\}$ so that $\bar{\nu}_j$ is supported on $\bar{Y}_j$. Notice that $\bar{T}^\ell(y, z, i) = (Ty, Tz, i)$ so that $h_{\bar{T}^l}(\bar{\nu}_j) = h_T(\zeta) + h_T(\nu_j) = b\ell + 2^{-j}\log 2$. It follows that $h_{\bar{T}}(\bar{\nu}_j) = b + \log 2/(2^j \ell)$ so that $\bar{\nu}_j$ is a measure of maximal entropy on $\bar{Y}_j$.

If the alphabet of Y is $\mathcal{W} = \{uw_0, \ldots, uw_{N-1}\}$, it is convenient to enlarge it to $\overline{\mathcal{W}} = \mathcal{W} \cup \{vw_0, \ldots, vw_{N-1}\}$. We now define a map Φ from the $\bar{Y}_j$'s to X by requiring that $\Phi(y, z, i)_n = \Phi(\bar{T}^n(y, z, i))_0$ and specifying $\Phi(y, z, i)_0$:

$$\Phi(y, z, i)_0 = \begin{cases} (y_0)_i & \text{if } i \geq |u| \text{ or } z_0 = 0; \\ v_i & \text{if } i < |u| \text{ and } z_0 = 1. \end{cases}$$

A simple description of Φ is that it concatenates the words forming the symbols in y, replacing the initial u's with v's in those coordinates where z has a 1. To see that $\Phi(y, z, i)$ lies in X, recall that an arbitrary finite concatenation of elements of $\mathcal{W}$ lies in $\mathcal{L}(X)$. Since $v \in \mathcal{L}(X)$ and has the same first and last symbols as u, replacing any number of u's by v's in an element of $\mathcal{L}(X)$ gives another element of $\mathcal{L}(X)$. It follows that arbitrary finite concatenations of elements of $\overline{\mathcal{W}}$ lie in $\mathcal{L}(X)$.

We let $X_j = \Phi(\bar{Y}_j)$ for all $1 \leq j \leq \infty$ and $\mu_j = \bar{\nu}_j \circ \Phi^{-1}$. We now claim that Φ is at most ℓ-to-one and hence entropy preserving. We will write Φ as a composition $\kappa \circ \Psi$. Firstly we introduce the space $\overline{\mathcal{W}}^{\mathbb{Z}} \times \{0, \ldots, \ell - 1\}$ with the suspension map $\bar{T}$ defined exactly like (21). Then $\Psi : \bar{Y}_j \to \overline{\mathcal{W}}^{\mathbb{Z}} \times \{0, \ldots, \ell - 1\}$ is defined by $\Psi(y, z, i) =$

$(s(y_n, z_n)_{n\in\mathbb{Z}}, i)$, where

$$s(y,z)_n = \begin{cases} y_n & \text{if } z_n = 0; \\ v(y_n)_{|u|} \dots (y_n)_{\ell-1} & \text{if } z_n = 1. \end{cases}$$

That is, s replaces the initial u's with v's in those coordinates where $z_n = 1$. This mapping is invertible as one can infer the sequence z from looking at $s(y,z)$. Secondly, the map $\kappa\colon \overline{\mathcal{W}}^{\mathbb{Z}} \times \{0,\dots,\ell-1\} \to X$ is a concatenation map. $\kappa(y,i)_n = \kappa(\bar{T}^n(y,i))_0$ and $\kappa(y,i)_0 = (y_0)_i$. The map κ is at most ℓ-to-1 as given $\kappa(y,i)$ and the value of i, one may recover y. Since Φ is entropy-preserving, we see

$$h_T(\mu_j) = h_{\bar{T}}(\nu_j) = b + \frac{\log 2}{2^j \ell} \text{ for all } 1 \le j \le \infty. \tag{22}$$

To finish the argument, we define $\phi(x) = -d(x, X_\infty)$. Clearly any maximizing measure μ_∞ is supported on X_∞ so that $\int \phi \, d\mu_\infty = 0$ and $h_T(\mu_\infty) = h_{\text{top}}(X_\infty) = h_{\text{top}}(\bar{Y}_\infty) = b$. This ensures that $p_\phi(t)$ is asymptotic to b as required.

