The Variational Principle for Gibbs States Fails on Trees

Robert M. Burton
Charles-Ed. Pfister
Jeffrey E. Steif


Supported by Federal Ministry of Science and Research, Austria
Available via http://www.esi.ac.at
The Variational Principle for Gibbs States Fails on Trees

Robert M. Burton *  Charles-Ed. Pfister  Jeffrey E. Steif

Abstract

We show how the variational principle for Gibbs States (which says that for the \(d\)-dimensional cubic lattice, the set of translation invariant Gibbs States is the same as the set of translation invariant measures which maximize entropy minus energy and moreover that this quantity corresponds to the pressure) fails for nearest neighbor finite state statistical mechanical systems on the homogeneous 3-ary tree. Given an interaction there is a unique measure \(\mu\) maximizing entropy minus energy, and we give necessary and sufficient conditions so that it is a Gibbs State for that interaction, and that the maximum is equal to the pressure. In the case of a 2-state system, these conditions define a 2-dimensional manifold of the natural 3-dimensional parameter space of interactions, so that generically in the interactions the entropy minus energy for \(\mu\) is strictly less than the pressure and \(\mu\) is not a Gibbs State (for those parameter values).

AMS 1991 subject classification. Primary 60G60; secondary 82B20 82B26.

Key words and phrases. Trees, Gibbs States, variational principle.

1 Introduction

In this paper, we consider nearest neighbor statistical mechanical systems on \(E^T\) where \(T\) is the 3-ary homogeneous tree and \(E\) an arbitrary finite set. \(T\) is therefore the unique connected graph which has countably infinitely many vertices, contains no loops, and in which every vertex is adjacent to exactly 3 vertices. The mathematical study of such systems was initiated in [7] and [9]; see [4] for more recent results and references. Our main goal is to study the variational problem

\[
\sup_{\mu}\{\mathcal{H}(\mu) - \epsilon(\mu)\}
\]

where the supremum is taken over all homogeneous probability measures on \(E^T\); here \(\mathcal{H}(\mu)\) is the specific entropy of \(\mu\) and \(\epsilon(\mu)\) the specific energy. By definition, a homogeneous measure is a measure which is invariant under all automorphisms of \(T\). (An automorphism of \(T\) is a bijection of the vertex set which preserves adjacencies.) We show that generically the variational principle for Gibbs States fails on trees, in the sense that the measure which gives the maximum for the above variational problem is not a Gibbs State for the corresponding energy. All our results extend immediately to the case of a \(k\)-ary homogeneous tree with \(k \geq 3\).

We first mention an important property of trees which is the source of some of their difficulties. Fix an element of \(T\) arbitrarily and call it 0. There is an obvious distance function, which we denote by \(d(\cdot, \cdot)\), on the vertices of \(T\), which is simply the number of
edges needed to go from one vertex to another. Let \( T_n \) denote the set of all vertices of \( T \) whose distance from 0 is at most \( n \). A simple calculation shows that
\[
|T_n| = 3(2^n) - 2.
\]
The key property of trees referred to above is that \( |T_n|/|T_{n-1}| \) does not tend to 0 as \( n \to \infty \). This easy observation (that “surface to volume ratio” does not tend to 0) is the exact source of the difficulties of dealing with trees and what causes among other things the failure of the variational principle. (Of course for \( \Z^d \), the analogue of \( |T_n|/|T_{n-1}| \) (which is \( |\Lambda_n|/|\Lambda_{n-1}| \) with \( \Lambda_n \) being \([-n,n]^d \cap \Z^d \)) does tend to 0 as \( n \to \infty \), a property referred to as amenability in the context of groups.)

While we will briefly explain the concept of Gibbs States for certain interactions, the reader is referred to [2] and [4] for more details. We will discuss these concepts both for the \( d \)-dimensional cubic lattice \( \Z^d \) and for the homogeneous 3-ary tree \( T \). This will allow us to describe the contrast in the variational principle in these two settings.

We first describe these concepts for \( \Z^d \). For simplicity, we will only deal with nearest neighbor interactions which are invariant under all automorphisms of \( \Z^d \). Such interactions are specified by a pair of real-valued functions \((\chi, \varphi)\), where \( \chi \), the one-body interaction, is defined on \( E \) and \( \varphi \), the two-body interaction, is a symmetric function on \( E \times E \). For example, if \( E = \{-1,1\} \), then the space of interactions is a 5-dimensional parameter space; however it is sufficient to consider a 3-dimensional subspace, since adding a constant to \( \chi \) or \( \varphi \) does not change the theory (see comment after Theorem 1.5). We can therefore restrict our attention to the 3-dimensional subspace of interactions parametrized by \((J_1, J_2, h)\),
\[
\chi(-1) = -h, \quad \chi(1) = h
\]
and
\[
\varphi(-1,-1) = J_1, \quad \varphi(-1,1) = \varphi(1,-1) = 0, \quad \varphi(1,1) = J_2
\]

Our state space will be \( X = E^{\Z^d} \). For \( x = (x_1, x_2, \ldots, x_d) \in \Z^d \), we let \( \|x\| \) be \(|x_1| + |x_2| + \ldots + |x_d| \). For any finite set \( A \subseteq \Z^d \), let \( B(A) = \{ x \notin A : \exists y \in A \text{ with } \|x-y\| = 1 \} \). We also let \( \Lambda_n \) be \([-n,n]^d \cap \Z^d \).

Given \( \delta \in E^{B(\Lambda_n)} \), we let \( \mu_{n}^{\delta} \) be the measure on \( E^{\Lambda_n} \) given by
\[
\mu_{n}^{\delta}(\eta) = \frac{e^{-H_{n}^{\delta}(\eta)}}{Z_{n}^{\delta}}
\]
where
\[
H_{n}^{\delta}(\eta) = -\sum_{x \in \Lambda_n} \chi(\eta(x)) - \frac{1}{2} \sum_{x,y \in \Lambda_n \atop \|x-y\|=1} \varphi(\eta(x),\eta(y)) - \sum_{x \in \Lambda_n, y \in B(\Lambda_n) \atop \|x-y\|=1} \varphi(\eta(x),\delta(y))
\]
and \( Z_{n}^{\delta} \) is a normalization making \( \mu_{n}^{\delta} \) a probability measure.

