Equivalence of relative Gibbs and relative equilibrium measures for actions of countable amenable groups

Sebastián Barbieri   Ricardo Gómez   Brian Marcus   Siamak Taati

Abstract

We formulate and prove a very general *relative* version of the Dobrushin–Lanford–Ruelle theorem which gives conditions on constraints of configuration spaces over a finite alphabet such that for every absolutely summable relative interaction, every translation-invariant relative Gibbs measure is a relative equilibrium measure and vice versa. Neither implication is true without some assumption on the space of configurations. We note that the usual finite type condition can be relaxed to a much more general class of constraints. By "relative" we mean that both the interaction and the set of allowed configurations are determined by a random environment. The result includes many special cases that are well known. We give several applications including 1) Gibbsian properties of measures that maximize pressure among all those that project to a given measure via a topological factor map from one symbolic system to another; 2) Gibbsian properties of equilibrium measures for group shifts defined on arbitrary countable amenable groups; 3) A Gibbsian characterization of equilibrium measures in terms of equilibrium condition on lattice slices rather than on finite sets; 4) A relative extension of a theorem of Meyerovitch, who proved a version of the Lanford–Ruelle theorem which shows that every equilibrium measure on an arbitrary subshift satisfies a Gibbsian property on interchangeable patterns.

Keywords: Equilibrium measures, Gibbs measures, relative systems, disordered systems, random environments, thermodynamic formalism.

MSC2010: 37B10, 82B44, 37D35, 82B20, 60K35.

Last update: December 5, 2019

1 Introduction

The starting point of Gibbs's approach to equilibrium statistical physics is the postulate that the macroscopic state of a system at thermal equilibrium is appropriately described by a probability distribution that minimizes the free energy. An equivalent formulation is obtained by maximizing the *pressure*, that is, the difference between the entropy and a constant times the expected energy. In a lattice model in which the microscopic states are configurations of symbols on an infinite lattice (e.g., the Ising model), there are two interpretations of this hypothesis:

(i) *Local* maximization: the conditional pressure for every finite region of the lattice is maximized, so that every finite region is in equilibrium with its surrounding. This leads to the concept of *Gibbs measures.*

(ii) *Global* maximization: the average pressure per site (i.e., Kolmogorov–Sinai entropy minus expected energy per site) is maximized. The maximizing measures in this interpretation are referred to as the *equilibrium measures.*

The celebrated theorem of Dobrushin [9], Lanford and Ruelle [24] says that under broad conditions, equilibrium measures and shift-invariant Gibbs measures coincide (see [39]).

Theorem 1.1. *Let Σ be a finite set of symbols. Let $X \subseteq \Sigma^{\mathbb{Z}^d}$ be a d-dimensional subshift, Φ an absolutely summable interaction on X, and f_Φ an associated energy observable.*

(a) (Dobrushin theorem)
Assume that X is D-mixing. Then, every shift-invariant Gibbs measure for Φ is an equilibrium measure for f_Φ.

(b) (Lanford–Ruelle theorem)
Assume that X is a subshift of finite type (SFT). Then, every equilibrium measure for f_Φ is a Gibbs measure for Φ.

Here, X is the space of allowed configurations on the d-dimensional lattice. Neither direction is true in general and so some kind of restrictions, such as D-mixing in part (a) and SFT in part (b), on X, are required. Terminology used in the statement of Theorem 1.1, as well as other terminology used in this section, will be given in Section §2.

We generalize this theorem in several directions. First, we allow the lattice to be any countable amenable group. Second, we allow the presence of a random environment that imposes constraints on the allowed configurations and affects the energy, and prove the equivalence of local and global maximization *relative to* this environment. Third, we relax the "finite type" hypothesis in the Lanford–Ruelle direction to the much weaker topological Markov property, and discuss the relationship between this and related properties. We also give several applications.

To be specific, let $\mathbb{G}$ be a countable amenable group (e.g., $\mathbb{G} = \mathbb{Z}^d$ with $d = 1, 2, \ldots$), Σ a finite alphabet, and Θ a measurable space on which $\mathbb{G}$ acts via measurable maps. The group $\mathbb{G}$ also acts on $\Sigma^{\mathbb{G}}$ by translations. For each $\theta \in \Theta$, let $X_\theta \subseteq \Sigma^{\mathbb{G}}$ be a non-empty closed set such that $\Omega \triangleq \{(\theta, x) : \theta \in \Theta \text{ and } x \in X_\theta\}$ is measurable and $X_{g\theta} = \{gx : x \in X_\theta\}$ for every $\theta \in \Theta$ and $g \in \mathbb{G}$. We think of $x \in \Sigma^{\mathbb{G}}$ as a microscopic configuration of a physical system and $\theta \in \Theta$ as the external environment. The fact that X_θ is not required to be the entire $\Sigma^{\mathbb{G}}$ indicates the possibility of "hard" (or combinatorial) constraints that the environment can impose on the system. We refer to Ω as a *relative system.*

With suitably formulated relative versions of the hypotheses in Theorem 1.1, our generalization is as follows.

Theorem 1.2. *Let the environment Θ and relative system Ω be as formulated above. Let ν be a $\mathbb{G}$-invariant probability measure on Θ. Let Φ be an absolutely summable relative interaction on Ω and f_Φ an associated energy observable.*

(a) (Relative Dobrushin theorem)
Assume that Ω is D-mixing relative to ν. Then, every $\mathbb{G}$-invariant relative Gibbs measure for Φ with marginal ν is an equilibrium measure for f_Φ relative to ν.

(b) (Relative Lanford–Ruelle theorem)
Assume that Θ is a standard Borel space. Assume further that Ω has the topological Markov property relative to ν. Then, every equilibrium measure for f_Φ relative to ν is a relative Gibbs measure for Φ with marginal ν.

The concepts of relative Gibbs measure, relative equilibrium measure, relative interaction, absolute summability, relative D-mixing and relative topological Markov property are natural analogues of the corresponding non-relative concepts in the relative setting. The proof of Theorem 1.2 is given in Section §3. In the non-relative setting, that is, when $\Theta \triangleq \{\theta\}$ and $\nu \triangleq \delta_\theta$, we recover a generalization of Theorem 1.1.

Many aspects of the above generalization are not new.

- Seppäläinen [42, Section §8] proved a relative version of the Dobrushin–Lanford–Ruelle theorem on the hyper-cubic lattice $\mathbb{Z}^d$. In his result, the alphabet is allowed to also be a complete separable metric space. On the other hand, the environment space in his setting is required to be a complete separable metric space, with $\mathbb{Z}^d$ acting by homeomorphisms, and the interaction is assumed to be continuous as a function of θ. Moreover, this setting does not allow hard constraints, i.e., $X_\theta = \Sigma^{\mathbb{Z}^d}$ for every θ (equivalently, $\Omega = \Theta \times \Sigma^{\mathbb{Z}^d}$). See also the paper by Zegarlinski [45].
- Moulin Ollagnier and Pinchon [34] (see Moulin Ollagnier [33, Thm. 7.2.5]) and Tempelman [43, Section §8]) have extended the (non-relative) Dobrushin–Lanford–Ruelle theorem to countable amenable groups, but again these results do not allow hard constraints. On the other hand, Templeman allows the alphabet to be an arbitrary σ-finite measure space.
- In the case where the acting group is $\mathbb{Z}$, very strong results are known even in the relative setting. In fact, in this case, if the variations of the energy observable decay rapidly enough, there is a unique equilibrium measure which coincides with a unique Gibbs measure, and this measure can be described explicitly as the unique fixed point of a Ruelle–Perron–Frobenius operator; see the survey by Kifer and Liu [22] (Theorem 4.1.1 and the following paragraph) and Kifer [21]. In this setting, these systems are known as random dynamical systems.
- In the case where the acting group is $\mathbb{Z}^d$, the framework of a relative system, much as we have formulated it above, is given in Kifer [19]. In this work the assumptions on Ω are in some ways more general and in some ways less general than ours.

For a given continuous observable f, an equilibrium measure achieves the supremum, over all $\mathbb{G}$-invariant measures μ, of the difference of the entropy of μ and the expected value of f with respect to μ. In the standard setting of a continuous $\mathbb{Z}^d$-action on a compact metric space, this supremum is characterized as an intrinsically defined notion of topological pressure for f. Similar variational principles have been established in the contexts of the above-mentioned results (see e.g. [35, 26, 20]). In our paper, we do not consider such variational principles (see however Prop. 3.2 for a special case); rather we focus on conditions which guarantee that every Gibbs measure is an equilibrium measure and that every equilibrium measure is a Gibbs measure. Also, the papers [42, 19] include, and are motivated by, large deviations principles, which is another topic that we do not consider.

Seppäläinen [42] gave several examples to which his result applied. This includes the Ising model with random external field and the Edwards–Anderson spin glass model in which the coupling parameters for neighboring spins are i.i.d. random. In these models, there are no hard constraints on the configurations. Below we give two examples in which there are hard constraints. In both of them we assume that $\mathbb{G}$ is finitely generated, and we consider a fixed finite symmetric generating set S with $S \not\ni 1_{\mathbb{G}}$. We consider the Cayley graph of $\mathbb{G}$ generated by S as a simple undirected graph with vertex set $\mathbb{G}$ and edge set $\mathbb{E} \triangleq \{\{a, b\} : a^{-1}b \in S\}$.

Example 1.3 (Ising model on percolation clusters)**.** Let $\Theta \triangleq \{\mathtt{0},\mathtt{1}\}^{\mathbb{G}}$, and let ν be a $\mathbb{G}$-invariant measure on Θ, for instance the Bernoulli measure with parameter $p \in (0,1)$. Let $\Sigma \triangleq \{-\mathtt{1},\mathtt{0},+\mathtt{1}\}$, and for $\theta \in \Theta$, let X_θ be the set of configurations $x \in \Sigma^{\mathbb{G}}$ for which $x_k = \mathtt{0}$ if and only if $\theta_k = \mathtt{0}$. Let $h \in \mathbb{R}$ and consider the relative interaction Φ defined by

$$\Phi_{\{k\}}(\theta, x) \triangleq -hx_k$$
$$\Phi_{\{i,j\}}(\theta, x) \triangleq -x_i x_j \qquad \text{if } \{i,j\} \in \mathbb{E},$$

and $\Phi_A(\theta, x) \triangleq 0$ whenever A is neither a singleton nor an edge in $\mathbb{G}$. So, the effect of the environment is simply to constrain the configurations. Observe that $\Phi(\theta,\cdot)$ is essentially the Ising interaction with external magnetic field h on the subgraph induced by $\{k \in \mathbb{G} : \theta_k = \mathtt{1}\}$.

This system has been studied as a model of a binary alloy consisting of a ferromagnetic and a non-magnetic metal [15, 25, 13, 16]. Each site is chosen at random to carry either a magnetic or a non-magnetic atom. The magnetic atoms interact with one another as in the Ising model, while the non-magnetic atoms do not interact.

In Section §2.5 we will see that the relative system Ω corresponding to this model satisfies the assumptions of both parts of Theorem 1.2 and thus the equilibrium measures for f_Φ relative to ν coincide with the $\mathbb{G}$-invariant relative Gibbs measures for Φ with marginal ν.

Let us point out that this example can be rephrased so that there are no constraints on the configurations and thus fits in the setting of Seppäläinen's result [42]. Let $\Theta \triangleq \{\mathtt{0},\mathtt{1}\}^{\mathbb{E}}$ be the set of bond configurations rather than site configurations. Let ν be the measure induced on Θ by a Bernoulli measure on the sites by letting a bond be open if and only if both its endpoints are open. Then we allow an Ising spin ($\pm\mathtt{1}$) at every site, except that the spins at closed sites will not interact with other spins, and therefore will be independent in both the relative equilibrium measures and the relative Gibbs measures. In this setting, the environment constrains the interaction but not the set of allowed configurations. ○

Example 1.4 (Random colorings of random graphs)**.** Let $\Theta \triangleq 2^{\mathbb{E}}$ denote the set of all subgraphs of $(\mathbb{G},\mathbb{E})$ that have the same vertex set $\mathbb{G}$. The group $\mathbb{G}$ acts on a subgraph $\theta \in \Theta$ by translation, that is, $g\theta \triangleq \{\{ga, gb\} : \{a,b\} \in \theta\}$. Let Σ be a finite set of colors. For a subgraph $\theta \in \Theta$, denote by X_θ the set of valid Σ-*colorings* of θ, that is, the configurations $x \in \Sigma^{\mathbb{G}}$ such that $x_a \neq x_b$ whenever $\{a,b\} \in \theta$. Clearly X_θ is closed and we have $X_{g\theta} = gX_\theta$ for each θ. Moreover, the set $\Omega \triangleq \{(\theta,x) : x \in X_\theta\}$ is measurable.

A $\mathbb{G}$-invariant measure ν on Θ may be viewed as the distribution of a stationary random subgraph $\boldsymbol{\theta}$ of $(\mathbb{G},\mathbb{E})$. We assume that $|\Sigma| > |S|$ and so $X_{\boldsymbol{\theta}}$ is almost surely non-empty. A *max-entropic random coloring* of $\boldsymbol{\theta}$ is a random configuration $\boldsymbol{x}$ from $\Sigma^{\mathbb{G}}$ defined in the same probability space as $\boldsymbol{\theta}$ such that $\boldsymbol{x} \in X_{\boldsymbol{\theta}}$ almost surely and the joint distribution μ of $(\boldsymbol{\theta},\boldsymbol{x})$ has maximum possible relative entropy $h_\mu(\Omega \,|\, \Theta)$. A *uniform-Gibbs coloring* is a random coloring of $\boldsymbol{\theta}$ such that for every finite set $A \subseteq \mathbb{G}$, the conditional distribution $\mathbb{P}(\boldsymbol{x}_A = \cdot \,|\, \boldsymbol{\theta}, \boldsymbol{x}_{A^\mathsf{c}})$ is almost surely uniform among all patterns $u \in \Sigma^A$ for which $u \vee \boldsymbol{x}_{A^\mathsf{c}}$ is a valid coloring of $\boldsymbol{\theta}$. We shall see in Section §2.5 that the assumptions of the relative Dobrushin–Lanford–Ruelle theorem hold and so a stationary random coloring is max-entropic if and only if it is uniform-Gibbs. ○

Now we consider some applications.

Equilibrium measures relative to a topological factor. Following Ledrappier and Walters [26, 44], a related notion of an equilibrium measure relative to an invariant measure on a topological factor has been studied, primarily in the context of one-dimensional symbolic dynamics.

Let $\eta : X \to Y$ be a topological factor map from a one-dimensional SFT X onto another subshift Y. Let ν be a fixed shift-invariant measure on Y. Consider an invariant measure μ on X that projects to ν and has maximal entropy within the fiber $\eta^{-1}(\nu)$. In [1, Thm. 3.3], Allahbakhshi and Quas proved that μ has the following Gibbsian property: for every finite set $A \Subset \mathbb{Z}$ and μ-almost every $x \in X$, the conditional distribution of the pattern on A given $\eta(x)$ and x_{A^c} is uniform among all patterns u on A that are consistent with x_{A^c} and $\eta(x)$ (i.e., u and x_{A^c} form a configuration that is in X and that maps to $\eta(x)$.)

As an immediate application of Theorem 1.2(b), the result of Allahbakhshi and Quas can be generalized in three directions. First, the SFT condition on X can be replaced by the more general topological Markov property. Second, we can allow for actions of arbitrary countable amenable groups.

Third, we may include an absolutely summable interaction on X, and obtain a similar Gibbsian property for measures that maximize pressure in the fiber. The precise statement and further details are given in Section §4.

Equilibrium measures on group shifts. Let $\mathbb{G}$ be a countable group and $\mathbb{H}$ a finite group. The full shift $\mathbb{H}^{\mathbb{G}}$ is itself a group with respect to the pointwise operation $(x \cdot y)_g \triangleq x_g \cdot y_g$ (for $x, y \in \mathbb{H}^{\mathbb{G}}$ and $g \in \mathbb{G}$). A group shift is a closed shift-invariant subset $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ which is also a subgroup of $\mathbb{H}^{\mathbb{G}}$. Kitchens and Schmidt [23] showed that every group shift over $\mathbb{G} \triangleq \mathbb{Z}$ or $\mathbb{G} \triangleq \mathbb{Z}^2$ is an SFT. More generally, any polycyclic-by-finite group has this property [41, Thms. 3.8 and 4.2]. However, this does not hold in general. For instance, if $\mathbb{G}$ is a countable group that is not finitely generated and $\mathbb{H} \triangleq \mathbb{Z}/2\mathbb{Z}$, then the subshift $\mathbb{X} \triangleq \{0^{\mathbb{G}}, 1^{\mathbb{G}}\}$ is a group shift but not an SFT. More generally, over any countable group $\mathbb{G}$ that contains a non-finitely generated subgroup, there are group shifts that are not SFTs [40].

Nevertheless, in Section §5 we show that every group shift over a countable group has the topological Markov property. The extended version of the non-relative Lanford–Ruelle theorem (Theorem 1.2(b) with trivial environment) thus gives the following result.

Theorem 1.5 (Equilibrium on group shifts)**.** *Let $\mathbb{G}$ be a countable amenable group and $\mathbb{H}$ a finite group, and let $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ be a group shift. Let Φ be an absolutely summable interaction on $\mathbb{X}$ with an associated observable f_Φ. Then every equilibrium measure on $\mathbb{X}$ for f_Φ is a Gibbs measure for Φ.*

Note that the special case of the above theorem with $\mathbb{G} \triangleq \mathbb{Z}^d$ follows from the classical Lanford–Ruelle theorem (Theorem 1.1(b)) and the fact that every group shift on $\mathbb{Z}^d$ is an SFT. The general case requires not only the extension to countable amenable groups but also the relaxation of the SFT condition to the topological Markov property.

In Section §5, we use Theorem 1.5 to give a sufficient condition for the Haar measure to be the unique measure of maximal entropy on a group shift.

Relative equilibrium measures on lattice slices. Our original motivation to develop a relative Dobrushin–Lanford–Ruelle theorem was to characterize equilibrium measures on two-dimensional subshifts in terms of equilibrium conditions on finite-height horizontal strips.

More specifically, let $Y \subseteq \Sigma^{\mathbb{Z}^2}$ be a two-dimensional subshift. Given a positive integer N, we can view Y as a relative system Ω_N with respect to horizontal shift, by thinking of each $y \in Y$ as a configuration $x \triangleq y_{\mathbb{Z}\times[0,N-1]}$ on the horizontal strip $\mathbb{Z} \times [0, N-1]$ together with the configuration $\theta \triangleq y_{\mathbb{Z}\times[0,N-1]^c}$ on the complement of the strip as the environment. In analogy with the Lanford–Ruelle theorem, one may expect that every equilibrium measure on Y (with respect to $\mathbb{Z}^2$-shift) is a relative equilibrium measure on Ω_N (with respect to horizontal shift). Conversely, analogy with the Dobrushin theorem suggests that if a $\mathbb{Z}^2$-invariant measure μ is a relative equilibrium measure on Ω_N (with respect to horizontal shift) for each positive N, then μ must be an equilibrium measure on Y (with respect to $\mathbb{Z}^2$-shift).

We now state a version of this characterization. Let Π_N denote the projection $y \mapsto y_{\mathbb{Z}\times[0,N-1]^c}$ on the complement of the strip $\mathbb{Z} \times [0, N-1]$. We assume that Y satisfies topological strong spatial mixing (TSSM), defined in Section §2.5, which implies, in this setting, hypotheses of both parts of the relative Dobrushin–Lanford–Ruelle theorem. Examples of subshifts with TSSM include the hard-core subshift and the subshift of 5-colorings on $\mathbb{Z}^2$ (see [3]).

Theorem 1.6 (Equilibrium vs. relative equilibrium on strips)**.** *Let Y be a $\mathbb{Z}^2$-subshift that satisfies TSSM. Let Φ be an absolutely summable interaction on Y and μ a $\mathbb{Z}^2$-invariant measure on Y. Then μ is an equilibrium measure for Φ (with respect to $\mathbb{Z}^2$-shift) if and only if for each positive integer N, μ is an equilibrium measure for Φ relative to its projection $\Pi_N\mu$ (with respect to horizontal shift).*

This theorem can be seen as an in-between characterization, being local in one direction and global in the other. In Section §6, we prove a more general statement for subshifts on countable amenable groups. In that setting, finite-width horizontal strips are replaced by finite-width slices, which are unions of finitely many cosets of a fixed subgroup. Interestingly, when the subgroup is the trivial subgroup, we recover the Dobrushin–Lanford–Ruelle theorem. In principle, Theorem 1.6 and its generalization may enable better understanding of an equilibrium measure for a $\mathbb{G}$-action by a relative equilibrium measure for an action of a subgroup.

Relative version of Meyerovitch's theorem. The Dobrushin–Lanford–Ruelle theorem is not valid on arbitrary subshifts, and the conditions of D-mixing and topological Markov property seem to be the appropriate hypotheses. Meyerovitch [31] has generalized the Lanford–Ruelle theorem by removing the assumption on the subshift while weakening the conclusion. To state his theorem, we need to introduce some terminology. Two finite patterns $u, v \in \Sigma^A$ are said to be *interchangeable* in a subshift $X \subseteq \Sigma^{\mathbb{G}}$ if for every $w \in \Sigma^{\mathbb{G}\setminus A}$, we have $u \vee w \in X$ if and only if $v \vee w \in X$. For example, in the golden mean shift, the words `010` and `000` are interchangeable, and in the even shift, the words `001` and `100` are interchangeable. Given $B \subseteq \mathbb{Z}^d$, we denote by ξ^B the σ-algebra on X consisting of the events that depend only on the pattern seen on B. Meyerovitch's result can be restated as follows.

Theorem 1.7 (Meyerovitch's theorem)**.** *Let $X \subseteq \Sigma^{\mathbb{Z}^d}$ be an arbitrary d-dimensional subshift. Let Φ be an absolutely summable interaction on X and μ an equilibrium measure for an associated observable f_Φ. Then, every two finite patterns $u, v \in \Sigma^A$ that are interchangeable in X satisfy*

$$\frac{\mu\big([u] \,\big|\, \xi^{A^c}\big)(x)}{\mathrm{e}^{-E_{A|A^c}(u \vee x_{A^c})}} = \frac{\mu\big([v] \,\big|\, \xi^{A^c}\big)(x)}{\mathrm{e}^{-E_{A|A^c}(v \vee x_{A^c})}}$$

for μ-almost every $x \in [u] \cup [v]$.

The conclusion of Meyerovitch's theorem becomes equivalent to the Gibbs property when the subshift has the topological Markov property: roughly speaking, the topological Markov property means that every two patterns that share the same sufficiently thick margin are interchangeable. It turns out that (the countable amenable group version of) Meyerovitch's result follows from the relative Lanford–Ruelle theorem by means of an appropriate encoding. In fact, we prove a relative version of Meyerovitch's result using this approach. The notion of interchangeability extends naturally to relative systems: we say that two finite patterns $u, v \in \Sigma^A$ are *interchangeable in X_θ* if for every $w \in \Sigma^{\mathbb{G}\setminus A}$, we have $u \vee w \in X_\theta$ if and only if $v \vee w \in X_\theta$. The *interchangeability set* of two finite patterns $u, v \in \Sigma^A$ is the set $\Theta_{u,v}$ of all environments $\theta \in \Theta$ for which u and v are interchangeable. Extending our earlier notation, we write ξ^B for the σ-algebra on Ω generated by the projection $(\theta, x) \mapsto x_B$. The σ-algebra on Ω generated by the environment will be denoted by $\mathscr{F}_\Theta$. We prove the following theorem for an arbitrary relative system Ω on a countable amenable group $\mathbb{G}$.

Theorem 1.8 (Relative version of Meyerovitch's theorem)**.** *Assume that Θ is a standard Borel space. Let ν be a $\mathbb{G}$-invariant probability measure on Θ and Φ a relative absolutely summable interaction on Ω. Let μ be an equilibrium measure for f_Φ relative to ν. Then, every two finite patterns $u, v \in \Sigma^A$ satisfy*

$$\frac{\mu\big([u] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, u \vee x_{A^c})}} = \frac{\mu\big([v] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, v \vee x_{A^c})}}$$

for μ-almost every $(\theta, x) \in [u] \cup [v]$ such that $\theta \in \Theta_{u,v}$.

We prove Theorem 1.8 in Section §7. As in the non-relative case, the conclusion of Theorem 1.8 becomes equivalent to the relative Gibbs property when the system has the topological Markov property relative to ν, and we recover the relative Lanford–Ruelle theorem. This means that, up to relatively simple reductions, Theorem 1.2(b) and Theorem 1.8 are equivalent!