Let $x \in X_j$ so that $x = \Phi(y,z,i)$ with $y \in Y$, $z \in Z_j$ and $0 \le i < \ell$. By Lemma 12, there exists $z' \in Z_\infty$ with $d(z,z') < 2^{-n_{j+1}}$. By definition $\Phi(y,z',i) \in X_\infty$. From the definition of Φ, one sees that $x_n = \Phi(y,z,i)_n = \Phi(y,z',i)_n$ for all $|n| < n_{j+1}\ell$ so that

$$d(x, X_\infty) \le 2^{-n_{j+1}\ell} \text{ for all } x \in X_j. \tag{23}$$

We claim that for $t_j \le t \le t_{j+1}$, $P(t\phi) \ge f(t)$. Combining (22) and (23) for $t_j \le t \le t_{j+1}$ we obtain

$$\begin{aligned} P(t\phi) &\ge h_T(\mu_j) + t\int_X \phi \, d\mu_j \\ &= b + \frac{1}{\ell 2^j} - t\int_{S_j} \text{dist}(x,S)\, d\mu_j \\ &\ge b + \frac{\log 2}{\ell 2^j} - \frac{t_{j+1}}{\ell t_{j+1} 2^{j+2}} \\ &\ge b + \frac{1}{\ell 2^{j+2}} \\ &\ge f(t). \end{aligned}$$

□

6. Genericity of upper bounds

First we give examples of two potentials on the full shift on two symbols for which the behaviour of the pressure functions is very different even though they share the same maximizing measure.

Example 13. *There is a potential ϕ on the full shift $\{0,1\}^{\mathbb{Z}}$ satisfying*

- *the unique ground state of ϕ is a point-mass measure at $\bar{0}$;*
- $p_\phi(t) - \ell_\infty(t) \sim e^{-t}$.

Proof. Let $X = \{0,1\}^{\mathbb{Z}}$ and let $\phi(x) = -x_0$. Then the unique maximizing measure is δ_0, the measure supported on the fixed point $\bar{0}$ so that $h(\delta_0) = 0$ and $\int \phi\, d\delta_0 = 0$. Hence $\ell_\infty(t) = 0$. We see from Example 10 that $P_{\mathrm{top}}(t\phi) = \log(1 + e^{-t}) = e^{-t} + O(e^{-2t})$. In particular, $p_\phi(t)$ approaches $\ell_\infty(t)$ exponentially fast. □

Example 14. *There is a potential ϕ on the full shift $\{0,1\}^{\mathbb{Z}}$ satisfying*

- *the unique ground state of ϕ is a point-mass measure at $\bar{0}$;*
- $p_\phi(t) - \ell_\infty(t) \gtrsim \log\log t/\log t$.

Proof. Let $X = \{0,1\}^{\mathbb{Z}}$ and let

$$S = \{x \in X : x \text{ has at most one 1 symbol}\}.$$

Then S is a countable closed subshift of X, sometimes known as the "sunny side up" system. Let $\phi(x) = -d(x, S)$. Clearly the only invariant measure with support lying in S is δ_0, so that $\ell_\infty(t) = 0$ again. We show, however, that in this case $p_\phi(t)$ does not approach ℓ_∞ exponentially fast. Indeed let μ_n be the measure on X where gaps are uniform and equally likely in the range $2n, \ldots, 3n$. Then a calculation using Abramov's theorem shows that $h(\mu_n) = (\log(n+1))/(\frac{5n}{2})$ (one chooses between $n+1$ equally likely choices on average once every $\frac{5n}{2}$ steps). For μ_n-a.e. x, $x_{-(n-1)}, \ldots x_{n-1}$ contains at most one 1, so that $d(x, S) \le 2^{-n}$. It follows that $\int \phi\, d\mu_n \ge -2^{-n}$. We now have the lower bound for $p_\phi(t) = P(t\phi) \ge \max_n(\frac{2}{5}\log n/n - t2^{-n})$. Taking $n = \lceil \log_2 t + \log_2(\log_2 t)\rceil$, we see that for large t, $p_\phi(t) \ge \frac{1}{5}\log(\log_2 t)/\log_2 t - 1/(\log_2 t)$. So that $p_\phi(t) - \ell_\infty(t)$ converges to zero *much* slower than exponentially. □

We conclude the paper with the proof of Theorem 3, which we restate for convenience.