**Definition 1.1:** A measure \( \mu \) on \( X \) is called a Gibbs State (with respect to the interaction \((\chi, \varphi)\)) if for each \( n \), the conditional distribution on \( \Lambda_n \) given the configuration \( \delta \) on \( \Lambda_n^c \) is given by \( \mu_{n}^{\delta'} \) above where \( \delta' \) is the restriction of \( \delta \) to \( B(\Lambda_n) \).

Although it will play a small role in this paper, it is of much importance and interest that for \( d \geq 2 \), there are interactions for which there is more than one Gibbs State.

The following theorem is well known, with a much more general formulation being found in [2] and [4].
Theorem 1.2:
\[ \lim_{n \to \infty} \frac{\ln(Z_n^\delta)}{|\Lambda_n|} \]
exists and is independent of \( \delta \) (where \( \delta \) can also be taken to be free boundary conditions which means that we delete the last sum in the definition of \( H_n^\delta(\eta) \)). This limit is called the pressure and denoted by \( P = P(\chi, \varphi) \).

By a translation invariant probability measure on \( X \), we will mean a probability measure on \( X \) which is invariant under the canonical \( \mathbb{Z}^d \) action of translations.

**Definition 1.3:** Suppose that \( \mu \) is a translation invariant probability measure on \( X \). Then the entropy of \( \mu \), denoted by \( H(\mu) \), is
\[ \lim_{n \to \infty} \frac{1}{|\Lambda_n|} \sum_{\eta \in E^\Lambda_n} \mu_n(\eta) \ln \mu_n(\eta) \]
where by \( \mu_n(\eta) \) is meant the measure of \( \eta \) with respect to \( \mu \) restricted to \( \Lambda_n \).

It is an easy consequence of subadditivity that the above limit exists (see [2] or [4]).

**Definition 1.4:** Suppose that \( \mu \) is a translation invariant probability measure on \( X \). Then the energy of \( \mu \) (with respect to the interaction \((\chi, \varphi))\,\), denoted by \( e(\mu) \), is
\[ \lim_{n \to \infty} \frac{1}{|\Lambda_n|} \sum_{\eta \in E^\Lambda_n} \mu_n(\eta) H_n^\delta(\eta) \]
where \( \mu_n(\eta) \) is as in Definition 1.3.

It is shown in [2] and [4] that this limit exists and is independent of \( \delta \) where \( \delta \) can again be taken to be free boundary conditions.

Note that the pressure depends only on the interaction, that entropy depends only on the measure (and not on the interaction) and that energy depends both on the measure and on the interaction.

The following theorem is a special case of a theorem which is one of the cornerstones of the theory of Gibbs States in statistical mechanics. It is proved in [2] and [4]. It is this theorem which will turn out to fail on trees.

**Theorem 1.5 (The Variational Principle):** Fix the interaction \((\chi, \varphi))\,\). Suppose that \( \mu \) is a translation invariant probability measure on \( X \). Then
\[ H(\mu) - e(\mu) \leq P \]
with equality holding if and only if \( \mu \) is a Gibbs State (for the interaction).

One can show that for all interactions, there is at least one translation invariant Gibbs State. Note that adding a constant \( c_1 \) to \( \chi \) and a constant \( c_2 \) to \( \varphi \) changes the value of \( e(\mu) \) and \( P \) by the same constant \( c_1 + c_2 \); this is why it is sufficient to consider a 3-dimensional subspace of the interactions in the 2-state case.

We now need to give all of the above definitions when \( \mathbb{Z}^d \) is replaced by the homogeneous 3-ary tree \( T \). Fix an element of \( T \) arbitrarily and call it 0. The choice of the vertex 0 defines uniquely an orientation of the edges of \( T \), so that all oriented edges incident to 0 have 0 as initial point, and any vertex different from 0 is the final point of exactly one incident edge. An oriented edge is denoted by \( \gamma = \{x, y\} \), with \( x \) the initial vertex of \( \gamma \).
Our state space will now be \( Y = E^T \). For any finite set \( A \subseteq T \), let \( B(A) = \{ x \notin A : \exists y \in A \text{ with } d(x, y) = 1 \} \). We also let \( \partial(A) = \{ x \in A : \exists y \notin A \text{ with } d(x, y) = 1 \} \). We will be using both of these types of boundaries throughout the paper.

Given an interaction \((\chi, \varphi)\) and \( \delta \in E^{B(T_n)} \), we let \( \mu_n^{\delta} \) be the measure on \( E^{T_n} \) given by
\[
\mu_n^{\delta}(\eta) = \frac{e^{-H_n^{\delta}(\eta)}}{Z_n^{\delta}}
\]
where
\[
H_n^{\delta}(\eta) = -\sum_{x \in T_n} \chi(\eta(x)) - \frac{1}{2} \sum_{\{x, y\} \in \partial(T_n), |x-y|=1} \varphi(\eta(x), \eta(y)) - \sum_{\{x, y\} \in \partial(T_n), |x-y|=1} \varphi(\eta(x), \delta(y))
\]
and \( Z_n^{\delta} \) is a normalization making \( \mu_n^{\delta} \) a probability measure. If the last sum is deleted in the definition of \( H_n^{\delta}(\eta) \) (in which case the dependence on \( \delta \) is removed), we write \( H_n^{\delta}(\eta) \) instead and write \( Z_n^{\delta} \) and \( \mu_n^{\delta} \) for the corresponding normalization and probability measure respectively.

