



Free Energy, Gibbs Measures, and Glauber Dynamics for Nearest-Neighbor Interactions

Christopher Shriver 

Department of Mathematics, The University of Texas at Austin, 2515 Speedway, Austin, TX 78712, USA.
E-mail: christopher.shriver@math.utexas.edu

Received: 18 November 2021 / Accepted: 16 September 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract: We extend results of Richard Holley beyond the integer lattice to a large class of countable groups which includes free groups and all amenable groups: for nearest-neighbor interactions on the Cayley graphs of such groups, we show that a shift-invariant measure is Gibbs if and only if it is Glauber-invariant. Moreover, any shift-invariant measure converges weakly to the set of Gibbs measures when evolved under the corresponding Glauber dynamics. These results are proven using a notion of free energy density relative to a sofic approximation by homomorphisms, which avoids the boundary problems which appear when applying a standard free energy method in a nonamenable setting. We also show that any measure which minimizes this free energy density is Gibbs.

1. Introduction, Main Results

A prototypical example of the type of system we study here is the Ising model, a model of magnetism first studied by Ising in 1925 [1]. In the classical version of this model we have a rectangular grid of particles, each of which can have ‘spin’ either $+1$ or -1 . Each particle interacts only with its nearest neighbors: neighboring particles with opposite spins increase the energy and neighboring particles with the same spin decrease the energy. Statistically, the system “prefers” to be in a low-energy state, but a temperature parameter introduces randomness. The notion that a system in equilibrium minimizes free energy formalizes the competing preferences of low energy and randomness.

The rectangular grid is natural for modeling an arrangement of particles in euclidean space. However, it is also natural to study systems with similar pairwise interactions between particles but other dependence structures. These arise, for example, in some models of social networks and disordered systems in statistical physics. One important feature of the rectangular lattice \mathbb{Z}^d that we would like to keep is its notion of ‘translation.’ It is also important that each vertex has only finitely many neighbors. To preserve these features we generalize by replacing the rectangular lattice with the Cayley graph of a

finitely-generated group. Some of our results will require the additional assumption that the group has “property PA” (also known as MD; a definition is given in Sect. 2.4).

In this setting we can model an infinite regular tree as the graph of the free group \mathbb{F}_r (which produces a $2r$ -regular tree) or of the free product of cyclic groups $(\mathbb{Z}_2)^{*r}$ (which produces an r -regular tree). Note that these result in different notions of translation.

In the present paper we focus on extending results of Holley in [2]. He studied a notion of free energy density for systems with sites indexed by \mathbb{Z}^r , and used it to relate Gibbs measures and Glauber dynamics. His approach does not seem to work for nonamenable groups due to non-negligibility of the boundary of large finite subsystems.

In its place we use an “extrinsic” approach to free energy density which is based on more recent work on the entropy theory of nonamenable group actions, initiated by Lewis Bowen [3] to solve similar problems which appear in that area.

1.1. Related work. In one respect, Holley [2] worked in slightly more generality than we do here: he considered finite-range interactions, not just nearest-neighbor interactions. Higuchi and Shida [4] extended his results to spin systems on \mathbb{Z}^r which may have infinite-range interactions, as long as the strength of the interaction decays sufficiently quickly.

The method of the present paper may be compatible with such generalizations, but for the sake of simplicity we choose not to pursue them here. We also omit for the sake of simplicity interactions with hard constraints, such as the hard core model supported on independent sets.

More recently, Jahnke and Külske [5] have extended the free energy density approach to non-reversible dynamics on integer-lattice systems.

Caputo and Martinelli [6] have shown that if the product of plus-biased Bernoulli measures on an infinite tree is evolved under Ising Glauber dynamics, then it converges weakly to the “plus boundary conditions” Gibbs measure (and similarly the minus-biased Bernoulli shift converges to the “minus boundary conditions” Gibbs measure). The present paper does not establish any particular limits, but shows that all limit points are Gibbs under much more general conditions.

There has been other work on notions of free energy density on nonamenable groups, but these notions do not appear to have the properties we want for our present purposes. For example, Dembo and Montanari [7] consider, as we do below, a sequence of finite graphs that locally converge to an infinite tree. Each of these finite graphs has a unique measure which is Gibbs for the given interaction, and this measure has its own free energy density. The limit of these free energy densities is called the free energy density of the model. In the present paper, we define the free energy density of any shift-invariant measure μ as a limit of free energy densities of measures on the finite graphs which are locally consistent with μ .

1.2. Precise statements of basic definitions and main theorems. Let Γ be a countable group with r generators s_1, \dots, s_r and arbitrary relations. Let $e \in \Gamma$ denote the identity. Assume that $e \notin \{s_1, \dots, s_r\}$. We will identify Γ with its left Cayley graph, which has vertex set Γ and an i -labeled directed edge $(\gamma, s_i \gamma)$ for every $i \in [r]$ and $\gamma \in \Gamma$.

For some finite alphabet A , we define the action of Γ on A^Γ by

$$(\beta \mathbf{y})(\gamma) = \mathbf{y}(\gamma \beta)$$

for $\beta, \gamma \in \Gamma$. We can think of this as moving the center of the labeling to β^{-1} . We say that a measure $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$ is shift-invariant if $\beta_*\mu = \mu$ for any $\beta \in \Gamma$, where β_* denotes the pushforward. We denote the set of shift-invariant probability measures by $\text{Prob}^\Gamma(\mathbb{A}^\Gamma)$.

If V is a finite set, we can consider the set $\text{Hom}(\Gamma, \text{Sym}(V))$ of homomorphisms from Γ to the group of permutations of V . It is possible for this set to be empty. Given $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, we write the permutation which is the image of $\gamma \in \Gamma$ by σ^γ . We can associate to σ a directed graph with vertex set V and an i -labeled edge $(v, \sigma^{s_i}(v))$ for each $i \in [r]$ and $v \in V$.

Either Γ or the graph of some σ can be endowed with a natural graph distance: the distance between a pair of vertices is defined to be the minimal number of edges in a path between them, ignoring edge directions. Let $B^\sigma(v, R)$ denote the closed radius- R ball centered at $v \in V$, and similarly define $B^\Gamma(\gamma, R)$ for $\gamma \in \Gamma$.

The correspondence between finite and infinite systems is established using *empirical distributions*, which we now define. Let $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$ and $\mathbf{x} \in \mathbb{A}^V$. For any $v \in V$ there is a natural way to lift \mathbf{x} to a labeling $\Pi_v^\sigma \mathbf{x} \in \mathbb{A}^\Gamma$, starting by lifting \mathbf{x}_v to the root e . Specifically,

$$(\Pi_v^\sigma \mathbf{x})(\gamma) = \mathbf{x}(\sigma^\gamma(v)).$$

The empirical distribution of \mathbf{x} is defined by

$$P_{\mathbf{x}}^\sigma = (v \mapsto \Pi_v^\sigma \mathbf{x})_* \text{Unif}(V) = \frac{1}{|V|} \sum_{v \in V} \delta_{\Pi_v^\sigma \mathbf{x}} \in \text{Prob}(\mathbb{A}^\Gamma).$$

This captures the ‘local statistics’ of \mathbf{x} . This notation was used in the approach to sofic entropy in [8].

To state our results we use the following way of measuring local similarity of a finite graph σ to Γ : for $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, we say $B^\sigma(v, R) \cong B^\Gamma(e, R)$ if there is an isomorphism of the induced subgraphs $B^\sigma(v, R), B^\Gamma(e, R)$ which respects both edge labels and directions. Define

$$\delta_R^\sigma = \frac{1}{|V|} |\{v \in V : B^\sigma(v, R) \not\cong B^\Gamma(e, R)\}|$$

$$\Delta^\sigma = \inf_R (9 \cdot (2/3)^R + 6\delta_R^\sigma).$$

The constants which appear here are connected to the choice of metric d in Sect. 2 below. If Δ^σ is small, then σ looks like Γ to a large radius near most vertices. This is a way of saying that the action of σ is approximately free.

Note that a sequence $\Sigma = \{\sigma_n\}$ is a sofic approximation to the group Γ (or, equivalently, Benjamini-Schramm converges to the Cayley graph) if and only if $\Delta^{\sigma_n} \rightarrow 0$. Although the exact quantity Δ^σ defined here is somewhat ad hoc, this is essentially a standard definition of sofic approximation in this context.

As mentioned above, our central tool is a notion of “free energy density” $a_\Sigma(\mu)$ of a measure $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$ with respect to a nearest-neighbor potential. This free energy density is defined relative to a choice of sofic approximation $\Sigma = (\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$. It may be $+\infty$, but if it is finite then it is nonincreasing as μ evolves according to the Glauber dynamics of the relevant potential (Proposition 3.2). Moreover if μ is not Gibbs then it is strictly decreasing; Proposition 3.3 gives a stronger version of this claim. For every choice of Σ there exist measures with finite free energy density, so this implies the following:

Theorem A. *For any choice of Σ , every $\mu \in \text{Prob}(\mathcal{A}^\Gamma)$ minimizing $a_\Sigma(\cdot)$ is Gibbs.*

Note that “ a_Σ ” and “Gibbs” here refer to the same arbitrary nearest-neighbor potential, but we will keep these dependencies implicit.

The converse is false, since a Gibbs measure may have free energy density $+\infty$ with respect to some Σ . It is unclear whether a Gibbs measure may have finite but non-minimal free energy density.

We also use the same framework to show that every limit point of Glauber dynamics is Gibbs, as long as at some point on the trajectory the free energy density is finite with respect to some Σ :

Theorem B. *Suppose $\mu \in \text{Prob}^\Gamma(\mathcal{A}^\Gamma)$, and let μ_t denote its evolution under Glauber dynamics. If there exist $s \geq 0$ and Σ such that $a_\Sigma(\mu_s) < +\infty$, then μ_t converges weakly to the set of Gibbs measures as $t \rightarrow \infty$.*

It is possible to avoid the degenerate case of infinite free energy density by an appropriate choice of Σ when Γ has a property called “property PA”; see Sect. 2.4 below for a definition and the relevant result (Proposition 2.4). All amenable groups have property PA, and the property is preserved under free products. Hence we have the following:

Corollary 1.1. *If Γ has property PA, then Glauber dynamics converges to the set of Gibbs measures from any starting point. In particular, a shift-invariant measure is Gibbs if and only if it is Glauber-invariant.*

The assumption that the group has property PA may also not be necessary for particular starting points of the dynamics. For example, if μ is a Bernoulli shift then $a_\Sigma(\mu)$ is finite for any Σ , so Theorem B implies that all limit points of μ_t are Gibbs without additional assumptions on Γ .

2. Definitions

For $\gamma \in \Gamma$, let $|\gamma|$ denote the graph distance between γ and the identity $e \in \Gamma$.

Give \mathcal{A}^Γ the metric

$$d(\mathbf{x}, \mathbf{y}) = \sum_{\gamma \in \Gamma} (3r)^{-|\gamma|} \mathbf{1}_{\mathbf{x}(\gamma) \neq \mathbf{y}(\gamma)};$$

the factor 3 is chosen to ensure convergence. Let \bar{d} denote the corresponding transportation metric on $\text{Prob}(\mathcal{A}^\Gamma)$: specifically, with $\text{Lip}_1(\mathcal{A}^\Gamma)$ denoting the set of 1-Lipschitz real-valued functions, we define

$$\bar{d}(\mu, \nu) = \sup \{ |\mu f - \nu f| : f \in \text{Lip}_1(\mathcal{A}^\Gamma) \}.$$

Here μf denotes the integral of f with respect to μ . Note that d generates the product topology (which is compact), and \bar{d} generates the weak topology induced by the pairing with continuous functions (which is also compact).