Acknowledgements. We thank Raimundo Briceño for helpful discussions on the lattice slices application, Nishant Chandgotia for useful remarks and many discussions concerning the relation between interactions and observables, and François Ledrappier for raising a question that led to the application on lattice slices and in turn motivated this paper. We wish to express our gratitude to PIMS and the Department of Mathematics of the University of British columbia for their support. S. Barbieri was supported by the ANR project CoCoGro (ANR-16-CE40-0005). R. Gómez was supported by the PAPIIT-UNAM project IN107718 and by PASPA-UNAM.

2 Preliminaries

2.1 Setting

Throughout this article, we work in a general setting in which the underlying lattice is a countable amenable group. A countable group $(\mathbb{G}, \cdot)$ is *amenable* if there is a sequence of non-empty finite subsets

$F_n \subseteq \mathbb{G}$ that are approximately (right) $\mathbb{G}$-invariant, in the sense that for every $g \in \mathbb{G}$, $|F_n \triangle F_n g| = o(|F_n|)$ as $n \to \infty$. Such a sequence is called a (right) *Følner sequence*. For instance, the hypercubic lattice $(\mathbb{Z}^d, +)$ (with $d = 1, 2, \ldots$) is amenable with the boxes $F_n \triangleq [-n, n]^d \cap \mathbb{Z}^d$ forming a Følner sequence. For many results in ergodic theory including the pointwise ergodic theorem [29], the Shannon–McMillan–Breiman theorem [37, 29], and the Ornstein isomorphism theorem [38], the setting of amenable groups seems to be the right level of generality. A reader who is not concerned with this level of generality is welcome to take $\mathbb{G} \triangleq \mathbb{Z}^d$.

Let $\mathbb{G}$ be a countable amenable group and Σ a finite alphabet. We will use the notation $A \Subset \mathbb{G}$ to indicate that A is a finite subset of $\mathbb{G}$. We think of $x \in \Sigma^{\mathbb{G}}$ as a microscopic configuration of a physical system, with x_g representing the local state of the system at spatial position $g \in \mathbb{G}$. The group $\mathbb{G}$ acts on $\Sigma^{\mathbb{G}}$ by translations: the action of an element $g \in \mathbb{G}$ on a configuration $x \in \Sigma^{\mathbb{G}}$ is the shifted configuration gx where $(gx)_k \triangleq x_{g^{-1}k}$ for every $k \in \mathbb{G}$. Given $A, B \subseteq \mathbb{G}$, $u \in \Sigma^A$ and $v \in \Sigma^B$ such that $u_{A\cap B} = v_{A\cap B}$, define $u \vee v \in \Sigma^{A\cup B}$ by $(u \vee v)_g = u_g$ for $g \in A$ and $(u \vee v)_g = v_g$ for $g \in B$.

The system interacts with an external environment. The space of all possible states of the environment is a measurable space Θ on which $\mathbb{G}$ acts via measurable maps. For each $\theta \in \Theta$, let $X_\theta \subseteq \Sigma^{\mathbb{G}}$ be a non-empty closed set, representing the configurations that are consistent with environment θ. We impose two assumptions on the family $(X_\theta : \theta \in \Theta)$:

(i) (measurability) $\Omega \triangleq \{(\theta, x) : \theta \in \Theta \text{ and } x \in X_\theta\}$ is measurable in the product σ-algebra,

(ii) (translation symmetry) $X_{g\theta} = gX_\theta$ for each $\theta \in \Theta$ and $g \in \mathbb{G}$.

We call Ω a *relative system*. As an alternative interpretation, if ν is a probability measure on Θ, then $\theta \mapsto X_\theta$ is a random set in the probability space (Θ, ν).

Recall that the σ-algebra on Ω generated by projection on Θ is denoted by $\mathscr{F}_\Theta$. We denote by ξ the finite partition of Ω generated by the projection $(\theta, x) \mapsto x_{1_{\mathbb{G}}}$. A *cylinder* set is a set of the form $[q] \triangleq \{(\theta, x) \in \Omega : x_A = q\}$ where $q \in \Sigma^A$ is a pattern with (finite) support $A \Subset \mathbb{G}$. The set A is called the *base* of $[q]$. Given a subset $B \subseteq \mathbb{G}$, we write $\xi^B \triangleq \bigvee_{k\in B} \xi^k$ (with $\xi^k \triangleq k\xi$) for the σ-algebra on Ω generated by cylinder sets whose bases are included in B.

We call a measurable function $f\colon \Omega \to \mathbb{R}$ an *observable*. An observable is said to be *relatively local* if it is $(\mathscr{F}_\Theta \vee \xi^A)$-measurable for some $A \Subset \mathbb{G}$. An observable f is *relatively continuous* if the family $(f(\theta, \cdot) : \theta \in \Theta)$ is equicontinuous, that is, for every $\varepsilon > 0$, there exists a set $A \Subset \mathbb{G}$ such that for every $\theta \in \Theta$ and $x, y \in X_\theta$ satisfying $x_A = y_A$, we have $|f(\theta, x) - f(\theta, y)| < \varepsilon$. Every relatively local observable is clearly relatively continuous. The set of bounded relatively continuous observables, denoted by $C_\Theta(\Omega)$, is a Banach space with the uniform norm. The bounded relatively local observables form a dense linear subspace of $C_\Theta(\Omega)$.

For a closed subset $Y \subseteq \Sigma^{\mathbb{G}}$ and a finite set $A \Subset \mathbb{G}$, we write $L_A(Y)$ for the set of all patterns $q \in \Sigma^A$ such that $y_A = q$ for some $y \in Y$. We define $L(Y) \triangleq \bigcup_{A \Subset \mathbb{G}} L_A(Y)$.

2.2 Relative interactions and Hamiltonians

A (relative) *interaction* on Ω is a collection $\Phi = (\Phi_A : A \Subset \mathbb{G})$ of bounded measurable maps $\Phi_A\colon \Omega \to \mathbb{R}$ such that

(i) (relative locality) Φ_A is $(\mathscr{F}_\Theta \vee \xi^A)$-measurable,

(ii) (translation symmetry) $\Phi_{gA}(\theta, x) = \Phi_A(g^{-1}\theta, g^{-1}x)$.

A relative interaction is *absolutely summable* if

$$\|\Phi\| \triangleq \sum_{\substack{A \Subset \mathbb{G} \\ A \ni 1_{\mathbb{G}}}} \|\Phi_A\| < \infty ,$$

where $\|\Phi_A\|$ denotes the uniform norm of Φ_A.

Given an interaction Φ, the *energy content* of a configuration $x \in X$ in a finite set A relative to an environment $\theta \in \Theta$ is

$$E_A(\theta, x) \triangleq \sum_{C \subseteq A} \Phi_C(\theta, x) .$$

The collection $E = (E_A : A \Subset \mathbb{G})$ is called the (relative) *Hamiltonian* defined by Φ. The *conditional* energy content of x inside $A \Subset \mathbb{G}$ in the *context* of $B \Subset \mathbb{G}$ and environment $\theta \in \Theta$ is

$$\begin{aligned} E_{A|B}(\theta, x) &\triangleq E_{A\cup B}(\theta, x) - E_B(\theta, x) \\ &= \sum_{\substack{C \subseteq A\cup B \\ C\cap(A\setminus B)\neq\varnothing}} \Phi_C(\theta, x) \ . \end{aligned}$$

Observe that E_A is relatively continuous with $\|E_A\| \leq |A| \, \|\Phi\|$, and similarly, $\left\|E_{A|B}\right\| \leq |A \setminus B| \, \|\Phi\|$. The absolute summability of Φ ensures that the limit

$$\begin{aligned} E_{A|A^{\mathsf{c}}}(\theta, x) &\triangleq \lim_{B \nearrow \mathbb{G}} E_{A|(B\setminus A)}(\theta, x) \\ &= \sum_{\substack{C \Subset \mathbb{G} \\ C\cap A\neq\varnothing}} \Phi_C(\theta, x) \end{aligned} \tag{1}$$

exists along the finite subsets of $\mathbb{G}$ directed by inclusion. Moreover, the convergence is uniform in (θ, x), hence $E_{A|A^{\mathsf{c}}}(\theta, x)$ is bounded (namely, $\left\|E_{A|A^{\mathsf{c}}}\right\| \leq |A| \, \|\Phi\|$) and relatively continuous.

It is easy to see that

$$\left\|E_{A|A^{\mathsf{c}}} - E_A\right\| \leq \sum_{\substack{C \Subset \mathbb{G} \\ C\cap A\neq\varnothing \\ C\cap A^{\mathsf{c}}\neq\varnothing}} \|\Phi_C\| \ . \tag{2}$$

Suppose that $(F_n)_{n\in\mathbb{N}}$ is a Følner sequence in $\mathbb{G}$. It follows from the absolute summability of Φ that if we choose $A \triangleq F_n$, the right-hand side of (2) becomes of order $o(|F_n|)$ as $n \to \infty$, hence

$$\left\|E_{F_n|F_n^{\mathsf{c}}} - E_{F_n}\right\| = o(|F_n|) \tag{3}$$

as $n \to \infty$ (see Sec. §A.2.1). Another useful inequality is

$$\left\|E_{B|B^{\mathsf{c}}} - E_{A|A^{\mathsf{c}}}\right\| \leq |B \setminus A| \, \|\Phi\| \ , \tag{4}$$

which holds whenever A and B are finite and $A \subseteq B$ (see Sec. §A.2.2).

The value $\Phi_A(\theta, x)$ is interpreted as the energy resulting from the interaction between the symbols at sites in A and the environment. In models from physics, the interaction values are often physically meaningful values, either being prescribed by the microscopic physics behind the model, or representing rough microscopic tendencies for alignment or misalignment of the physical quantities at different locations. The contribution of a single site to the energy can be measured, for instance, by the following bounded relatively continuous observable

$$f_\Phi(\theta, x) \triangleq \sum_{\substack{A \Subset \mathbb{G} \\ A \ni 1_{\mathbb{G}}}} \frac{1}{|A|} \Phi_A(\theta, x) \ .$$

There are many other choices to distribute the energy contributions between sites; see [39, Sec. §3.2] for some other choices. The key relationship between Φ and f_Φ is that for every Følner sequence $(F_n)_{n\in\mathbb{N}}$,

$$\left|E_{F_n}(\theta, x) - \sum_{g\in F_n} f_\Phi(g^{-1}\theta, g^{-1}x)\right| = o(|F_n|) \tag{5}$$

as $n \to \infty$, uniformly in $(\theta, x) \in \Omega$ (see Sec. §A.2.3). As a consequence,

$$\mu(f_\Phi) = \lim_{n\to\infty} \frac{\mu\big(E_{F_n}\big)}{|F_n|} \tag{6}$$

for every $\mathbb{G}$-invariant measure μ on Ω. From (3) it follows that the above equality remains valid if we replace E_{F_n} with $E_{F_n|F_n^{\mathsf{c}}}$.

2.3 Relative pressure

Let E be the relative Hamiltonian associated to a relative absolutely summable interaction Φ and μ a $\mathbb{G}$-invariant probability measure on Ω. For every $A \Subset \mathbb{G}$, we define

$$\Psi_\mu(A) \triangleq H_\mu(\xi^A \,|\, \mathscr{F}_\Theta) - \mu(E_A) ,$$

where $H_\mu(\xi^A \,|\, \mathscr{F}_\Theta)$ denotes the conditional entropy of ξ^A given $\mathscr{F}_\Theta$ under μ. This is the *relative pressure* on A under μ.

The *relative pressure per site* under μ is given by

$$\psi(\mu) \triangleq \lim_{n\to\infty} \frac{1}{|F_n|} \Psi_\mu(F_n)$$

where $(F_n)_{n\in\mathbb{N}}$ is an arbitrary Følner sequence in $\mathbb{G}$. It can be verified using [18, Sec. §4.7] and (6) that the limit exists, is independent of the choice of the Følner sequence, and coincides with $h_\mu(\Omega \,|\, \Theta) - \mu(f_\Phi)$, where $h_\mu(\Omega \,|\, \Theta)$ denotes the conditional entropy per site (i.e., the conditional Kolmogorov–Sinai entropy for the $\mathbb{G}$-action) of μ given the $\mathbb{G}$-invariant σ-algebra $\mathscr{F}_\Theta$, and f_Φ is the energy observable associated to the interaction Φ.

The *conditional* relative pressure on $A \Subset \mathbb{G}$ given $B \Subset \mathbb{G}$ under μ is

$$\begin{aligned}\Psi_\mu(A \,|\, B) &\triangleq \Psi_\mu(A \cup B) - \Psi_\mu(B) \\ &= H_\mu(\xi^A \,|\, \xi^B \vee \mathscr{F}_\Theta) - \mu(E_{A|B}) .\end{aligned}$$

The advantage of this definition is that it has formal properties similar to those of conditional entropy $H_\mu(\xi^A \,|\, \xi^B)$ and conditional energy $E_{A|B}$. Most importantly, the conditional relative pressure satisfies the chain rule

$$\Psi_\mu(A \cup B \,|\, C) = \Psi_\mu(B \,|\, C) + \Psi_\mu(A \,|\, B \cup C) .$$

Observe that $\Psi_\mu(A \,|\, B)$ depends only on the restriction of μ to $\mathscr{F}_\Theta \vee \xi^{A\cup B}$. Moreover,

$$\Psi_\mu(A \,|\, B) \leq (\log|\Sigma| + \|\Phi\|)\, |A \setminus B| .$$

The martingale convergence theorem, the monotonicity of conditional entropy on the condition, the absolute summability of the interaction (in particular, the existence of the limit in (1)) and the bounded convergence theorem imply the existence of the limit

$$\begin{aligned}\Psi_\mu(A \,|\, A^{\mathsf{c}}) &\triangleq \lim_{B\nearrow\mathbb{G}} \Psi\big(A \,\big|\, (B \setminus A)\big) \\ &= H_\mu(\xi^A \,|\, \xi^{A^{\mathsf{c}}} \vee \mathscr{F}_\Theta) - \mu(E_{A|A^{\mathsf{c}}}) .\end{aligned}$$

Let us remark that for a fixed $A \Subset \mathbb{G}$ and a measure ν on Θ, the conditional entropy $H_\mu(\xi^A \,|\, \mathscr{F}_\Theta)$ and as a result the relative pressure $\Psi_\mu(A)$ are concave as functions of μ when μ runs over measures with marginal ν. In turn, the conditional entropy per site $h_\mu(\Omega \,|\, \Theta)$ and the relative pressure per site $\psi(\mu)$ are affine when restricted to measures μ with marginal ν.

2.4 Relative Gibbs measures and relative equilibrium measures

According to a fundamental hypothesis of equilibrium statistical mechanics, the macroscopic states of a system at thermal equilibrium are suitably described by probability distributions maximizing the pressure. Identifying the equilibrium measures thus amounts to solving an optimization problem, where the pressure is interpreted as the gain.

On a finite space, the optimization problem is solved by the Boltzmann distribution. The uniqueness of the solution is a consequence of the strict concavity of the entropy.

Proposition 2.1 (Finitary variational principle)**.** *Let M be a finite set and $U\colon M \to \mathbb{R}$ a real-valued function. Given a probability distribution $p\colon M \to [0,1]$, define*

$$\Psi(p) \triangleq H(p) - p(U) .$$

Then, $\Psi(p)$ takes its maximum if and only if $p(a) = \mathrm{e}^{-U(a)}/Z$ for each $a \in M$, where $Z \triangleq \sum_{a\in M} \mathrm{e}^{-U(a)}$ is the normalizing constant. The maximum value is $\log Z$.

This is well known and easily follows from Jensen's inequality.

A relative Gibbs measure for an absolutely summable relative interaction Φ is a probability measure on Ω that is *locally optimal*, in the sense that it maximizes the pressure on every finite region of the lattice $\mathbb{G}$ conditioned on the configuration outside the region and the environment. In other words, a probability measure μ on Ω is a relative Gibbs measure for Φ if for every $A \Subset \mathbb{G}$, the conditional probability according to μ of seeing a pattern u on A given a configuration x_{A^c} outside A and an environment θ is the Boltzmann distribution associated to the energy function $U(u) \triangleq E_{A|A^c}(\theta, x_{A^c} \vee u)$, where E is the Hamiltonian associated to Φ.

More specifically, for every $A \Subset \mathbb{G}$, the prescribed distribution of the pattern on A given a boundary condition θ, x_{A^c} is the Boltzmann distribution

$$\pi_{\theta, x_{A^c}}(u) \triangleq \begin{cases} \dfrac{1}{Z_{A|A^c}(\theta, x)} \mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee u)} & \text{if } x_{A^c} \vee u \in X_\theta, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $Z_{A|A^c}(\theta, x)$ is the normalizing constant known as the *partition function*. Given $(\theta, x) \in \Omega$, the distribution $\pi_{\theta, x_{A^c}}(u)$ extends to a probability measure $K_A\big((\theta, x), \cdot\big)$ on Ω by setting

$$K_A\big((\theta, x), [u] \cap W\big) \triangleq \mathbb{1}_W(\theta, x) \pi_{\theta, x_{A^c}}(u)$$

for each $u \in \Sigma^A$ and $W \in \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A}$. It can be verified that given a set $W \in \mathscr{F}_\Omega$, the function $K_A(\cdot, W)$ is measurable. A probability measure μ on Ω is a *relative Gibbs* measure for Φ if for every $A \Subset \mathbb{G}$ and each measurable set $W \in \mathscr{F}_\Omega$,

$$\mu\big(W \,\big|\, \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A}\big)(\cdot) = K_A(\cdot, W)$$

μ-almost surely. Notice that in order for μ to be a relative Gibbs measure, it is sufficient that the above equality holds for every $W \in \xi^A$.

We refer to the function $K_A(\cdot, \cdot)$ as the *Gibbs kernel* for set A. Every Gibbs kernel K_A naturally defines a linear operator $\nu \mapsto \nu K_A$ on probability measures on Ω by

$$(\nu K_A)(W) \triangleq \nu\big(K_A(\cdot, W)\big) = \int K_A(\cdot, W) \, \mathrm{d}\nu \; ,$$

and its adjoint operator $f \mapsto K_A f$ on bounded measurable observables on Ω by

$$(K_A f)(\omega) \triangleq K_A(\omega, f) = \int f(\omega') \, K_A(\omega, \mathrm{d}\omega') \; .$$

If $(\boldsymbol{\theta}, \boldsymbol{x})$ is a random point chosen according to ν, the measure νK_A can be interpreted as the distribution of $(\boldsymbol{\theta}, \boldsymbol{x})$ after resampling the pattern on A according to $\pi_{\boldsymbol{\theta}, \boldsymbol{x}_{A^c}}$.

With the above definition, one can see that a measure μ is relative Gibbs for Φ if and only if $\mu K_A = \mu$ for every $A \Subset \mathbb{G}$. The collection $K = (K_A : A \Subset \mathbb{G})$ of the Gibbs kernels for all $A \Subset \mathbb{G}$ is referred to as the *relative Gibbs specification* associated to Φ.

The local optimality of relative Gibbs measures is an immediate consequence of Proposition 2.1. For future reference, let us spell this out as a corollary in the following specific way.

Corollary 2.2 (Local optimality of relative Gibbs measures)**.** *Let Φ be an absolutely summable relative interaction on Ω. Let Ψ denote the pressure associated to Φ, and K_A the Gibbs kernel associated to Φ for a set $A \Subset \mathbb{G}$. Then, for every probability measure μ on Ω and μ-almost every $(\theta, x) \in \Omega$, we have*

$$\begin{aligned} &H_{\mu(\cdot | \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A})(\theta, x)}(\xi^A) - \mu\big(E_{A|A^c} \,|\, \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A}\big)(\theta, x) \\ &\qquad \leq H_{K_A((\theta, x), \cdot)}(\xi^A) - K_A\big((\theta, x), E_{A|A^c}\big) \quad \big(= \log Z_{A|A^c}(\theta, x) \ \big) \end{aligned}$$

with equality if and only if $\mu(\cdot \,|\, \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A})(\theta, x) = K_A((\theta, x), \cdot)$. In particular,

$$\Psi_\mu(A \,|\, A^c) \leq \Psi_{\mu K_A}(A \,|\, A^c) \; ,$$

with equality if and only if $\mu K_A = \mu$.

Proof. To obtain the first inequality, apply the finitary variational principle (Prop. 2.1) with $M \triangleq \{u \in \Sigma^A : x_{A^c} \vee u \in X_\theta\}$, $U(u) \triangleq E_{A|A^c}(\theta, x_{A^c} \vee u)$, and $p(u) \triangleq \mu([u] \,|\, \mathscr{F}_\Theta \vee \xi^{\mathbb{G} \setminus A})(\theta, x)$. The second inequality follows from the first inequality by integrating both sides with respect to μ. □

An alternative way to think about the above corollary is that applying a Gibbs kernel K_A on a measure locally optimizes that measure on the set A.

A relative equilibrium measure on Ω is a $\mathbb{G}$-invariant measure that is *globally optimal* among all $\mathbb{G}$-invariant measures with the same marginal on Θ. More generally, let ν be a $\mathbb{G}$-invariant measure on Θ and $f \in C_\Theta(\Omega)$ an arbitrary bounded relatively continuous observable (i.e., not necessarily one associated to an absolutely summable interaction). An *equilibrium measure* for f *relative to* ν is a $\mathbb{G}$-invariant measure on Ω with marginal ν on Θ which maximizes the relative pressure $h_\mu(\Omega \mid \Theta) - \mu(f)$ among all $\mathbb{G}$-invariant measures with the same marginal ν on Θ. A measure that is an equilibrium measure relative to its marginal on Θ is simply said to be a relative equilibrium measure.

2.5 Types of constraints on configurations

In this section, we define various classes of constraints on configuration spaces that are sufficient for the relative Dobrushin–Lanford–Ruelle theorem.

A *subshift* on $\mathbb{G}$ (or a $\mathbb{G}$-*subshift*) is a closed $\mathbb{G}$-invariant subset of $\Sigma^{\mathbb{G}}$. A subshift $X \subseteq \Sigma^{\mathbb{G}}$ is *of finite type* (*SFT*) if there exists a finite set $F \Subset \mathbb{G}$ and a subset $\mathcal{F} \subseteq \Sigma^F$ such that $x \in X$ if and only if $(g^{-1}x)_F \notin \mathcal{F}$ for all $g \in \mathbb{G}$. The elements of $\mathcal{F}$ are called the *forbidden* patterns defining X.

2.5.1 Conditions for the Lanford–Ruelle direction.

The classical constraint on the set of configurations X which enables the Lanford–Ruelle direction in Theorem 1.1 is that X is an SFT. However, this is used only in the form of a Markovian property: the possible configurations that may appear in a finite support $A \Subset \mathbb{G}$ given a fixed configuration in $\mathbb{G} \setminus A$ do not depend upon the whole complement but only on a finite subset $B \supseteq A$.

We say that a closed set $X \subseteq \Sigma^{\mathbb{G}}$ satisfies the *topological Markov property* (*TMP*) if for all $A \Subset \mathbb{G}$ there exists a finite set $B \supseteq A$ such that whenever $x, x' \in X$ satisfy $x_{B\setminus A} = x'_{B\setminus A}$, then $x_B \vee x'_{\mathbb{G}\setminus A} \in X$. We call such a set B a *memory set* for A in X. Equivalently, one may think of this property as follows: if $x_{B\setminus A} = x'_{B\setminus A}$, then x_B and x'_B are interchangeable in the sense that every appearance of x_B may be replaced by x'_B and vice-versa.

In the relative setting, similar notions of relative SFT and relative TMP can be formulated as follows. Let $\Omega \subseteq \Theta \times \Sigma^{\mathbb{G}}$ be a relative system. We say that Ω is a *relative SFT* (or is an *SFT relative to* Θ) if there is a finite set $F \Subset \mathbb{G}$ and a family of subsets $\mathcal{F}_\theta \subseteq \Sigma^F$ (for $\theta \in \Theta$) such that for each $\theta \in \Theta$, we have $x \in X_\theta$ if and only if $(g^{-1}x)_F \notin \mathcal{F}_{g^{-1}\theta}$ for all $g \in \mathbb{G}$. Similarly, we say that Ω satisfies the *topological Markov property relative to* Θ (*relative TMP*) if all the sets X_θ (for $\theta \in \Theta$) satisfy the TMP with common choices of the memory sets. In other words, Ω has relative TMP if for every $A \Subset \mathbb{G}$, there is a finite set $B \supseteq A$ such that whenever $\theta \in \Theta$ and $x, x' \in X_\theta$ satisfy $x_{B\setminus A} = x'_{B\setminus A}$, then $x_B \vee x'_{\mathbb{G}\setminus A} \in X_\theta$.