**Definition 1.6:** A measure \( \mu \) on \( Y \) is called a **Gibbs State** (with respect to the interaction \((\chi, \varphi)\)) if for each \( n \), the conditional distribution on \( T_n \) given the configuration \( \delta \) on \( T_n^c \) is given by \( \mu_n^{\delta} \) above where \( \delta' \) is the restriction of \( \delta \) to \( B(T_n) \).

We now need analogues of pressure, entropy and energy in the setting of trees.

**Theorem 1.7:** The limit
\[
\lim_{n \to \infty} \frac{\ln(Z_n^{\delta})}{|T_n|}
\]
exists. This limit, denoted by \( P = P(\chi, \varphi) \), is called the **pressure**.

The following theorem states that entropy can be defined for homogeneous measures on trees. This result is proved in [1].

**Theorem 1.8:** Suppose that \( \mu \) is a homogeneous probability measure on \( Y \). Then the limit
\[
\lim_{n \to \infty} \frac{1}{|T_n|} \sum_{\eta \in E^{T_n}} \mu_n(\eta) \ln \mu_n(\eta)
\]
exists where by \( \mu_n(\eta) \) is meant the measure of \( \eta \) with respect to \( \mu \) restricted to \( T_n \). This limit will (in analogy with \( Z^d \)) be denoted by \( \mathcal{H}(\mu) \) and be called the **entropy** of \( \mu \).

Our next result tells us that one can define the energy of a homogeneous measure. The proof is trivial and left to the reader.

**Theorem 1.9:** Suppose that \( \mu \) is a homogeneous probability measure on \( Y \). Then the limit
\[
\lim_{n \to \infty} \frac{1}{|T_n|} \sum_{\eta \in E^{T_n}} \mu_n(\eta) H_n^{\delta}(\eta)
\]
exists and is equal to \( \mu(\epsilon_{\gamma}) \), where
\[
\epsilon_{\gamma}(\eta) = -\varphi(\eta(x), \eta(y)) - \chi(\eta(y))
\]
with \( \gamma = \{x, y\} \) an arbitrary oriented edge and \( \mu_n(\eta) \) is as in Theorem 1.8. This limit will (in analogy with \( Z^d \)) be denoted by \( \epsilon(\mu) \) and be called the **energy** of \( \mu \) (with respect to the interaction \((\chi, \varphi)\)).
The main result is formulated in terms of properties of the interaction matrix. For each oriented edge $\gamma = \{x, y\}$ the interaction matrix $I_\gamma$ is defined as

\[
I_\gamma(\eta) = \exp(-\epsilon_\gamma(\eta)) = (\exp\{\varphi(\eta(x), \eta(y)) + \chi(\eta(y))\}).
\]

(This is of course an $|E|$ by $|E|$ matrix). The interaction matrix is positive. It has therefore a largest positive eigenvalue $\lambda$ and positive right and left eigenvectors $r$ and $l$,

\[
I_\gamma r = \lambda r, \quad lI_\gamma = \lambda l
\]

which are normalized so that

\[
\sum_{\eta(x)} l(\eta(x)) \cdot r(\eta(x)) = 1.
\]

(These are of course independent of $\gamma$.)

We now state the main result of this paper.

**Theorem 1.10:** If the interaction $(\chi, \varphi)$ is such that $\sum_{\eta_x} I_\gamma(\eta_x, \eta_y)$ is not independent of $\eta_x$, then there is a homogeneous measure $\mu$ such that

1. $\mathcal{H}(\mu) - \epsilon(\mu) > \mathcal{H}(\nu) - \epsilon(\nu)$ for all homogeneous measures $\nu \neq \mu$
2. $\mathcal{H}(\mu) - \epsilon(\mu) < P(\chi, \varphi)$ and
3. $\mu$ is not a Gibbs State for $(\chi, \varphi)$.

On the other hand, if $\sum_{\eta_x} I_\gamma(\eta_x, \eta_y)$ is independent of $\eta_x$, then there is a homogeneous measure $\mu$ such that

1'. $\mathcal{H}(\mu) - \epsilon(\mu) > \mathcal{H}(\nu) - \epsilon(\nu)$ for all homogeneous measures $\nu \neq \mu$
2'. $\mathcal{H}(\mu) - \epsilon(\mu) = P(\chi, \varphi)$ and
3'. $\mu$ is the Gibbs State with free boundary conditions for $(\chi, \varphi)$.

In the case $E = \{-1, 1\}$, the set of interactions $(J_1, J_2, h)$ which satisfy the property that $\sum_{\eta_x} I_\gamma(\eta_x, \eta_y)$ is independent of $\eta_x$, is

\[
S = \{(J_1, J_2, h) \in \mathbb{R}^3 : e^{-J_1 h} + e^{-J_2 h} - e^{-h} = e^{-h} + e^{J_1 h + J_2 h}\}.
\]

The function $f(x, y, z) = e^{x+y} + e^{x-z} - e^{y+z}$ has a nonzero gradient everywhere and so $S$ is a smooth 2-dimensional submanifold of $\mathbb{R}^3$. Using Theorem 1 in [9] together with simple algebra gives that the two parameters $(J_1, J_2, h)$ and $(J_1', J_2', h')$ yield the same family of Gibbs States if and only if $J_1 + J_2 = J_1' + J_2'$ and $3(J_1 - J_2) - 4h = 3(J_1' - J_2') - 4h'$. (One would then say, using the terminology of [4], that $(J_1, J_2, h)$ and $(J_1', J_2', h')$ are equivalent interactions; this notion of equivalence is different from the notion introduced by Ruelle [8].) One can also easily check that given $(J_1, J_2, h) \in \mathbb{R}^3$, there exists $(J_1', J_2', h') \in S$ which is equivalent to $(J_1, J_2, h)$. Since $S$ is a 2-dimensional manifold in $\mathbb{R}^3$ and the set of interactions equivalent to a given interaction is a 1-dimensional subspace of $\mathbb{R}^3$ (being the intersection of 2 nonparallel planes), this is of course expected. While 2 parameters equivalent in the above sense have the same set of Gibbs States, it is possible that the variational principle holds for one of these parameters and not for the other and it would therefore be inappropriate to identify these parameters when one is interested in the variational principle; it is natural to use our 3-parameter space of interactions.