For any set V and any $\mathbf{x} \in \mathcal{A}^V$, $v \in V$, $a \in \mathcal{A}$ we let $\mathbf{x}^{v \rightarrow a} \in \mathcal{A}^V$ be given by

$$\mathbf{x}^{v \rightarrow a}(w) = \begin{cases} \mathbf{x}(w), & w \neq v \\ a, & w = v. \end{cases}$$

Below, an element of \mathcal{A}^V will be referred to as a *microstate* and an element of $\text{Prob}(\mathcal{A}^V)$ as a *state*.

2.1. Glauber dynamics. Let V be an at most countable set and fix $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$. We will apply this in two cases: when V is finite, and when $V = \Gamma$ and σ is the action of Γ on itself by right multiplication. We will distinguish between these cases by giving notation superscripts of σ or Γ respectively (e.g. Ω^σ versus Ω^Γ).

For simplicity we consider only nearest-neighbor interactions. Fix a symmetric function $J : \mathbb{A}^2 \rightarrow \mathbb{R}$ and a function $h : \mathbb{A} \rightarrow \mathbb{R}$, and let $S = \{s_1, \dots, s_r, s_1^{-1}, \dots, s_r^{-1}\}$. For $v \in V, \mathbf{x} \in \mathbb{A}^V$, let

$$\Phi_v(\mathbf{x}) = h(\mathbf{x}(v)) + \sum_{s \in S} J(\mathbf{x}(v), \mathbf{x}(\sigma^s v)),$$

and for $\mathbf{a} \in \mathbb{A}$ let

$$c_v(\mathbf{x}, \mathbf{a}) = Z_v(\mathbf{x})^{-1} \exp\{-\Phi_v(\mathbf{x}^{v \rightarrow \mathbf{a}})\},$$

where $Z_v(\mathbf{x})$ is the normalizing factor which makes $c_v(\mathbf{x}, \cdot)$ a probability measure on \mathbb{A} . Note that $c_v(\mathbf{x}, \mathbf{a})$ only depends on \mathbf{x} through its values at vertices adjacent to v . A standard stochastic Ising model has $\mathbb{A} = \{-1, 1\}$ and $J(\mathbf{a}, \mathbf{b}) = \beta \mathbf{a} \mathbf{b}$ for some $\beta \geq 0$ (the inverse temperature), and h represents an ambient electric field.

We study here the continuous-time Markov process with state space \mathbb{A}^V and generator given by

$$\Omega f(\mathbf{x}) = \sum_{v \in V} \sum_{\mathbf{a} \in \mathbb{A}} c_v(\mathbf{x}, \mathbf{a}) [f(\mathbf{x}^{v \rightarrow \mathbf{a}}) - f(\mathbf{x})].$$

If V is finite then this gives a well-defined linear operator on $C(\mathbb{A}^V)$. Otherwise we need to first define Ω using the above formula on a particular ‘core’ of ‘smooth’ functions for which the sum converges, then take the closure of Ω ; see Liggett’s book [9] for details. We denote the induced Markov semigroup by $\{S(t) : t \geq 0\}$.

For any continuous function $f : \mathbb{A}^V \rightarrow \mathbb{R}$ and $\mathbf{x} \in \mathbb{A}^V$, we interpret $[S(t)f](\mathbf{x})$ as expected value of $f(\mathbf{x}_t)$, where the random variable \mathbf{x}_t is the evolution of \mathbf{x} to time t . The action of the semigroup on a probability measure is denoted $\mu S(t)$; this is interpreted the evolution of a probability measure $\mu \in \text{Prob}(\mathbb{A}^V)$ to time t . The evolved measure is defined via the formula $[\mu S(t)]f := \mu[S(t)f]$.

Further details of the construction of the dynamics will only be needed for proofs of the following two results. The relevant details are contained in Sect. 5.

The first result may be thought of as an approximate equivariance between the Glauber semigroups and the empirical distribution:

Theorem 2.1. *There is a constant $M > 0$ such that for any $\mathbf{x} \in \mathbb{A}^V$, $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, and $t \geq 0$*

$$\bar{d}(S^\sigma(t)P_{\mathbf{x}}^\sigma, P_{\mathbf{x}}^\sigma S^\Gamma(t)) \leq \Delta^\sigma \cdot t e^{Mt}.$$

The constant M depends only on the Cayley graph Γ and on the interaction.

It may be helpful to clarify that the first term on the left, $S^\sigma(t)P_{\mathbf{x}}^\sigma$, is the evolution to time t of the function $P_{\bullet}^\sigma : \mathbb{A}^V \rightarrow \text{Prob}(\mathbb{A}^\Gamma)$ evaluated at $\mathbf{x} \in \mathbb{A}^V$. The second term is the evolution of the empirical distribution $P_{\mathbf{x}}^\sigma \in \text{Prob}(\mathbb{A}^\Gamma)$. So this theorem says that the expected empirical distribution after running the finitary dynamics for time t is close to the result of evolving the original empirical distribution for time t , as long as σ locally looks like Γ .

We also use the following Lipschitz bound on the Markov semigroup:

Lemma 2.2. *If $\mu, \nu \in \text{Prob}(A^\Gamma)$ then*

$$\bar{d}(\mu S^\Gamma(t), \nu S^\Gamma(t)) \leq \exp(Mt) \bar{d}(\mu, \nu).$$

2.2. Gibbs measures. For a finite graph with vertex set V and edges induced by $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, we define the total energy

$$U(\mathbf{x}) = \sum_{v \in V} h(\mathbf{x}(v)) + \sum_{v \in V} \sum_{i \in [r]} J(\mathbf{x}(v), \mathbf{x}(\sigma^{s_i} v)).$$

Note that each directed edge appears in the double sum exactly once; in particular, this is not the same as $\sum_v \Phi_v(\mathbf{x})$, in which directed edges will be counted twice. We can, however, write

$$U(\mathbf{x}) = \sum_{v \in V} U_v(\mathbf{x})$$

where

$$U_v(\mathbf{x}) := h(\mathbf{x}(v)) + \frac{1}{2} \sum_{s \in \{s_1^{\pm 1}, \dots, s_r^{\pm 1}\}} J(\mathbf{x}(v), \mathbf{x}(\sigma^s v)).$$

This definition of U_v also makes sense if V is infinite; below we will make use of the case $V = \Gamma, v = e$.

Note that if we define

$$u^{\max} = \max_{a \in A} \left(h(a) + r \max_{b \in A} J(a, b) \right)$$

and

$$u^{\min} = \min_{a \in A} \left(h(a) + r \min_{b \in A} J(a, b) \right)$$

then for any V, σ and any $\mathbf{x} \in A^V$ we have

$$u^{\min} \leq \frac{1}{|V|} U(\mathbf{x}) \leq u^{\max}.$$

The finitary Gibbs measure $\xi_V \in \text{Prob}(A^V)$ can now be defined by

$$\xi_V\{\mathbf{x}\} = Z_V^{-1} \exp\{-U(\mathbf{x})\}$$

where Z_V is the normalizing constant. Note that

$$\xi_V(\mathbf{y}(v) = a \mid \mathbf{y}(w) = \mathbf{x}(w) \forall w \neq v) = \frac{\exp\{-U(\mathbf{x}^{v \rightarrow a})\}}{\sum_{b \in A} \exp\{-U(\mathbf{x}^{v \rightarrow b})\}} = c_v(\mathbf{x}, a),$$

since

$$\frac{\exp\{-U(\mathbf{x}^{v \rightarrow a})\}}{\exp\{-U(\mathbf{x}^{v \rightarrow b})\}} = \frac{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow a})\}}{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow b})\}}.$$

On the infinite graph Γ we must use a different approach, since the sum defining the total energy will not converge. We use a natural generalization of [9, Definition IV.1.5]; see also [10] for a much more general treatment of infinite-volume Gibbs measures.

Let \mathcal{T}_γ denote the σ -algebra on \mathbb{A}^Γ generated by all vertices except for γ . We call $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$ a Gibbs measure if for each $\gamma \in \Gamma$ and $\mathbf{a} \in \mathbb{A}$, the function $\mathbf{y} \mapsto c_\gamma(\mathbf{y}, \mathbf{a})$ is a version of the conditional expectation $\mu(\{\mathbf{x} : \mathbf{x}(\gamma) = \mathbf{a} \} \mid \mathcal{T}_\gamma)(\mathbf{y})$. This means that for every integrable $f : \mathbb{A}^\Gamma \rightarrow \mathbb{R}$ and $\gamma \in \Gamma$ we have

$$\int \sum_{\mathbf{a} \in \mathbb{A}} c_\gamma(\mathbf{x}, \mathbf{a}) f(\mathbf{x}^{\gamma \rightarrow \mathbf{a}}) \mu(d\mathbf{x}) = \int f(\mathbf{x}) \mu(d\mathbf{x}).$$

We may also describe this relation by saying that μ is invariant under re-randomizing the spin at γ using the kernel c_γ .

In particular, if μ is Gibbs then for any ‘smooth’ f

$$\mu \Omega^\Gamma f = \int \left(\sum_{\gamma, \mathbf{a}} c_\gamma(\mathbf{x}, \mathbf{a}) [f(\mathbf{x}^{\gamma \rightarrow \mathbf{a}}) - f(\mathbf{x})] \right) \mu(d\mathbf{x}) = 0.$$

It follows that $\mu \Omega^\Gamma = 0$, which means Gibbs measures are Glauber-invariant.

2.3. Good models for measures on \mathbb{A}^Γ . Let V be a finite set and let $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$.

A labeling $\mathbf{x} \in \mathbb{A}^V$ is said to be a *good model* for $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$ over σ if the empirical distribution $P_{\mathbf{x}}^\sigma$ is close to μ in the weak topology. More precisely, we can say \mathbf{x} is \mathcal{O} -good if $P_{\mathbf{x}}^\sigma \in \mathcal{O}$ for some weak-open neighborhood $\mathcal{O} \ni \mu$. The set of such \mathbf{x} is denoted $\Omega(\sigma, \mathcal{O})$. An interpretation of this relationship is that average local quantities of the finite system are consistent with μ .

We define the empirical distribution of a state $\zeta \in \text{Prob}(\mathbb{A}^V)$ by

$$P_\zeta^\sigma := \zeta P_{\mathbf{x}}^\sigma \left(= \int P_{\mathbf{x}}^\sigma \zeta(d\mathbf{x}) \right)$$

and say that ζ is \mathcal{O} -consistent with μ (for some neighborhood $\mathcal{O} \ni \mu$) over σ if $P_\zeta^\sigma \in \mathcal{O}$. We can still interpret this in terms of averages of local quantities: now the average also involves a random microstate \mathbf{x} with law ζ . We denote the set of such states by $\Omega(\sigma, \mathcal{O})$. This way of lifting a finitary state is used in [11]; it also essentially appears in the notion of ‘local convergence on average’ introduced in [12, Definition 2.3]. See also [13, Sect. 4] and [14, Sect. 5].