Theorem. *Let X be a mixing shift of finite type and for $\alpha < 1$, let $\mathcal{H}_\alpha$ denote the set of potentials on X for which there is a c such that $|\phi(x) - \phi(y)| \le c\alpha^{n(x,y)}$ for all $x, y \in X$. Then there is a dense open subset $\mathcal{U}$ of potentials in $\mathcal{H}_\alpha$ such that for all $\phi \in \mathcal{U}$, there exist $C_1 > 0$ and $C_2 > 0$ such that $p_\phi(t) - \ell_\infty(t) \le C_1 e^{-C_2 t}$ for all t.*

Proof. Note that $\mathcal{H}_\alpha$ is the class of functions that is Lipschitz with respect to d_α where $d_\alpha(x, y) = \alpha^{\min\{|n| : x_n \ne y_n\}}$. Let Per denote the collection of periodic points in X. For a point $p \in \mathsf{Per}$, let $\mathcal{H}_\alpha(p)$ denote the set potentials ϕ in $\mathcal{H}_\alpha$ such that $\int \phi\, d\mu_p \le \int \phi\, d\nu$ for all invariant measures $\nu \neq \mu_p$ (where μ_p denotes the unique invariant measure supported on the orbit of p).

Recall that by a theorem of Contreras [10], $\bigcup_{p\in\mathsf{Per}}\mathcal{H}_\alpha(p)$ contains a dense open subset of $\mathcal{H}_\alpha$. For a periodic point p, let

$$d_p(x) = \min_j d_\alpha(x, T^j p).$$

The set $\mathcal{U}$ that we consider is defined by

$$\mathcal{U} = \bigcup_{p\in\mathsf{Per}} \{\psi - ad_p \colon \psi \in \mathcal{H}_\alpha(p), a > 0\}.$$

Since $\bigcup_{p\in\mathsf{Per}}\mathcal{H}_\alpha(p)$ is in the closure of $\mathcal{U}$, we see that $\mathcal{U}$ is dense. That $\mathcal{U}$ is open follows from the proof of Proposition 4.1 in Yuan and Hunt [29].

We now let $\phi = \psi - ad_p$ where $p \in \mathsf{Per}$, $\psi \in \mathcal{H}_\alpha(p)$ and $a > 0$. Since $a > 0$, we see that $\int \phi\, d\nu < \int \phi\, d\mu_p$ for all invariant measures $\nu \neq \mu_p$. It follows that $\ell_\infty(t)$, the tangent line to $p_\phi(t)$ at infinity, is given by $\ell_\infty(t) = \beta t$ where $\beta = \int \phi\, d\mu_p$. By the Mañé lemma, there exists a Hölder continuous function $\tilde\psi$ cohomologous to ψ such that $\tilde\psi(x) \leq \beta$ for all $x \in X$. Let $\tilde\phi = \tilde\psi - ad_p$ and notice that $t\tilde\phi$ is cohomologous to $t\phi$ for all t so that

$$p_\phi(t) = P_{\text{top}}(t\tilde\phi) = P_{\text{top}}(t\tilde\psi - ad_p) \leq P_{\text{top}}(\beta t - atd_p) = \beta t + P_{\text{top}}(-atd_p).$$

Hence it suffices to show that $P_{\text{top}}(-atd_p)$ decreases exponentially to 0. Since μ_p is the unique maximizing measure for $-atd_p$, we already know that $P_{\text{top}}(-atd_p) \geq 0$ for all t and it converges to 0 as $t \to \infty$. We give a crude bound showing the exponential convergence. First, we enlarge the space X to the full shift $\bar X$ whose alphabet is the alphabet $\mathcal{A}$ of X. Then define a potential $\bar f$ on $\bar X$ by

$$\bar f(x) = \begin{cases} -ad_p(x) & \text{if } x_{-1}x_0 \text{ is legal in } X; \\ -\infty & \text{otherwise;} \end{cases}$$

so that $P_{\text{top}}(X, -atd_p) = P_{\text{top}}(\bar X, t\bar f)$. Let k be the period of p. We then observe that $\bar f(x) \leq \bar g(x)$ for all x where

$$\bar g(x) = \begin{cases} -a/2^k & \text{if } x_0 \neq x_{-k}; \\ 0 & \text{otherwise.} \end{cases}$$

Hence it follows that $P_{\text{top}}(\bar X, t\bar f) \leq P_{\text{top}}(\bar X, t\bar g)$. Summing $e^{S_{kn}t\bar g(x)}$ over words of length nk shows that $P_{\text{top}}(\bar X, t\bar g) = \log(1 + (|\mathcal{A}| - 1)e^{-at/2^k}))$. As noted above $P_{\text{top}}(X, t\phi) = \beta t + P_{\text{top}}(X, -atd_p) \leq \beta t + P_{\text{top}}(\bar X, t\bar g)$. Combining this with the inequality that we just derived shows that $p_\phi(t)$ approaches the asymptote ℓ_∞ exponentially fast. □

References

[1] L. Barreira, *Dimension and Recurrence in Hyperbolic Dynamics*, (Progress in Mathematics, 272). Birkhäuser, Basel, 2008.