The fact that we can find interactions $(\chi, \varphi)$ and $(\chi', \varphi')$ which have the same set of Gibbs states, but such that the first part of Theorem 1.10 holds for $(\chi, \varphi)$, and the second part of Theorem 1.10 holds for $(\chi', \varphi')$, is also true in the general case. Before stating the result we recall the notion of a homogeneous Markov chain on $T$ [9], which plays an
important role below. Let \( M \) be a reversible stochastic matrix defined on \( E \times E \), with stationary distribution \( \pi \), \( \pi M = \pi \). Let \( A \subseteq T \) be any finite connected set, and \( \delta : A \to E \) any map from \( A \) to \( E \). We order \( A \) as \( \{x_1, x_2, \ldots, x_k\} \) in such a way that each \( x_j \) for \( j > 1 \) is the neighbor of exactly one \( x_i \in \{x_1, x_2, \ldots, x_{j-1}\} \). Let \( i = i(j) \) denote this neighbor (in which case \( i(2) = 1 \)). We then let

\[
\mu_M(\eta = \delta \text{ on } A) = \pi(\delta(x_1)) \prod_{j=2}^k M(\delta(x_{i(j)}), \delta(x_j)).
\]

Because of the reversibility of \( M \) this value is independent of the ordering of \( x_1, x_2, \ldots, x_k \); it is also easy to verify the necessary consistency conditions which must be satisfied in order that the \( \mu_M(\eta = \delta \text{ on } A) \) define an homogeneous measure \( \mu_M \) on \( Y \), which is called the **Markov chain** associated to \( M \).

**Theorem 1.11:** Let \( \mu \) be a Gibbs State for the interaction \((\chi, \varphi)\). If \( \mu \) is also a homogeneous Markov chain, then there exists an equivalent interaction \((\chi', \varphi')\) such that \( \mu \) is a Gibbs State for \((\chi', \varphi')\), and \( \mu \) is the unique measure such that

\[
P(\chi', \varphi') = \mathcal{H}(\mu) - e'(\mu) > \mathcal{H}(\nu) - e'(\nu)
\]

for all homogeneous measures \( \nu \neq \mu \).

We mention other relevant papers on this subject. In [9], the problem of computing the number of extremal Gibbs States on trees is studied but there are some mistakes in this paper (see [6]). We mention [3] as a complementary paper. In [3], it is mentioned that certain properties of Gibbs States fails on trees, in particular that 2 Gibbs States for the same interaction parameters do not necessarily have relative entropy per site equal to 0, a fact known to hold on \( \mathbb{Z}^d \). In [3], a substitute for the usual variational principle (Theorem 1.5 here), a type of inner variational principle, is also obtained which is valid for a large class of graphs, including trees. In order to deal with the difficulty mentioned at the beginning of the introduction, that the “surface to volume ratio” does not tend to 0, the authors split off boundary effects and replace “bulk” quantities by “inner quantities”. (This is similar to the strategy used in Statistical Physics.) In [12], a variational principle for Gibbs States on amenable groups is also obtained and a distinction is made between lower entropy (which is more or less that given in [3]) and upper entropy (which is the definition of entropy given here). The setting in the present paper is of course a nonamenable situation. We finally mention three other papers dealing with related topics. [5] deals with Gibbs States and phase transitions on countable not necessarily amenable groups. [10] deals with extending the variational principle for topological entropy to amenable groups. Finally, [11] shows that this variational principle can be violated for the action of a free group with \( n > 1 \) generators.

For the benefit of the reader, we describe briefly and in vague terms what the above notion of amenability is. Sticking to countable groups, the amenable groups are the groups for which classical ergodic theory can be extended (at least to some extent). Loosely speaking, a group is amenable if there exists a sequence of sets whose boundary to volume ratios go to 0. For example, for the \( d \)-dimensional cubic lattice, the sets \( \Lambda_n \) form such a sequence. Equivalently, these groups have the property that whenever they act on a compact metric space, there exists an invariant probability measure for this action.

Although the proof of the above result does not rely on having an exact formula for the pressure \( P \), it turns out that on trees, one can obtain an explicit formula for the pressure. We give two results in this direction for the 2-state case \( E = \{-1, 1\} \).
Theorem 1.12: \( \lim_{n \to \infty} \frac{\ln(Z_n)}{|T_n|} \) exists and is given by

\[
\frac{\ln(e^{-h} + e^{h})}{2} + \sum_{n=0}^{\infty} \frac{\ln(\psi^n(\frac{e^{-h}}{e^{-h} + e^{h}})))}{2n+2}
\]

where

\[\phi(x) = \frac{w_1(x)}{w_1(x) + w_2(x)}\]

and

\[\psi(x) = w_1(x) + w_2(x)\]

where

\[w_1(x) = e^{-h+2J_1}x^2 + 2e^{-h+J_2}x(1-x) + e^{-h}(1-x)^2\]

and

\[w_2(x) = e^{h+2J_2}(1-x)^2 + 2e^{h+J_2}x(1-x) + e^{h}x^2.\]

Theorem 1.13: Let \(Z_n^\delta\) be the normalization for \(\mu_n^\delta\) above where the boundary condition \(\delta\) is taken to be identically \(-1\). Then

\[
\lim_{n \to \infty} \frac{\ln(Z_n^\delta)}{|T_n|} = \frac{\ln(e^{-3h+2J_1} + e^{-h})}{2} + \sum_{n=0}^{\infty} \frac{\ln(\psi^n(\frac{e^{-3h+2J_1}}{e^{-3h+2J_1} + e^{-h}})))}{2n+2}
\]

where \(\phi(x)\) and \(\psi(x)\) are as in Theorem 1.12.