This consistency is stable under Glauber dynamics in the following sense:

Proposition 2.3. *Suppose $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, $\zeta \in \text{Prob}(\mathbb{A}^V)$, and $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$. Let ζ_t, μ_t denote their evolutions under Glauber dynamics on σ, Γ respectively. Then for any $t \geq 0$*

$$\bar{d}(P_{\zeta_t}^\sigma, \mu_t) \leq [\Delta^\sigma t + \bar{d}(P_\zeta^\sigma, \mu)] \exp(Mt).$$

Proof. For any $f \in \text{Lip}_1(\mathbb{A}^\Gamma)$, the triangle inequality gives

$$\begin{aligned} |P_{\zeta_t}^\sigma f - \mu_t f| &\leq |P_{\zeta_t}^\sigma f - P_\zeta^\sigma S^\Gamma(t) f| + |P_\zeta^\sigma S^\Gamma(t) f - \mu_t f| \\ &= |\zeta[S^\sigma(t) P_{\mathbf{x}}^\sigma f - P_{\mathbf{x}}^\sigma S^\Gamma(t) f]| + |P_\zeta^\sigma S^\Gamma(t) f - \mu S^\Gamma(t) f|. \end{aligned}$$

Using that ζ is a probability measure and the definition of \bar{d} , this implies the bound

$$|P_{\zeta_t}^\sigma f - \mu_t f| \leq \max_{\mathbf{x} \in \mathbb{A}^V} \bar{d}(S^\sigma(t) P_{\mathbf{x}}^\sigma, P_{\mathbf{x}}^\sigma S^\Gamma(t)) + \bar{d}(P_\zeta^\sigma S^\Gamma(t), \mu S^\Gamma(t)).$$

The first term may be controlled with Theorem 2.1 and the second with Lemma 2.2 to get

$$|P_{\zeta_t}^\sigma f - \mu_t f| \leq [\Delta^\sigma t + \bar{d}(P_\zeta^\sigma, \mu)] \exp(Mt).$$

The result then follows by taking the supremum over $f \in \text{Lip}_1(\mathbb{A}^\Gamma)$. \square

2.4. Free energy density. Given $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$, the *free energy* of $\zeta \in \text{Prob}(\mathbb{A}^V)$ is defined by

$$A(\zeta) = \zeta(U) - H(\zeta),$$

where $\zeta(U) = \int U(\mathbf{x}) \zeta(d\mathbf{x})$ is the average energy and $H(\zeta)$ is the Shannon entropy.

Given a sofic approximation $\Sigma = (\sigma_n)_{n \in \mathbb{N}}$ with $\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n))$ such that $|V_n| \rightarrow \infty$, we define the *free energy density* of $\mu \in \text{Prob}^\Gamma(\mathbb{A}^\Gamma)$ relative to Σ by

$$a_\Sigma(\mu) = \lim_{\mathcal{O} \downarrow \mu} \limsup_{n \rightarrow \infty} \inf_{\zeta \in \Omega(\sigma_n, \mathcal{O})} \frac{1}{|V_n|} A(\zeta).$$

We follow the convention that the infimum of the empty set is $+\infty$.

The outermost limit is over the net of weak-open neighborhoods of μ , partially ordered by inclusion. Note that for each n the expression $\inf_{\zeta \in \Omega(\sigma_n, \mathcal{O})} \frac{1}{|V_n|} A(\zeta)$ is non-decreasing as $\mathcal{O} \downarrow \mu$, so the limit exists and is equal to the supremum over $\mathcal{O} \ni \mu$.

It is straightforward to check from the definitions that $\zeta(U)/|V| = P_\zeta^\sigma(U_e)$. Consequently we have

$$a_\Sigma(\mu) = \mu(U_e) - h_\Sigma^{\text{mod}}(\mu),$$

where h^{mod} is the ‘modified sofic entropy’ in [11], except with a \liminf instead of \limsup . Other very similar definitions have been studied, for example \bar{h} in [13] (in the residually finite case) and ‘model-measure sofic entropy’ in [8, Sect. 6] (with other notions of local convergence). Since this connection will not be used below, we omit the proof. It may also be interesting to investigate other sofic free energy densities with h^{mod} replaced by a different type of sofic entropy.

Since the map $\mu \mapsto a_\Sigma(\mu)$ is defined in terms of a supremum over neighborhoods of μ , it is lower semi-continuous. Consequently, it attains its minimum on $\text{Prob}^\Gamma(\mathbb{A}^\Gamma)$.

Note also that as long as $\Omega(\sigma_n, \mathcal{O})$ is nonempty, for every ζ in this set we have

$$u^{\min} - \log |A| \leq \frac{1}{|V_n|} [\zeta(U) - H(\zeta)] \leq u^{\max}.$$

In particular,

$$a_\Sigma(\mu) \in [u^{\min} - \log |A|, u^{\max}] \cup \{+\infty\}.$$

The case $a_\Sigma(\mu) = +\infty$ can actually occur, for example if μ is ergodic and the sofic entropy relative to Σ is $-\infty$. This is because, in the ergodic case, ζ being consistent

with μ is the same as being mostly supported on labelings which are good models for μ , but if the sofic entropy is $-\infty$ then there are no good models.

Note, however, that the function $\mu \mapsto a_\Sigma(\mu)$ is not identically $+\infty$ for any choice of Σ , since the point mass at a constant labeling in A^Γ always has good models.

If μ were not shift-invariant then the expression defining $a_\Sigma(\mu)$ would still make sense, but would take the value $+\infty$ for any Σ . This is because empirical distributions are always shift-invariant and $\text{Prob}^\Gamma(A^\Gamma)$ is closed so, no matter what Σ we choose, any small enough neighborhood of μ contains no empirical distributions.

Following [15], a shift-invariant measure in $\text{Prob}^\Gamma(A^\Gamma)$ is called *periodic* if it has finite support, and a group Γ is said to have *property PA* if the set of periodic measures is dense in $\text{Prob}^\Gamma(A^\Gamma)$ for every finite alphabet A ; in other words, if every shift-invariant probability measure on A^Γ has periodic approximations.

In Sect. 4 we prove the following:

Proposition 2.4. *A group Γ has property PA if and only if for every finite alphabet A and every $\mu \in \text{Prob}^\Gamma(A^\Gamma)$ there exists a sofic approximation $\Sigma = (\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$ such that $a_\Sigma(\mu) < +\infty$.*

Property PA was proved to hold for free groups by Bowen in [15, Theorem 3.4]. Kechris later studied another equivalent property in [16] which he called “MD”; see also the survey [17] for more recent information on which groups are known to have this property. In particular (all from [17, Sect. 5]):

- Amenable groups have property PA.
- If two nontrivial groups are both either finite or property-PA, then their free product has property PA.
- The recent negative solution to Connes’ embedding conjecture implies that the direct product $\mathbb{F}_2 \times \mathbb{F}_2$ does *not* have property PA.

2.4.1. Connection with a related notion of free energy density Fix a sofic approximation Σ , and for each n let $\xi_n \in \text{Prob}(A^{V_n})$ denote the unique finitary Gibbs measure. Suppose that $P_{\xi_n}^{\sigma_n} \rightarrow \mu$. Then for any $\mathcal{O} \ni \mu$ we have $\xi_n \in \Omega(\sigma_n, \mathcal{O})$ for all large enough n . But ξ_n , being the Gibbs measure, has minimal free energy among all probability measures on A^{V_n} , so

$$\inf_{\zeta \in \Omega(\sigma_n, \mathcal{O})} A(\zeta) = A(\xi_n) = -\log Z_n$$

where Z_n is the normalizing constant appearing in the definition of ξ_n . It follows that

$$a_\Sigma(\mu) = -\liminf_{n \rightarrow \infty} \frac{1}{|V_n|} \log Z_n.$$

This quantity (without the minus sign) is often called the asymptotic free energy density.

[7] shows that for the Ising model with any positive temperature and positive external field, the weak limit of $P_{\xi_n}^{\sigma_n}$ is μ_+ , and that $\frac{1}{|V_n|} \log Z_n$ converges to a limit predicted by the “cavity method” from statistical physics. In particular, this gives a formula for $a_\Sigma(\mu_+)$.

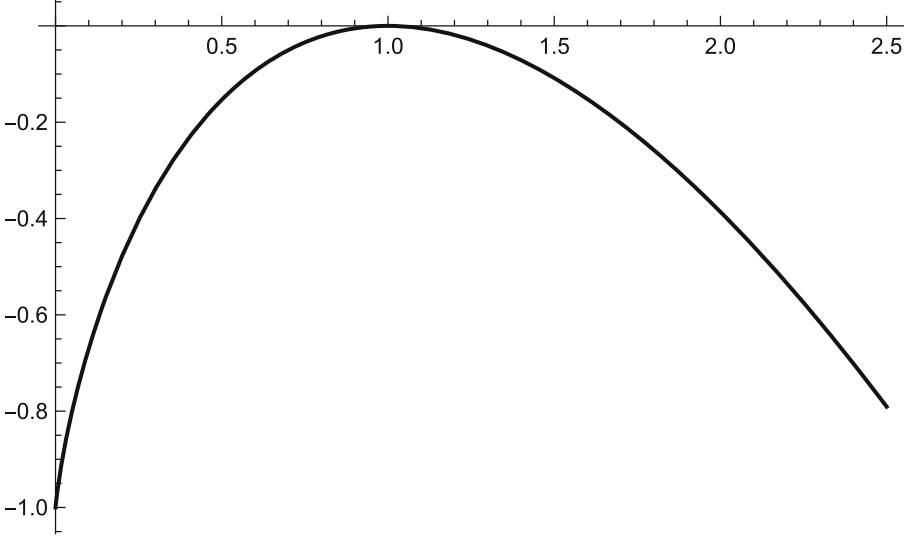


Fig. 1. A graph of the function F_0

2.5. Measuring non-Gibbs-ness. As in [2], we make use of the function

$$F_0(s) = \begin{cases} s - s \log s - 1, & s > 0 \\ -1, & s = 0 \end{cases}$$

which appears in an expression for the time derivative of free energy [Proposition 3.1]. This function is concave, nonpositive, and equal to 0 if and only if $s = 1$. A graph is included in Fig. 1. If $\mu_R \in \text{Prob}(\mathcal{A}^{\mathcal{B}(e,R)})$ has full support, we define

$$\Delta_a^R(\mu_R) = \sum_{\mathbf{y} \in \mathcal{A}^{\mathcal{B}(e,R)}} \mu_R\{\mathbf{y}\} \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{\mu_R\{\mathbf{y}^{e \rightarrow a}\}}{\mu_R\{\mathbf{y}\}} \right).$$

This measures the average failure of μ_R to be consistent with the Gibbs specification.