Given a $\mathbb{G}$-invariant measure ν on the environment space Θ, we can also consider the more relaxed conditions of SFT *relative to* ν and TMP *relative to* ν under which the corresponding conditions are satisfied for ν-almost every $\theta \in \Theta$ rather than for all $\theta \in \Theta$. However, by removing a null set from Θ, we can always turn the system into one that satisfies the condition surely.

Observe that for a subshift X that satisfies TMP, if B is a memory set for A in X, then for all $g \in \mathbb{G}$, gB is a memory set for gA in X. Similarly, for a relative system that satisfies TMP, it follows that if B is a memory set for A, then gB is a memory set for gA.

A closed set $X \subseteq \Sigma^{\mathbb{G}}$ satisfies the *strong topological Markov property* (*strong TMP*) if there is a finite set $F \Subset \mathbb{G}$ with $1_{\mathbb{G}} \in F$ such that for every finite set $A \Subset \mathbb{G}$, the set AF is a memory set for A in X. The notion of relative strong TMP is defined analogously. We remark that for subshifts, TMP and strong TMP are topological conjugacy invariants and that strong TMP is the conjugacy invariant class generated by the class of topological Markov fields as defined in [6, 7].

Clearly, every SFT satisfies the strong TMP, and the strong TMP implies the TMP. Moreover, these collections are all distinct. The class of subshifts with strong TMP, is much larger than the class of SFTs in the sense that the latter is countable while the former are uncountable: if X is any $\mathbb{Z}$-subshift over Σ, then the set of all configurations on $\mathbb{Z}^2$ whose rows are elements of X and whose columns are constant satisfies strong TMP but is not necessarily an SFT. See also [6] (bottom of page 233) for a simple example of a $\mathbb{Z}$-subshift which satisfies the strong TMP but is not an SFT. In Section §2.5.4 below, we provide an example of a $\mathbb{Z}^2$-subshift that satisfies TMP but not strong TMP.

2.5.2 Conditions for the Dobrushin direction.

The notion of D-mixing used in Theorem 1.1(a) was introduced by Ruelle [39, Sec. §4.1] for SFTs on $\mathbb{Z}^d$ but remains meaningful in a more general setting.

Let Y be a closed subset of $\Sigma^{\mathbb{G}}$. Given $A \Subset \mathbb{G}$, we say that a finite set $B \supseteq A$ is a *mixing set* for A in Y if for every $y, y' \in Y$, there exists $z \in Y$ satisfying $z_A = y_A$ and $z_{\mathbb{G}\setminus B} = y'_{\mathbb{G}\setminus B}$. In other words, we can paste a pattern on A from a configuration $y \in Y$ into any other configuration $y' \in Y$ provided that we modify the annulus $B \setminus A$. We say that Y is *Dobrushin-mixing* (or *D-mixing*) *with respect to* a Følner sequence $(F_n)_{n\in\mathbb{N}}$ if for each n, there is a mixing set $\overline{F}_n$ for F_n in Y such that $\left|\overline{F}_n \setminus F_n\right| = o(|F_n|)$ as $n \to \infty$. We say that Y satisfies *D-mixing* if it satisfies D-mixing with respect to some Følner sequence $(F_n)_{n\in\mathbb{N}}$.

A relative version of the D-mixing property suitable for our purposes is the following. Let Ω be a relative system and ν a $\mathbb{G}$-invariant measure on its environment space Θ. We say that Ω satisfies *D-mixing relative to ν with respect to a Følner sequence* $(F_n)_{n\in\mathbb{N}}$ if for ν-almost every $\theta \in \Theta$ and each $n \in \mathbb{N}$, there is a mixing set F_n^θ for F_n in X_θ such that $\left|F_n^\theta \setminus F_n\right|$ is measurable and $\int \left|F_n^\theta \setminus F_n\right| \mathrm{d}\nu(\theta) = o(|F_n|)$ as $n \to \infty$. We say that Ω satisfies *D-mixing relative to* ν if it satisfies D-mixing relative to ν with respect to some Følner sequence $(F_n)_{n\in\mathbb{N}}$.

There are two stronger notions which imply D-mixing and are better known in the symbolic dynamics community. We say that Y satisfies the *uniform filling property* (*UFP*) with respect to a Følner sequence $(F_n)_{n\in\mathbb{N}}$ if there exists a finite set $F \Subset \mathbb{G}$ such that F_nF is a mixing set for F_n. We say that Y satisfies the *UFP* if Y satisfies the UFP with respect to some Følner sequence. We say that Y is *strongly irreducible* (*SI*) if there exists $F \Subset \mathbb{G}$ such that for every two finite sets $A, B \Subset \mathbb{G}$ satisfying $AF \cap BF = \varnothing$ and every two configurations $y, y' \in Y$ there is a configuration $z \in Y$ such that $z_A = y_A$ and $z_B = y'_B$.

The UFP can be regarded as a uniform version of D-mixing: the fact that $(F_n)_{n\in\mathbb{N}}$ is Følner ensures that $|F_nF \setminus F_n| = o(|F_n|)$. In turn, a compactness argument shows that SI implies the UFP. An example of a subshift satisfying UFP but not SI is given in [17]. We do not know of any example of a D-mixing subshift which does not satisfy the UFP.

The relative versions of SI and the UFP are defined analogously.

2.5.3 Conditions implying both directions of the theorem

A natural condition that implies both directions of the theorem in the non-relative setting is that X is the full $\mathbb{G}$-shift $\Sigma^{\mathbb{G}}$. In the terminology of TMP, a $\mathbb{G}$-subshift is a full $\mathbb{G}$-shift if and only if every set $A \Subset \mathbb{G}$ is a memory set for itself. In other words, the symbol at each site can be changed *independently* of the rest of the configuration. The only $\mathbb{G}$-subshift with alphabet Σ satisfying that property is $\Sigma^{\mathbb{G}}$. However, the relative version of this notion turns out to be more interesting.

We say that a relative system Ω has the *relative independence property* (or *independence property relative to* Θ) if every finite set is a memory set for itself, that is, if for every $\theta \in \Theta$, every finite set $A \Subset \mathbb{G}$ and every pair $x, x' \in X_\theta$, we have $x_A \vee x'_{\mathbb{G}\setminus A} \in X_\theta$. Equivalently, Ω has the relative independence property if for each $\theta \in \Theta$ and $A \Subset \mathbb{G}$, any two elements of $L_A(X_\theta)$ are interchangeable in X_θ. Independence property *relative to* a measure ν on Θ is defined accordingly. Note that, as in the case of a relative SFT, there is no need for all the sets X_θ to be the same.

Every relatively independent system satisfies the TMP, and moreover, is relatively D-mixing (with $F_n^\theta \triangleq F_n$). Therefore, in a relatively independent system, both hypotheses of the relative Dobrushin–Lanford–Ruelle theorem hold. In Section §7, we shall show that under simple reductions, the Lanford–Ruelle theorem for relatively independent systems implies Theorem 1.2 (b) for relative systems with relative TMP.

There is a notion introduced by Briceño [3] that is less restrictive than independence but still implies both conditions of the Dobrushin–Lanford–Ruelle theorem. A closed subset $Y \subseteq \Sigma^{\mathbb{G}}$ is *topologically strong spatial mixing* (*TSSM*) if there exists a set $F \Subset \mathbb{G}$ such that for any finite disjoint sets $A, B, S \Subset \mathbb{G}$ such that $AF \cap BF = \varnothing$, and for every $u \in L_A(Y)$, $v \in L_B(Y)$ and $w \in L_S(Y)$ such that $u \vee w \in L_{A\cup S}(Y)$ and $v \vee w \in L_{B\cup S}(Y)$, we have $u \vee v \vee w \in L_{A\cup B\cup S}(Y)$. In fact, TSSM implies both SI and SFT (see [3]).

Proposition 2.3 (TSSM $\Longrightarrow$ SI SFT)**.** *Let X be a TSSM $\mathbb{G}$-subshift. Then X is a strongly irreducible SFT.*

Proof. The set S in the definition of TSSM can be chosen to be empty, hence X is SI. In order to show that X is an SFT, let F be the set appearing in the definition of TSSM, and without loss of generality, assume that $F \ni 1_{\mathbb{G}}$. Let $\mathcal{F} \subseteq \Sigma^{FF^{-1}}$ be the set of all patterns on FF^{-1} that do not occur on the elements of X, that is, $\mathcal{F} \triangleq \{q \in \Sigma^{FF^{-1}} : x_{FF^{-1}} \neq q \text{ for all } x \in X\}$. Let X' be the SFT defined by $\mathcal{F}$ as the set of forbidden patterns. We show that $X = X'$.

Clearly, $X \subseteq X'$. Suppose that there exists a configuration $y \in X' \setminus X$. Let $D \subseteq \mathbb{G}$ be a minimal set containing FF^{-1} such that $x_D \neq y_D$ for all $x \in X$. By compactness, D is finite and thus $y_D \notin L_D(X)$. By the definition of X', we have $y_{FF^{-1}} \in L_{FF^{-1}}(X)$, hence there exists an element $g \in D \setminus FF^{-1}$. Now, set $A \triangleq \{1_{\mathbb{G}}\}$, $B \triangleq \{g\}$ and $S \triangleq D \setminus \{1_{\mathbb{G}}, g\}$. Observe that $AF \cap BF = F \cap gF = \varnothing$. By the choice of D, we have $y_{A\cup S} \in L_{A\cup S}(X)$ and $y_{B\cup S} \in L_{B\cup S}(X)$ but $y_D \notin L_D(X)$. But this contradicts the TSSM property of X. □

The relative version of TSSM demands the existence of a set F for which the above condition holds on X_θ for all (or ν-almost every) $\theta \in \Theta$. It can be verified that relative TSSM implies both relative SI and relative SFT. In Figure 1, we summarize the relationships between all of the conditions introduced in this section. The same relations hold for their relative counterparts.

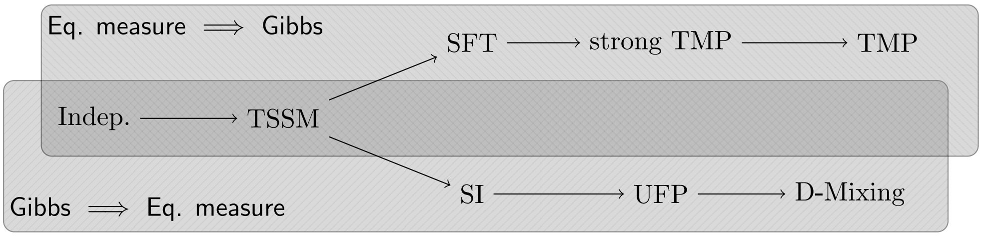


Figure 1: Sufficient conditions for both directions of the (relative) Dobrushin–Lanford–Ruelle theorem.

2.5.4 Examples

This last subsection is dedicated to examples that illustrate the conditions introduced in the previous subsections. We begin by examining the two examples given in the Introduction.

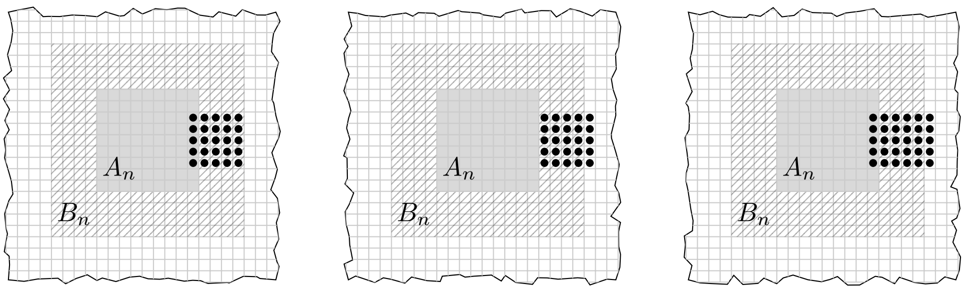
In Example 1.3 the environment θ completely determines the allowed symbols at each site, independently from site to site. Thus, the Ising model on percolation clusters satisfies the relative independence property and so the relative Dobrushin–Lanford–Ruelle theorem holds. Note that in this case, all the sets X_θ are disjoint.

In Example 1.4 the coloring condition is a nearest neighbor condition, and so the relative system is a relative SFT and therefore the relative Lanford–Ruelle theorem holds. Moreover, we claim that the assumption $|\Sigma| > |S|$ is sufficient to ensure that the system is relative SI, and thus the relative Dobrushin theorem holds. Indeed, let $F \triangleq S$ be the set of generators. Given a subgraph θ and two valid colorings $x, x' \in X_\theta$, if $AF \cap BF = \varnothing$ then no vertex in A is adjacent to a vertex in B. The partial configuration $w \triangleq x_A \vee x'_B$ can be inductively extended to a valid coloring of $(\mathbb{G}, \mathbb{E})$ by filling each position in $\mathbb{G} \setminus A \cup B$ with a color not already taken by any of its $|S|$ neighbors along the generators. In fact, the same argument shows that the system is indeed relatively TSSM (see [3]).

Example 2.4 (A $\mathbb{Z}^2$-subshift with TMP but not strong TMP)**.** Consider the set Y of all configurations $y \in \{\square, \boxdot\}^{\mathbb{Z}^2}$ such that every 8-connected component of sites with symbol $\boxdot$ is a finite square, that is, every 8-connected component of $y^{-1}(\boxdot)$ is a set of the form $\vec{u} + [0, n-1]^2 \cap \mathbb{Z}^2$ for some $\vec{u} \in \mathbb{Z}^2$ and $n \geq 1$. Let $X \subseteq \{\square, \boxdot\}^{\mathbb{Z}^2}$ be the closure of Y. We claim that X has TMP but not strong TMP.

In order to see that X does not have strong TMP, let $A_n \triangleq [-2n, 2n]^2$ and consider x to be equal to $\boxdot$ in the support $(3n, 0) + [-n, n]^2$ and $\square$ everywhere else. Similarly, consider y equal to $\boxdot$ in $(3n+1, 0) + [-n, n]^2$ and $\square$ in the complement. Let $B_n \triangleq [-4n, 4n]^2$. For any fixed $F \Subset \mathbb{Z}^2$ we have that $A_n F \subseteq B_n$ for all large enough n. However, we have that $x_{B_n \setminus A_n} = y_{B_n \setminus A_n}$ but $z \triangleq x_{B_n} \vee y_{A_n^c}$ is not an element of X as $z^{-1}(\boxdot) = (2n, 0) + ([0, n+1] \times [0, n])$ is not a square. This is illustrated in Figure 2.

To see that X has TMP, without loss of generality let $A = [-n, n]^2$. We claim that $B \triangleq [-10n, 10n]^2$ is a memory set for A for every $n \geq 1$. Indeed, let $x, y \in X$ be such that $x_{B \setminus A} = y_{B \setminus A}$. We must show that $z \triangleq x_A \vee y_{A^c} \in X$.



$x \in X$ $\qquad$ $y \in X$ $\qquad$ $x_{A_n} \vee y_{A_n^{\mathsf{c}}} \notin X$

Figure 2: Two configurations x and y which coincide in $B_n \setminus A_n$ but cannot be put together.

First, we claim that we can reduce to the case where $x, y \in Y$, that is, where every square is finite. Indeed, assume that $x, y \in X \setminus Y$. By definition, there are sequences $(x^{(m)})_{m \in \mathbb{N}}$ and $(y^{(m)})_{m \in \mathbb{N}}$ in Y converging to x and y respectively. For m large enough, we have $x^{(m)}_{B \setminus A} = y^{(m)}_{B \setminus A}$. The sequence of configurations $z^{(m)} \triangleq x^{(m)}_B \vee y^{(m)}_{A^{\mathsf{c}}}$ converges to z. Thus if we know that each $z^{(m)}$ is in Y, then we can conclude that $z \in X$. Now, observe that for m large enough, $x^{(m)}_B = x_B$ and $y^{(m)}_B = y_B$, and in particular $x^{(m)}_{B \setminus A} = y^{(m)}_{B \setminus A}$.

So, let $x, y \in Y$ and suppose that $z \notin Y$. Then there exists an 8-connected region $R_z \subseteq z^{-1}(\blacksquare)$ which is either an infinite set, or a finite set that is not a square. The first case cannot occur, as every connected component of $y^{-1}(\blacksquare)$ is bounded and z differs from y only on the finite set A. Therefore, R_z is finite but is not a square.

As $y \in Y$, we have that $R_z \cap A \neq \varnothing$. Let $R_x, R_y \Subset \mathbb{Z}^2$ be the finite squares in $x^{-1}(\blacksquare)$ and $y^{-1}(\blacksquare)$ which contain $R_z \cap A$ and $R_z \cap A^{\mathsf{c}}$ respectively. Denote the 8-boundary of A by ∂A, that is, $\partial A \triangleq [-n-1, n+1]^2 \setminus [-n, n]^2$. Let $N \triangleq [-n-1, n+1] \times \{n+1\}$, $S \triangleq [-n-1, n+1] \times \{-n-1\}$, $W \triangleq \{-n-1\} \times [-n-1, n+1]$ and $E \triangleq \{n+1\} \times [-n-1, n+1]$ be respectively the north, south, west and east 8-boundaries of A, so that $\partial A = N \cup E \cup S \cup W$. There are three possibilities on how R_x can intersect ∂A.

- If $R_x \cap \partial A = \varnothing$, then necessarily $R_x \subseteq A$, and since $x_{B \setminus A} = y_{B \setminus A}$, we conclude that $R_z = R_x$, which is a square. This contradicts the assumption.
- If R_x intersects only one of the sets N, S, W and E, then the size of R_x can be at most $(2n+1) \times (2n+1)$, the size of A. As R_x intersects A, we deduce that $R_x \cup \partial R_x \subseteq [-10n, 10n] = B$. Again, this implies that $R_z = R_x$, contradicting the assumption.
- If R_x intersects two (or more) of the boundaries N, S, W and E, such boundaries must themselves intersect, otherwise R_x is not a square. Observe that if $\blacksquare$ appears on two diagonally adjacent sites in any configuration from Y, that is, either the pattern $\begin{smallmatrix} ? & \blacksquare \\ \blacksquare & ? \end{smallmatrix}$ or the pattern $\begin{smallmatrix} \blacksquare & ? \\ ? & \blacksquare \end{smallmatrix}$ appears, then the said pattern is necessarily $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ (i.e. **?** $= \blacksquare$). Thus, in the current case, the information on $x_{\partial A}$ uniquely determines $A \cap (R_z \cup \partial R_z)$. As $\partial A \subseteq B$ and $x_{B \setminus A} = y_{B \setminus A}$ we conclude that $x_{A \cap R_z} = y_{A \cap R_z} \equiv \blacksquare$ and $x_{A \cap \partial R_z} = y_{A \cap \partial R_z} \equiv \square$. Thus $A \cap R_z = A \cap R_x = A \cap R_y$ and hence $R_z = R_y$ which is a square. This contradicts the assumption.

As this example satisfies TMP, the conclusion of the Lanford–Ruelle direction holds even though X is not an SFT nor satisfies strong TMP. ○

In Section §5, we will see another example of a subshift that satisfies TMP but not strong TMP.

2.6 A topology on measures and the Feller property of Gibbs kernels

An important consequence of the relative TMP is the following continuity property of the Gibbs kernels.

Proposition 2.5 (Relative Feller property of Gibbs kernels)**.** *Let Ω be a relative system and K_A be the Gibbs kernel for $A \Subset \mathbb{G}$ associated to a relative absolutely summable interaction. The following are equivalent.*

(a) Ω *has the topological Markov property relative to* Θ.

(b) *For every* $A \Subset G$ *and* $p \in \Sigma^A$, *the function* $K_A(\cdot, [p])$ *is in* $C_\Theta(\Omega)$.

(c) *For every* $A \Subset G$ *we have* $K_A f \in C_\Theta(\Omega)$ *whenever* $f \in C_\Theta(\Omega)$.

Proof.

(a) $\implies$ (c) By definition of the Gibbs kernel, we have

$$(K_A f)(\theta, x) = \frac{1}{Z_{A|A^c}(\theta, x)} \sum_{u \in \Sigma^A} \mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee u)} \mathbb{1}_{X_\theta}(x_{A^c} \vee u) f(\theta, x_{A^c} \vee u) . \tag{8}$$

The map $E_{A|A^c}$ is relatively continuous and $x \mapsto x_{A^c} \vee u$ is uniformly continuous. Let $B \supseteq A$ be a memory set for A witnessing the relative TMP. Then, for every $\theta \in \Theta$, we have $\mathbb{1}_{X_\theta}(x_{A^c} \vee u) = \mathbb{1}_{L(X_\theta)}(x_{B \setminus A} \vee u)$, from which we can see that the maps $x \mapsto \mathbb{1}_{X_\theta}(x_{A^c} \vee u)$ are equicontinuous for $\theta \in \Theta$. The partition function $Z_{A|A^c}(\theta, x)$ has the same form as the sum in (8) with f replaced with constant 1. Therefore, if $f \in C_\Theta(\Omega)$, the functions $x \mapsto (K_A f)(\theta, x)$ are equicontinuous for $\theta \in \Theta$, which means $K_A f \in C_\Theta(\Omega)$.

(c) $\implies$ (b) The function $f \triangleq \mathbb{1}_{[p]}$ is in $C_\Theta(\Omega)$, therefore $K_A(\cdot, [p])$ is in $C_\Theta(\Omega)$.

(b) $\implies$ (a) Let $\varepsilon \triangleq \frac{1}{2} \inf_{(\theta,x) \in \Omega} K_A((\theta, x), [x_A])$. Since $E_{A|A^c}$ is bounded, we deduce that $\varepsilon > 0$. By (b) we know that for every $p \in \Sigma^A$ we have $K_A(\cdot, [p]) \in C_\Theta(\Omega)$. Since $K_A((\theta, x), [p])$ depends only on θ, x_{A^c} and p, we can find $B \supseteq A$ such that for all $\theta \in \Theta$, $p \in \Sigma^A$ and $x, y \in X_\theta$, if $x_{B \setminus A} = y_{B \setminus A}$, we have

$$\left| K_A((\theta, x), [p]) - K_A((\theta, y), [p]) \right| < \varepsilon.$$

In particular, we obtain that if $x, y \in X_\theta$ and $x_{B \setminus A} = y_{B \setminus A}$, then

$$\left| K_A((\theta, x), [x_A]) - K_A((\theta, y), [x_A]) \right| < \varepsilon \leq \frac{1}{2} K_A((\theta, x), [x_A])$$

and so $K_A((\theta, y), [x_A]) \geq \varepsilon > 0$. This shows that $x_A \vee y_{A^c} \in X_\theta$. As the choice of B does not depend upon θ or $x, y \in X_\theta$ we deduce that Ω satisfies relative TMP. □

Let $\mathcal{P}_\nu(\Omega)$ denote the space of probability measures on Ω with marginal ν on Θ. The above proposition suggests topologizing $\mathcal{P}_\nu(\Omega)$ by declaring the integration $\mu \mapsto \mu(f)$ continuous for each $f \in C_\Theta(\Omega)$. The operator $\mu \mapsto \mu K_A$ would then become continuous whenever Ω has the TMP relative to Θ.

Recall that $C_\Theta(\Omega)$ is a Banach space with the uniform norm. When the environment space Θ is a standard Borel space, one can identify $\mathcal{P}_\nu(\Omega)$ with a closed subset of the dual space $C^*_\Theta(\Omega)$ (Proposition A.1). Alaoglu's theorem then implies that the space $\mathcal{P}_\nu(\Omega)$ is compact. We will use the compactness of $\mathcal{P}_\nu(\Omega)$ only at one point in the proof of Theorem 1.2(b) to argue that if ν is $\mathbb{G}$-invariant and $\mu \in \mathcal{P}_\nu(\Omega)$, then the sequence of averages $|F_n|^{-1} \sum_{g \in F_n} g^{-1}\mu$, with $(F_n)_{n \in \mathbb{N}}$ a Følner sequence, has a ($\mathbb{G}$-invariant) cluster point.

At the more fundamental level, the compactness of $\mathcal{P}_\nu(\Omega)$ together with the relative Feller property of the Gibbs kernels can be used to give a direct proof of the existence of (invariant) relative Gibbs measures.