By a similar argument one can see that the pressure with “plus” boundary conditions exists and that (for most parameters), it is different than that obtained in Theorem 1.13. This implies, among other things, that pressure cannot exist for arbitrary boundary conditions since we could alternate plus and minus boundary conditions on different layers obtaining a sequence which has two limit points.

The rest of this paper is devoted to the proofs of Theorems 1.10, 1.11, 1.12 and 1.13. Theorems 1.10 and 1.11 will be proved in \$2\ while Theorems 1.12 and 1.13 will be proved in \$3.

We thank Olle Håggström for reading an early draft of this paper and pointing out an improvement in our parameterization.

2 Proof of Theorems 1.10 and 1.11

Given the interaction matrix \(I_\gamma\), there are two natural ways of forming an \(|E|\) by \(|E|\) stochastic matrix: either we define \(K_\gamma\) by

\[K_\gamma(\eta(x), \eta(y)) = \frac{I_\gamma(\eta(x), \eta(y))r(\eta(y))}{\lambda r(\eta(x))}\]

or \(M_\gamma\) by

\[M_\gamma(\eta(x), \eta(y)) = \frac{I_\gamma(\eta(x), \eta(y))}{\sum_{\eta(y)} I_\gamma(\eta(x), \eta(y))}\]

The next lemma is elementary and left to the reader.
Lemma 2.1: The matrices $K_\gamma$ and $M_\gamma$ are reversible and the stationary distribution of $K_\gamma$ is $\pi(\eta(x))=i(\eta(x))\cdot r(\eta(x))$. Moreover, the following four statements are equivalent:

a) The right eigenvector $r$ of $I_\gamma$ has all entries equal.

b) $\sum_{\eta_x} L_\gamma(\eta_x, \eta_y)$ is independent of $\eta_x$. (The common value is the eigenvalue $\lambda$ of $I_\gamma$.)

c) $l(\eta_x) = \exp(\chi(\eta_x))$ is a left eigenvector of $I_\gamma$.

d) $K_\gamma = M_\gamma$.

Proof of Theorem 1.10: Let $(\chi, \varphi)$ be an interaction, and $I_\gamma$ be the corresponding interaction matrix. We first find the value of the supremum of

$$\mathcal{H}(\mu) - \epsilon(\mu)$$

over all homogeneous probability measures on $T$. The first step consists in showing that the supremum is equal to the supremum over all homogeneous Markov chains: for every homogeneous measure $\mu$ there exists a Markov chain $\mu_M$ such that

$$\mathcal{H}(\mu_M) - \epsilon(\mu_M) \geq \mathcal{H}(\mu) - \epsilon(\mu).$$

Indeed, let $M$ be the stochastic matrix

$$M(\eta(x), \eta(y)) = \left(\frac{\mu(\eta(x), \eta(y))}{\mu(\eta(x))}\right)$$

where $x$ and $y$ are two nearest neighbors. (If $\mu(\eta(x)) = 0$, then the corresponding row is defined by setting $M(\eta(x), \eta(y)) = 0$ for $x \neq y$ and $M(\eta(x), \eta(x)) = 1.$) This stochastic matrix has $\mu(\eta(x))$ as stationary distribution and is reversible since $\mu$ is homogeneous. By definition of $\mu_M$,

$$\epsilon(\mu) = \mu(\epsilon_\gamma) = \mu_M(\epsilon_\gamma) = \epsilon(\mu_M)$$

and it is a standard result that

$$\mathcal{H}(\mu_M) \geq \mathcal{H}(\mu).$$

Let therefore $\mu$ be a homogeneous Markov chain which is characterized by the reversible stochastic matrix $M$ and stationary distribution $\rho$. We have

$$(2.1) \quad \mu[\{\omega : \omega(x) = \eta(x), x \in T_n\}] := \rho(\eta(0)) \prod_{\gamma = \{x,y\} \in T_n} M(\eta(x), \eta(y)).$$

An elementary computation gives

$$(2.2) \quad \mathcal{H}(\mu) - \epsilon(\mu) =$$

$$\sum_{\eta(x) \in E} \rho(\eta(x)) \sum_{\eta(y) \in E} M(\eta(x), \eta(y)) \{-\ln M(\eta(x), \eta(y)) - \epsilon_\gamma(\eta(x), \eta(y))\}$$

The supremum is therefore not greater than the supremum over all stationary Markov chains on $\mathbb{Z}$. Since the variational principle holds for the model defined on $\mathbb{Z}$ (Theorem 1.5), the supremum is attained by the Gibbs State of the model defined on $\mathbb{Z}$, which is the Markov chain on $\mathbb{Z}$ associated to $(K_\gamma, \pi)$. The supremum is easily computed from formula (2.2), and is equal to $\ln \lambda$, where $\lambda$ is the positive maximal eigenvalue of the interaction matrix $I_\gamma$. Since there is a unique Gibbs state for the model on $\mathbb{Z}$, for any homogeneous measure $\mu$ on $T$, which is different from the Markov chain $\nu$ on $T$ associated to $(K_\gamma, \pi)$, we have

$$\mathcal{H}(\mu) - \epsilon(\mu) < \ln \lambda.$$
This proves (1) and (1').

There are two cases to consider. If one of the equivalent conditions of Lemma 2.1 is verified, then the Markov chain \( \nu \) on \( T \) defined above is the Gibbs measure with free boundary conditions, since on \( T_n \) the measure \( \mu_n^f \) can be expressed as

\[
\mu_n^f := \frac{\exp(\chi(\eta(0))) \prod_{\gamma = \{x,y\} \in T_n} \gamma(x, \eta(y))}{Z_n^f}
\]

Comparing (2.1) and (2.3) one finds immediately that

\[
P(\chi, \varphi) = \lim_n \frac{\ln Z_n^f}{|T_n|} = \ln \lambda.
\]

This proves (2') and (3').