Lemma 2.5. *Suppose $\mu \in \text{Prob}(\mathcal{A}^\Gamma)$ is a translation-invariant measure such that for every $R \geq 1$ the marginal $\mu_R \in \text{Prob}(\mathcal{A}^{\mathcal{B}(e,R)})$ has full support and $\Delta_a^R(\mu_R) = 0$ for every $a \in \mathcal{A}$. Then μ is Gibbs.*

Proof. Fix $R \geq 1$, and let $\mathcal{S}_{\mathcal{B}(e,R) \setminus \{e\}}$ denote the σ -algebra generated by sites in $\mathcal{B}(e, R) \setminus \{e\}$. Then by definition of conditional expectation, for any $a \in \mathcal{A}$

$$\mu(\{\mathbf{x} \in \mathcal{A}^\Gamma : \mathbf{x}(e) = a\} \mid \mathcal{S}_{\mathcal{B}(e,R) \setminus \{e\}})(\mathbf{y}) = \frac{\mu_R\{\mathbf{y}^{e \rightarrow a}\}}{\sum_{b \in \mathcal{A}} \mu_R\{\mathbf{y}^{e \rightarrow b}\}},$$

where on the right-hand side we use the shorthand $\mu_R\{\mathbf{y}\} = \mu_R\{\mathbf{y} \mid_{\mathcal{B}(e,R)}\}$ for $\mathbf{y} \in \mathcal{A}^\Gamma$. Our assumption that μ_R has full support and $\Delta_a^R(\mu_R) = 0$ implies that

$$\frac{\mu_R\{\mathbf{y}^{e \rightarrow a}\}}{\sum_{b \in \mathcal{A}} \mu_R\{\mathbf{y}^{e \rightarrow b}\}} = \left(\sum_{b \in \mathcal{A}} \frac{\mu_R\{\mathbf{y}^{e \rightarrow b}\}}{\mu_R\{\mathbf{y}^{e \rightarrow a}\}} \right)^{-1} = \left(\sum_{b \in \mathcal{A}} \frac{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow b})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \right)^{-1} = c_e(\mathbf{y}, a)$$

Taking R to infinity, by martingale convergence we get

$$\mu(\cdot \mid \mathcal{T}_e)(\mathbf{y}) = c_e(\mathbf{y}, \cdot).$$

By translation invariance, it follows that μ is Gibbs. \square

3. Proof of Theorem B

Proposition 3.1. *Let $\zeta_0 \in \text{Prob}(\mathbb{A}^V)$, and let ζ_t denote its evolution under Glauber dynamics. Then for all $t > 0$, ζ_t has full support and*

$$\frac{d}{dt} A(\zeta_t) = \sum_{\mathbf{x}, v, a} F_0 \left(\frac{\exp\{-\Phi_v(\mathbf{x})\}}{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow a})\}} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \zeta_t\{\mathbf{x}\} c_v(\mathbf{x}, a).$$

Our proof of this proposition is based on the proof of the analogous result in Holley's paper [2], with some minor changes.

For $\mathbf{x} \in \mathbb{A}^V$, write

$$P(\mathbf{x}) = \exp\{-U(\mathbf{x})\}.$$

This is just the Gibbs measure on V , except without the normalizing factor. It is easy to see that

$$A(\zeta) = \sum_{\mathbf{x}} \zeta\{\mathbf{x}\} \log \frac{\zeta\{\mathbf{x}\}}{P(\mathbf{x})}.$$

Proof of proposition. A calculation using the Markov generator shows that

$$\frac{d}{dt} \zeta_t\{\mathbf{x}\} = \sum_{v, a} [\zeta_t\{\mathbf{x}^{v \rightarrow a}\} c_v(\mathbf{x}^{v \rightarrow a}, \mathbf{x}(v)) - \zeta_t\{\mathbf{x}\} c_v(\mathbf{x}, a)].$$

Now if ζ_t does not have full support, there is some \mathbf{x} such that $\zeta_t\{\mathbf{x}\} = 0$ but there exist v, a such that $\zeta_t\{\mathbf{x}^{v \rightarrow a}\} > 0$ (if we think of \mathbb{A}^V as a graph with edges between labelings that differ at a single vertex, we are picking an \mathbf{x} in the outer boundary of the support of ζ_t). For such an \mathbf{x} , by the formula above we would have $\frac{d}{dt} \zeta_t\{\mathbf{x}\} > 0$. Unless $t = 0$ this would imply the existence of times where ζ_s gives negative mass to $\{\mathbf{x}\}$. Therefore ζ_t has full support for all $t > 0$.

Therefore

$$\begin{aligned} \frac{d}{dt} A(\zeta_t) &= \sum_{\mathbf{x} \in \mathbb{A}^V} \frac{d}{dt} \left[\zeta_t\{\mathbf{x}\} \log \frac{\zeta_t\{\mathbf{x}\}}{P(\mathbf{x})} \right] \\ &= \sum_{\mathbf{x} \in \mathbb{A}^V} \frac{d}{dt} [\zeta_t\{\mathbf{x}\}] \log \frac{\zeta_t\{\mathbf{x}\}}{P(\mathbf{x})} \\ &= \sum_{\mathbf{x} \in \mathbb{A}^V} \sum_{v, a} [\zeta_t\{\mathbf{x}^{v \rightarrow a}\} c_v(\mathbf{x}^{v \rightarrow a}, \mathbf{x}(v)) - \zeta_t\{\mathbf{x}\} c_v(\mathbf{x}, a)] \log \frac{\zeta_t\{\mathbf{x}\}}{P(\mathbf{x})}. \end{aligned}$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{A}^V$, define

$$\mathfrak{A}(\mathbf{x}, \mathbf{y}) = \begin{cases} - \sum_{\substack{\mathbf{a}, v \\ \mathbf{a} \neq \mathbf{x}(v)}} c_v(\mathbf{y}, \mathbf{a}), & \mathbf{x} = \mathbf{y} \\ c_v(\mathbf{y}, \mathbf{a}), & \mathbf{x} \neq \mathbf{y}, \mathbf{x} = \mathbf{y}^{v \rightarrow \mathbf{a}} \text{ for some } v \in V, \mathbf{a} \in \mathbb{A} \\ 0, & \text{else.} \end{cases}$$

This has the following useful properties:

$$\sum_{\mathbf{x}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) = 0 \quad \forall \mathbf{y}. \quad (1)$$

□

Proof. For any \mathbf{y} ,

$$\begin{aligned} \sum_{\mathbf{x}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) &= \mathfrak{A}(\mathbf{y}, \mathbf{y}) + \sum_{v \in V} \sum_{\mathbf{a} \neq \mathbf{y}(v)} \mathfrak{A}(\mathbf{y}^{v \rightarrow \mathbf{a}}, \mathbf{y}) + 0 = 0. \\ \sum_{\mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) P(\mathbf{y}) &= 0 \quad \forall \mathbf{x}. \end{aligned} \quad (2)$$

□

Proof. For any \mathbf{x} ,

$$\begin{aligned} \sum_{\mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) P(\mathbf{y}) &= \mathfrak{A}(\mathbf{x}, \mathbf{x}) P(\mathbf{x}) + \sum_{v \in V} \sum_{\mathbf{a} \neq \mathbf{x}(v)} \mathfrak{A}(\mathbf{x}, \mathbf{x}^{v \rightarrow \mathbf{a}}) P(\mathbf{x}^{v \rightarrow \mathbf{a}}) + 0 = 0 \\ &= \sum_{\substack{\mathbf{a}, v \\ \mathbf{a} \neq \mathbf{x}(v)}} \left[-c_v(\mathbf{x}, \mathbf{a}) \exp\{-U(\mathbf{x})\} + c_v(\mathbf{x}^{v \rightarrow \mathbf{a}}, \mathbf{x}(v)) \exp\{-U(\mathbf{x}^{v \rightarrow \mathbf{a}})\} \right] \\ &= 0. \end{aligned}$$

In fact every term of this last sum is zero because

$$\frac{c_v(\mathbf{x}, \mathbf{a}) \exp\{-U(\mathbf{x})\}}{c_v(\mathbf{x}^{v \rightarrow \mathbf{a}}, \mathbf{x}(v)) \exp\{-U(\mathbf{x}^{v \rightarrow \mathbf{a}})\}} = \frac{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow \mathbf{a}})\} \exp\{-U(\mathbf{x})\}}{\exp\{-\Phi_v(\mathbf{x})\} \exp\{-U(\mathbf{x}^{v \rightarrow \mathbf{a}})\}} = 1.$$

Using these two properties of \mathfrak{A} , and the fact that ζ_t has full support, we see that

$$\begin{aligned} \frac{d}{dt} A(\zeta_t) &= \sum_{\mathbf{x}, \mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) \zeta_t\{\mathbf{y}\} \log \frac{\zeta_t\{\mathbf{x}\}}{P(\mathbf{x})} \\ &= \sum_{\mathbf{x}, \mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{y})}{P(\mathbf{x})} \frac{P(\mathbf{x})}{P(\mathbf{y})} \frac{\zeta_t\{\mathbf{y}\}}{\zeta_t\{\mathbf{x}\}} \log \frac{\zeta_t\{\mathbf{x}\}}{P(\mathbf{x})} \\ &\quad + \sum_{\mathbf{x}, \mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{y})}{P(\mathbf{x})} \frac{P(\mathbf{x})}{P(\mathbf{y})} \frac{\zeta_t\{\mathbf{y}\}}{\zeta_t\{\mathbf{x}\}} \log \frac{P(\mathbf{y})}{\zeta_t\{\mathbf{y}\}} \quad (= 0 \text{ by Equation 1}) \\ &\quad + \sum_{\mathbf{x}, \mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{y})}{P(\mathbf{x})} \left(-1 + \frac{P(\mathbf{x})}{P(\mathbf{y})} \frac{\zeta_t\{\mathbf{y}\}}{\zeta_t\{\mathbf{x}\}} \right) \quad (= 0 \text{ by Equation 2}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\mathbf{x}, \mathbf{y}} \mathfrak{A}(\mathbf{x}, \mathbf{y}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{y})}{P(\mathbf{x})} F_0 \left(\frac{P(\mathbf{x})}{P(\mathbf{y})} \frac{\zeta_t\{\mathbf{y}\}}{\zeta_t\{\mathbf{x}\}} \right) \quad (\text{definition of } F_0) \\
&= \sum_{\mathbf{x}} \left(\mathfrak{A}(\mathbf{x}, \mathbf{x}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{x})}{P(\mathbf{x})} \overbrace{F_0 \left(\frac{P(\mathbf{x})}{P(\mathbf{x})} \frac{\zeta_t\{\mathbf{x}\}}{\zeta_t\{\mathbf{x}\}} \right)}^0 \right. \\
&\quad \left. + \sum_{\substack{v, a \\ a \neq \mathbf{x}(v)}} \mathfrak{A}(\mathbf{x}, \mathbf{x}^{v \rightarrow a}) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{x}^{v \rightarrow a})}{P(\mathbf{x})} F_0 \left(\frac{P(\mathbf{x})}{P(\mathbf{x}^{v \rightarrow a})} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \right) \\
&= \sum_{\mathbf{x}} \sum_{v, a} c_v(\mathbf{x}^{v \rightarrow a}, \mathbf{x}(v)) \zeta_t\{\mathbf{x}\} \frac{P(\mathbf{x}^{v \rightarrow a})}{P(\mathbf{x})} F_0 \left(\frac{P(\mathbf{x})}{P(\mathbf{x}^{v \rightarrow a})} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \\
&\quad (\text{terms where } a = \mathbf{x}(v) \text{ are 0 anyway}) \\
&= \sum_{\mathbf{x}, v, a} F_0 \left(\frac{P(\mathbf{x})}{P(\mathbf{x}^{v \rightarrow a})} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \zeta_t\{\mathbf{x}\} c_v(\mathbf{x}, a).
\end{aligned}$$

The desired formula follows from the identity $\frac{P(\mathbf{x})}{P(\mathbf{x}^{v \rightarrow a})} = \frac{\exp\{-\Phi_v(\mathbf{x})\}}{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow a})\}}$, also used above. \square

We first use the previous result to show that free energy density is nonincreasing.

Proposition 3.2. *Suppose $\mu \in \text{Prob}(\mathbb{A}^\Gamma)$, and let $\Sigma = (\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$ be a sofic approximation.*

Then $a_\Sigma(\mu_0) \geq a_\Sigma(\mu_t)$ for all $t \geq 0$.