Proposition 2.6 (Existence of invariant relative Gibbs measures)**.** *Assume that* Θ *is a standard Borel space and* Ω *satisfies TMP relative to* Θ. *Let* ν *be a* $\mathbb{G}$*-invariant probability measure on* Θ. *Then there exists a* $\mathbb{G}$*-invariant relative Gibbs measure with marginal* ν.

Proof. Since Θ is a standard Borel space, $\mathcal{P}_\nu(\Omega)$ is compact. Since Ω satisfies relative TMP, the Gibbs kernels have the relative Feller property. Let $A_1 \subseteq A_2 \subseteq \ldots$, be a nested sequence of finite subsets that exhaust $\mathbb{G}$. Let $\mu_0 \in \mathcal{P}_\nu(\Omega)$. For $n \geq 1$, set $\mu_n \triangleq \mu_0 K_{A_n}$. Then for all $B \subseteq A_n$, $\mu_n K_B = \mu_0 K_{A_n} K_B = \mu_0 K_{A_n} = \mu_n$. So, using the topology on $P_\nu(\Omega)$ and the relative Feller property of the Gibbs kernels, any accumulation point μ of the sequence μ_n is a relative Gibbs measure. It follows that for any $g \in \mathbb{G}$, $g^{-1}\mu$ is also a relative Gibbs measure. Thus, each $(1/|A|)\sum_{g\in A} g^{-1}\mu$ is a relative Gibbs measure. For any Følner sequence F_n, any accumulation point of

$$(1/|F_n|)\sum_{g\in F_n} g^{-1}\mu$$

is a $\mathbb{G}$-invariant relative Gibbs measure with marginal ν. The existence of such accumulation points is guaranteed by the compactness of $P_\nu(\Omega)$. □

An example of a subshift on which Gibbs measures (invariant or not) do not exist is the *sunny-side up* shift $X \subseteq \{\mathtt{0},\mathtt{1}\}^{\mathbb{Z}}$, which is defined as the set of all configurations with at most one occurrence of symbol $\mathtt{1}$.

The following crude notion of closeness between measures will be sufficient for our purposes.

Proposition 2.7 (Closeness of measures)**.** *Let ν be a probability measure on Θ and let $f \in C_\Theta(\Omega)$. For every $\varepsilon > 0$, there exists $B \Subset \mathbb{G}$ such that $|\mu'(f) - \mu(f)| < \varepsilon$ whenever $\mu, \mu' \in \mathcal{P}_\nu(\Omega)$ satisfy $\mu'|_{\mathscr{F}_\Theta \vee \xi^B} = \mu|_{\mathscr{F}_\Theta \vee \xi^B}$ (i.e., μ and μ' have the same marginal on (θ, x_B)).*

Proof. Since $f \in C_\Theta(\Omega)$, the family $(f(\theta,\cdot) : \theta \in \Theta)$ is equicontinuous. Let $B \Subset \mathbb{G}$ be such that $|f(\theta,x') - f(\theta,x)| < \varepsilon$ whenever $x'_B = x_B$. We have

$$\begin{aligned}
|\mu'(f) - \mu(f)| &= \left|\int \mu'(f \,|\, \mathscr{F}_\Theta \vee \xi^B)\mathrm{d}\mu' - \int \mu(f \,|\, \mathscr{F}_\Theta \vee \xi^B)\mathrm{d}\mu\right| \\
&= \left|\int \Big[\mu'(f \,|\, \mathscr{F}_\Theta \vee \xi^B) - \mu(f \,|\, \mathscr{F}_\Theta \vee \xi^B)\Big]\mathrm{d}\mu\right| \\
&< \varepsilon\ .
\end{aligned}$$

□

3 Proof of the main theorem

3.1 Relative Gibbs measures are relative equilibrium

Proof of Theorem 1.2(a). Let μ be a $\mathbb{G}$-invariant measure on Ω that projects to ν and is relative Gibbs for Φ. Let μ' be another $\mathbb{G}$-invariant measure that projects to ν. We show that $\psi(\mu') \leq \psi(\mu)$.

Let $K = (K_A : A \Subset \mathbb{G})$ be the relative Gibbs specification associated to Φ. Let $(F_n)_{n\in\mathbb{N}}$ be a Følner sequence in $\mathbb{G}$ with respect to which Ω is D-mixing relative to ν, and denote by F_n^θ the mixing set corresponding to F_n and θ that witnesses the D-mixing condition. Fix $n \in \mathbb{N}$. For ν-almost every $\theta \in \Theta$ and every $x \in X_\theta$, let $\mu''_{\theta,x}$ be a measure on Ω which has the same distribution as $\mu'(\cdot|\mathscr{F}_\Theta)(\theta,x)$ on ξ^{F_n} and is supported on $\{(\theta,y) \in \Omega : y_{(F_n^\theta)^\mathsf{c}} = x_{(F_n^\theta)^\mathsf{c}}\}$. We choose $\mu''_{\theta,x}$ in such a way that for every cylinder $[u]$, the value $\mu''_{\theta,x}([u])$ is measurable as a function of (θ,x).

Observe that

$$H_{\mu'(\cdot|\mathscr{F}_\Theta)(\theta,x)}(\xi^{F_n}) - \mu'(E_{F_n} \,|\, \mathscr{F}_\Theta)(\theta,x) = H_{\mu''_{\theta,x}}(\xi^{F_n}) - \mu''_{\theta,x}(E_{F_n})\ ,$$

and hence

$$\overbrace{H_{\mu'}(\xi^{F_n} \,|\, \mathscr{F}_\Theta) - \mu'(E_{F_n})}^{\Psi_{\mu'}(F_n)} = \int \big[\overbrace{H_{\mu''_{\theta,x}}(\xi^{F_n}) - \mu''_{\theta,x}(E_{F_n})}^{\circledcirc}\big]\mathrm{d}\mu(\theta,x)\ , \tag{9}$$

where we have used the fact that μ and μ' have the same marginals on Θ and that $\circledcirc$ does not depend on x. On the other hand, by the finitary variational principle (Prop. 2.1, or Cor. 2.2), we have

$$\underbrace{H_{\mu''_{\theta,x}}(\xi^{F_n^\theta}) - \mu''_{\theta,x}(E_{F_n^\theta|(F_n^\theta)^\mathsf{c}})}_{\odot} \leq \underbrace{H_{K_{F_n^\theta}((\theta,x),\cdot)}(\xi^{F_n^\theta}) - K_{F_n^\theta}\big((\theta,x), E_{F_n^\theta|(F_n^\theta)^\mathsf{c}}\big)}_{\circledast}\ . \tag{10}$$

Here, we are applying this variational principle to the set $M \triangleq \{u \in \Sigma^{F_n^\theta} : x_{(F_n^\theta)^\mathsf{c}} \vee u \in X_\theta\}$, the energy function $U(u) \triangleq E_{F_n^\theta|(F_n^\theta)^\mathsf{c}}(\theta, x_{(F_n^\theta)^\mathsf{c}} \vee u)$, and the distribution $p(u) \triangleq \mu''_{\theta,x}([u])$.

Combining (3) and (4), we have

$$\left\| E_{F_n^\theta|(F_n^\theta)^c} - E_{F_n} \right\| \leq \left| F_n^\theta \setminus F_n \right| \|\Phi\| + o(|F_n|) \tag{11}$$

as $n \to \infty$, with the $o(|F_n|)$ term not depending on (θ, x). Therefore,

$$|\circledbullet - \circledcirc| \leq \left| F_n^\theta \setminus F_n \right| (\log|\Sigma| + \|\Phi\|) + o(|F_n|) \; .$$

Integrating with respect to μ and using the relative D-mixing condition, we get

$$\begin{aligned} \left| \int \circledbullet \, \mathrm{d}\mu - \int \circledcirc \, \mathrm{d}\mu \right| &\leq (\log|\Sigma| + \|\Phi\|) \int \left| F_n^\theta \setminus F_n \right| \mathrm{d}\mu(\theta, x) + o(|F_n|) \\ &= o(|F_n|) \end{aligned} \tag{12}$$

as $n \to \infty$.

For $\circledast$, on the other hand, we have

$$\begin{aligned} \int \circledast \, \mathrm{d}\mu &= \int \mu(\circledast \,|\, \mathscr{F}_\Theta)(\theta, x) \, \mathrm{d}\mu(\theta, x) && (13) \\ &= \int \Big[H_{\mu(\cdot|\mathscr{F}_\Theta)(\theta,x)}\big(\xi^{F_n^\theta} \,\big|\, \xi^{(F_n^\theta)^c}\big) - \mu\big(E_{F_n^\theta|(F_n^\theta)^c} \,\big|\, \mathscr{F}_\Theta\big)(\theta, x) \Big] \mathrm{d}\mu(\theta, x) && (14) \\ &\leq \int \Big[H_{\mu(\cdot|\mathscr{F}_\Theta)}\big(\xi^{F_n} \,\big|\, \xi^{F_n^c}\big) - \mu\big(E_{F_n} \,\big|\, \mathscr{F}_\Theta\big) \Big] \mathrm{d}\mu \\ &\qquad + (\log|\Sigma| + \|\Phi\|) \int \left| F_n^\theta \setminus F_n \right| \mathrm{d}\mu(\theta, x) + o(|F_n|) \\ &= H_\mu\big(\xi^{F_n} \,\big|\, \xi^{F_n^c} \vee \mathscr{F}_\Theta\big) - \mu(E_{F_n}) + o(|F_n|) && (15) \\ &\leq \underbrace{H_\mu\big(\xi^{F_n} \,\big|\, \mathscr{F}_\Theta\big) - \mu(E_{F_n})}_{\Psi_\mu(F_n)} + o(|F_n|) && (16) \end{aligned}$$

as $n \to \infty$, where we have again used (11). The equality between $\mu(\circledast \,|\, \mathscr{F}_\Theta)(\theta, x)$ and the integrand on the right-hand side of (14) can be seen by partitioning Ω into countably many $\mathscr{F}_\Theta$-measurable subsets over each of which F_n^θ is constant.

Putting together (9), (10), (12) and (16), we obtain

$$\Psi_{\mu'}(F_n) \leq \Psi_\mu(F_n) + o(|F_n|)$$

as $n \to \infty$. Dividing by $|F_n|$ and letting $n \to \infty$ yields $\psi(\mu') \leq \psi(\mu)$ as desired. □

Remark 3.1 (Relative inner entropy for Gibbs measures)**.** A closer look at the proof of Theorem 1.2(a), namely (15), shows that in fact

$$\Psi_{\mu'}(F_n) \leq H_\mu\big(\xi^{F_n} \,\big|\, \xi^{F_n^c} \vee \mathscr{F}_\Theta\big) - \mu(E_{F_n}) + o(|F_n|)$$

as $n \to \infty$. Choosing $\mu' = \mu$, we obtain

$$\psi(\mu) = \lim_{n \to \infty} \frac{H_\mu\big(\xi^{F_n} \,\big|\, \xi^{F_n^c} \vee \mathscr{F}_\Theta\big)}{|F_n|} - \mu(f_\Phi) \; .$$

In particular, every $\mathbb{G}$-invariant Gibbs measure relative to ν satisfies

$$h_\mu(\Omega \,|\, \Theta) = \lim_{n \to \infty} \frac{H_\mu\big(\xi^{F_n} \,\big|\, \xi^{F_n^c} \vee \mathscr{F}_\Theta\big)}{|F_n|} \; ,$$

as long as Ω is D-mixing relative to ν. The corresponding equality in the non-relative setting is observed by Föllmer and Snell [12] and Tempelman [43, Sec. §5.3]. ◇

When both relative D-mixing and relative TMP are satisfied, we can obtain an explicit expression for the maximum pressure in terms of partition functions, generalizing the similar expression in the non-relative setting (see e.g. [39, Thm. 3.12]). Recall the definition of the partition function $Z_{A|A^c}(\theta, x)$ for environment θ and boundary condition x in (7). Given $A \Subset \mathbb{G}$ and $\theta \in \Theta$, we may also define the partition function with *free* boundary condition as

$$Z_A(\theta) \triangleq \sum_{u \in L_A(X_\theta)} \mathrm{e}^{-E_A(\theta, u)} \; ,$$

where $E_A(\theta, u)$ is understood as $E_A(\theta, x)$ for any $x \in [u] \cap X_\theta$.

Proposition 3.2 (Variational principle)**.** *Let Ω be a relative system and ν a $\mathbb{G}$-invariant probability measure on its environment space Θ. Let Φ be an absolutely summable relative interaction on Ω and f_Φ its associated energy observable. Assume that Θ is a standard Borel space. Assume further that Ω satisfies TMP and D-mixing relative to ν. Then,*

$$\sup_{\mu\in\mathcal{P}_\nu(\Omega)} \big[h_\mu(\Omega\,|\,\Theta) - \mu(f_\Phi)\big] = \lim_{n\to\infty} \int \frac{\log Z_{F_n}(\theta)}{|F_n|}\mathrm{d}\nu(\theta)\ , \tag{17}$$

where $(F_n)_{n\in\mathbb{N}}$ is a Følner sequence with respect to which the D-mixing condition holds. Moreover, every relative $\mathbb{G}$-invariant Gibbs measure for Φ with marginal ν achieves the supremum in the left hand side of (17).

Proof. Let μ be a $\mathbb{G}$-invariant relative Gibbs measure for Φ with marginal ν. The existence of relative Gibbs measures is guaranteed by Proposition 2.6. By the relative Dobrushin theorem, μ achieves the supremum on the left-hand side of (17). It remains to show that the pressure of μ coincides with the right-hand side of (17).

The observation made in Remark 3.1 together with (3) gives the expression

$$\psi(\mu) = \lim_{n\to\infty} \frac{\Psi_\mu(F_n\,|\,F_n^\mathsf{c})}{|F_n|}$$

for the relative pressure of μ. Let $K = (K_A : A \Subset \mathbb{G})$ be the relative Gibbs specification associated to Φ. Since μ is a relative Gibbs measure, we have

$$\begin{aligned}\Psi_\mu(F_n\,|\,F_n^\mathsf{c}) &= \int \Big[H_{K_{F_n}((\theta,x),\cdot)}(\xi^{F_n}) - K_{F_n}\big((\theta,x), E_{F_n|F_n^\mathsf{c}}\big)\Big]\mathrm{d}\mu(\theta,x) \\ &= \int \log Z_{F_n|F_n^\mathsf{c}}(\theta,x)\,\mathrm{d}\mu(\theta,x)\ .\end{aligned}$$

Thus, we only need to show that

$$\left|\int \log Z_{F_n|F_n^\mathsf{c}}(\theta,x)\,\mathrm{d}\mu(\theta,x) - \int \log Z_{F_n}(\theta)\,\mathrm{d}\mu(\theta,x)\right| = o(|F_n|) \tag{18}$$

as $n\to\infty$.

Let F_n^θ be the mixing set for F_n witnessing the D-mixing condition relative to ν. In order to prove (18), it is enough to show that

$$\log Z_{F_n|F_n^\mathsf{c}}(\theta,x) \leq \log Z_{F_n}(\theta) + o(|F_n|)\ , \tag{19}$$

$$\log Z_{F_n}(\theta) \leq \log Z_{F_n^\theta|(F_n^\theta)^\mathsf{c}}(\theta,x) + \|\Phi\|\,\big|F_n^\theta\setminus F_n\big|\ , \tag{20}$$

$$\int \log Z_{F_n^\theta|(F_n^\theta)^\mathsf{c}}(\theta,x)\,\mathrm{d}\mu(\theta,x) \leq \int \log Z_{F_n|F_n^\mathsf{c}}(\theta,x)\,\mathrm{d}\mu(\theta,x) + o(|F_n|) \tag{21}$$

and use the fact that $\int\big|F_n^\theta\setminus F_n\big|\,\mathrm{d}\mu(\theta,x) = o(|F_n|)$ by the D-mixing condition.

Inequalities (19) and (20) can be verified by a straightforward calculation using the fact that F_n^θ is a mixing set for F_n (see Sec. §A.2.4). Inequality (21) follows from the fact that the left hand side of (13) is lesser or equal to the right hand side of (15) once we recall that $\log Z_{F_n^\theta|(F_n^\theta)^\mathsf{c}}(\theta,x)$ is the same as $\circledast$, and that the integral of $\log Z_{F_n|F_n^\mathsf{c}}(\theta,x)$ is nothing but $\Psi_\mu(F_n\,|\,F_n^\mathsf{c}) = H_\mu\big(\xi^{F_n}\,\big|\,\mathscr{F}_\Theta\vee\xi^{F_n^\mathsf{c}}\big) - \mu(E_{F_n|F_n^\mathsf{c}})$, which differs from the the right-hand side of (15) by no more than $o(|F_n|)$. □

3.2 Relative equilibrium measures are relative Gibbs

The idea of the proof of Theorem 1.2(b) is as follows: if a measure μ on Ω is not relative Gibbs, then the conditional relative pressure $\Psi_\mu(A\,|\,A^\mathsf{c})$ has to be sub-optimal for some $A \Subset \mathbb{G}$ (Corollary 2.2). Therefore, applying the Gibbs kernel K_A on μ would locally increase the pressure. In order to increase the relative pressure per site ψ, we apply the Gibbs kernels on a positive-density set of translations of A, one after another. The translations of A should be sufficiently far apart so that the applications of the different kernels do not significantly interfere with one another. The final step is to do the standard averaging procedure to make the new measure $\mathbb{G}$-invariant.

This strategy for proving a result of this type is not entirely new. The fundamental idea of making a local improvement in a positive density set in order to achieve a global gain has been used many times in the literature. This idea is explicit in the works of Föllmer [11] and Burton and Steif [4] (see also [14, Sec. §15.4] and the bibliographic notes therein). Similar ideas have appeared in other contexts, for instance in the proof of the Garden-of-Eden theorem [32, 36] (see [5, Chap. 5]).

To follow the above strategy, we need three lemmas. The first provides a sufficient condition for the uniform convergence of a certain type of martingale. The second lemma complements Corollary 2.2 by stating that the improvement achieved by applying a Gibbs kernel is truly local. The last lemma ensures the existence of a non-overlapping packing of copies of a given finite set with strictly positive uniform lower density. Without loss of generality, by removing a ν-null set from Θ if necessary, we will assume that Ω has the TMP relative to the entire Θ.

Let $f\colon \Sigma^{\mathbb{G}} \to \mathbb{R}$ be a bounded measurable function and μ a probability measure on $\Sigma^{\mathbb{G}}$. According to the martingale convergence theorem, the conditional expectations $\mu(f \,|\, \xi^B)$ converge μ-almost surely to f as B grows to $\mathbb{G}$ along any co-final sequence of finite subsets of $\mathbb{G}$. Marcus and Pavlov [30] observed that if f has a continuous version modulo μ (i.e., $f = g$ μ-almost surely for a continuous map $g\colon \Sigma \to \mathbb{R}$), then the convergence of $\mu(f \,|\, \xi^B)$ is uniform over a set of full measure and holds in the net sense, along the family of finite subsets of $\mathbb{G}$ directed by inclusion. The following lemma is a relative version of the Marcus–Pavlov lemma.

Lemma 3.3 (Relative uniform martingale convergence)**.** *Let $f \in C_\Theta(\Omega)$ and let ν be a probability measure on Θ. Then, for every probability measure $\mu \in \mathcal{P}_\nu(\Omega)$, there is a set of full measure on which $\mu(f \,|\, \xi^B \vee \mathscr{F}_\Theta)$ converges uniformly to f as $B \nearrow \mathbb{G}$ along the family of finite subsets of $\mathbb{G}$ directed by inclusion. Furthermore, the convergence is also uniform over the choice of μ.*

Proof. Let $\varepsilon > 0$. Choose a finite set $B_0 \Subset \mathbb{G}$ large enough so that $|f(\theta,x) - f(\theta,y)| < \varepsilon$ whenever $x_{B_0} = y_{B_0}$. For every $B \Subset \mathbb{G}$ we have

$$\mu(f \,|\, \xi^B \vee \mathscr{F}_\Theta)(\theta,x) = \frac{1}{\mu([x_B] \,|\, \mathscr{F}_\Theta)(\theta,x)} \int_{[x_B]} f \,\mathrm{d}\mu(\cdot \,|\, \mathscr{F}_\Theta)(\theta,x)$$

for μ-almost every $(\theta,x) \in \Omega$ (see Sec. §A.2.5). It follows that when $B \supseteq B_0$,

$$\big|\mu(f \,|\, \xi^B \vee \mathscr{F}_\Theta)(\theta,x) - f(\theta,x)\big| < \varepsilon$$

for μ-almost every $(\theta,x) \in \Omega$. This shows the uniform convergence. Observe that B_0 does not depend on μ. Hence the convergence is also uniform in μ. □

Along with Corollary 2.2, the next lemma constitutes the main ingredient for proving Theorem 1.2(b). It allows to see the improvement predicted by Corollary 2.2 at the level of finite sets.

Lemma 3.4 (Local enhancement)**.** *Suppose that $\Psi_\mu(A \,|\, A^{\mathsf{c}}) < \Psi_{\mu K_A}(A \,|\, A^{\mathsf{c}})$ for some $A \Subset \mathbb{G}$. Then, there exists an $\varepsilon > 0$ and a finite set $B_0 \supseteq A$ such that*

$$\Psi_\mu\big(A \,\big|\, (B \setminus A)\big) \leq \Psi_{\mu' K_A}\big(A \,\big|\, (B \setminus A)\big) - \varepsilon$$

for every measure μ' with $\mu'|_{\mathscr{F}_\Theta \vee \xi^{B_0}} = \mu|_{\mathscr{F}_\Theta \vee \xi^{B_0}}$ and every finite set $B \supseteq B_0$.

Proof. Let $\delta \triangleq \Psi_{\mu K_A}(A \,|\, A^{\mathsf{c}}) - \Psi_\mu(A \,|\, A^{\mathsf{c}})$ and set $\varepsilon \triangleq \delta/7$. We make six separate approximations, and choose $B_0 \supseteq A$ large enough so that the error in each approximation is less than $\delta/7$.

Recall that the convergence $E_{A|(B\setminus A)} \to E_{A|A^{\mathsf{c}}}$ is uniform. Therefore, if we choose B_0 large enough, we can make sure that

$$\big|E_{A|(B\setminus A)}(\theta,x) - E_{A|A^{\mathsf{c}}}(\theta,x)\big| < \delta/7$$

for every $(\theta,x) \in \Omega$, whenever $B \supseteq B_0$. With such choice of B_0, we have

$$\big|\mu(E_{A|(B\setminus A)}) - \mu(E_{A|A^{\mathsf{c}}})\big| < \delta/7\,, \tag{22}$$

$$\big|(\mu' K_A)(E_{A|(B\setminus A)}) - (\mu' K_A)(E_{A|A^{\mathsf{c}}})\big| < \delta/7 \tag{23}$$

whenever $B \supseteq B_0$. Since Ω has the TMP and $E_{A|A^{\mathsf{c}}}$ is in $C_\Theta(\Omega)$, Proposition 2.5 implies that the function $K_A E_{A|A^{\mathsf{c}}}$ is in $C_\Theta(\Omega)$. Therefore, if we choose B_0 large enough, then we have

$$\big|(\mu' K_A)(E_{A|A^{\mathsf{c}}}) - (\mu K_A)(E_{A|A^{\mathsf{c}}})\big| < \delta/7 \tag{24}$$

whenever μ' has the same marginal on (θ, x_{B_0}) as μ (Proposition 2.7). Combining (23) and (24), for sufficiently large $B_0 \supseteq A$ we get

$$\left|(\mu' K_A)(E_{A|(B\setminus A)}) - (\mu K_A)(E_{A|A^c})\right| < (2/7)\delta \tag{25}$$

whenever $B \supseteq B_0$ and μ' has the same marginal on (θ, x_{B_0}) as μ.