If one of the equivalent conditions of Lemma 2.1 is not verified, then the Markov chain \( \nu \) on \( T \) is not a Gibbs State. Indeed, let \( x \neq 0 \) and \( \gamma_1 = \{y_1, x\} \), \( \gamma_2 = \{x, y_2\} \) and \( \gamma_3 = \{x, y_3\} \) be the three incident edges to \( x \); we have

\[
\nu(\eta(x)|\eta(y), y \neq x) = \frac{K_{\gamma_1}(\eta(y_1), \eta(x)) \prod_{i=2,3} K_{\gamma_i}(\eta(x), \eta(y_i))}{\sum_{\eta(x)} K_{\gamma_1}(\eta(y_1), \eta(x)) \prod_{i=2,3} K_{\gamma_i}(\eta(x), \eta(y_i))}
\]

since \( r(\eta(x)) \) is not constant. This proves (3).

It remains to show that

\[
\lim_n \frac{\ln Z_n^f}{|T_n|} > \ln \lambda.
\]

This is done with the help of Lemma 2.1, which should be viewed as a variational principle for finite sets. Its proof which we skip is a simple application of Jensen’s inequality and is given in [13].

**Lemma 2.1:** Let \( a_1, a_2, \ldots, a_n \) be real numbers and \( p_1, p_2, \ldots, p_n \) be a probability vector. Then

\[-\sum_i p_i \ln p_i - \sum_i a_i p_i \leq \ln(\sum_i e^{-a_i})
\]

with equality if and only if \( p_i = e^{-a_i}/\sum_i e^{-a_i} \) for all \( i \).

Let us consider the second sum in (2.2) when \( \eta(x) \) is fixed. The maximum of this sum over the measures \( M(\eta(x), \cdot) \) is attained, according to Lemma 2.1, by the measure \( M_\gamma(\eta(x), \cdot) \), where \( M_\gamma \) is the stochastic matrix of Lemma 2.1. Let \( \theta(\eta(x)) \) be the difference between this maximum and the value obtained for the measure \( K_\gamma(\eta(x), \cdot); \) since \( K_\gamma \neq M_\gamma \),

\[
\sum_{\eta(x)} \pi(\eta(x)) \theta(\eta(x)) \equiv \alpha > 0.
\]

On \( T_n \) we define a measure \( \nu_n^\ast \) by modifying on the boundary \( \partial T_n \) the Markov chain \( \nu \) associated to \( K_\gamma \) and \( \pi \):

\[
\nu_n^\ast[\{\omega : \omega(x) = \eta(x), x \in T_n\}] :=
\]
\[ \pi(\eta(0)) \prod_{\gamma = \{x, y\} \in T_n \gamma \cap \partial T_n = \emptyset} K_\gamma(\eta(x), \eta(y)) \prod_{\gamma = \{x, y\} \in T_n \gamma \cap \partial T_n \neq \emptyset} M_\gamma(\eta(x), \eta(y)) \]

Using Lemma 2.1 we have
\[
\ln Z_n^d = - \sum_{\eta \in E^{T_n}} \mu_n^d(\eta) \ln \mu_n^d(\eta) - \sum_{\eta \in E^{T_n}} \mu_n^d(\eta) H_n^d(\eta)
\geq - \sum_{\eta \in E^{T_n}} \nu_n^*(\eta) \ln \nu_n^*(\eta) - \sum_{\eta \in E^{T_n}} \nu_n^*(\eta) H_n^d(\eta)
= - \sum_{\eta \in \partial T_n} \pi(\eta(0)) \{\ln \pi(\eta(0)) - \chi(\eta(0))\} + T_n |\ln \lambda + |\partial T_n| \alpha.
\]

Since |\partial T_n| and |T_n| are of the same order this proves (2). ∎

**Proof of Theorem 1.11:** Let \( M \) be the transition matrix associated to the Markov chain \( \mu \) and \( \rho \) the corresponding stationary distribution. The stochastic matrix \( M \) is strictly positive and reversible. Therefore the matrix
\[ A(\eta(x), \eta(y)) = \frac{M(\eta(x), \eta(y))}{\rho(\eta(y))} \]
is symmetric. We set
\[ \chi'(\eta(y)) := \ln \rho(\eta(y)) \]
and
\[ \varphi'(\eta(x), \eta(y)) := \ln A(\eta(x), \eta(y)) \]
Thus \( M \) is the interaction matrix for the interaction \( (\chi', \varphi') \). (2.3) now immediately gives that \( \mu \) is the Gibbs State with free boundary conditions with respect to the interaction \( (\chi', \varphi') \). This proves that the interaction \( (\chi', \varphi') \) is equivalent to the interaction \( (\chi, \varphi) \). Since \( M \) is stochastic, it has a positive right eigenvector (corresponding to the maximum eigenvalue of 1) which is constant and therefore Lemma 2.1 and Theorem 1.10 now imply that
\[ P(\chi', \varphi') = H(\mu) - \epsilon'(\mu) > H(\nu) - \epsilon'(\nu) \]
for all homogeneous measures \( \nu \neq \mu \). ∎

### 3 Proof of Theorem 1.12

In order to compute the pressure explicitly, it is easier to compute the pressure first on the binary tree (where one has the advantage of obtaining a recursion which allows the calculation) and then to extend this result to the homogeneous tree \( T \).

We have a fixed origin 0 in \( T \) and we let \( T_1, T_2 \) and \( T_3 \) denote the 3 components of \( T \setminus \{0\} \). Let \( B = \{0\} \cup T_1 \cup T_2 \) which we call the **binary tree**. Clearly all the vertices have 3 neighbors except 0 which has only 2 and which we think of as a root. Let \( B_n \) denote the set of all vertices of \( B \) whose distance from 0 is at most \( n \) (the distance function being the restriction of \( d(\cdot, \cdot) \) to \( B \)).