Proof. If $a_\Sigma(\mu_0) = +\infty$ then the result is trivial, so suppose $a_\Sigma(\mu_0) < +\infty$. This means that for any $\mathcal{O} \ni \mu_0$ we have $\Omega(\sigma_n, \mathcal{O}) \neq \emptyset$ for all large enough n .

Let \mathcal{U}_t be an arbitrary weak-open neighborhood of μ_t . By Proposition 2.3 there exists $\mathcal{U}_0 \ni \mu_0$ such that, for all large enough n , we have $\zeta_t \in \Omega(\sigma_n, \mathcal{U}_t)$ whenever $\zeta_0 \in \Omega(\sigma_n, \mathcal{U}_0)$.

Suppose n is large enough that $\Omega(\sigma_n, \mathcal{U}_0) \neq \emptyset$. Since $F_0 \leq 0$, the previous proposition implies that for any $\zeta_0 \in \text{Prob}(\mathbb{A}^{V_n})$ and any $t > 0$

$$\frac{d}{dt} A(\zeta_t) \leq 0.$$

Therefore for any $t \geq 0$

$$A(\zeta_0) \geq A(\zeta_t),$$

and hence

$$\inf_{\zeta \in \Omega(\sigma_n, \mathcal{U}_0)} A(\zeta) \geq \inf_{\zeta \in \Omega(\sigma_n, \mathcal{U}_t)} A(\zeta).$$

Now by definition of a_Σ we have

$$\begin{aligned}
a_\Sigma(\mu_0) &= \sup_{\mathcal{O} \ni \mu_0} \limsup_{n \rightarrow \infty} \frac{1}{|V_n|} \inf_{\zeta \in \Omega(\sigma_n, \mathcal{O})} A(\zeta) \\
&\geq \limsup_{n \rightarrow \infty} \frac{1}{|V_n|} \inf_{\zeta \in \Omega(\sigma_n, \mathcal{U}_0)} A(\zeta)
\end{aligned}$$

$$\geq \limsup_{n \rightarrow \infty} \frac{1}{|V_n|} \inf_{\zeta \in \Omega(\sigma_n, \mathcal{U}_t)} A(\zeta).$$

Taking the supremum over \mathcal{U}_t gives the result. \square

By a more careful analysis we can get the following proposition, the second part of which may be interpreted as semicontinuity of the time derivative of $a_\Sigma(\mu_t)$ as a function of the measure:

Proposition 3.3. *Suppose $\mu \in \text{Prob}^\Gamma(A^\Gamma)$ is not Gibbs. Then there exist $c, T > 0$ and an open neighborhood $\mathcal{O} \ni \mu$ such that if $\mu_0 \in \mathcal{O}$ then*

1. *There exists $\delta > 0$ such that for any $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$ with $\Delta^\sigma < \delta$ and any $\zeta_0 \in \Omega(\sigma, \mathcal{O})$ we have $A(\zeta_t) \geq A(\zeta_0) + ct|V|$ for all $t \in [0, T]$.*
2. *If $\Sigma = (\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$ is a sofic approximation then $a_\Sigma(\mu_0) \geq a_\Sigma(\mu_t) + ct$ for all $t \in [0, T]$.*

Here we take the convention that $+\infty + ct = +\infty$.

Proof. Since μ is not Gibbs, there exists R such that either μ_R does not have full support or $\Delta_a^R(\mu_R) < 0$ for some $a \in A$. We will come back to these two cases in a moment, but for now let R be fixed so that one of them occurs. We may assume $R \geq 1$.

Fix $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$. Let

$$s = \min \left\{ \frac{\exp\{-\Phi_e(\mathbf{x})\}}{\sum_{b \in A} \exp\{-\Phi_e(\mathbf{x}^{e \rightarrow b})\}} : \mathbf{x} \in A^{B(e,1)} \right\} > 0,$$

and call $v \in V$ *good* if $B^\sigma(v, R) \cong B^\Gamma(e, R)$, and let V' be the set of such v . Then

$$\frac{d}{dt} A(\zeta_t) \leq s \sum_{\substack{\mathbf{x}, v, a \\ v \text{ good}}} F_0 \left(\frac{\exp\{-\Phi_v(\mathbf{x})\}}{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow a})\}} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \zeta_t\{\mathbf{x}\}.$$

Let $\widetilde{P_\zeta^{\sigma, R}} \in \text{Prob}(A^{B^\Gamma(e, R)})$ be given by

$$\widetilde{P_\zeta^{\sigma, R}}\{\mathbf{y}\} = \frac{1}{|V'|} \sum_{\substack{\mathbf{x}, v \\ v \text{ good} \\ \mathbf{x}|_{B(v, R)} = \mathbf{y}}} \zeta\{\mathbf{x}\}$$

Note that $\widetilde{P_\zeta^{\sigma, R}}$ is close to the $B^\Gamma(e, R)$ -marginal of P_ζ^σ in total variation distance if most vertices are good. Then, applying Jensen's inequality,

$$\begin{aligned} \frac{d}{dt} A(\zeta_t) &\leq s \sum_{a \in A} \sum_{\mathbf{y} \in A^{B^\Gamma(e, R)}} \sum_{\substack{\mathbf{x}, v \\ v \text{ good} \\ \mathbf{x}|_{B(v, R)} = \mathbf{y}}} F_0 \left(\frac{\exp\{-\Phi_v(\mathbf{x})\}}{\exp\{-\Phi_v(\mathbf{x}^{v \rightarrow a})\}} \frac{\zeta_t\{\mathbf{x}^{v \rightarrow a}\}}{\zeta_t\{\mathbf{x}\}} \right) \zeta_t\{\mathbf{x}\} \\ &\leq s |V'| \sum_{a \in A} \sum_{\mathbf{y} \in A^{B^\Gamma(e, R)}} \widetilde{P_{\zeta_t}^{\sigma, R}}\{\mathbf{y}\} \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{\widetilde{P_{\zeta_t}^{\sigma, R}}\{\mathbf{y}^{e \rightarrow a}\}}{\widetilde{P_{\zeta_t}^{\sigma, R}}\{\mathbf{y}\}} \right). \end{aligned}$$

We now consider the two cases mentioned above. First, if μ_R does not have full support, there exist $\mathbf{y} \in \mathbb{A}^{B(e, R)}$, $v \in B(e, R)$ and $a \in \mathbb{A}$ such that $\mu_R\{\mathbf{y}\} \neq 0$ but $\mu_R\{\mathbf{y}^{v \rightarrow a}\} = 0$ (this is the same principle used at the beginning of the proof of Proposition 3.1). Using translation-invariance of μ , we may assume $v = e$, increasing R if necessary: if R' is large enough that $B(e, R) \subset B(v, R')$ then there is some extension \mathbf{y}' of \mathbf{y} such that $\mu_{B(v, R')}\{\mathbf{y}'\} \neq 0$ and $\mu_{B(v, R')}\{\mathbf{y}'^{v \rightarrow a}\} = 0$. We could then take R to be R' and \mathbf{y} to be a translate of \mathbf{y}' with v shifted to e . Then

$$\mu_R\{\mathbf{y}\} \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{\mu_R\{\mathbf{y}^{e \rightarrow a}\}}{\mu_R\{\mathbf{y}\}} \right) = -\mu_R\{\mathbf{y}\} < 0.$$

By continuity of F_0 , there exists $\varepsilon > 0$ such that

$$|a - \mu_R\{\mathbf{y}\}| < \varepsilon \text{ and } 0 \leq b < \varepsilon \Rightarrow a \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{b}{a} \right) < -\frac{\mu_R\{\mathbf{y}\}}{2}.$$

In particular, if $\|\widetilde{P_{\zeta_t}^{\sigma, R}} - \mu_R\|_{\text{TV}} < \varepsilon$ then

$$\frac{d}{dt} A(\zeta_t) < -s|V'| \frac{\mu_R\{\mathbf{y}\}}{2}.$$

In this case we will take $c = s \mu_R\{\mathbf{y}\}/4$.

Now consider the other case, in which $\Delta_a^R(\mu_R) < 0$ for some a . By definition of $\Delta_a^R(\mu_R)$, we can pick \mathbf{y} such that

$$C := \mu_R\{\mathbf{y}\} \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{\mu_R\{\mathbf{y}^{e \rightarrow a}\}}{\mu_R\{\mathbf{y}\}} \right) < 0.$$

and proceed in the same way, picking $\varepsilon > 0$ such that

$$|a - \mu_R\{\mathbf{y}\}| < \varepsilon \text{ and } |b - \mu_R\{\mathbf{y}^{e \rightarrow a}\}| < \varepsilon \Rightarrow a \cdot F_0 \left(\frac{\exp\{-\Phi_e(\mathbf{y})\}}{\exp\{-\Phi_e(\mathbf{y}^{e \rightarrow a})\}} \frac{b}{a} \right) < -\frac{C}{2}.$$

In particular, if $\|\widetilde{P_{\zeta_t}^{\sigma, R}} - \mu_R\|_{\text{TV}} < \varepsilon$ then

$$\frac{d}{dt} A(\zeta_t) < -s|V'| \frac{C}{2}.$$

In this case we will take $c = s C/4$.

In either case, we now have chosen $\varepsilon > 0$ such that if $\|\widetilde{P_{\zeta_t}^{\sigma, R}} - \mu_R\|_{\text{TV}} < \varepsilon$ then

$$\frac{d}{dt} A(\zeta_t) < -2|V'|c.$$

Let $\mathcal{O}_1 = \{v \in \text{Prob}(\mathbb{A}^\Gamma) : \|v_R - \mu_R\|_{\text{TV}} < \varepsilon\}$. By continuity of $(\mu_0, t) \mapsto \mu_t$, we can pick \mathcal{O}, T such that if $\mu_0 \in \mathcal{O}$ then $\mu_t \in \mathcal{O}_1$ for all $t \in [0, T]$.

Fix $\mu_0 \in \mathcal{O}$. By Proposition 2.3, for any $\eta > 0$ there exists $\mathcal{U} \ni \mu_0$ and $\delta > 0$ such that if $\zeta_0 \in \Omega(\sigma_n, \mathcal{U})$ and $\Delta^\sigma < \delta$ then $d(P_{\zeta_t}^\sigma, \mu_t) < \eta$ for all $t \in [0, T]$. If we pick η small enough, this implies $\zeta_t \in \Omega(\sigma, \mathcal{O}_1)$ for all $t \in [0, T]$. Hence if $\zeta_0 \in \Omega(\sigma_n, \mathcal{U})$ and $\Delta^\sigma < \delta$ then for any $t \in [0, T]$

$$A(\zeta_t) - A(\zeta_0) \leq -2|V'|ct.$$

Now by definition we have

$$|V'| = (1 - \delta_R^\sigma) |V|.$$

If δ is small enough then $\Delta^\sigma < \delta$ implies $\delta_R^\sigma < 1/2$, so that $2|V'| > |V|$. This completes the proof of the first part.