Using the martingale convergence theorem and the monotonicity of conditional entropy with respect to the condition, we know that

$$H_\mu(\xi^A \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta) \to H_\mu(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta)$$

as $B \nearrow \mathbb{G}$ along the finite subsets of $\mathbb{G}$ directed by inclusion. Therefore, choosing $B_0 \supseteq A$ large enough, we get

$$\left|H_\mu(\xi^A \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta) - H_\mu(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta)\right| < \delta/7 \tag{26}$$

whenever $B \supseteq B_0$. Note that

$$H_{\mu' K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta) = -\int \overbrace{\sum_{p\in\Sigma^A} \mathbb{1}_{[p]} \cdot \log K_A(\cdot, [p])}^{\gamma(\cdot)} \,\mathrm{d}(\mu' K_A) = \mu' K_A \gamma.$$

Since Ω has TMP relative to Θ, Proposition 2.5 implies that the integrand γ and as a result $K_A\gamma$ are in $C_\Theta(\Omega)$. Therefore, if we choose B_0 large enough, we can make sure, using Proposition 2.7, that

$$\left|H_{\mu' K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta) - H_{\mu K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta)\right| < \delta/7 \tag{27}$$

whenever μ' has the same marginal on (θ, x_{B_0}) as μ. Lastly, by the martingale convergence theorem, we know that for $(\mu' K_A)$-almost every $(\theta, x) \in \Omega$ and every $p \in \Sigma^A$,

$$\underbrace{(\mu' K_A)\big([p] \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta\big)(\theta, x)}_{\mu'\left(K_A(\cdot,[p]) \,\middle|\, \xi^{B\setminus A}\vee\mathscr{F}_\Theta\right)(\theta,x)} \to \underbrace{(\mu' K_A)\big([p] \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}_{K_A((\theta,x),[p])}$$

as B grows to $\mathbb{G}$ along any co-final sequence of finite subsets of $\mathbb{G}$. Since the limit has a version $K_A\big((\theta, x), [p]\big)$ which is in $C_\Theta(\Omega)$ (Proposition 2.5), Lemma 3.3 ensures that the convergence is uniform both in (θ, x) (on a set of full μ'-measure) and in μ'. Since $E_{A|A^c}$ is bounded, for each $p \in \Sigma^A$ the function $K_A(\cdot, [p])$ is bounded away from 0 on $[p]$. It follows that the convergence of

$$H_{\mu' K_A}(\xi^A \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta) = -\int \sum_{p\in\Sigma^A} \mathbb{1}_{[p]} \cdot \log\Big[(\mu' K_A)\big([p] \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta\big)\Big] \,\mathrm{d}(\mu' K_A)$$

to

$$H_{\mu' K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta) = -\int \sum_{p\in\Sigma^A} \mathbb{1}_{[p]} \cdot \log K_A(\cdot, [p]) \,\mathrm{d}(\mu' K_A)$$

is uniform among all $\mu' \in \mathcal{P}_\nu(\Omega)$. In particular, choosing $B_0 \supseteq A$ large enough, we can ensure that

$$\left|H_{\mu' K_A}(\xi^A \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta) - H_{\mu' K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta)\right| < \delta/7 \tag{28}$$

for every $\mu' \in \mathcal{P}_\nu(\Omega)$, whenever $B \supseteq B_0$. Combining (27) and (28), for sufficiently large $B_0 \supseteq A$ we get

$$\left|H_{\mu' K_A}(\xi^A \,|\, \xi^{B\setminus A} \vee \mathscr{F}_\Theta) - H_{\mu K_A}(\xi^A \,|\, \xi^{A^c} \vee \mathscr{F}_\Theta)\right| < (2/7)\delta \tag{29}$$

whenever $B \supseteq B_0$ and μ' has the same marginal on (θ, x_{B_0}) as μ.

Putting (22), (25), (26) and (29) together with the hypothesis $\delta = \Psi_{\mu K_A}(A \,|\, A^c) - \Psi_\mu(A \,|\, A^c) > 0$, the result follows. □

In the course of the proof, we will need to pack copies of a finite set $P \Subset \mathbb{G}$ on $\mathbb{G}$ in a non-overlapping fashion in such a way that the uniform density of the copies is strictly positive. On a hyper-cubic lattice $\mathbb{G} = \mathbb{Z}^d$, a periodic packing does the job. On a general countable amenable group, a positive-density non-overlapping packing is achieved by a Delone set.

Lemma 3.5 (Existence of Delone sets)**.** *Let $\mathbb{G}$ be a group, and $P, C \subseteq \mathbb{G}$ subsets satisfying $C \supseteq PP^{-1}$. Then, there exists a set $D \subseteq \mathbb{G}$ satisfying the following two conditions:*

(i) (Packing) *$dP \cap d'P = \varnothing$ for every two distinct elements $d, d' \in D$,*

(ii) (Covering) *$gC \cap D \neq \varnothing$ for every $g \in \mathbb{G}$.*

Proof. Let $\mathscr{D}$ denote the family of all subsets of $\mathbb{G}$ that satisfy the packing condition. This family is partially ordered by inclusion. Furthermore, every chain in $\mathscr{D}$ has an upper bound in $\mathscr{D}$, namely the union of its elements. By Zorn's lemma, $\mathscr{D}$ has a maximal element, which we call D. We claim that D also satisfies the covering condition. For if D does not satisfy the covering condition, there must exist an element $g \in \mathbb{G}$ such that $gPP^{-1} \cap D = \varnothing$, or equivalently, $gP \cap dP = \varnothing$ for every $d \in D$. It follows that $\{g\} \cup D$ is in $\mathscr{D}$, contradicting the maximality of D. □

We leave it to the reader to show that when $\mathbb{G}$ is countable and P and C are finite, the existence of Delone sets can be established without resorting to the axiom of choice. Let us remark that in the case that $\mathbb{G}$ is a countable amenable group and P and C are finite, the covering condition in the above lemma ensures that D has positive *uniform lower density*

$$\underline{d}(D) \triangleq \liminf_{n\to\infty} \inf_{g\in\mathbb{G}} \frac{|gD \cap F_n|}{|F_n|}$$

with respect to every Følner sequence $(F_n)_{n\in\mathbb{N}}$. Indeed, for every $g \in \mathbb{G}$ and $h \in F_n$ there exists at least one $c \in C$ such that $hc \in gD$. It follows that $|gD \cap F_nC|\,|C| \geq |F_n|$. On the other hand, $|gD \cap F_n| \geq |gD \cap F_nC| - |F_n \triangle F_nC|$. Hence, $\underline{d}(D) \geq |C|^{-1}$. We are now ready to prove the relative Lanford–Ruelle theorem.

Proof of Theorem 1.2(b). Let μ be a $\mathbb{G}$-invariant measure on Ω with marginal ν and suppose that μ is not a relative Gibbs measure for Φ. We show that μ is not an equilibrium measure for f_Φ relative to ν by constructing another $\mathbb{G}$-invariant measure $\overline{\mu}_+$ with marginal ν that has strictly larger relative pressure per site. Let $K = (K_A : A \Subset \mathbb{G})$ be the relative Gibbs specification associated to Φ, and Ψ_μ the relative pressure under μ.

Since μ is not relative Gibbs, there exists a set $A \Subset \mathbb{G}$ such that $\mu K_A \neq \mu$. According to Corollary 2.2, this implies that

$$\Psi_\mu(A \,|\, A^{\mathsf{c}}) < \Psi_{\mu K_A}(A \,|\, A^{\mathsf{c}}) \,.$$

Let $\varepsilon > 0$, and take $B \triangleq B_0 \supseteq A$ as guaranteed by Lemma 3.4. Thus,

$$\Psi_\mu\big(A \,|\, (B' \setminus A)\big) \leq \Psi_{\mu' K_A}\big(A \,|\, (B' \setminus A)\big) - \varepsilon \tag{30}$$

whenever $B' \supseteq B$ and μ' has the same marginal on (θ, x_B) as μ.

Let $D \subseteq \mathbb{G}$ be a Delone set with packing shape B and covering shape BB^{-1} (Lemma 3.5). Let $k_1, k_2, \ldots$ be an arbitrary enumeration of the elements of D. Let $A_i \triangleq k_iA$ and $B_i \triangleq k_iB$ for $i = 1, 2, \ldots$. Define $\mu^{(0)} \triangleq \mu$ and $\mu^{(i)} \triangleq \mu^{(i-1)}K_{A_i}$ for $i \geq 1$. From the facts that the sets A_i are disjoint and the kernels K_{A_i} are proper (i.e., K_{A_i} keeps the marginal on $(\theta, x_{A_i^{\mathsf{c}}})$ intact) it follows that the limit $\mu_+ \triangleq \lim_{i\to\infty} \mu^{(i)}$ exists. Note however that μ_+ may depend on the enumeration of D, and more importantly, is not necessarily $\mathbb{G}$-invariant.

Let $(F_n)_{n\in\mathbb{N}}$ be a fixed Følner sequence. We average over the $\mathbb{G}$-orbit of μ_+ to construct a $\mathbb{G}$-invariant measure $\overline{\mu}_+$. More specifically, let $\overline{\mu}_+$ be an accumulation point of the sequence

$$\overline{\mu}_+^{(m)} \triangleq \frac{1}{|F_m|} \sum_{g\in F_m} g^{-1}\mu_+$$

as $m \to \infty$. Any such accumulation point will be a $\mathbb{G}$-invariant measure. The existence of accumulation points is guaranteed by the compactness of $\mathcal{P}_\nu(\Omega)$, whose argument relies on Θ being a standard Borel space (see Appendix §A.1).

To show that $\overline{\mu}_+$ has strictly larger pressure per site than μ, we compare the pressure of $g^{-1}\mu_+$ and $g^{-1}\mu$ on F_n for arbitrary $g \in \mathbb{G}$ and show that uniformly in g, there is a gap of at least $\varepsilon\,\underline{d}(D)\,|F_n| + o(|F_n|)$ between them, that is,

$$\inf_{g\in\mathbb{G}} \Psi_{g^{-1}\mu_+}(F_n) = \inf_{g\in\mathbb{G}} \Psi_{\mu_+}(gF_n) \geq \Psi_\mu(F_n) + \varepsilon\,\underline{d}(D)\,|F_n| + o(|F_n|)$$

as $n \to \infty$. By the concavity of the relative pressure, for each m, we have

$$\Psi_{\overline{\mu}_+^{(m)}}(F_n) \geq \inf_{g\in\mathbb{G}} \Psi_{g^{-1}\mu_+}(F_n) \ .$$

Taking the limit as $m \to \infty$ and using the continuity of the pressure gives

$$\Psi_{\overline{\mu}_+}(F_n) \geq \inf_{g\in\mathbb{G}} \Psi_{g^{-1}\mu_+}(F_n) \geq \Psi_\mu(F_n) + \varepsilon\,\underline{d}(D)\,|F_n| + o(|F_n|)$$

as $n \to \infty$. Dividing by $|F_n|$ and letting $n \to \infty$ will then yield the result.

For $g \in \mathbb{G}$, let $D_n^g \triangleq \{k \in D : kB \subseteq gF_n\}$ and $\widehat{D}_n^g \triangleq \{k \in D : kB \cap gF_n \neq \varnothing\}$. Note that $|D_n^g| \geq \underline{d}(D)\,|F_n| + o(|F_n|)$ and $\big|\widehat{D}_n^g \setminus D_n^g\big| = o(|F_n|)$ as $n \to \infty$ uniformly in g. Let $k_{\ell_1}, k_{\ell_2}, \ldots, k_{\ell_m}$ be the elements of D_n^g ordered according to the previously fixed enumeration of D. Let $R_n^g \triangleq \bigcup_{k\in(\widehat{D}_n^g\setminus D_n^g)}(kA \cap gF_n)$ be the union of A-neighborhoods of the elements of D that intersect gF_n but are not entirely included in gF_n. Using the chain rule, we decompose $\Psi_\mu(gF_n)$ and $\Psi_{\mu_+}(gF_n)$ as follows:

$$\begin{aligned}
\Psi_\mu(gF_n) &= \Psi_\mu\Big(gF_n \setminus \big[R_n^g \cup \bigcup_{i=1}^m A_{\ell_i}\big]\Big) + \sum_{i=1}^m \Psi_\mu\Big(A_{\ell_i}\,\Big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\Big) \\
&\quad + \Psi_\mu\big(R_n^g \,\big|\, gF_n \setminus R_n^g\big) && (31)\\
\Psi_{\mu_+}(gF_n) &= \Psi_{\mu_+}\Big(gF_n \setminus \big[R_n^g \cup \bigcup_{i=1}^m A_{\ell_i}\big]\Big) + \sum_{i=1}^m \Psi_{\mu_+}\big(A_{\ell_i}\,\big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\big) \\
&\quad + \Psi_{\mu_+}\big(R_n^g \,\big|\, gF_n \setminus R_n^g\big) \ . && (32)
\end{aligned}$$

Observe that the first terms on the right-hand sides of (31) and (32) are identical, because the two measures μ and μ_+ have the same marginals on $(\theta, x_{\mathbb{G}\setminus\bigcup_{k\in D} kA})$, and in particular on $(\theta, x_{gF_n\setminus[R_n^g\cup\bigcup_{i=1}^m A_{\ell_i}]})$. On the other hand, the last terms in (31) and (32) are each bounded by

$$\big(\log|\Sigma| + \|\Phi\|\big)\,|R_n^g| \leq \big(\log|\Sigma| + \|\Phi\|\big)\,|A|\,\big|\widehat{D}_n^g \setminus D_n^g\big| \ ,$$

which is $o(|F_n|)$ as $n \to \infty$ uniformly in g. To compare the middle terms, observe that on $(\theta, x_{A_{\ell_i}\cup(gF_n\setminus[R_n^g\cup\bigcup_{j=i}^m A_{\ell_j}])})$, the measure μ_+ has the same marginal as $\mu^{(\ell_i)} = \mu^{(\ell_i-1)}K_{A_{\ell_i}}$. Therefore,

$$\Psi_{\mu_+}\Big(A_{\ell_i}\,\Big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\Big) = \Psi_{\mu^{(\ell_i-1)}K_{A_{\ell_i}}}\Big(A_{\ell_i}\,\Big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\Big) \ .$$

Since $\mu^{(\ell_i-1)}$ and μ have the same marginals on B_{ℓ_i}, from (30) we get

$$\Psi_{\mu_+}\Big(A_{\ell_i}\,\Big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\Big) \geq \Psi_\mu\Big(A_{\ell_i}\,\Big|\,gF_n \setminus \big[R_n^g \cup \bigcup_{j=i}^m A_{\ell_j}\big]\Big) + \varepsilon \ .$$

It follows that, uniformly in g,

$$\begin{aligned}
\Psi_{\mu^+}(gF_n) &\geq \Psi_\mu(gF_n) + \varepsilon m + o(|F_n|) \\
&= \Psi_\mu(F_n) + \varepsilon\,\underline{d}(D)\,|F_n| + o(|F_n|)
\end{aligned}$$

as claimed. □

4 Equilibrium measures relative to a topological factor

In the setting of topological factor maps between subshifts, we have the following extension of the result of Allahbakhshi and Quas [1, Thm. 3.3] as a corollary of Theorem 1.2(b).

Theorem 4.1 (Gibbs property for equilibrium measures relative to a topological factor)**.** *Let X and Y be $\mathbb{G}$-subshifts, η a topological factor map from X onto Y, ν a $\mathbb{G}$-invariant measure on Y, and Φ an absolutely summable interaction on X. Assume that X satisfies the TMP. Then, every invariant* *measure μ projecting to ν that maximizes the pressure for f_Φ within the fiber $\eta^{-1}(\nu)$ satisfies the following Gibbs property: for every $A \Subset \mathbb{G}$ and $u \in L_A(X)$, and μ-almost every $x \in X$, we have*

$$\mu\big([u] \,\big|\, \xi^{A^c} \vee \eta^{-1}(\mathscr{F}_Y)\big)(x) = \begin{cases} \dfrac{1}{Z^\eta_{A|A^c}(x)} \mathrm{e}^{-E_{A|A^c}(x_{A^c} \vee u)} & \text{if } x_{A^c} \vee u \in X \text{ and } \eta(x_{A^c} \vee u) = \eta(x), \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

where $\mathscr{F}_Y$ is the σ-algebra on Y and $Z^\eta_{A|A^c}(x)$ is the appropriate normalizing constant.

To see how topological factor maps fit in the setting of relative systems, let X be a $\mathbb{G}$-subshift, Y a compact metric space with a continuous $\mathbb{G}$-action and $\eta\colon X \to Y$ a topological factor map, that is, a $\mathbb{G}$-equivariant continuous surjection from X onto Y. Regarding $\Theta \triangleq Y$ as an environment space and setting $X_y \triangleq \eta^{-1}(y)$, we obtain a relative system $\Omega \triangleq \{(\eta(x), x) : x \in X\}$, which is nothing other than the graph of η.

Let ν be a $\mathbb{G}$-invariant measure on Y. Via the natural topological conjugacy $X \to \Omega$, $x \mapsto (\eta(x), x)$, there is a one-to-one correspondence between $\mathbb{G}$-invariant measures μ on Ω that project to ν and $\mathbb{G}$-invariant measures on X that project to ν. Let Φ be an absolutely summable interaction on X, and note that Φ can be considered, via the same conjugacy, as an absolutely summable relative interaction on Ω. (Note however that the class of absolutely summable relative interactions on Ω is larger than those obtained in this fashion.)

With the above correspondence, the invariant measures μ (on X) that maximize pressure for f_Φ among all invariant measures projecting to ν are identified with the equilibrium measures (on Ω) for f_Φ relative to ν. Indeed, the pressure of μ can be written as

$$h_\mu(X) - \mu(f_\Phi) = h_\nu(Y) + h_\mu(\Omega \,|\, Y) - \mu(f_\Phi) \;.$$

Since $h_\nu(Y)$ is independent of μ, maximizing the pressure $h_\mu(X) - \mu(f_\Phi)$ is equivalent to maximizing the relative pressure $h_\mu(\Omega \,|\, Y) - \mu(f_\Phi)$. Likewise, relative Gibbs measures on Ω for Φ with marginal ν correspond precisely to measures on X that project to ν and satisfy (33).

Proof of Theorem 4.1. Following the above discussion, it is sufficient to show that the TMP on X implies the TMP on Ω relative to ν. The result then follows from Theorem 1.2(b).

Every continuous shift-commuting map between subshifts can be expressed as a sliding factor map (see e.g. [27]). Hence, denoting the alphabet of Y by Γ, there exists a set $F \Subset \mathbb{G}$ and a map $M\colon L_F(X) \to \Gamma$ such that

$$\eta(x)_g = M\big((g^{-1}x)_F\big)$$

for every $x \in X$ and $g \in \mathbb{G}$.

Let $A \Subset \mathbb{G}$ and let $B \supseteq A$ be a memory set for A witnessing the TMP of X. Note that every $\tilde{B} \Subset \mathbb{G}$ with $\tilde{B} \supseteq B$ is also a memory set for A. We claim that if we choose $\tilde{B}$ large enough such that $AF \cap \tilde{B}^c F = \varnothing$ (in particular, if we set $\tilde{B} \triangleq B \cup AFF^{-1}$), then $\tilde{B}$ is also a memory set for A witnessing the TMP of Ω relative to Y. Indeed, let $y \in Y$ and $x, x' \in X$ be such that $\eta(x) = \eta(x') = y$ and $x_{\tilde{B}\setminus A} = x'_{\tilde{B}\setminus A}$. By the TMP of X, the configuration $w \triangleq x_{\tilde{B}} \vee x'_{A^c}$ is in X. On the other hand, it is easy to see that if the condition $AF \cap \tilde{B}^c F = \varnothing$ is satisfied, then we also have $\eta(w) = y$. □

5 Equilibrium measures on group shifts

As stated in the Introduction, not all group shifts are SFTs [40]. In fact, a group shift may not even satisfy the strong TMP. For instance, if $\mathbb{G} \triangleq \bigoplus_{n\in\mathbb{N}} \mathbb{Z}/2\mathbb{Z}$ is the direct sum of countably many copies of $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{H} \triangleq \mathbb{Z}/2\mathbb{Z}$, then the group shift $\mathbb{X} \triangleq \{\mathtt{0}^{\mathbb{G}}, \mathtt{1}^{\mathbb{G}}\}$ does not satisfy the strong TMP. Indeed, note that the subgroup $\langle F\rangle \subseteq \mathbb{G}$ generated by each finite subset $F \Subset \mathbb{G}$ is finite. Suppose that F is such that AF is a memory set for A. Choosing $A \triangleq \langle F\rangle$ yields that $AF = A$ is a memory set for A, which is absurd.

However, all group shifts satisfy TMP as long as they are defined on a countable group $\mathbb{G}$. The proof is a straightforward adaptation of Lemma 2.2 in [23].

Proposition 5.1 (Group shifts have TMP)**.** *Let $\mathbb{G}$ be a countable group and $\mathbb{H}$ a finite group. Then every group shift $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ satisfies the TMP.*

Proof. For disjoint $A, B \Subset \mathbb{G}$ and $x \in \mathbb{X}$, let us define $L_{A|B}(x)$ as the set of all patterns $p \in L_A(\mathbb{X})$ such that $p \vee x_B$ is in $L_{A \cup B}(\mathbb{X})$. Denote by $L_{A|B}$ the set $L_{A|B}((1_{\mathbb{H}})^{\mathbb{G}})$.

Observe that $L_{A|B}$ is a subgroup of $L_A(\mathbb{X})$. Let us verify that for every $x \in \mathbb{X}$, the set $L_{A|B}(x)$ is a left coset of $L_{A|B}$. Clearly, $x_A \cdot L_{A|B} \subseteq L_{A|B}(x)$. Conversely, let $u \in L_{A|B}(x)$. Let $z \in \mathbb{X}$ be such that $z_{A \cup B} = u \vee x_B$. Then $(x^{-1} \cdot z)_B = 1_{\mathbb{H}}^B$ and hence $(x^{-1} \cdot z)_A = x_A^{-1} \cdot u$ is in $L_{A|B}$. It follows that $u \in x_A \cdot L_{A|B}$. Therefore, $L_{A|B}(x) = x_A \cdot L_{A|B}$.

Now let $g_0, g_1, g_2, \dots$ be an enumeration of $\mathbb{G}$ and $B_n \triangleq \{g_0, \dots, g_n\} \setminus A$. Clearly,

$$L_A(\mathbb{X}) \supseteq L_{A|B_0} \supseteq L_{A|B_1} \supseteq L_{A|B_2} \supseteq \cdots .$$

As $L_A(\mathbb{X})$ is finite, this chain eventually stabilizes, and thus there exists an $N \in \mathbb{N}$ such that $L_{A|B_{N+m}} = L_{A|B_N}$ for all $m \geq 0$. It follows that $L_{A|B_{N+m}}(x) = L_{A|B_N}(x)$ for every $x \in \mathbb{X}$ and all $m \geq 0$. But this is equivalent to saying that $C \triangleq A \cup B_N$ is a memory set for A. We conclude that $\mathbb{X}$ satisfies the TMP. □

Theorem 1.5 follows immediately from Proposition 5.1 and the extended version of the non-relative Lanford–Ruelle theorem (Theorem 1.2(b) on the system $\Omega \triangleq \Theta \times \mathbb{X}$ in which $\Theta \triangleq \{\theta\}$ is singleton and $\nu \triangleq \delta_\theta$).

We now give an algebraic interpretation of Theorem 1.5 in the case $\Phi \equiv 0$ and as a corollary, find a sufficient condition for the uniqueness of the measure of maximal entropy on group shifts.

More generally, let $\mathbb{X}$ be a compact metric group on which a countable group $\mathbb{G}$ acts by continuous automorphisms. A point $z \in \mathbb{X}$ is said to be *homoclinic* (or *asymptotic*) if for every open neighborhood $U \ni 1_{\mathbb{X}}$, there is a finite set $F \Subset \mathbb{G}$ such that $gz \in U$ for all $g \in \mathbb{G} \setminus F$. The homoclinic points of $\mathbb{X}$ form a subgroup of $\mathbb{X}$ denoted by $\Delta(\mathbb{X})$. The homoclinic points in a group shift are precisely the *finitary* configurations, that is, the configurations in which all but at most finitely many of the sites have the identity symbol.

Let us call a probability measure μ on $\mathbb{X}$ an *almost Haar* measure if it is invariant under the action of the homoclinic subgroup of $\mathbb{X}$ by left-translations, that is, if $\mu(z^{-1}U) = \mu(U)$ for every measurable $U \subseteq \mathbb{X}$ and each $z \in \Delta(\mathbb{X})$. Clearly, the Haar measure is almost Haar, but in general, there can be many other almost Haar measures. For instance, when $\mathbb{H} \triangleq \mathbb{Z}/2\mathbb{Z}$ and $\mathbb{G}$ is an arbitrary countable group, every probability measure on the group shift $\mathbb{X} \triangleq \{0^{\mathbb{G}}, 1^{\mathbb{G}}\}$ is almost Haar, simply because $\mathbb{X}$ has no homoclinic point other than its identity element $0^{\mathbb{G}}$.

The almost Haar measures on a group shift are precisely the Gibbs measures for the trivial interaction $\Phi \equiv 0$.