We let \( Z_n^d = \sum_{\eta \in \{-1, 1\}^B_n} e^{-H_n^d(\eta)} \) where
\[
H_n^d(\eta) = -J_2 \sum_{\{x, y\} \in B_n} I_{\eta(x) = 1, \eta(y) = 0} - J_1 \sum_{\{x, y\} \in B_n} I_{\eta(x) = 1, \eta(y) = 0} - h \sum_{x \in B_n} \eta(x).
\]
We will now explicitly compute $\lim_{n \to \infty} \ln (Z'_n)/|B_n|$.

**Proposition 3.1:**

$$\lim_{n \to \infty} \frac{\ln (Z'_n)}{|B_n|} = \frac{\ln (e^{-h} + e^{h})}{2} + \sum_{n=0}^{\infty} \frac{\ln (\phi^n (x^{-h} + x^h))}{2^{n+2}}$$

where

$$\phi(x) = \frac{w_1(x)}{w_1(x) + w_2(x)}$$

and

$$\psi(x) = w_1(x) + w_2(x)$$

where

$$w_1(x) = e^{-h+2J_1} x^2 + 2e^{-h} J_1 x (1-x) + e^{-h}(1-x)^2$$

and

$$w_2(x) = e^{h+2J_2}(1-x)^2 + 2e^{h} J_2 x (1-x) + e^{h}x^2.$$ 

**Proof:** Let $L : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ be the so-called Schur product defined by

$$L(u, v) = (u_1 v_1, u_2 v_2)$$

where $u = (u_1, u_2)$ and $v = (v_1, v_2)$, $M$ be the matrix

$$\begin{pmatrix} e^{J_1} & 1 \\ 1 & e^{J_2} \end{pmatrix},$$

$B$ be the mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by

$$B(v) = L(Mv, Mv),$$

and $F$ be the mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by

$$F(v) = L((e^{-h}, e^{h}), B(v)).$$

We next let for $i = \pm 1$

$$f_n(i) = \sum_{\eta \in \{-1, 1\}^{\mu_n, \nu_n}} e^{-H_n(v)}$$

where we clearly have $Z'_n = f_n(-1) + f_n(1)$. Let $f_n$ denote the vector $(f_n(-1), f_n(1))$.

The following key lemma which we prove later tells us why the operator $F$ is useful in computing entropy.

**Lemma 3.2:** $f_n = F^n(e^{-h}, e^{h})$.

Clearly $F$ preserves the positive quadrant $\{(x,y), x \geq 0, y \geq 0\}$ and preserves rays in the sense that for all $\alpha > 0$, $F(\alpha v) = \alpha^2 F(v)$. $F$ induces a map $G$ on the 1-simplex $\{(x,y), x \geq 0, y \geq 0, x+y = 1\}$ given by

$$G(v) = \frac{F(v)}{\|F(v)\|_1}.$$
where \( \| \cdot \|_1 \) denotes the \( L^1 \) norm on \( \mathbb{R}^2 \) given by \( \| (x, y) \| = |x| + |y| \). We can view \( G \) as just acting on the first coordinate of the simplex giving us the mapping \( \phi \) from \([0, 1] \to [0, 1] \) given by

\[
\phi(x) = G(x, 1 - x) \cdot (1, 0)
\]

which by an easy computation is the map \( \psi \) given in the statement of Proposition 3.1.

Note also that the function \( \psi \) given in the statement of Proposition 3.1 is simply

\[
\psi(x) = \| F(x, 1 - x) \|_1.
\]

An induction argument (using \( F(\alpha v) = \alpha^2 F(v) \)) now gives us that for any \((x, 1 - x)\) in the 1-simplex

\[
F^n(x, 1 - x) = (\psi(x))^{2n-1} (\psi(\phi(x)))^{2n-2} (\psi(\phi^2(x)))^{2n-3} \ldots
\]

\[
(\psi(\phi^{n-1}(x)))^{2n} (\phi^n(x), 1 - \phi^n(x)).
\]

It follows (again using \( F(\alpha v) = \alpha^2 F(v) \)) that

\[
F^n(\varepsilon^{-h}, e^h) = F^n((\varepsilon^{-h} + e^h)(\varepsilon^{-h} + e^h)) =
\]

\[
(e^{-h} + e^h)^{2n} (\psi(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-1} (\psi(\phi(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-2} (\psi(\phi^2(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-3} \ldots (\psi(\phi^{n-1}(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n} (\phi^n(\frac{\varepsilon^{-h}}{e^{-h} + e^h}), 1 - \phi^n(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))
\]

which implies by Lemma 3.2 that

\[
Z_n' = F^n(\varepsilon^{-h}, e^h) \cdot (1, 1) =
\]

\[
(e^{-h} + e^h)^{2n} (\psi(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-1} (\psi(\phi(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-2} (\psi(\phi^2(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n-3} \ldots (\psi(\phi^{n-1}(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))^{2n} (\phi^n(\frac{\varepsilon^{-h}}{e^{-h} + e^h}), 1 - \phi^n(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))
\]

where \( |B_n| = 2^{n+1} - 1 \)

\[
\frac{\ln Z_n'}{|B_n|} =
\]

\[
\frac{2^n}{2^{n+1} - 1} \ln(\varepsilon^{-h} + e^h) + \frac{2^{n-1}}{2^{n+1} - 1} \ln(\psi(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))) + \]

\[
\frac{2^{n-2}}{2^{n+1} - 1} \ln(\psi(\phi(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))) + \frac{2^{n-3}}{2^{n+1} - 1} \ln(\psi(\phi^2(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))) + \ldots +
\]

\[
\frac{2^0}{2^{n+1} - 1} \ln(\psi(\phi^{n-1}(\frac{\varepsilon^{-h}}{e^{-h} + e^h}))).
\]