For the second part, since Σ is a sofic approximation, for all large enough n we have $\Delta^{\sigma_n} < \delta$ and therefore

$$\begin{aligned} A(\zeta_0) &\geq A(\zeta_t) + |V_n|ct \\ &\geq \inf_{\zeta \in \Omega(\sigma_n, \mathcal{B}(\mu_t, \eta))} A(\zeta) + |V_n|ct \end{aligned}$$

for all $t \in [0, T]$. Then, since the first limit in the definition of a_Σ is a supremum

$$\begin{aligned} a_\Sigma(\mu_0) &\geq \limsup_{n \rightarrow \infty} \inf_{\zeta_0 \in \Omega(\sigma_n, \mathcal{U})} \frac{1}{|V_n|} A(\zeta_0) \\ &\geq \limsup_{n \rightarrow \infty} \inf_{\zeta \in \Omega(\sigma_n, \mathcal{B}(\mu_t, \eta))} \frac{1}{|V_n|} A(\zeta) + ct. \end{aligned}$$

Taking η to zero gives the result. \square

Proof of Theorem. Suppose for the sake of contradiction that μ_t does not converge to the set of Gibbs measures; then we can pick some $\nu \in \text{Prob}^\Gamma(\mathbb{A}^\Gamma)$ which is not a Gibbs measure but for every open $\mathcal{O} \ni \nu$ we have $\mu_t \in \mathcal{O}$ for arbitrarily large t .

By Proposition 3.3, we can pick $\delta, T > 0$ and an open neighborhood $\mathcal{O} \ni \nu$ such that for every t with $\mu_t \in \mathcal{O}$ we have $a_\Sigma(\mu_t) \geq a_\Sigma(\mu_{t+T}) + \delta T$.

On the other hand, under the assumption that $a_\Sigma(\mu_s) < +\infty$, the set $\{a_\Sigma(\mu_t) : t \geq s\}$ is bounded: an upper bound is $a_\Sigma(\mu_s)$ by Proposition 3.2, and a lower bound is $u^{\min} - \log |A|$. This is a contradiction, so the theorem follows. \square

4. Connection to Property PA

We first prove the following:

Proposition 4.1. *A group Γ has property PA if and only if for any $\mu \in \text{Prob}^\Gamma(\mathbb{A}^\Gamma)$ there exists a sofic approximation $\Sigma = (\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$ and a sequence $(\mathbf{x}_n \in A^{V_n})_{n \in \mathbb{N}}$ with $P_{\mathbf{x}_n}^{\sigma_n} \xrightarrow{wk} \mu$.*

Some ideas for this proof were shared with me by Lewis Bowen.

Proof. The ‘if’ direction is clear, since each $P_{\mathbf{x}_n}^{\sigma_n}$ is periodic.

For the other direction, suppose Γ has property PA and let $\mu \in \text{Prob}^\Gamma(\mathbb{A}^\Gamma)$. By definition of property PA, we can pick a sequence of periodic measures $(\mu_n)_{n \in \mathbb{N}}$ converging to μ .

Fix $n \in \mathbb{N}$. The support of μ_n consists of finitely many orbits under Γ ; let $\{\mathbf{y}_1, \dots, \mathbf{y}_k\} \subset \mathbb{A}^\Gamma$ be a set which contains exactly one element of each orbit, and denote the (finite) orbits by $\Gamma \mathbf{y}_i$. Then we can write

$$\mu_n = \sum_{i=1}^k a_i \text{Unif}(\Gamma \mathbf{y}_i).$$

Pick natural numbers m_1, \dots, m_k , and let V_n be the disjoint union of m_i copies of $\Gamma \mathbf{y}_i$ for each i . Let $\sigma_n \in \text{Hom}(G, \text{Sym}(V_n))$ act separately on each copy of each orbit. Let $\mathbf{x}_n \in \mathbb{A}^{V_n}$ be given by

$$\mathbf{x}_n(v) = v(e).$$

Then

$$P_{\mathbf{x}_n}^{\sigma_n} = \sum_{i=1}^k \frac{m_i}{\sum_{j=1}^k m_j} \text{Unif}(\Gamma \mathbf{y}_i),$$

so if m_1, \dots, m_k are chosen appropriately then we can ensure $P_{\mathbf{x}_n}^{\sigma_n} \rightarrow \mu$.

But $\Sigma = \{\sigma_n\}$ as just constructed is not necessarily a sofic approximation. We can fix this using the following trick: let $\nu \in \text{Prob}(\{0, 1\}^\Gamma)$ be the product measure with uniform base. Then then above argument, applied to $\mu \times \nu$ instead of μ , implies the existence of sequences $(\sigma_n \in \text{Hom}(\Gamma, \text{Sym}(V_n)))_{n \in \mathbb{N}}$ and $\mathbf{z}_n \in (\mathbb{A} \times \{0, 1\})^{V_n}$ with $P_{\mathbf{z}_n}^{\sigma_n} \rightarrow \mu \times \nu$. If we write $\mathbf{z}_n = (\mathbf{x}_n, \mathbf{y}_n)$ with $\mathbf{x}_n \in \mathbb{A}^{V_n}$ and $\mathbf{y}_n \in \{0, 1\}^{V_n}$, then $P_{\mathbf{x}_n}^{\sigma_n} \rightarrow \mu$ and $P_{\mathbf{y}_n}^{\sigma_n} \rightarrow \nu$. We will show that the latter fact implies $\Delta^{\sigma_n} \rightarrow 0$ (i.e. Σ is a sofic approximation). Suppose $v \in V_n$, $\gamma \in \Gamma$ are such that $\sigma_n^\gamma v = v$. Then for any $\beta \in \Gamma$,

$$(\Pi_v^{\sigma_n} \mathbf{y}_n)(\beta\gamma) = \mathbf{y}_n(\sigma_n^{\beta\gamma} v) = \mathbf{y}_n(\sigma_n^\beta v) = (\Pi_v^{\sigma_n} \mathbf{y}_n)(\beta).$$

In particular, for any finite set $D \subset \Gamma$ we have

$$P_{\mathbf{y}_n}^{\sigma_n} \{\mathbf{w} \in \{0, 1\}^\Gamma : \mathbf{w}(\beta\gamma) = \mathbf{w}(\beta) \forall \beta \in D\} \geq \frac{1}{|V_n|} |\{v \in V_n : \sigma_n^\gamma v = v\}|.$$

But by assumption, as $n \rightarrow \infty$ the left-hand side converges to

$$\begin{aligned} \nu\{\mathbf{w} \in \{0, 1\}^\Gamma : \mathbf{w}(\beta\gamma) = \mathbf{w}(\beta) \forall \beta \in D\} &\leq \nu\{\mathbf{w} : \mathbf{w}(\beta\gamma) = \mathbf{w}(\beta) \forall \beta \in D \text{ s.t. } \beta\gamma \notin D\} \\ &= 2^{-|D\gamma \setminus D|}, \end{aligned}$$

and hence

$$\limsup_{n \rightarrow \infty} \frac{1}{|V_n|} |\{v \in V_n : \sigma_n^\gamma v = v\}| \leq 2^{-|D\gamma \setminus D|}.$$

As long as $\gamma \neq e$, the set D can be chosen to make $|D\gamma \setminus D|$ arbitrarily large, so that

$$\lim_{n \rightarrow \infty} \frac{1}{|V_n|} |\{v \in V_n : \sigma_n^\gamma v = v\}| = 0.$$

For any $R \in \mathbb{N}$, it can be checked that if $\sigma_n^\gamma v \neq v$ for all $\gamma \neq e$ such that $|\gamma| \leq 2R+1$ then the map

$$\begin{aligned} B^\Gamma(e, R) &\rightarrow B^\sigma(v, R) \\ \gamma &\mapsto \sigma^\gamma v \end{aligned}$$

is an isomorphism of the (labeled, directed) induced subgraphs. Therefore

$$\delta_R^{\sigma_n} \leq \sum_{\gamma \in B^\Gamma(e, 2R+1) \setminus \{e\}} \frac{1}{|V_n|} |\{v \in V_n : \sigma_n^\gamma v = v\}| \rightarrow 0,$$

which, since R is arbitrary, implies $\Delta^{\sigma_n} \rightarrow 0$. □

4.1. Proof of Proposition 2.4. Note $a_\Sigma(\mu) = +\infty$ if and only if there exists an open neighborhood $\mathcal{O} \ni \mu$ such that $\Omega(\sigma_n, \mathcal{O})$ is empty for infinitely many n . Therefore if $a_\Sigma(\mu) < +\infty$ there exists a sequence $\zeta_n \in \text{Prob}(\mathbb{A}^{V_n})$ with $P_{\zeta_n}^{\sigma_n} \rightarrow \mu$. Since each $P_{\zeta_n}^{\sigma_n}$ is periodic, this shows that if for every μ there exists Σ with $a_\Sigma(\mu) < +\infty$ then Γ has property PA.

Conversely, if Γ has property PA and $\mu \in \text{Prob}^\Gamma(\mathbb{A}^\Gamma)$ is given, then by the above proposition we can pick a sofic approximation Σ and $(\mathbf{x}_n)_{n \in \mathbb{N}}$ with $P_{\mathbf{x}_n}^{\sigma_n} \rightarrow \mu$. But then for any open $\mathcal{O} \ni \mu$ we have $\delta_{\mathbf{x}_n} \in \Omega(\sigma_n, \mathcal{O})$ for all large enough n , so $a_\Sigma(\mu) < +\infty$.

5. Proofs of Statements Involving Infinitary Dynamics

We first give some additional setup regarding the Glauber dynamics on \mathbb{A}^Γ . First, recall that on an infinite graph we must first define the Markov generator on a ‘core’ of ‘smooth’ functions. Let $C(\mathbb{A}^\Gamma)$ denote the space of continuous real-valued functions on \mathbb{A}^Γ , with the supremum norm $\|\cdot\|_\infty$. The smooth functions are defined as follows: given $f \in C(\mathbb{A}^\Gamma)$ and $v \in \Gamma$, let

$$\Delta_f(v) = \sup\{|f(\eta_1) - f(\eta_2)| : \eta_1(u) = \eta_2(u) \ \forall u \neq v\},$$

$$\|f\| = \sum_{v \in \Gamma} \Delta_f(v),$$

and

$$D(\mathbb{A}^\Gamma) = \{f \in C(\mathbb{A}^\Gamma) : \|f\| < \infty\}.$$

Every function which depends on only finitely many coordinates is in $D(\mathbb{A}^\Gamma)$, so $D(\mathbb{A}^\Gamma)$ is dense in $C(\mathbb{A}^\Gamma)$. For every $f \in D(\mathbb{A}^\Gamma)$ the series defining Ωf converges absolutely, and $\Omega f \in C(\mathbb{A}^\Gamma)$.

Note that the condition $\|f\| < \infty$ does not imply that f is continuous; in fact for every tail-measurable f we have $\|f\| = 0$.

Continuing to follow mostly the notation from Liggett’s book, let

$$c_u(v) = \sup\{\|c_u(\eta_1, \cdot) - c_u(\eta_2, \cdot)\|_{TV} : \eta_1(\gamma) = \eta_2(\gamma) \ \forall \gamma \neq v\}.$$

Then

$$\Theta\beta(u) = \sum_{v \in \Gamma} \beta(v) c_u(v)$$

defines a bounded linear operator on $\ell^1(\Gamma)$.

The closure $\bar{\Omega}^\Gamma$ is a Markov generator, so its domain is a dense subset of the continuous functions $C(\mathbb{A}^\Gamma)$ and the range of $I - \lambda \bar{\Omega}^\Gamma$ is all of $C(\mathbb{A}^\Gamma)$ for all $\lambda \geq 0$. We also have $\|f\| \leq \|(I - \lambda \bar{\Omega}^\Gamma)f\|$ for all $\lambda \geq 0$ [9, comment after Definition 2.1 of Chapter 1]. In particular $I - \lambda \bar{\Omega}^\Gamma$ is injective. An important consequence is that we have a contraction $(I - \lambda \bar{\Omega}^\Gamma)^{-1} : C(\mathbb{A}^\Gamma) \rightarrow C(\mathbb{A}^\Gamma)$.