Proposition 5.2 (almost Haar $\equiv$ uniform Gibbs)**.** *Let $\mathbb{G}$ be a countable group and $\mathbb{H}$ a finite group, and let $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ be a group shift. A probability measure μ on $\mathbb{X}$ is almost Haar if and only if it is Gibbs for the interaction $\Phi \equiv 0$.*

Proof. First, suppose that μ is a Gibbs measure for $\Phi \equiv 0$. Let z be a homoclinic point. Let $A \Subset \mathbb{G}$ be the support of z, that is, $A \triangleq z^{-1}(\mathbb{H} \setminus \{1_{\mathbb{H}}\})$, and set $w \triangleq z_A$. Let $u \in L_A(\mathbb{X})$ and $Q \in \xi^{A^{\mathsf{c}}}$. By the (uniform) Gibbs property of μ, we have

$$\mu\big([u] \,|\, \xi^{A^{\mathsf{c}}}\big) = \mu\big([w^{-1}u] \,|\, \xi^{A^{\mathsf{c}}}\big)$$

μ-almost surely. Integrating over Q gives

$$\mu([u] \cap Q) = \mu([w^{-1}u] \cap Q) = \mu\big(z^{-1}([u] \cap Q)\big)$$

which implies that μ is invariant under left-translation by z. Since z was arbitrary, we find that μ is almost Haar.

Conversely, suppose that μ is almost Haar. Let $A \Subset \mathbb{G}$ be a finite set and $u, v \in L_A(\mathbb{X})$. If there is no configuration $x \in \mathbb{X}$ for which both $x_{A^{\mathsf{c}}} \vee u$ and $x_{A^{\mathsf{c}}} \vee v$ are in $\mathbb{X}$, there is nothing to show. So, suppose that there exists a configuration $\hat{x} \in \mathbb{X}$ such that $\hat{x}_{A^{\mathsf{c}}} \vee u, \hat{x}_{A^{\mathsf{c}}} \vee v \in \mathbb{X}$. Since $\mathbb{X}$ is a group shift, $z \triangleq (\hat{x}_{A^{\mathsf{c}}} \vee u)(\hat{x}_{A^{\mathsf{c}}} \vee v)^{-1}$ is in $\mathbb{X}$. Note that z is a homoclinic point with support A and $w \triangleq z_A = uv^{-1}$. By the almost Haar property, for every $Q \in \xi^{A^{\mathsf{c}}}$ we have

$$\mu([u] \cap Q) = \mu\big(z^{-1}([u] \cap Q)\big) = \mu([w^{-1}u] \cap Q) = \mu([v] \cap Q) .$$

This implies, by the definition of conditional probability, that $\mu([u] \,|\, \xi^{A^c}) = \mu([v] \,|\, \xi^{A^c})$ μ-almost surely. We conclude that μ is a Gibbs measure for $\Phi \equiv 0$. $\square$

As a corollary, we have the following restatement of the special case of Theorem 1.5 with $\Phi \equiv 0$.

Corollary 5.3 (Maximal entropy $\Longrightarrow$ almost Haar)**.** *Let $\mathbb{G}$ be a countable amenable group and $\mathbb{H}$ a finite group, and let $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ be a group shift. Then every measure of maximal entropy on $\mathbb{X}$ (with respect to the action of $\mathbb{G}$) is almost Haar.*

Observe that when $\Delta(\mathbb{X})$ is dense, the Haar measure is the unique almost Haar measure on $\mathbb{X}$. Therefore, we find the following corollary. See [8, Thm. 8.6] and [28, 2] for closely related results.

Corollary 5.4 (Uniqueness of measure of maximal entropy)**.** *Let $\mathbb{G}$ be a countable amenable group and $\mathbb{H}$ a finite group. Let $\mathbb{X} \subseteq \mathbb{H}^{\mathbb{G}}$ be a group shift and suppose that its homoclinic subgroup $\Delta(\mathbb{X})$ is dense in $\mathbb{X}$. Then, the Haar measure on $\mathbb{X}$ is the unique measure of maximal entropy on $\mathbb{X}$ (with respect to the action of $\mathbb{G}$).*

6 Relative equilibrium measures on lattice slices

Recall from the Introduction that a two-dimensional subshift $Y \subseteq \Sigma^{\mathbb{Z}^2}$ can be viewed as a one-dimensional relative system Ω_N (for N an arbitrary positive integer) in which the environment space Θ_N consists of the configurations on the complement of the horizontal strip $\mathbb{Z} \times [0, N-1]$ that are admissible in Y, and for each $\theta \in \Theta$, the set X_θ consists of all configurations x of the strip $\mathbb{Z} \times [0, N-1]$ that are consistent with θ in that $\theta \vee x \in Y$. Note that $\mathbb{Z}$ acts on Ω_N by horizontal shifts.

In this section, we shall prove Theorem 1.6, which states that under suitable conditions on Y, the equilibrium measures on Y are precisely the $\mathbb{Z}^2$-invariant measures that are relative equilibrium on Ω_N for each N. In fact, we prove this in a more general setting in which $\mathbb{Z}^2$ is replaced with an arbitrary countable amenable group $\mathbb{G}$, the horizontal strip is replaced with a union of a finite number of cosets of a fixed subgroup $\mathbb{H} \subseteq \mathbb{G}$ (called a *slice* of $\mathbb{G}$), and the horizontal $\mathbb{Z}$-action is replaced with the action of $\mathbb{H}$.

Before introducing the general setting, let us give a few examples to show why the above-mentioned equivalence cannot hold without some assumption on Y.

Example 6.1 (Equilibrium but not relative equilibrium I)**.** Let Y be the $\mathbb{Z}^2$-subshift over the alphabet $\Sigma \triangleq \{\square, \blacksquare, \boxtimes\}$ consisting of all configurations in which the two symbols $\blacksquare$ and $\boxtimes$ appear in at most one horizontal row (see Fig. 3), that is,

$$Y \triangleq \left\{ y \in \{\square, \blacksquare, \boxtimes\}^{\mathbb{Z}^2} : y_{(u,v)}, y_{(u',v')} \in \{\blacksquare, \boxtimes\} \text{ implies } v' = v \right\} .$$

The only non-wandering point in Y is the uniform configuration $\square^{\mathbb{Z}^2}$. Thus the atomic measure μ supported at $\square^{\mathbb{Z}^2}$ is the only $\mathbb{Z}^2$-invariant measure on Y. In particular, μ is the unique measure of maximal entropy on Y. However, given its marginal on Θ_1, μ does not maximize relative entropy on Ω_1. Namely, consider the measure μ' under which each site outside the strip $\mathbb{Z} \times \{0\}$ has almost surely the symbol $\square$, while the symbols inside the strip $\mathbb{Z} \times \{0\}$ are chosen independently uniformly at random from Σ. Note that μ' is invariant under horizontal shift and has the same marginal as μ on Θ_1. On the other hand, $h_\mu(\Omega_1 \,|\, \Theta_1) = 0$ while $h_{\mu'}(\Omega_1 \,|\, \Theta_1) = \log 3$. Let us observe that Y does not satisfy TMP, but the relative system Ω_1 is relatively D-mixing (even more, it has the relative independence property). $\bigcirc$

Example 6.2 (Equilibrium but not relative equilibrium II)**.** Let us consider a variant of the subshift from the previous example in which there is an additional constraint that the symbol $\square$ cannot occur in the same row as the symbols $\blacksquare$ and $\boxtimes$ (see Fig. 3). Namely, let

$$Y \triangleq \left\{ y \in \{\square, \blacksquare, \boxtimes\}^{\mathbb{Z}^2} : y_{(u,v)}, y_{(u',v')} \in \{\blacksquare, \boxtimes\} \text{ and } y_{(u'',v'')} = \square \text{ imply } v'' \neq v' = v \right\} .$$

As in the previous example, the atomic measure μ supported at $\square^{\mathbb{Z}^2}$ is the unique measure of maximal entropy on Y, but given its marginal on Θ_1, μ does not maximize relative entropy on Ω_1. In this case, the maximum relative entropy is achieved by the measure under which the sites outside $\mathbb{Z} \times \{0\}$ are almost surely given the symbol $\square$ and the sites in $\mathbb{Z} \times \{0\}$ are given random symbols chosen independently and uniformly from $\{\blacksquare, \boxtimes\}$. In contrast to the previous example, in this example, Y does have TMP (even strong TMP) but Ω_1 is not relatively D-mixing. $\bigcirc$

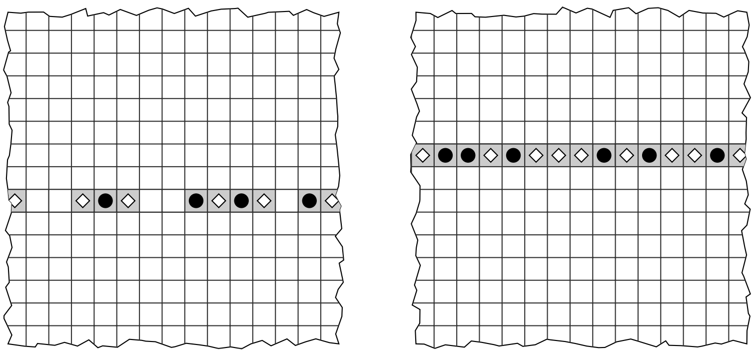


Figure 3: Configurations from the subshifts in Example 6.1 (left) and Example 6.2 (right).

Example 6.3 (Relative equilibrium but not equilibrium)**.** Let $Y \triangleq \{0\}^{\mathbb{Z}^2} \cup \{1,2\}^{\mathbb{Z}^2}$. Consider the atomic measure μ supported at $0^{\mathbb{Z}^2}$. For every positive N, μ maximizes relative entropy given its marginal on Θ_N. Nevertheless, μ is not a measure of maximal entropy for Y. Note that Y is an SFT, thus has TMP. It however does not satisfy D-mixing. ○

Let us now introduce the general setting. Let $\mathbb{G}$ be a countable amenable group and $\mathbb{H}$ a subgroup of $\mathbb{G}$. A union of finitely many right cosets of $\mathbb{H}$ in $\mathbb{G}$ is called an $\mathbb{H}$-*slice* of $\mathbb{G}$. Symbolic configurations on an $\mathbb{H}$-slice can be viewed in a natural way as configurations with a larger alphabet on $\mathbb{H}$. Namely, given an $\mathbb{H}$-slice S, we choose a collection $F \triangleq \{a_1, \ldots, a_k\}$ of representatives of distinct right cosets of $\mathbb{H}$ participating in S, so that $S = \mathbb{H}F$. With some abuse of notation, we will identify the configurations $x \in \Sigma^S$ on the slice S with the configurations $\hat{x} \in (\Sigma^F)^{\mathbb{H}}$ on $\mathbb{H}$ via the natural bijection between $\Sigma^{\mathbb{H}F}$ and $(\Sigma^F)^{\mathbb{H}}$ given by $(\hat{x}_h)_f = x_{hf}$ for $h \in \mathbb{H}$ and $f \in F$. Likewise, for $A \Subset \mathbb{H}$ we identify Σ^{AF} with $(\Sigma^F)^A$. For $d \in \mathbb{H}F$, we define $p_S(d)$ as the unique element $h \in \mathbb{H}$ such that $d \in hF$.

Let Y be a subshift on $\mathbb{G}$. Each $\mathbb{H}$-slice of $\mathbb{G}$ defines a relative system on which $\mathbb{H}$ acts. Namely, let $S \triangleq \mathbb{H}F$ be an $\mathbb{H}$-slice. For $B \subseteq \mathbb{G}$, let Π_B denote the projection $y \mapsto y_B$. We introduce a relative system Ω_S by considering $\Theta_S \triangleq \Pi_{S^{\mathsf{c}}}(Y)$ as the environment space and defining $X_\theta \triangleq \{x \in (\Sigma^F)^{\mathbb{H}} : \theta \vee x \in Y\}$ as the set of configurations consistent with $\theta \in \Theta_S$. Note that $\mathbb{H}$ acts on Ω_S by translations, and that this action is topologically conjugate to the action of $\mathbb{H}$ on Y.

An interaction Φ on Y induces a relative interaction $\widehat{\Phi}$ on Ω_S. Namely, for every finite subset $A \subseteq \mathbb{H}$ and every $u \in (\Sigma^F)^A$ and $\theta \in \Theta_S$, let

$$\widehat{\Phi}_A(\theta, u) \triangleq \sum_{\substack{B \Subset \mathbb{G} \\ p_S(B\cap S)=A}} \Phi_B(\theta_{B\setminus S} \vee u_{B\cap S}) \,.$$

Note that $\widehat{\Phi}$ is absolutely summable if Φ is (in fact, $\|\widehat{\Phi}\| \leq |F|\,\|\Phi\|$).

Proposition 6.4 (Gibbs kernels vs. relative Gibbs kernels)**.** *Let $S \triangleq \mathbb{H}F$ be an $\mathbb{H}$-slice of $\mathbb{G}$. Let Φ be an absolutely summable interaction on Y and $\widehat{\Phi}$ the corresponding relative interaction on Ω_S. Let K be the Gibbs specification on Y associated to Φ and $\widehat{K}$ the relative Gibbs specification on Ω_S associated to $\widehat{\Phi}$. Let $y \in Y$, and set $\theta \triangleq y_{S^{\mathsf{c}}}$ and $x \triangleq y_S$. Then, for every $A \Subset \mathbb{H}$ and $u \in (\Sigma^F)^A$, we have*

$$\widehat{K}_A\big((\theta, x), [u]\big) = K_{AF}(y, [u]) \,.$$

Proof. Let E denote the Hamiltonian on Y associated to Φ, and let $\widehat{E}$ denote the relative Hamiltonian on Ω_S associated to $\widehat{\Phi}$. Clearly, $x_{\mathbb{H}\setminus A} \vee u \in X_\theta$ if and only if $y_{\mathbb{G}\setminus AF} \vee u \in Y$. If either of the latter conditions is satisfied, we have

$$\begin{aligned}
\widehat{E}_{A|A^{\mathsf{c}}}(\theta, x_{\mathbb{H}\setminus A} \vee u) &= \sum_{\substack{C \Subset \mathbb{H} \\ C\cap A\neq\varnothing}} \widehat{\Phi}_C(\theta, x_{\mathbb{H}\setminus A} \vee u) \\
&= \sum_{\substack{C \Subset \mathbb{H} \\ C\cap A\neq\varnothing}} \sum_{\substack{B \Subset \mathbb{G} \\ p_S(B\cap S)=C}} \Phi_B(\theta_{B\setminus S} \vee (x_{\mathbb{H}\setminus A} \vee u)_{B\cap S}) \\
&= E_{(AF)|(\mathbb{G}\setminus AF)}(y_{\mathbb{G}\setminus AF} \vee u) \,.
\end{aligned}$$

The result then follows from the definitions of the Gibbs kernels. □

Before stating the main result, let us verify that TMP on a subshift implies relative TMP with respect to slices.

Proposition 6.5 (TMP $\implies$ relative TMP)**.** *Let $S \triangleq \mathbb{H}F$ be an $\mathbb{H}$-slice of $\mathbb{G}$. If Y satisfies TMP, then Ω_S satisfies relative TMP.*

Proof. Let $A \Subset \mathbb{H}$ and let $B \supseteq AF$ be a memory set for AF witnessing the TMP of Y. Since any finite superset of a memory set is also a memory set, we may assume that $B \cap S = CF$ for some $C \Subset \mathbb{H}$. We claim that C is a memory set for A in the relative system Ω_S.

Let $\theta \in \Theta_S$ and $x, x' \in X_\theta$ be such that $x_{C\setminus A} = x'_{C\setminus A}$. Let $y, y' \in Y$ be such that $y_{\mathbb{G}\setminus S} = y'_{\mathbb{G}\setminus S} = \theta$, $y_S = x$ and $y'_S = x'$. Since B is a memory set for AF in Y, there is a configuration $\tilde{y} \in Y$ that agrees with y on B, and thus on $CF = B \cap S$ and with y' on $\mathbb{G} \setminus AF$. In other words, $\tilde{y}_{\mathbb{G}\setminus S} = \theta$ and $\tilde{y}$ agrees with y on CF and with y' on $S \setminus AF$. In particular, if we set $w \triangleq \tilde{y}_S$, then $w \in X_\theta$ and w agrees with x on C and with x' on $\mathbb{H} \setminus A$. This means that C is a memory set for A in Ω_S. □

Now we can state the main general result of this section.

Theorem 6.6. *Let Y be a subshift on a countable amenable group $\mathbb{G}$. Let Φ an absolutely summable interaction on Y and μ a $\mathbb{G}$-invariant probability measure on Y. Let $\mathbb{H}$ be a subgroup of $\mathbb{G}$.*

(a) (Lanford–Ruelle theorem for slices)
Assume that Y satisfies TMP. Assume further that μ is an equilibrium measure on Y for Φ. Let $S \triangleq \mathbb{H}F$ be an $\mathbb{H}$-slice of $\mathbb{G}$, and denote by $\widehat{\Phi}$ the relative interaction on Ω_S corresponding to Φ. Then, μ is an equilibrium measure on Ω_S for $\widehat{\Phi}$ relative to $\Pi_{S^c}\mu$, provided that Ω_S is D-mixing relative to $\Pi_{S^c}\mu$.

(b) (Dobrushin theorem for slices)
Assume that for every $\mathbb{H}$-slice $S \triangleq \mathbb{H}F$ of $\mathbb{G}$, Ω_S satisfies relative TMP. Assume further that for every $\mathbb{H}$-slice S, μ is an equilibrium measure on Ω_S for $\widehat{\Phi}$ relative to $\Pi_{S^c}\mu$, where $\widehat{\Phi}$ denotes the relative interaction on Ω_S corresponding to Φ. Then, μ is an equilibrium measure on Y for Φ, provided that Y is D-mixing.

Proof. Let K denote the Gibbs specification on Y for Φ.

(a) Let $\widehat{K}$ denote the relative Gibbs specification on Ω_S for $\widehat{\Phi}$. Since Y satisfies the TMP, μ is a Gibbs measure for Φ by the (non-relative) Lanford–Ruelle theorem (Theorem 1.2(b) with trivial environment). By Proposition 6.4, for every $A \Subset \mathbb{H}$ and $u \in (\Sigma^F)^A$, and μ-almost every $(\theta, x) \in \Omega_S$,

$$\begin{aligned}\mu\big([u] \,\big|\, \mathscr{F}_\Theta \vee (\xi^F)^{\mathbb{H}\setminus A}\big)(\theta, x) &= \mu\big([u] \,\big|\, \xi^{\mathbb{G}\setminus AF}\big)(\theta \vee x) \\ &= K_{AF}(\theta \vee x, [u]) \\ &= \widehat{K}_A((\theta, x), [u])\end{aligned}$$

and so μ is a relative Gibbs measure on Ω_S. Now, assuming that Ω_S is D-mixing relative to $\Pi_{S^c}\mu$, by the relative Dobrushin theorem (Theorem 1.2(a)), μ is a relative equilibrium measure on Ω_S for $\widehat{\Phi}$ relative to $\Pi_{S^c}\mu$.

(b) Let $S \triangleq \mathbb{H}F$ be an arbitrary $\mathbb{H}$-slice in $\mathbb{G}$. Let $\widehat{K}$ denote the relative Gibbs specification on Ω_S for $\widehat{\Phi}$. Since μ is a relative equilibrium measure for $\widehat{\Phi}$, we can apply the relative Lanford–Ruelle theorem (Theorem 1.2(b)) to get that μ is relative Gibbs for $\widehat{\Phi}$. Using Proposition 6.4, it follows that for every $A \Subset \mathbb{H}$ and $u \in \Sigma^{AF}$ and μ-almost every $y \in Y$, we have

$$\begin{aligned}\mu\big([u] \,\big|\, \xi^{\mathbb{G}\setminus AF}\big)(y) &= \mu\big([u] \,\big|\, \mathscr{F}_\Theta \vee (\xi^F)^{\mathbb{H}\setminus A}\big)(y_{S^c}, y_S) \\ &= \widehat{K}_A((y_{S^c}, y_S), [u]) \\ &= K_{AF}(y, [u]).\end{aligned}$$

Thus, μ satisfies the Gibbs condition for sets of the form AF, with $A \subseteq \mathbb{H}$. Since the collection of sets of the form AF, for all such A and F, forms a cofinal subset of the collection of finite subsets of $\mathbb{G}$, μ is a Gibbs measure for Φ (see Remark 1.24 in [14]). Since μ is $\mathbb{G}$-invariant and Gibbs, it is an equilibrium measure by the (non-relative) Dobrushin theorem (Theorem 1.2(a) with trivial environment). □

Note that in part (b) of the above theorem, we can use Proposition 6.5 to replace the condition of relative TMP for every slice with the condition that Y satisfies TMP.

Remark 6.7 (Recovering Dobrushin–Lanford–Ruelle theorem)**.** When $\mathbb{H}$ is the trivial subgroup, the statement of Theorem 6.6 recovers the statement of the Dobrushin–Lanford–Ruelle theorem (Theorem 1.2 with trivial environment). Indeed, in this case, $\mathbb{H}$-slices of $\mathbb{G}$ are precisely the finite subsets of $\mathbb{G}$ and thus the conditions of relative TMP and relative D-mixing become trivial. Note that according to Corollary 2.2, μ is a Gibbs measure if and only if it is a relative equilibrium measure for every $\mathbb{H}$-slice of $\mathbb{G}$. ○

Remark 6.8 (The missing counter-example)**.** Examples 6.1 and 6.2 show that neither of the two conditions in Theorem 6.6(a), namely TMP and relative D-mixing, can be dropped. On the other hand, Example 6.3 shows that Theorem 6.6(b) would not hold if we dropped the D-mixing condition. This begs the question of whether the remaining condition of TMP can be dropped in Theorem 6.6(b).

However, a counter-example in which Theorem 6.6(b) fails in absence of TMP would be more complicated to construct. In fact, as the following argument suggests, such an example may require Y to satisfy D-mixing but not the UFP (see Sec. §2.5.2), at least when $\mathbb{G} \triangleq \mathbb{Z}^2$ and $\mathbb{H} \triangleq \mathbb{Z}$. We do not know if such a subshift exists.

Consider the basic case of horizontal strips on two-dimensional subshifts, thus $\mathbb{G} \triangleq \mathbb{Z}^2$ and $\mathbb{H} \triangleq \mathbb{Z}$. Suppose that $Y \subseteq \Sigma^{\mathbb{Z}^2}$ has the UFP with respect to the sequence of boxes $F_n \triangleq [-n, n]^2$. Let us sketch an argument showing that if a $\mathbb{Z}^2$-invariant measure μ on Y has maximal relative entropy (with respect to horizontal shift) on every horizontal strip, it also maximizes entropy on Y (with respect to two-dimensional shift).

Indeed, let μ' be any other $\mathbb{Z}^2$-invariant measure on Y and suppose that the $\mathbb{Z}^2$-entropy of μ' is larger than the $\mathbb{Z}^2$-entropy of μ. Then there exists $\varepsilon > 0$ such that $H_{\mu'}(\xi^{F_n}) \geq H_\mu(\xi^{F_n}) + \varepsilon\,|F_n|$ for all sufficiently large n. By the UFP, there exists a non-negative integer r such that for every $y, y' \in Y$ and $n \in \mathbb{N}$, there exists a configuration $\tilde{y} \in Y$ that agrees with y on F_n and with y' outside F_{n+r}. Now, consider the strip $S \triangleq \mathbb{Z} \times [-n-r, n+r]$ and the sequence $\ldots, B_{-1}, B_0, B_1, \ldots$ of translates of F_n contained in $\mathbb{Z} \times [-n, n]$ in such a way that each B_k is at distance $r+1$ from B_{k-1} and B_{k+1}. Let us draw a random configuration $\boldsymbol{y}$ from $\Sigma^{\mathbb{Z}^2}$ by choosing $\boldsymbol{y}_{B_k}$ (for $k \in \mathbb{Z}$) according to μ', and $\boldsymbol{y}_{\mathbb{Z}^2 \setminus S}$ according to μ, all independently of one another. By the UFP, the remaining symbols can be chosen in such a way that $\boldsymbol{y}$ is (almost surely) in Y. Let $\tilde{\mu}_0$ be the distribution of $\boldsymbol{y}$. This is not necessarily horizontally invariant, so let $\tilde{\mu}$ be a horizontally invariant measure obtained from $\tilde{\mu}_0$ by the standard averaging procedure. One can now verify that when n is large enough, the relative entropy of $\tilde{\mu}$ on S given its complement is larger than that of μ, contradicting the assumption. ○

In concrete examples, the conditions of Theorem 6.6 (TMP, D-mixing and relative D-mixing) can be cumbersome to verify. Clearly, these conditions are satisfied if Y is a full shift. A more relaxed condition covering important examples such as the hard-core model is the notion of TSSM introduced in Section §2.5. The following corollary (which contains Theorem 1.6 as a special case) is a handy version of Theorem 6.6 in which generality is traded for simplicity.