[2] L. Barreira, B. Saussol and J. Schmeling *Higher-dimensional multifractal analysis*, J. Math. Pures Appl. **9** (2002), 67—91.

[3] L. Barreira, Ya. Pesin, and J. Schmeling, *On a general concept of multifractality: multifractal spectra for dimensions, entropies, and Lyapunov exponents. Multifractal rigidity*, Chaos, **7(1)** (1997), 27–38.

[4] T. Bousch, *Un lemme de Mañé bilatéral*, C. R. Math. Acad. Sci. Paris **335** (2002), 533–536.

[5] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Mathematics **470**, Springer, 1975.

[6] R. Bowen, *Hausdorff dimension of quasi-circles*, Inst. Hautes E´tudes Sci. Publ. Math. **50** (1979), 259–273.

[7] J. Brémont, *Gibbs measures at temperature zero*, Nonlinearity 16 (2003), 419–426.

[8] K. Burns and K. Gelfert, *Lyapunov spectrum for geodesic flows of rank 1 surfaces*, Discrete Contin. Dyn. Syst., **34(5)** (2014), 1841–1872.

[9] V. Climenhaga, *Topological pressure of simultaneous level sets*, Nonlinearity **26** (2013), 241-268.

[10] G. Contreras, *Ground states are generically a periodic orbit*, Inventiones Mathematicae **25** (2016), 383–412.

[11] G. Contreras, A.O. Lopes and P. Thieullen, *Lyapunov minimizing measures for expanding maps of the circle*, Ergodic Theory and Dynamical Systems **21** (2001), 1379–1409.

[12] J. R. Chazottes and M. Hochman, *On the Zero-Temperature Limit of Gibbs States*, Communications in Mathematical Physics **297** (2010), no. 1, 265–281.

[13] Z. Coelho and A. Quas, *Criteria for $\bar{d}$-continuity*, Trans. Amer. Math. Soc. **350** (1998), 3257–3268.

[14] O. Jenkinson, *Ergodic optimization*, Discrete and Continuous Dynamical Systems **15** (2006), 197–224.

[15] T. Kucherenko and A. Quas, *Flexibility of the Pressure Function*, Commun. Math. Phys., **395** (2022), 1431–1461.

[16] M. Lothaire, *Algebraic Combinatorics on Words* (Encyclopedia of Mathematics and its Applications). Cambridge: Cambridge University Press (2002).

[17] L. Ma and M. Pollicott, *Rigidity of pressures of Hölder potentials and the fitting of analytic functions via them*, preprint arXiv:2210.16794

[18] W. Parry, *On the β-expansions of real numbers*, Act Math. Acad. Sci. Hungar., **11** (1960), 401-416.

[19] Y. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*, Chicago Lectures in Mathematics, Chicago University Press, Chicago, 1997.

[20] F. Przytycki and M. Urbanski, *Conformal fractals: ergodic theory methods*, London Mathematical Society Lecture Note Series, 371. Cambridge University Press, Cambridge, 2010. x+354 pp.

[21] A. Quas, *Coupling and Splicing*, Unpublished lecture notes available at `http:www.math.uvic.ca/faculty/aquas/CoupleSplice.pdf`.

[22] A. Rothstein, *Versik processes: First steps*, Israel J. Math. **36** (1980), 205-223.

[23] A. Van Enter, R. Fernáandez and A. Sokal, *Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory*, Journal of Statistical Physics **72** (1993), 879–1167.

[24] D. Ruelle, *Statistical mechanics on a compact set with* $\mathbb{Z}^\nu$ *action satisfying expansiveness and specification*, Trans. Amer. Math. Soc. **185** (1973), 237–251.

[25] D. Ruelle, *Thermodynamic formalism: The mathematical structures of equilibrium statistical mechanics*, Second edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2004.

[26] D. Ruelle, *Repellers for real analytic maps*, Ergodic Theory and Dynamical Systems **2** (1982), 99–107.

[27] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics 79, Springer, 1981.

[28] P. Walters, *Differentiability Properties of the Pressure of a Continuous Transformation on a Compact Metric Space*, J. Lond. Math. Soc. **46** (1992), 471–481.

[29] G.C. Yuan and B.R. Hunt, *Optimal orbits of hyperbolic systems*, Nonlinearity **12** (1999), no. 4, 1207-1224.

Department of Mathematics, The City College of New York, New York, NY, 10031, USA

Email address: tkucherenko@ccny.cuny.edu

Department of Mathematics and Statistics, University of Victoria, Victoria, BC Canada

Email address: aquas@uvic.ca