5.1. Approximate Equivariance (Proof of Theorem 2.1). For $\beta \in \mathbb{R}^\Gamma$, let

$$\|\beta\| = \sup_{\gamma \in \Gamma} |\beta(\gamma)| (3r)^{|\gamma|}.$$

Lemma 5.1. For any continuous $g: A^\Gamma \rightarrow \mathbb{R}$,

$$\frac{1}{3} \|g\| \leq \| * \| \Delta_g = |g|_{\text{Lip}}.$$

In particular, every Lipschitz function is in $D(A^\Gamma)$.

Proof. For the inequality:

$$\|g\| = \sum_{\gamma} \Delta_g(\gamma) \leq \left(\sup_{\gamma} \Delta_g(\gamma) (3r)^{|\gamma|} \right) \sum_{\gamma} (3r)^{-|\gamma|} \leq 3 \sup_{\gamma} \Delta_g(\gamma) (3r)^{|\gamma|}.$$

Now similarly, for any $\mathbf{x}, \mathbf{y} \in A^\Gamma$

$$\begin{aligned} |g(\mathbf{x}) - g(\mathbf{y})| &\leq \sum_{\gamma: \mathbf{x}(\gamma) \neq \mathbf{y}(\gamma)} \Delta_g(\gamma) \quad (\text{using continuity}) \\ &\leq \sup_{\gamma} \Delta_g(\gamma) (3r)^{|\gamma|} \sum_{\gamma: \mathbf{x}(\gamma) \neq \mathbf{y}(\gamma)} (3r)^{-|\gamma|} \\ &= \sup_{\gamma} \Delta_g(\gamma) (3r)^{|\gamma|} \cdot d(\mathbf{x}, \mathbf{y}), \end{aligned}$$

so

$$|g|_{\text{Lip}} \leq \sup_{\gamma} \Delta_g(\gamma) (3r)^{|\gamma|} = \|\Delta_g\|.$$

The converse inequality follows from the fact that for any $\gamma \in \Gamma$

$$\Delta_g(\gamma) \leq (3r)^{-|\gamma|} |g|_{\text{Lip}}.$$

□

Lemma 5.2. With respect to the norm $\|\cdot\|$ on \mathbb{R}^Γ , Θ has operator norm at most

$$M := \sup_{\gamma} \sum_{h \in \Gamma} c_{\gamma}(h) (3r)^{d(h, \gamma)} < \infty.$$

Proof. For any $\gamma \in \Gamma$,

$$\begin{aligned} |[\Theta\beta](\gamma) \cdot (3r)^{|\gamma|}| &\leq \sum_{h \in \Gamma} |\beta(h)| c_{\gamma}(h) (3r)^{|\gamma|} \\ &\leq \sum_{h \in \Gamma} |\beta(h)| c_{\gamma}(h) (3r)^{|h| + d(h, \gamma)} \\ &\leq \|\beta\| \sum_{h \in \Gamma} c_{\gamma}(h) (3r)^{d(h, \gamma)}, \end{aligned}$$

so, taking the supremum over γ , we see that $\|\Theta\beta\| \leq \|\beta\| M$ and hence the operator norm is bounded by M .

Finiteness of M follows from the fact that always $c_{\gamma}(h) \leq 2$, and $c_{\gamma}(h) = 0$ if h, γ are not adjacent. So for any γ

$$\sum_{h \in \Gamma} c_{\gamma}(h) (3r)^{d(h, \gamma)} \leq 2 \cdot 2r \cdot (3r)^1 = 12r^2.$$

□

We can now give a proof of Lemma 2.2:

Proof of Lemma 2.2. By [9, Theorem 3.9(c) of Chapter 1], we have the pointwise inequality

$$\Delta_{S^\Gamma(t)f} \leq \exp(t\Theta)\Delta_f \quad \forall f \in D(\mathbb{A}^\Gamma).$$

Taking $\|\cdot\|$ norms of both sides gives, by Lemma 5.1,

$$\|S^\Gamma(t)f\|_{\text{Lip}} \leq \exp(Mt)\|f\|_{\text{Lip}}.$$

The result follows from this and the definition of \bar{d} . \square

Proposition 5.3. *For all small enough $\lambda > 0$, for all $g \in D(\mathbb{A}^\Gamma)$ we have*

$$\|(I - \lambda\bar{\Omega})^{-k}g\|_{\text{Lip}} \leq [1 - \lambda M]^{-k}\|g\|_{\text{Lip}}.$$

Proof. Recall from [9, proof of Theorem 3.9 of Chapter 1] that for all small enough $\lambda > 0$

$$\Delta_{(I - \lambda\bar{\Omega})^{-k}g} \leq [(1 + \lambda\varepsilon)I - \lambda\Theta]^{-k}\Delta_g \quad (\text{pointwise})$$

for any $g \in D(\mathbb{A}^\Gamma)$. If we apply the $\|\cdot\|$ norm to both sides we get, by Lemma 5.1,

$$\|(I - \lambda\bar{\Omega})^{-k}g\|_{\text{Lip}} \leq [1 - \lambda(M - \varepsilon)]^{-k}\|g\|_{\text{Lip}}.$$

The stated bound follows after dropping ε , which is positive. \square

Define

$$\|f\|_R = \sum_{|\gamma| \geq R} \Delta_f(\gamma).$$

If $f \in D(\mathbb{A}^\Gamma)$ then $\lim_{R \rightarrow \infty} \|f\|_R = 0$; if f is Lipschitz then for any $R \geq 0$

$$\|f\|_R \leq 3\|f\|_{\text{Lip}}(2/3)^R.$$

The following result establishes an approximate equivariance of $P_{\mathbf{x}}^\sigma$ with the generator:

Proposition 5.4. *Let V be a finite set and $\sigma \in \text{Hom}(\Gamma, \text{Sym}(V))$. For any $f \in D(\mathbb{A}^\Gamma)$, $R \in \mathbb{N}$, and $\mathbf{x} \in \mathbb{A}^V$,*

$$\|\Omega^\sigma P_{\mathbf{x}}^\sigma f - P_{\mathbf{x}}^\sigma \Omega^\Gamma f\| \leq 3\|f\|_R + 2\delta_R^\sigma \|f\|.$$

Proof. From the definitions of Ω^σ and $P_{\mathbf{x}}^\sigma$,

$$\begin{aligned} \Omega^\sigma P_{\mathbf{x}}^\sigma f &= \sum_{v \in V} \sum_{\mathbf{a} \in \mathbb{A}} c_v(\mathbf{x}, \mathbf{a}) [P_{\mathbf{x}^{v \rightarrow \mathbf{a}}}^\sigma f - P_{\mathbf{x}}^\sigma f] \\ &= \frac{1}{|V|} \sum_{v, w \in V} \sum_{\mathbf{a} \in \mathbb{A}} c_v(\mathbf{x}, \mathbf{a}) [f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow \mathbf{a}}) - f(\Pi_w^\sigma \mathbf{x})]. \end{aligned}$$

We can compare this to if the sum is restricted to pairs v, w which are nearby in the graph σ :

$$\begin{aligned} & \left| \Omega^\sigma P_{\mathbf{x}}^\sigma f - \frac{1}{|V|} \sum_{w \in V} \sum_{v \in B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) [f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right| \\ & \leq \frac{1}{|V|} \sum_{w \in V} \sum_{v \notin B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) |f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})| \end{aligned}$$

Now since the labelings $\mathbf{x}^{v \rightarrow a}$ and \mathbf{x} differ only at v , their lifted labelings $\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}$ and $\Pi_w^\sigma \mathbf{x}$ differ only at preimages of v under the map $\gamma \mapsto \sigma^\gamma w$. Let $\Pi_w^\sigma \{v\} \subset \Gamma$ denote the set of these preimages. Then the above is bounded by

$$\frac{1}{|V|} \sum_{w \in V} \sum_{v \notin B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) \sum_{\gamma \in \Pi_w^\sigma \{v\}} \Delta f(\gamma) = \frac{1}{|V|} \sum_{w \in V} \sum_{v \notin B^\sigma(w, R-1)} \sum_{\gamma \in \Pi_w^\sigma \{v\}} \Delta f(\gamma).$$

Since for each w the sets in the collection $\{\Pi_w^\sigma \{v\} : v \notin B^\sigma(w, R-1)\}$ are disjoint and contained in the complement of $B^\Gamma(e, R-1)$, we can bound this by

$$\frac{1}{|V|} \sum_{w \in V} \sum_{\gamma \notin B^\Gamma(e, R-1)} \Delta f(\gamma) = \|f\|_R.$$

Now suppose $w \in V$ is such that $B^\sigma(w, R) \cong B^\Gamma(e, R)$: then for each $v \in B^\sigma(w, R-1)$ the intersection $B^\Gamma(e, R-1) \cap \Pi_w^\sigma \{v\}$ consists of a single point, which we call γ^v . We then have $c_v(\mathbf{x}, a) = c_{\gamma^v}(\Pi_w^\sigma \mathbf{x}, a)$. From this we can get

$$\begin{aligned} & \left| \sum_{v \in B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) [f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right. \\ & \quad \left. - \sum_{\gamma \in B^\Gamma(e, R-1)} \sum_{a \in A} c_\gamma(\Pi_w^\sigma \mathbf{x}, a) [f((\Pi_w^\sigma \mathbf{x})^{\gamma \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right| \\ & = \left| \sum_{v \in B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) [f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}) - f((\Pi_w^\sigma \mathbf{x})^{\gamma^v \rightarrow a})] \right| \\ & \leq \sum_{v \in B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) |f(\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}) - f((\Pi_w^\sigma \mathbf{x})^{\gamma^v \rightarrow a})| \quad (*) \end{aligned}$$

Now our construction also guarantees that the labelings $\Pi_w^\sigma \mathbf{x}^{v \rightarrow a}$ and $(\Pi_w^\sigma \mathbf{x})^{\gamma^v \rightarrow a}$ differ only at sites in $\Pi_w^\sigma \{v\}$ other than γ^v , all of which lie outside $B^\Gamma(e, R-1)$. Therefore we can continue

$$\begin{aligned} (*) & \leq \sum_{v \in B^\sigma(w, R-1)} \sum_{a \in A} c_v(\mathbf{x}, a) \sum_{\substack{\gamma \in \Pi_w^\sigma \{v\} \\ |\gamma| \geq R}} \Delta f(\gamma) \\ & \leq \sum_{|\gamma| \geq R} \Delta f(\gamma) = \|f\|_R. \end{aligned}$$

In the second-to-last line, we have again used that $\Pi_w^\sigma\{v_1\}$ and $\Pi_w^\sigma\{v_2\}$ are disjoint if $v_1 \neq v_2$.

For other ‘bad’ w where $B^\sigma(w, R) \not\cong B^\Gamma(e, R)$, approximating the sum over v by the sum over γ in this way may be inaccurate, but the fraction of $w \in V$ which are bad is only δ_R^σ . For these w we note that the magnitudes of both sums in $(*)$ can be bounded by $\sum_{\gamma \in \Gamma} \Delta_f(\gamma) = \|f\|$.