Corollary 6.9 (Dobrushin–Lanford–Ruelle theorem for slices: handy version)**.** *Let Y be a subshift on a countable amenable group $\mathbb{G}$, and assume that Y satisfies TSSM. Let Φ be an absolutely summable interaction on Y. Let $\mathbb{H}$ be a subgroup of $\mathbb{G}$. Let μ be a $\mathbb{G}$-invariant probability measure on Y. Then μ is an equilibrium measure for Φ if and only if for every $\mathbb{H}$-slice S of $\mathbb{G}$, μ is an equilibrium measure on Ω_S for $\widehat{\Phi}$ relative to $\Pi_{S^c}\mu$, where $\widehat{\Phi}$ denotes the relative interaction corresponding to Φ on Ω_S.*

Proof. By Theorem 6.6, it suffices to show that Y satisfies TMP and is D-mixing, and that for every $\mathbb{H}$-slice $S \triangleq \mathbb{H}F$, the relative system Ω_S is relatively D-mixing. From Proposition 2.3, we know that Y is an SI SFT, in particular, it satisfies TMP and is D-mixing. Thus, it remains to show that Ω_S is relatively D-mixing. We shall in fact show that Ω_S is relatively SI.

Indeed, let $R \Subset \mathbb{G}$ be a finite set that certifies the TSSM property of Y. Fix $\theta \in \Theta$ and let $x, y \in X_\theta$. Let $A, B \Subset \mathbb{H}$ be such that $(AF)R \cap (BF)R = \varnothing$. Let $g_0, g_1, \ldots$ be an enumeration of the elements of $\mathbb{G} \setminus S$ and set $M_n \triangleq \{g_0, g_1, \ldots, g_n\}$ for $n \in \mathbb{N}$. Note that $\theta_{M_n} \vee x_A$ and $\theta_{M_n} \vee y_B$ are in $L(Y)$. Therefore, by TSSM, there is a configuration $z^{(n)} \in Y$ such that $z^{(n)}_{M_n} = \theta_{M_n}$, $z^{(n)}_{AF} = x_A$ and $z^{(n)}_{BF} = y_B$. Let z be an accumulation point of $z^{(n)}$ as $n \to \infty$. Since Y is closed, $z \in Y$. On the other hand, $z_{\mathbb{G}\setminus S} = \theta$, $z_{AF} = x_A$ and $z_{BF} = y_B$. Note that if we define $D \triangleq (FR)(FR)^{-1} \cap \mathbb{H}$ then whenever $A(FR) \cap B(FR) \neq \varnothing$ we have that $AD \cap BD \neq \varnothing$. This shows that X_θ is strongly irreducible with the finite set D as a witness. Since D does not depend upon θ, we find that Ω_S is relatively SI. □

7 Relative version of Meyerovitch's theorem

Before proving Theorem 1.8 we need to introduce two technical tools. One is the concept of non-overlapping patterns and the second one is a subshift which separates shapes. Let $A \subseteq \mathbb{G}$ be a finite set. We say that two patterns $u, v \in \Sigma^A$ are *non-overlapping* in Ω if

$$g_1([u] \cup [v]) \cap g_2([u] \cup [v]) = \varnothing$$

whenever $g_1, g_2 \in \mathbb{G}$ are two distinct elements with $g_1 A \cap g_2 A \neq \varnothing$. The *hard-core* shift with *shape* A is defined as

$$Y \triangleq \left\{ y \in \{\mathtt{0}, \mathtt{1}\}^{\mathbb{G}} : y_k = y_\ell = \mathtt{1} \text{ implies either } k = \ell \text{ or } kA \cap \ell A = \varnothing \right\} .$$

If we think of symbol $\mathtt{1}$ as a particle with shape A, then Y consists of all configurations of particles whose volumes do not overlap.

We will proceed through the proof in two steps. First, we treat the simpler case in which u and v are non-overlapping in Ω. We encode the relative system Ω into another relative system $\widehat{\Omega}$ in which the symbolic part contains only the information about the occurrences of u and v wherever they are interchangeable. This new system will have the relative TMP, even more, it will have the relatively independence property, and thus the relative Lanford–Ruelle theorem will yield the result. In the second step, we treat the general case where u and v might overlap. This time we use an auxiliary subshift Y (namely, the hard-core shift with shape A) and construct a new relative system $\widetilde{\Omega} \triangleq \Omega \times Y$ in which the symbolic part has an extra layer $y \in Y$ chosen independently of x and θ. The auxiliary subshift Y consists of configurations of particles on $\mathbb{G}$ that are sufficiently far apart. Associated to u and v, there are two non-overlapping patterns $\tilde{u}$ and $\tilde{v}$, which are simply u and v with a particle on top. Since $\tilde{u}$ and $\tilde{v}$ are non-overlapping, the result of the first step will hold. The general result for u and v will then follow from the independence of the auxiliary layer.

Proof of Theorem 1.8. Let $u, v \in \Sigma^A$ be non-overlapping in Ω. Without loss of generality, we shall assume that $A \ni 1_{\mathbb{G}}$; if not, we reduce to this case by shifting x, θ, u and v appropriately. Let $Z \triangleq \{⬡, ⓤ, ⓥ\}^{\mathbb{G}}$ and consider the map

$$\phi \colon \Omega \to (\Theta \times (\Sigma \cup \{☼, \star\})^{\mathbb{G}}) \times Z$$

where $\phi(\theta, x) \triangleq ((\theta, \widehat{x}), \widehat{z})$ is defined by leaving θ unchanged and setting

$$\widehat{x}_k \triangleq \begin{cases} ☼ & \text{if } k^{-1}(\theta, x) \in [u] \cup [v] \text{ and } k^{-1}\theta \in \Theta_{u,v}, \\ \star & \text{if } k \in \ell A \setminus \ell \text{ for some } \ell \in \mathbb{G} \text{ where } \ell^{-1}(\theta, x) \in [u] \cup [v] \text{ and } \ell^{-1}\theta \in \Theta_{u,v}, \\ x_k & \text{otherwise.} \end{cases}$$

$$\widehat{z}_k \triangleq \begin{cases} ⓤ & \text{if } k^{-1}(\theta, x) \in [u] \text{ and } k^{-1}\theta \in \Theta_{u,v}, \\ ⓥ & \text{if } k^{-1}(\theta, x) \in [v] \text{ and } k^{-1}\theta \in \Theta_{u,v}, \\ ⬡ & \text{otherwise.} \end{cases}$$

In other words, $\widehat{x}$ is obtained from x by erasing the appearances of u and v wherever they are interchangeable (i.e., at positions k such that u and v are interchangeable for $k^{-1}\theta$). Each erased pattern is replaced by the symbols $\star$ and ☼, where ☼ indicates the reference point of the occurrence. The information regarding the erased occurrences of u and v is then recorded in $\widehat{z}$.

The map ϕ is clearly $\mathbb{G}$-equivariant, bijective and measurable. Furthermore, given $((\theta, \widehat{x}), \widehat{z}) = \phi(\theta, x)$, one can recover x from $\widehat{x}$ and $\widehat{z}$ alone, by means of a block map. More precisely, each symbol x_k can be recovered by looking at the restrictions of $\widehat{x}$ and $\widehat{z}$ to kA^{-1} using the local rule

$$\Xi \colon (\Sigma \cup \{☼, \star\})^{A^{-1}} \times \{⬡, ⓤ, ⓥ\}^{A^{-1}} \to \Sigma$$

given by

$$\Xi(p, q) \triangleq \begin{cases} u_a & \text{if } p_{a^{-1}} = ☼ \text{ and } q_{a^{-1}} = ⓤ \text{ for some } a \in A, \\ v_a & \text{if } p_{a^{-1}} = ☼ \text{ and } q_{a^{-1}} = ⓥ \text{ for some } a \in A, \\ p_{1_{\mathbb{G}}} & \text{otherwise.} \end{cases}$$

The local rule Ξ is well-defined because u and v are non-overlapping.

Consider the system $\widehat{\Omega} \triangleq \phi(\Omega)$ where the environment $\widehat{\Theta}$ is the set of all $(\theta, \widehat{x})$ that appear in the projection of $\widehat{\Omega}$ on the first coordinate and $\widehat{X}_{(\theta,\widehat{x})}$ is the set of all $\widehat{z} \in Z$ that are consistent with $(\theta, \widehat{x})$ in $\widehat{\Omega}$. The new system $\widehat{\Omega}$ has the relative TMP — even more, it has the relative independence property. Let $\widehat{\mu} \triangleq \phi\mu$, and define $\widehat{\nu}$ as the projection of $\widehat{\mu}$ onto $\widehat{\Theta}$. Define a relative interaction $\widehat{\Phi}$ on $\widehat{\Omega}$ by

$$\begin{aligned}\widehat{\Phi}_B((\theta, \widehat{x}), \widehat{z}) &\triangleq \sum_{C: C\cdot A^{-1}=B} \Phi_C\big(\phi^{-1}((\theta, \widehat{x}), \widehat{z})\big) \\ &= \sum_{C: C\cdot A^{-1}=B} \Phi_C\Big(\theta, \big\{\Xi\big((c^{-1}\widehat{x})_{A^{-1}}, (c^{-1}\widehat{z})_{A^{-1}}\big)\big\}_{c\in C}\Big),\end{aligned}$$

and let $\widehat{E}$ denote the corresponding relative Hamiltonian. It is easy to verify that $\widehat{\Phi}$ is absolutely summable, and that, for every $\mathbb{G}$-invariant probability measure μ,

$$\mu(f_\Phi) = \widehat{\mu}(f_{\widehat{\Phi}})$$

(see Sec. §A.2.6).

We claim that $\widehat{\mu}$ is an equilibrium measure for $\widehat{\Phi}$ relative to $\widehat{\nu}$. Indeed, let $\underline{\widehat{\mu}}$ be any other $\mathbb{G}$-invariant measure that projects to $\widehat{\nu}$, and let $\underline{\mu}$ be the induced measure on Ω. Since μ is assumed to be an equilibrium measure for Φ relative to ν and $\underline{\mu}$ projects to ν, we have

$$h_{\underline{\mu}}(\Omega \mid \Theta) - \underline{\mu}(f_\Phi) \leq h_\mu(\Omega \mid \Theta) - \mu(f_\Phi) \ .$$

By the chain rule, $h_{\underline{\mu}}(\Omega \mid \Theta) = h_{\underline{\widehat{\mu}}}(\widehat{\Theta} \mid \Theta) + h_{\underline{\widehat{\mu}}}(\widehat{\Omega} \mid \widehat{\Theta})$ and $h_\mu(\Omega \mid \Theta) = h_{\widehat{\mu}}(\widehat{\Theta} \mid \Theta) + h_{\widehat{\mu}}(\widehat{\Omega} \mid \widehat{\Theta})$. As both $\underline{\widehat{\mu}}$ and $\widehat{\mu}$ project to $\widehat{\nu}$, we have $h_{\underline{\widehat{\mu}}}(\widehat{\Theta} \mid \Theta) = h_{\widehat{\mu}}(\widehat{\Theta} \mid \Theta)$. Putting this together with equation (7) yields

$$h_{\underline{\widehat{\mu}}}(\widehat{\Omega} \mid \widehat{\Theta}) - \underline{\widehat{\mu}}(f_{\widehat{\Phi}}) \leq h_{\widehat{\mu}}(\widehat{\Omega} \mid \widehat{\Theta}) - \widehat{\mu}(f_{\widehat{\Phi}}) \ ,$$

which establishes the claim.

Denote by $[\textcircled{u}]$ and $[\textcircled{v}]$ the cylinder set consisting of all points $((\theta, \widehat{x}), \widehat{z}) \in \widehat{\Omega}$ in which respectively $\textcircled{u}$ and $\textcircled{v}$ appear at position $1_{\mathbb{G}}$ of $\widehat{z}$. Recall that ξ denotes the partition of Ω induced by by the projection $(\theta, x) \mapsto x_{1_{\mathbb{G}}}$. Similarly, we denote by $\widehat{\xi}$ the partition of $\widehat{\Omega}$ induced by the projection $((\theta, \widehat{x}), \widehat{z}) \mapsto \widehat{z}_{1_{\mathbb{G}}}$, and write $\mathscr{F}_{\widehat{\Theta}}$ for the σ-algebra on $\widehat{\Omega}$ generated by $\widehat{\Theta}$. Applying Theorem 1.2(b), we know that $\widehat{\mu}$ is a relative Gibbs measure for $\widehat{\Phi}$, thus for $\widehat{\mu}$-almost every $((\theta, \widehat{x}), \widehat{z}) \in [\textcircled{u}] \cup [\textcircled{v}]$,

$$\begin{aligned}&\widehat{\mu}\big([\textcircled{u}] \,\big|\, \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}}\big)((\theta, \widehat{x}), \widehat{z}) \\ &\qquad = \frac{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{u})}}{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{u})} + \mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{v})}} \ , \end{aligned} \tag{34}$$

$$\begin{aligned}&\widehat{\mu}\big([\textcircled{v}] \,\big|\, \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}}\big)((\theta, \widehat{x}), \widehat{z}) \\ &\qquad = \frac{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{v})}}{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{u})} + \mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{v})}} \ . \end{aligned} \tag{35}$$

Putting equations (34) and (35) together, we obtain

$$\frac{\widehat{\mu}\big([\textcircled{u}] \,\big|\, \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}}\big)((\theta, \widehat{x}), \widehat{z})}{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{u})}} = \frac{\widehat{\mu}\big([\textcircled{v}] \,\big|\, \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}}\big)((\theta, \widehat{x}), \widehat{z})}{\mathrm{e}^{-\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee\textcircled{v})}} \tag{36}$$

for $\widehat{\mu}$-almost every $((\theta, \widehat{x}), \widehat{z}) \in [\textcircled{u}] \cup [\textcircled{v}]$.

On one hand, letting $(\theta, x) = \phi^{-1}((\theta, \widehat{x}), \widehat{z})$, we have

$$\begin{aligned}\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta, \widehat{x}), \widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \textcircled{u}) &= \sum_{B\ni 1_{\mathbb{G}}} \widehat{\Phi}_B((\theta, \widehat{x}), \widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \textcircled{u}) \\ &= \sum_{B\ni 1_{\mathbb{G}}} \sum_{C: C\cdot A^{-1}=B} \Phi_C(\theta, x_{A^{\mathsf{c}}} \vee u) \\ &= \sum_{C: C\cdot A^{-1}\ni 1_{\mathbb{G}}} \Phi_C(\theta, x_{A^{\mathsf{c}}} \vee u) \\ &= \sum_{C: C\cap A\neq\varnothing} \Phi_C(\theta, x_{A^{\mathsf{c}}} \vee u) \\ &= E_{A|A^{\mathsf{c}}}(\theta, x_{A^{\mathsf{c}}} \vee u) \ ,\end{aligned} \tag{37}$$

and by a similar argument

$$\widehat{E}_{\{1_{\mathbb{G}}\}|\{1_{\mathbb{G}}\}^{\mathsf{c}}}((\theta,\widehat{x}),\widehat{z}_{\{1_{\mathbb{G}}\}^{\mathsf{c}}}\vee \langle v\rangle) = E_{A|A^{\mathsf{c}}}(\theta, x_{A^{\mathsf{c}}}\vee v)\ . \tag{38}$$

On the other hand,

$$\widehat{\mu}([\langle u\rangle] \mid \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}})((\theta,\widehat{x}),\widehat{z}) = \mu([u] \mid (\xi_X)^{A^{\mathsf{c}}} \vee \mathscr{F}_{\Theta}, [u]\cup[v])(\theta,x)\ , \tag{39}$$

$$\widehat{\mu}([\langle v\rangle] \mid \widehat{\xi}^{\{1_{\mathbb{G}}\}^{\mathsf{c}}} \vee \mathscr{F}_{\widehat{\Theta}})((\theta,\widehat{x}),\widehat{z}) = \mu([v] \mid (\xi_X)^{A^{\mathsf{c}}} \vee \mathscr{F}_{\Theta}, [u]\cup[v])(\theta,x)\ . \tag{40}$$

Putting together equations (36), (37), (38), (39) and (40), we get that for μ-almost every $(\theta,x)\in[u]\cup[v]$ satisfying $\theta\in\Theta_{u,v}$,

$$\frac{\mu([u] \mid \xi^{A^{\mathsf{c}}} \vee \mathscr{F}_{\Theta})(\theta,x)}{\mathrm{e}^{-E_{A|A^{\mathsf{c}}}(\theta,x_{A^{\mathsf{c}}}\vee u)}} = \frac{\mu([v] \mid \xi^{A^{\mathsf{c}}} \vee \mathscr{F}_{\Theta})(\theta,x)}{\mathrm{e}^{-E_{A|A^{\mathsf{c}}}(\theta,x_{A^{\mathsf{c}}}\vee v)}}\ .$$

This concludes the proof in the case where u and v are non-overlapping.

We now consider the general case. If $u=v$, the result is immediate. Otherwise, let Y be the hard-core shift with shape A. We claim that there must exist a measure of maximal entropy π on Y such that $\pi([\mathtt{1}])>0$. This can be seen in various ways, for instance by verifying that Y has positive topological entropy, or by invoking the Lanford–Ruelle theorem. For a more direct argument, note that if π_0 is a $\mathbb{G}$-invariant measure such that $\pi_0([\mathtt{1}])=0$, then clearly $h_{\pi_0}(Y)=0$. Hence, it is enough to show that there exists a $\mathbb{G}$-invariant measure giving positive measure to $[\mathtt{1}]$. By Lemma 3.5, there exists a set $D\subseteq\mathbb{G}$ which is A-separated and has positive uniform lower density with respect to a Følner sequence $(F_n)_{n\in\mathbb{N}}$. Let $w\in\{\mathtt{0},\mathtt{1}\}^{\mathbb{G}}$ be the configuration with $w_k\triangleq\mathtt{1}$ if and only if $k\in D$, and define $\pi_n\triangleq|F_n|^{-1}\sum_{g\in F_n}g^{-1}\delta_w$. Any accumulation point of $(\pi_n)_{n\in\mathbb{N}}$ is a $\mathbb{G}$-invariant measure π that satisfies $\pi([\mathtt{1}])>0$.

Now consider the system $\widetilde{\Omega}\triangleq\Omega\times Y$ as a relative system with environment Θ and $\widetilde{X}_\theta\triangleq\{(x,y) : x\in X_\theta \text{ and } y\in Y\}$. Endow $\widetilde{\Omega}$ with the measure $\tilde{\mu}\triangleq\mu\times\pi$ and the interaction $\widetilde{\Phi}_C(\theta,(x,y))\triangleq\Phi_C(\theta,x)$. By construction, $\tilde{\mu}$ is an equilibrium measure for $\widetilde{\Phi}$ relative to ν. Consider now the patterns $\tilde{u},\tilde{v}\in(\Sigma\times\{\mathtt{0},\mathtt{1}\})^A$ defined by

$$\tilde{u}_a = \begin{cases}(u_a,\mathtt{1}) & \text{if } a=1_{\mathbb{G}},\\ (u_a,\mathtt{0}) & \text{otherwise,}\end{cases} \qquad\qquad \tilde{v}_a = \begin{cases}(v_a,\mathtt{1}) & \text{if } a=1_{\mathbb{G}},\\ (v_a,\mathtt{0}) & \text{otherwise.}\end{cases}$$

By the definition of Y and the fact that $u\neq v$, the patterns $\tilde{u},\tilde{v}$ are non-overlapping in $\widetilde{\Omega}$. We can thus apply the result for non-overlapping patterns to obtain that for $\tilde{\mu}$-almost every $(\theta,(x,y))\in[\tilde{u}]\cup[\tilde{v}]$ such that $\theta\in\Theta_{\tilde{u},\tilde{v}}$,

$$\frac{\tilde{\mu}\big([\tilde{u}]\,\big|\,\tilde{\xi}^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,(x,y))}{\mathrm{e}^{-\widetilde{E}_{A|A^{\mathsf{c}}}(\theta,\tilde{u}\vee(x,y)_{A^{\mathsf{c}}})}} = \frac{\tilde{\mu}\big([\tilde{v}]\,\big|\,\tilde{\xi}^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,(x,y))}{\mathrm{e}^{-\widetilde{E}_{A|A^{\mathsf{c}}}(\theta,\tilde{v}\vee(x,y)_{A^{\mathsf{c}}})}}\ , \tag{41}$$

where $\tilde{\xi}$ denotes the partition of $\widetilde{\Omega}$ induced by $(\theta,(x,y))\mapsto(x_{1_{\mathbb{G}}},y_{1_{\mathbb{G}}})$ and $\widetilde{E}$ is the relative Hamiltonian associated to $\widetilde{\Phi}$. With some abuse of notation, we write $\mathscr{F}_{\Theta}$ for the σ-algebras generated by Θ both in Ω and in $\widetilde{\Omega}$.

By the definition of $\widetilde{\Phi}$, we have that

$$\widetilde{E}_{A|A^{\mathsf{c}}}(\theta,(x,y)) = E_{A|A^{\mathsf{c}}}(\theta,x)\ . \tag{42}$$

Furthermore, as $\tilde{\mu}=\mu\times\pi$, we have

$$\tilde{\mu}\big([\tilde{u}]\,\big|\,\tilde{\xi}^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,(x,y)) = \mu\big([u]\,\big|\,\xi^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,x)\cdot\pi\big([\mathtt{1}]\,\big|\,\zeta^{A^{\mathsf{c}}}\big)(y)\ , \tag{43}$$

$$\tilde{\mu}\big([\tilde{v}]\,\big|\,\tilde{\xi}^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,(x,y)) = \mu\big([v]\,\big|\,\xi^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,x)\cdot\pi\big([\mathtt{1}]\,\big|\,\zeta^{A^{\mathsf{c}}}\big)(y)\ , \tag{44}$$

where ζ stands for the partition of Y generated by the symbol at the origin.

Substituting (42), (43) and (44) in equation (41), we get that for μ-almost every $(\theta,x)\in[u]\cup[v]$ satisfying $\theta\in\Theta_{u,v}$ and π-almost every $y\in[\mathtt{1}]$,

$$\begin{aligned}&\frac{\mu\big([u]\,\big|\,\xi^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,x)\cdot\pi\big([\mathtt{1}]\,\big|\,\zeta^{A^{\mathsf{c}}}\big)(y)}{\mathrm{e}^{-E_{A|A^{\mathsf{c}}}(\theta,x_{A^{\mathsf{c}}}\vee u)}}\\ &\qquad = \frac{\mu\big([v]\,\big|\,\xi^{A^{\mathsf{c}}}\vee\mathscr{F}_{\Theta}\big)(\theta,x)\cdot\pi\big([\mathtt{1}]\,\big|\,\zeta^{A^{\mathsf{c}}}\big)(y)}{\mathrm{e}^{-E_{A|A^{\mathsf{c}}}(\theta,x_{A^{\mathsf{c}}}\vee v)}}\ .\end{aligned} \tag{45}$$

Note that we may replace the condition "$y \in [1]$" by "$y_{\mathbb{G}\setminus\{1_{\mathbb{G}}\}} \vee 1 \in Y$" and the equality will still hold. Also, if we integrate the factor $\pi\big([1] \,\big|\, \zeta^{A^c}\big)(y)$ with respect to π, we obtain

$$\pi([1]) = \int_{y:y_{\{1_{\mathbb{G}}\}^c} \vee 1 \in Y} \pi\big([1] \,\big|\, \zeta^{A^c}\big)(y)\, \mathrm{d}\pi(y) + \int_{y:y_{\{1_{\mathbb{G}}\}^c} \vee 1 \notin Y} \pi\big([1] \,\big|\, \zeta^{A^c}\big)(y)\, \mathrm{d}\pi(y) \ ,$$

where the second term is 0. Thus, integrating (45) with respect to π, we obtain

$$\frac{\mu\big([u] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x) \cdot \pi\big([1]\big)}{\mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee u)}} = \frac{\mu\big([v] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x) \cdot \pi\big([1]\big)}{\mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee v)}} \ .$$

As $\pi([1]) > 0$, it follows that

$$\frac{\mu\big([u] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee u)}} = \frac{\mu\big([v] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, x_{A^c} \vee v)}} \ .$$

for μ-almost every $(\theta, x) \in [u] \cup [v]$ such that $\theta \in \Theta_{u,v}$. This concludes the proof of the theorem. □

We have used the relative Lanford–Ruelle theorem to prove Theorem 1.8. We now show the converse implication, so that the two theorems are really equivalent under fairly simple reductions. More specifically, we show that when Ω has the relative TMP, the conclusion of Theorem 1.8 becomes equivalent to saying that μ is a relative Gibbs measure for Φ with marginal ν.