Now letting \( n \to \infty \) in the above, we finally obtain

\[
\lim_{n \to \infty} \frac{\ln Z_n'}{|B_n|} = \frac{\ln(\varepsilon^{-h} + e^h)}{2} + \sum_{n=0}^{\infty} \frac{\ln(\psi(\phi^n(\frac{\varepsilon^{-h}}{e^{-h} + e^h})))}{2^{n+2}}
\]

as desired, proving Proposition 3.1. \( \square \)
Proof of Lemma 3.2: The main part of this proof is to prove that

$$f_n = L((e^{-h}, e^h), B(f_{n-1})) \forall n \geq 1$$

from which the result of the lemma follows easily by induction. In the following computation, we let

$$g(a, b) = \begin{cases} 
J_2 & \text{if } a = b = 1 \\
J_1 & \text{if } a = b = -1 \\
0 & \text{if } a = -b 
\end{cases}.$$ 

We also denote the two neighbors of the origin 0 by \(z_1\) and \(z_2\). Let \(L(R)\) (to be thought of as right and left) denote the component of \(B \setminus \{0\}\) containing \(z_1\) (\(z_2\)). If \(\eta \in \{-1, 1\}^{B_n \setminus B_i}\) and \(i, j, k \pm 1\), let \(\eta^{i,j,k}\) be the element in \(\{-1, 1\}^{B_n}\) which is \(\eta\) on \(B_n \setminus B_i\), \(i \neq 0, k \neq 0\) at \(z_1\) and \(\ell\) at \(z_2\). If \(\eta \in \{-1, 1\}^{(B_n \setminus B_i) \cap L}\) \(k \pm 1\), let \(\eta_k\) be the element in \(\{-1, 1\}^{(B_n \setminus B_i) \cap L}\) which is \(\eta\) on \((B_n \setminus B_i) \cap L\) and \(k \neq 0\) at \(z_1\). Similarly, if \(\delta \in \{-1, 1\}^{(B_n \setminus B_i) \cap R}\) \(\ell \pm 1\), let \(\delta\) be the element in \(\{-1, 1\}^{(B_n \setminus B_i) \cap R}\) which is \(\delta\) on \((B_n \setminus B_i) \cap R\) and \(\ell \neq 0\) at \(z_2\).

This proves Equation (3.1) and finally the proof of the lemma is completed by an easy induction left to the reader. \(\square\)

Proof of Theorem 1.12: Let \(g\) be defined as in the proof of Lemma 3.2 and let \(z_1\) be some neighbor of 0 in \(T\). Let \(c_1, c_2 > 0\) be such that \(c_1 \leq e^{g(x,y)} \leq c_2\) for all \(x\) and \(y\).
We then have
\[ Z_n^f = \sum_{\eta \in \{-1,1\}^T} e^{-H_n^f(\eta)} = \sum_{\eta \in \{-1,1\}^T} e^{\sum_{x,y \in T_n, v(x,y) = 1} g(\eta(x), \eta(y)) + h \sum_{x \in T_n} \eta(x)} \]

\[ c_2 \sum_{\eta \in \{-1,1\}^T} e^{\sum_{x,y \in T_n, v(x,y) = 1} g(\eta(x), \eta(y)) + h \sum_{x \in T_n} \eta(x)} \]

\[ c_2 \left( \sum_{\eta \in \{-1,1\}^T} e^{\sum_{x,y \in T_n, v(x,y) = 1} g(\eta(x), \eta(y)) + h \sum_{x \in T_n} \eta(x)} \right) \]

\[ c_2 Z_n^f Z_{n-1}^f \]

the second to last inequality obtained by looking separately at \( \eta \) on the component of \( T_n \setminus \{z_1\} \) containing 0 and on the component of \( T_n \setminus 0 \) containing \( z_1 \). We then get

\[ \frac{\ln(Z_n^f)}{|T_n|} \leq \frac{\ln(c_2)}{|T_n|} + \frac{\ln(Z_n^f)}{|T_n|} + \frac{\ln(Z_{n-1}^f)}{|T_n|} = \]

\[ \frac{\ln(c_2)}{|T_n|} + \frac{|B_n| \ln(Z_n^f)}{|T_n|} + \frac{|B_{n-1}| \ln(Z_{n-1}^f)}{|T_n|} \]

Using the fact that \( |T_n| = |B_n| + |B_{n-1}| \) and Proposition 3.1, this last quantity approaches the sum given in the statement of Theorem 1.12. Using \( c_1 \) instead of \( c_2 \), we can get an analogous lower bound, proving the result. □

We mention the following. It is possible to obtain an exact formula for the second derivative of \( P(J_1, J_2, h) \) at the point \((0,0,0)\) in certain directions. The fact that this is strictly larger than the second derivative for the pressure on the 1-dimensional lattice (and that their first derivatives are the same) together with the fact that the entropy and energy are the same for a Markov chain on \( T \) as they are for the corresponding 1-dimensional chain implies that in some directions near 0, the pressure cannot be obtained. Of course this is much weaker than Theorem 1.10 but it is another way to demonstrate the phenomenon.

**Proof of Theorem 1.13:** This is carried out exactly as in the proof of Theorem 1.10 with Lemma 3.2 being replaced by

\[ f_n = F^n(e^{-3h+2J_1}, e^{-h}) \]

where \( F \) is defined exactly as above and \( f_n \) is defined analogously. The details are left to the reader. □

**Acknowledgement.** The first and third authors thank the Erwin Schrödinger Institute for its hospitality during which a part of this paper was written.

**References**


Robert M. Burton  
Department of Mathematics  
Kidder Hall  
Oregon State University  
Corvallis, OR 97331  
burton@math.orst.edu

Charles-Ed. Pfister  
Department of Mathematics  
EPFL  
CH-1015 Lausanne  
Switzerland  
cpfister@eldp.epfl.ch

Jeffrey E. Steif  
Department of Mathematics  
Chalmers University of Technology  
S-412 96 Gothenburg  
Sweden  
steif@math.chalmers.se