So far we have shown that

$$\begin{aligned} & \left| \Omega^\sigma P_{\mathbf{x}}^\sigma f - \frac{1}{|V|} \sum_{w \in V} \sum_{|\gamma| < R} \sum_{a \in A} c_\gamma(\Pi_w^\sigma \mathbf{x}, a) [f((\Pi_w^\sigma \mathbf{x})^{\gamma \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right| \\ & \leq 2\|f\|_R + 2\delta_R^\sigma \|f\|. \end{aligned}$$

To finish, we compare the second term on the left-hand side to $P_{\mathbf{x}}^\sigma \Omega^\Gamma f$ using an approach similar to above:

$$\begin{aligned} & \left| \frac{1}{|V|} \sum_{w \in V} \sum_{|\gamma| < R} \sum_{a \in A} c_\gamma(\Pi_w^\sigma \mathbf{x}, a) [f((\Pi_w^\sigma \mathbf{x})^{\gamma \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right. \\ & \quad \left. - \frac{1}{|V|} \sum_{w \in V} \sum_{\gamma \in \Gamma} \sum_{a \in A} c_\gamma(\Pi_w^\sigma \mathbf{x}, a) [f((\Pi_w^\sigma \mathbf{x})^{\gamma \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})] \right| \\ & \leq \frac{1}{|V|} \sum_{w \in V} \sum_{|\gamma| \geq R} \sum_{a \in A} c_\gamma(\Pi_w^\sigma \mathbf{x}, a) |f((\Pi_w^\sigma \mathbf{x})^{\gamma \rightarrow a}) - f(\Pi_w^\sigma \mathbf{x})| \\ & \leq \frac{1}{|V|} \sum_{w \in V} \sum_{|\gamma| \geq R} \Delta_f(\gamma) \\ & = \|f\|_R. \end{aligned}$$

□

Lemma 5.5. *For all small enough $\lambda > 0$, for any $m \in \mathbb{N}$ and $g \in \text{Lip}(A^\Gamma)$ we have*

$$\|(I - \lambda\Omega^\sigma)^{-m} P_{\mathbf{x}}^\sigma g - P_{\mathbf{x}}^\sigma (I - \lambda\bar{\Omega}^\Gamma)^{-m} g\|_{\ell^\infty(A^V)} \leq \lambda \Delta^\sigma |g|_{\text{Lip}} \sum_{k=1}^m (1 - \lambda M)^{-k}.$$

Proof. We use induction on m , starting with the base case $m = 1$. Throughout, we assume λ is small enough for Proposition 5.3 to apply.

Given $g \in \text{Lip}(A^\Gamma)$, let $f = (I - \lambda\bar{\Omega}^\Gamma)^{-1}g$. Then for any $R \in \mathbb{N}$

$$\begin{aligned} \|P_{\mathbf{x}}^\sigma g - (I - \lambda\Omega^\sigma)[P_{\mathbf{x}}^\sigma f]\|_{\ell^\infty(A^V)} &= \|P_{\mathbf{x}}^\sigma [f - \lambda\bar{\Omega}^\Gamma f] - (I - \lambda\Omega^\sigma)[P_{\mathbf{x}}^\sigma f]\|_{\ell^\infty(A^V)} \\ &= \lambda \|\Omega^\sigma P_{\mathbf{x}}^\sigma f - P_{\mathbf{x}}^\sigma \bar{\Omega}^\Gamma f\|_{\ell^\infty(A^V)} \\ &\leq \lambda(3\|f\|_R + 2\delta_R^\sigma \|f\|) \quad (\text{Prop. 5.4}) \\ &\leq \lambda(9 \cdot (2/3)^R + 6\delta_R^\sigma) |f|_{\text{Lip}}. \end{aligned}$$

Taking the infimum over R gives

$$\begin{aligned} \|P_{\mathbf{x}}^{\sigma} g - (I - \lambda\Omega^{\sigma})[P_{\mathbf{x}}^{\sigma} f]\|_{\ell^{\infty}(\mathbb{A}^V)} &\leq \lambda\Delta^{\sigma} |f|_{\text{Lip}} \\ &\leq \lambda\Delta^{\sigma} (1 - \lambda M)^{-1} |g|_{\text{Lip}}. \end{aligned} \quad (\text{Prop. 5.3})$$

Since $(I - \lambda\Omega^{\sigma})^{-1}$ is a contraction on $\ell^{\infty}(\mathbb{A}^V)$,

$$\begin{aligned} &\|(I - \lambda\Omega^{\sigma})^{-1} P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-1} g\|_{\ell^{\infty}} \\ &= \|(I - \lambda\Omega^{\sigma})^{-1} [P_{\mathbf{x}}^{\sigma} g - (I - \lambda\Omega^{\sigma})[P_{\mathbf{x}}^{\sigma} f]]\|_{\ell^{\infty}} \\ &\leq \lambda\Delta^{\sigma} (1 - \lambda M)^{-1} |g|_{\text{Lip}}. \end{aligned}$$

This proves the base case.

Now assuming the m case and the base case, we prove the $m + 1$ case:

$$\begin{aligned} &\|(I - \lambda\Omega^{\sigma})^{-(m+1)} P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-(m+1)} g\|_{\ell^{\infty}} \\ &= \|(I - \lambda\Omega^{\sigma})^{-1} [(I - \lambda\Omega^{\sigma})^{-m} P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-m} g] \\ &\quad + [(I - \lambda\Omega^{\sigma})^{-1} P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-m} g - P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-1} (I - \lambda\bar{\Omega}^{\Gamma})^{-m} g]\|_{\ell^{\infty}} \\ &\leq \|(I - \lambda\Omega^{\sigma})^{-m} P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} (I - \lambda\bar{\Omega}^{\Gamma})^{-m} g\|_{\ell^{\infty}} \quad (\text{contraction}) \\ &\quad + \lambda\Delta^{\sigma} (1 - \lambda M)^{-1} |(I - \lambda\bar{\Omega}^{\Gamma})^{-m} g|_{\text{Lip}} \quad (\text{base case}) \\ &\leq \lambda\Delta^{\sigma} |g|_{\text{Lip}} \sum_{k=1}^{m+1} (1 - \lambda M)^{-k}. \quad (\text{inductive hyp., Prop. 5.3}) \end{aligned}$$

This completes the induction. \square

Proof of Theorem 2.1 Given $g \in \text{Lip}_1(\mathbb{A}^{\Gamma})$, for all large enough m we can apply the previous lemma with $\lambda = t/m$, which gives

$$\|(I - \frac{t}{m}\Omega^{\sigma})^{-m} P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} (I - \frac{t}{m}\bar{\Omega}^{\Gamma})^{-m} g\|_{\ell^{\infty}(\mathbb{A}^V)} \leq \frac{t}{m}\Delta^{\sigma} |g|_{\text{Lip}} \sum_{k=1}^m (1 - \frac{t}{m}M)^{-k}.$$

Let $m \rightarrow \infty$. The left-hand side converges (by Hille-Yosida; [9, Theorem 2.9(b) of Chapter 1]) to $\|S^{\sigma}(t)P_{\mathbf{x}}^{\sigma} g - P_{\mathbf{x}}^{\sigma} S^{\Gamma}(t)g\|_{\ell^{\infty}}$ while the lim sup of the right-hand side is bounded by $\Delta^{\sigma} t e^{Mt} |g|_{\text{Lip}}$, since

$$\limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m (1 - \frac{t}{m}M)^{-k} \leq \limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m (1 - \frac{t}{k}M)^{-k} = e^{Mt}.$$

Since $g \in \text{Lip}(\mathbb{A}^{\Gamma})$ was arbitrary, the inequality of transportation distance follows.

Acknowledgements. Thanks to Tim Austin for many helpful conversations and for providing feedback on a preprint, and to Lewis Bowen for telling me about property PA and its relation to sofic approximations and for sharing comments on a preprint. Thanks also to Raimundo Briceño for helpful correspondence regarding the Gibbness of free energy minimizers, and to an anonymous referee for many helpful comments on an earlier version.

Funding This material is based upon work completed at the University of California, Los Angeles and supported by the National Science Foundation under Grant No. DMS-1855694.

Declarations

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

References

1. Ising, E.: Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.* **31**(1), 253–258 (1925). <https://doi.org/10.1007/BF02980577>
2. Holley, R.: Free energy in a Markovian model of a lattice spin system. *Commun. Math. Phys.* **23**(2), 87–99 (1971). <https://doi.org/10.1007/BF01877751>
3. Bowen, L.: Measure conjugacy invariants for actions of countable sofic groups. *J. Am. Math. Soc.* **23**(1), 217–217 (2010). <https://doi.org/10.1090/S0894-0347-09-00637-7>
4. Higuchi, Y., Shiga, T.: Some results on Markov processes of infinite lattice spin systems. *J. Math. Kyoto Univ.* **15**(1), 211–229 (1975). <https://doi.org/10.1215/kjm/1250523126>
5. Jahnke, B., Külske, C.: Attractor properties for irreversible and reversible interacting particle systems. *Commun. Math. Phys.* **366**(1), 139–172 (2019). <https://doi.org/10.1007/s00220-019-03352-4>
6. Caputo, P., Martinelli, F.: Phase ordering after a deep quench: The stochastic Ising and hard core gas models on a tree. *Probab. Theory Relat. Fields* **136**(1), 37–80 (2006). <https://doi.org/10.1007/s00440-005-0475-y>
7. Dembo, A., Montanari, A.: Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20**(2), 565–592 (2010). <https://doi.org/10.1214/09-AAP627>. [arXiv:0804.4726](https://arxiv.org/abs/0804.4726)
8. Austin, T.: Additivity properties of sofic entropy and measures on model spaces. In: *Forum of Mathematics, Sigma*, vol. 4 (2016). <https://doi.org/10.1017/fms.2016.18>
9. Liggett, T.M.: *Interacting Particle Systems*. *Classics in Mathematics*. Springer, Berlin (2005)
10. Georgii, H.-O.: *Gibbs Measures and Phase Transitions*, 2nd edn. *De Gruyter Studies in Mathematics*, vol. 9. de Gruyter, Berlin (2011)
11. Alpeev, A.: The Entropy of Gibbs Measures on Sofic Groups. *J. Math. Sci.* **215**(6), 649–658 (2016). <https://doi.org/10.1007/s10958-016-2871-5>
12. Montanari, A., Mossel, E., Sly, A.: The weak limit of Ising models on locally tree-like graphs. *Probab. Theory Relat. Fields* **152**(1–2), 31–51 (2012). <https://doi.org/10.1007/s00440-010-0315-6>
13. Bowen, L.: Entropy for expansive algebraic actions of residually finite groups. *Ergod. Theory Dynam. Syst.* **31**(03), 703–718 (2011). <https://doi.org/10.1017/S0143385710000179>
14. Hayes, B.: Fuglede–Kadison determinants and sofic entropy. *Geom. Funct. Anal.* **26**(2), 520–606 (2016). <https://doi.org/10.1007/s00039-016-0370-y>
15. Bowen, L.: Periodicity and circle packings of the hyperbolic plane. *Geom. Dedicata*. **102**(1), 213–236 (2003). <https://doi.org/10.1023/B:GEOM.0000006580.47816.e9>
16. Kechris, A.S.: Weak containment in the space of actions of a free group. *Israel J. Math.* **189**(1), 461–507 (2012). <https://doi.org/10.1007/s11856-011-0182-6>
17. Burton, P.J., Kechris, A.S.: Weak containment of measure-preserving group actions. *Ergod. Theory Dyn. Syst.* **40**(10), 2681–2733 (2020). <https://doi.org/10.1017/etds.2019.26>

Communicated by J. Ding