Proof of Theorem 1.2(b) using Theorem 1.8. Let $A \Subset \mathbb{G}$ and let $B \supseteq A$ be a memory set for A witnessing the TMP of Ω relative to ν. Let $u, v \in \Sigma^A$ be arbitrary patterns. Then, for every $w \in \Sigma^{B\setminus A}$ and ν-almost every $\theta \in \Theta$, the patterns $w \vee u$ and $w \vee v$ are interchangeable for θ provided they are both in $L_B(X_\theta)$. From Theorem 1.8, it follows that for every $w \in \Sigma^{B\setminus A}$,

$$\frac{\mu\big([w \vee u] \,\big|\, \xi^{B^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{B|B^c}(\theta, w \vee u \vee x_{B^c})}} = \frac{\mu\big([w \vee v] \,\big|\, \xi^{B^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{B|B^c}(\theta, w \vee v \vee x_{B^c})}}$$

for μ-almost every $(\theta, x) \in [w]$ such that $w \vee u, w \vee v \in L_B(X_\theta)$. If we apply the chain rule to the numerators above and decompose the exponents in the denominators, and then cancel the common factor

$$\frac{\mu\big([w] \,\big|\, \xi^{B^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{(B\setminus A)|B^c}(\theta, w \vee x_{B^c \cup A})}} ,$$

then the resulting expression simplifies to

$$\frac{\mu\big([u] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, u \vee x_{A^c})}} = \frac{\mu\big([v] \,\big|\, \xi^{A^c} \vee \mathscr{F}_\Theta\big)(\theta, x)}{\mathrm{e}^{-E_{A|A^c}(\theta, v \vee x_{A^c})}}$$

for μ-almost every $(\theta, x) \in [w]$ such that $x_{A^c} \vee u, x_{A^c} \vee v \in X_\theta$. This is true for every $w \in \Sigma^{B\setminus A}$. The latter equality is equivalent to μ being a relative Gibbs measure. □

Considering the fact that in the proof of Theorem 1.8 we only applied the relative Lanford–Ruelle theorem on a relatively independent system, and that the relative Lanford–Ruelle theorem can be deduced from Theorem 1.8 as shown above, we obtain that the following three statements are essentially equivalent in the relative setting:

- The relative Lanford–Ruelle theorem for systems which satisfy relative independence.
- The relative Lanford–Ruelle theorem for systems satisfying the relative TMP.
- The relative version of Meyerovitch's theorem.

If we restrict exclusively to the non-relative setting, the Lanford–Ruelle theorem for subshifts with TMP (or even for SFTs) does not follow from the Lanford–Ruelle theorem for full shifts. Similarly, Meyerovitch's theorem cannot be deduced from Lanford–Ruelle through a simple recoding. The addition of an environment in the relative setting can be used as a tool to fix a measure on a restricted portion of a dynamical system and give information about measures which project to that portion and are optimal outside of it. Hence, the three statements become equally powerful in this setting. We see this as an indication that the relative setting is the appropriate level of generalization for these results.

References

[1] M. Allahbakhshi and A. Quas. Class degree and relative maximal entropy. *Transactions of the American Mathematical Society*, 365(3):1347–1368, 2013. doi:10.1090/S0002-9947-2012-05637-6.

[2] M. Boyle and M. Schraudner. $\mathbb{Z}^d$ group shifts and Bernoulli factors. *Ergodic Theory and Dynamical Systems*, 28(2):367–387, 2008. doi:10.1017/S0143385707000697.

[3] R. Briceño. The topological strong spatial mixing property and new conditions for pressure approximation. *Ergodic Theory and Dynamical Systems*, 38(5):1658–1696, 2017. doi:10.1017/etds.2016.107.

[4] R. Burton and J. E. Steif. Non-uniqueness of measures of maximal entropy for subshifts of finite type. *Ergodic Theory and Dynamical Systems*, 14(2):213–235, 1994. doi:10.1017/S0143385700007859.

[5] T. Ceccherini-Silberstein and M. Coornaert. *Cellular Automata and Groups*. Springer, 2010. doi:10.1007/978-3-642-14034-1.

[6] N. Chandgotia, G. Han, B. Marcus, T. Meyerovitch, and R. Pavlov. One-dimensional Markov random fields, Markov chains and topological Markov fields. *Proceedings of the American Mathematical Society*, 142(1):227–242, 2014. doi:10.1090/S0002-9939-2013-11741-7.

[7] N. Chandgotia and T. Meyerovitch. Markov random fields, Markov cocycles and the 3-colored chessboard. *Israel Journal of Mathematics*, 215(2):909–964, 2016. doi:10.1007/s11856-016-1398-2.

[8] N.-P. Chung and H. Li. Homoclinic groups, IE groups, and expansive algebraic actions. *Inventiones mathematicae*, 199(3):805–858, 2015. doi:10.1007/s00222-014-0524-1.

[9] R. L. Dobrushin. Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and Its Applications*, 2(4):292–301, 1968. doi:10.1007/BF01075681.

[10] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2004. doi:10.1017/CBO9780511755347.

[11] H. Föllmer. On entropy and information gain in random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 26(3):207–217, 1973. doi:10.1007/BF00532723.

[12] H. Föllmer and J. L. Snell. An "inner" variational principle for Markov fields on a graph. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 39(3):187–195, 1977. doi:10.1007/BF00535471.

[13] H.-O. Georgii. Spontaneous magnetization of randomly dilute ferromagnets. *Journal of Statistical Physics*, 25(3):369–396, 1981. doi:10.1007/BF01010795.

[14] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. Walter de Gruyter, 1988. doi:10.1515/9783110250329.

[15] R. B. Griffiths and J. L. Lebowitz. Random spin systems: Some rigorous results. *Journal of Mathematical Physics*, 9(8):1284–1292, 1968. doi:10.1063/1.1664710.

[16] O. Häggström, R. H. Schonmann, and J. E. Steif. The Ising model on diluted graphs and strong amenability. *The Annals of Probability*, 28(3):1111–1137, 2000. doi:10.1214/aop/1019160327.

[17] S. Kass and K. Madden. A note on the uniform filling property and strong irreducibility. *Acta Applicandae Mathematicae*, 126(1):253–262, 2013. doi:10.1007/s10440-013-9816-5.

[18] D. Kerr and H. Li. *Ergodic Theory: Independence and Dichotomies*. Springer, 2016. doi:10.1007/978-3-319-49847-8.

[19] Y. Kifer. Multidimensional random subshifts of finite type and their large deviations. *Probability Theory and Related Fields*, 103(2):223–248, 1995. doi:10.1007/bf01204216.

[20] Y. Kifer. On the topological pressure for random bundle transformations. In V. Turaev and A. Vershik, editors, *Topology, Ergodic Theory, Real Algebraic Geometry*, volume 202 of *American Mathematical Society Translations: Series 2*, pages 197–214. American Mathematical Society, 2001. doi:10.1090/trans2/202/14.

[21] Y. Kifer. Thermodynamic formalism for random transformations revisted. *Stochastics and Dynamics*, 8(1):77–102, 2008. doi:10.1142/S0219493708002238.

[22] Y. Kifer and P.-D. Liu. Random dynamics. In B. Hasselblatt and A. Katok, editors, *Handbook of Dynamical Systems*, volume 1B, chapter 5, pages 379–499. Elsevier, 2006. doi:10.1016/S1874-575X(06)80030-5.

[23] B. Kitchens and K. Schmidt. Periodic points, decidability and Markov subgroups. In J. C. Alexander, editor, *Dynamical Systems*, volume 1342 of *Lecture Notes in Mathematics*, pages 440–454. Springer, 1988. doi:10.1007/BFb0082845.

[24] O. E. Lanford III and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics*, 13(3):194–215, 1969. doi:10.1007/BF01645487.

[25] F. Ledrappier. Pressure and variational principle for random Ising model. *Communications in Mathematical Physics*, 56(3):297–302, 1977. doi:10.1007/BF01614214.

[26] F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *Journal of the London Mathematical Society*, s2-16(3):568–576, 1977. doi:10.1112/jlms/s2-16.3.568.

[27] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995. doi:10.1017/CBO9780511626302.

[28] D. Lind and K. Schmidt. Homoclinic points of algebraic $\mathbb{Z}^d$-actions. *Journal of the American Mathematical Society*, 12(4):953–980, 1999. doi:10.1090/S0894-0347-99-00306-9.

[29] E. Lindenstrauss. Pointwise theorems for amenable groups. *Inventiones Mathematicae*, 146(2):259–295, 2001. doi:10.1007/s002220100162.

[30] B. Marcus and R. Pavlov. An integral representation for topological pressure in terms of conditional probabilities. *Israel Journal of Mathematics*, 207(1):395–433, 2015. doi:10.1007/s11856-015-1178-4.

[31] T. Meyerovitch. Gibbs and equilibrium measures for some families of subshifts. *Ergodic Theory and Dynamical Systems*, 33(3):934–953, 2013. doi:10.1017/S0143385712000053.

[32] E. F. Moore. Machine models of self-reproduction. In *Proceedings of Symposia in Applied Mathematics*, pages 17–33. American Mathematical Society, 1962.

[33] J. Moulin Ollagnier. *Ergodic Theory and Statistical Mechanics*. Springer, 1985. doi:10.1007/bfb0101575.

[34] J. Moulin Ollagnier and D. Pinchon. Mesures de Gibbs invariantes et mesures d'equilibre. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(1):11–23, 1981. doi:10.1007/BF01013457.

[35] J. Moulin Ollagnier and D. Pinchon. The variational principle. *Studia Mathematica*, 72(2):151–159, 1982.

[36] J. Myhill. The converse of Moore's Garden-of-Eden theorem. *Proceedings of the American Mathematical Society*, 14(4):685–686, 1963. doi:10.2307/2034301.

[37] D. Ornstein and B. Weiss. The Shannon-McMillan-Breiman theorem for a class of amenable groups. *Israel Journal of Mathematics*, 44(1):53–60, 1983. doi:10.1007/bf02763171.

[38] D. Ornstein and B. Weiss. Entropy and isomorphism theorems for actions of amenable groups. *Journal d'Analyse Mathématique*, 48(1):1–141, 1987. doi:10.1007/bf02790325.

[39] D. Ruelle. *Thermodynamic Formalism.* Cambridge University Press, 2nd edition, 2004. doi:10.1017/CBO9780511617546.

[40] V. Salo. When are groups shifts of finite type? *Preprint*, 2018. arXiv:1807.01951.

[41] K. Schmidt. *Dynamical Systems of Algebraic Origin.* Birkhäuser, 1995. doi:10.1007/978-3-0348-0277-2.

[42] T. Seppäläinen. Entropy, limit theorems, and variational principles for disordered lattice systems. *Communications in Mathematical Physics*, 171(2):233–277, 1995. doi:10.1007/BF02099271.

[43] A. A. Tempelman. Specific characteristics and variational principle for homogeneous random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):341–365, 1984. doi:10.1007/BF00533741.

[44] P. Walters. Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts. *Transactions of the American Mathematical Society*, 296(1):1–31, 1986. doi:10.1090/S0002-9947-1986-0837796-8.

[45] B. Zegarlinski. Interactions and pressure functionals for disordered lattice systems. *Communications in Mathematical Physics*, 139(2):305–339, 1991. doi:10.1007/BF02352497.

A Appendix

A.1 Topology of $\mathcal{P}_\nu(\Omega)$

Let ν be a probability measure on Θ. Every measure $\mu \in \mathcal{P}_\nu(\Omega)$ defines a positive linear functional J on the Banach space $C_\Theta(\Omega)$. Every such functional is *ν-normalized* meaning that $J(\mathbb{1}_{A\times\Sigma^{\mathbb{G}}}) = \nu(A)$ for every measurable $A \subseteq \Theta$. When Θ is a standard Borel space (i.e., isomorphic, as a measurable space, to a Borel subset of a complete separable metric space), the converse is also true.

Proposition A.1 (Relative Riesz theorem)**.** *Assume that Θ is a standard Borel space. Then, for every ν-normalized positive linear functional J on $C_\Theta(\Omega)$ there corresponds a unique measure $\mu \in \mathcal{P}_\nu(\Omega)$ such that $\mu(f) = J(f)$ for all $f \in C_\Theta(\Omega)$.*

Proof. Without loss of generality (by passing through an isomorphism), we can assume that Θ is a compact metric space equipped with its Borel σ-algebra (see e.g. [10, Thm. 13.1.1]). Then the set $C(\Omega)$ of all continuous functions on Ω is a Banach subspace of $C_\Theta(\Omega)$. By the Riesz representation theorem, the restriction of J to $C(\Omega)$ identifies a unique probability measure μ on Ω such that $\mu(f) = J(f)$ for all $f \in C(\Omega)$.

Let $g\colon \Theta \to \mathbb{R}$ be a bounded measurable function and $[u]$ a cylinder set. Consider $f(\theta, x) \triangleq g(\theta)\mathbb{1}_{[u]}(\theta, x)$. Then f is relatively local. Furthermore, every relatively local function on Ω is a linear combination of functions of this form. Since the relatively local functions are dense in $C_\Theta(\Omega)$ and both J and μ are continuous on $C_\Theta(\Omega)$, it is enough to verify that $\mu(f) = J(f)$.

Let $\varepsilon > 0$. By Lusin's theorem, there exists a function $g_\varepsilon \in C(\Theta)$ and a closed set $E \subseteq \Theta$ such that $g = g_\varepsilon$ on E and $\nu(\Theta \setminus E) < \varepsilon$ and $\mu\big((\Theta \setminus E) \times \Sigma^{\mathbb{G}}\big) < \varepsilon$. Furthermore, we can choose g_ε such that $\|g_\varepsilon\| \leq \|g\|$. Define $f_\varepsilon(\theta, x) \triangleq g_\varepsilon(\theta)\mathbb{1}_{[u]}(\theta, x)$. Since $f_\varepsilon \in C(\Omega)$, we have $\mu(f_\varepsilon) = J(f_\varepsilon)$. Note that

$$|\mu(f_\varepsilon) - \mu(f)| \leq \mu(|f - f_\varepsilon|) \leq (\|f\| + \|f_\varepsilon\|)\mu\big((\Theta \setminus E) \times \Sigma^{\mathbb{G}}\big) < 2\,\|f\|\,\varepsilon\ .$$

Similarly, since J is positive linear, we have

$$|J(f_\varepsilon) - J(f)| \leq J(|f - f_\varepsilon|) \leq (\|f\| + \|f_\varepsilon\|)\nu(\Theta \setminus E) < 2\,\|f\|\,\varepsilon\ .$$

Therefore, $|\mu(f) - J(f)| < 4\,\|f\|\,\varepsilon$. Since ε is arbitrary, the claim follows. □

A consequence of the above proposition is that when Θ is a standard Borel space, the space $\mathcal{P}_\nu(\Omega)$ is compact. Indeed, as a set of linear functionals, $\mathcal{P}_\nu(\Omega)$ is a closed subset of the unit ball in the dual space $C^*_\Theta(\Omega)$, thus the compactness follows from Alaoglu's theorem. We do not know whether the assumption that Θ is standard Borel is necessary for the compactness of $\mathcal{P}_\nu(\Omega)$.

A.2 Omitted arguments

A.2.1 Verification of (3)

Let $B \Subset \mathbb{G}$ be a finite set and define $\partial_B^- F_n \triangleq \{g \in F_n : gB \cap F_n^{\mathsf{c}} \neq \varnothing\} = F_n \setminus \bigcap_{b\in B} F_n b^{-1}$. We have

$$\sum_{\substack{C\Subset\mathbb{G}\\ C\cap F_n\neq\varnothing\\ C\cap F_n^{\mathsf{c}}\neq\varnothing}} \|\Phi_C\| = \sum_{\substack{C\Subset\mathbb{G}\\ C\cap\partial_B^- F_n\neq\varnothing\\ C\cap F_n^{\mathsf{c}}\neq\varnothing}} \|\Phi_C\| + \sum_{\substack{C\Subset\mathbb{G}\\ C\cap F_n\neq\varnothing, C\cap F_n^{\mathsf{c}}\neq\varnothing\\ C\cap\partial_B^- F_n=\varnothing}} \|\Phi_C\|$$
$$\leq \left|\partial_B^- F_n\right| \|\Phi\| + \underbrace{\left|F_n \setminus \partial_B^- F_n\right|}_{\leq |F_n|} \sum_{\substack{C\Subset\mathbb{G}\\ C\ni 1_{\mathbb{G}}, C\not\subseteq B}} \|\Phi_C\| \ .$$

The first term is $o(|F_n|)$, whereas the second term is of the form $c_B\,|F_n|$ where $c_B \to 0$ as $B \nearrow \mathbb{G}$ along the finite subsets of $\mathbb{G}$ directed by inclusion. □

A.2.2 Verification of (4)

We have

$$\left\|E_{B|B^{\mathsf{c}}} - E_{A|A^{\mathsf{c}}}\right\| = \left\| \sum_{\substack{C\Subset\mathbb{G}\\ C\cap B\neq\varnothing}} \Phi_C - \sum_{\substack{C\Subset\mathbb{G}\\ C\cap A\neq\varnothing}} \Phi_C \right\|$$
$$\leq \sum_{\substack{C\Subset\mathbb{G}\\ C\cap B\neq\varnothing\\ C\cap A=\varnothing}} \|\Phi_C\|$$
$$\leq \sum_{\substack{C\Subset\mathbb{G}\\ C\cap(B\setminus A)\neq\varnothing}} \|\Phi_C\|$$
$$\leq \sum_{c\in(B\setminus A)} \sum_{C\ni c} \|\Phi_C\|$$
$$= |B\setminus A|\,\|\Phi\| \ .$$

□

A.2.3 Verification of (5)

Using the definition of f_Φ, for every finite set $A \Subset \mathbb{G}$, we have

$$\left|E_A(\theta,x) - \sum_{g\in A} f_\Phi(g^{-1}\theta, g^{-1}x)\right| \leq \sum_{\substack{C\Subset\mathbb{G}\\ C\cap A\neq\varnothing\\ C\cap A^{\mathsf{c}}\neq\varnothing}} \frac{|A\cap C|}{|C|} \|\Phi_C\| \ .$$

For $A \triangleq F_n$, the estimate (5) follows as in (3) (see Sec. §A.2.1). □

A.2.4 Verification of (19) and (20)

Inequality (19) follows by writing

$$Z_{F_n|F_n^{\mathsf{c}}}(\theta,x) = \sum_{\substack{u\in\Sigma^{F_n}\\ u\vee x_{F_n^{\mathsf{c}}}\in X_\theta}} \mathrm{e}^{-E_{F_n|F_n^{\mathsf{c}}}(\theta,u\vee x_{F_n^{\mathsf{c}}})} \leq \sum_{u\in L_{F_n}(X_\theta)} \mathrm{e}^{-E_{F_n}(\theta,u)+o(|F_n|)} = Z_{F_n}(\theta)\,\mathrm{e}^{o(|F_n|)} \ .$$

In order to verify (20), let us use the shorthand $\partial F_n^\theta \triangleq F_n^\theta \setminus F_n$. We can write

$$Z_{F_n^\theta|(F_n^\theta)^{\mathsf{c}}}(\theta,x) = \sum_{\substack{u\in\Sigma^{F_n}\\ u\vee\partial u\vee x_{(F_n^\theta)^{\mathsf{c}}}\in X_\theta}} \sum_{\partial u\in\Sigma^{\partial F_n^\theta}} \mathrm{e}^{-E_{F_n}(\theta,u)-E_{\partial F_n^\theta|(\partial F_n^\theta)^{\mathsf{c}}}(\theta,u\vee\partial u\vee x_{(F_n^\theta)^{\mathsf{c}}})}$$
$$= \sum_{u\in L_{F_n}(X_\theta)} \mathrm{e}^{-E_{F_n}(\theta,u)} \sum_{\substack{\partial u\in\Sigma^{\partial F_n^\theta}\\ u\vee\partial u\vee x_{(F_n^\theta)^{\mathsf{c}}}\in X_\theta}} \mathrm{e}^{-E_{\partial F_n^\theta|(\partial F_n^\theta)^{\mathsf{c}}}(\theta,u\vee\partial u\vee x_{(F_n^\theta)^{\mathsf{c}}})} \ .$$

Now observe that, since F_n^θ is a mixing set for F_n, the second sum in the latter inequality is non-empty. It follows that

$$Z_{F_n^\theta|(F_n^\theta)^c}(\theta,x) \geq \sum_{u\in L_{F_n}(X_\theta)} \mathrm{e}^{-E_{F_n}(\theta,u)}\mathrm{e}^{-|\partial F_n^\theta|\|\Phi\|} = Z_{F_n}(\theta)\,\mathrm{e}^{-|\partial F_n^\theta|\|\Phi\|}\ .$$

□

A.2.5 Verification of (3.2)

The right-hand side is $(\xi^B \vee \mathscr{F}_\Theta)$-measurable and for every $[u]\in\xi^B$ and $W\in\mathscr{F}_\Theta$ we have

$$\begin{aligned}
\int_{[u]\cap W} \frac{\mu(\mathbb{1}_{[x_B]}f\,|\,\mathscr{F}_\Theta)(\theta,x)}{\mu([x_B]\,|\,\mathscr{F}_\Theta)(\theta,x)}\,\mathrm{d}\mu(\theta,x) &= \mu\left(\mathbb{1}_{[u]}\mathbb{1}_W\frac{\mu(\mathbb{1}_{[u]}f\,|\,\mathscr{F}_\Theta)}{\mu([u]\,|\,\mathscr{F}_\Theta)}\right)\\
&= \mu\left(\mathbb{1}_{[u]}\frac{\mu(\mathbb{1}_{[u]}\mathbb{1}_W f\,|\,\mathscr{F}_\Theta)}{\mu([u]\,|\,\mathscr{F}_\Theta)}\right)\\
&= \mu\left(\mu\left(\mathbb{1}_{[u]}\frac{\mu(\mathbb{1}_{[u]}\mathbb{1}_W f\,|\,\mathscr{F}_\Theta)}{\mu([u]\,|\,\mathscr{F}_\Theta)}\,\middle|\,\mathscr{F}_\Theta\right)\right)\\
&= \mu\left(\frac{\mu(\mathbb{1}_{[u]}\mathbb{1}_W f\,|\,\mathscr{F}_\Theta)}{\cancel{\mu([u]\,|\,\mathscr{F}_\Theta)}}\cancel{\mu(\mathbb{1}_{[u]}\,|\,\mathscr{F}_\Theta)}\right)\\
&= \int_{[u]\cap W} f\,\mathrm{d}\mu
\end{aligned}$$

If two bounded measurable functions have equal integrals over each element of a generating semi-algebra, they are almost surely equal. □

A.2.6 Verification of (7)

For every $(\theta,x)\in\Omega$, we have

$$\begin{aligned}
f_{\widehat{\Phi}}(\phi(\theta,x)) &= \sum_{B\ni 1_{\mathbb{G}}}\frac{1}{|B|}\widehat{\Phi}(\phi(\theta,x))\\
&= \sum_{B\ni 1_{\mathbb{G}}}\frac{1}{|B|}\sum_{C:C\cdot A^{-1}=B}\Phi_C(\theta,x)\\
&= \sum_{C:C\cdot A^{-1}\ni 1_{\mathbb{G}}}\frac{1}{|C\cdot A^{-1}|}\Phi_C(\theta,x)\ .
\end{aligned}$$

Integrating with respect to a measure μ, we get

$$\widehat{\mu}(f_{\widehat{\Phi}}) = \sum_{C:C\cdot A^{-1}\ni 1_{\mathbb{G}}}\frac{1}{|C\cdot A^{-1}|}\,\mu(\Phi_C)\ . \tag{46}$$

Compare this with the expression

$$\mu(f_\Phi) = \sum_{C\ni 1_{\mathbb{G}}}\frac{1}{|C|}\,\mu(\Phi_C)\ , \tag{47}$$

and observe that when μ is $\mathbb{G}$-invariant, the right-hand sides of (46) and (47) coincide. □