UNIQUENESS OF MINIMIZER FOR COUNTABLE MARKOV SHIFTS

HIROKI TAKAHASI

ABSTRACT. We analyze large deviations rate functions for one-dimensional topological Markov shifts over infinite alphabet. We provide a sufficient condition on the potential which ensures that minimizers of the rate function are equilibrium states in the thermodynamic formalism. A combination of this result with Mauldin-Urbański’s and Sarig’s uniqueness theorems of Gibbs-equilibrium states yields a version of Dobrushin-Lanford-Ruelle’s variational principle: the Gibbs state, the equilibrium state, the minimizer are all unique and they coincide. A main technical ingredient is a certain hidden upper semi-continuity, first discovered by Fan-Jordan-Liao-Rams in a limited setting which we develop further and exploit. From the uniqueness of minimizer we deduce several conclusions.

1. Introduction

The theory of large deviations aims to characterize limit behaviors of measures in terms of rate functions. A sequence \( \{ \mu_n \}_{n=1}^{\infty} \) of Borel probability measures on a topological space \( \mathcal{X} \) satisfies the Large Deviation Principle (LDP) if there exists a lower semi-continuous function \( I: \mathcal{X} \to [0, \infty] \) such that for every Borel subset \( \mathcal{B} \) of \( \mathcal{X} \) the following holds:

\[
-\inf_{\mathcal{B}^o} I \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\mathcal{B}^o) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\overline{\mathcal{B}}) \leq -\inf_{\mathcal{B}} I,
\]

where \( \log 0 = -\infty \), \( \inf \emptyset = \infty \), \( \mathcal{B}^o \) and \( \overline{\mathcal{B}} \) denote the interior and the closure of \( \mathcal{B} \) respectively. The function \( I \) is called a rate function, and it is called a good rate function if the level set \( \{ x \in \mathcal{X} : I(x) \leq \alpha \} \) is compact for every \( \alpha \in (0, \infty) \).

We call \( x \in \mathcal{X} \) a minimizer if \( I(x) = 0 \). If \( \mathcal{B} \) is a closed set which is disjoint from the set of minimizers, the LDP ensures that \( \mu_n(\mathcal{B}) \) decays exponentially as \( n \to \infty \). If \( \mathcal{X} \) is a metric space and \( I \) is a good rate function, minimizers exist and the support of any accumulation point of \( \{ \mu_n \} \) is contained in the set of minimizers. Hence, it is important to determine this set. Also important is to determine the effective domain \( \{ x \in \mathcal{X} : I(x) < \infty \} \).

This paper addresses the problem of determining the set of minimizers and effective domain of the LDP for certain countable Markov shifts. The LDP was established in [31] under the assumption of the existence of a Gibbs state and a strong connectivity of the Markov shift. The assumption on potentials in [31] is rather general, and little is known about minimizers of the rate function in such cases.

\textit{Date:} April 11, 2019.

\textit{2010 Mathematics Subject Classification.} 37A45, 37A50, 37A60, 60F10.

\textit{Keywords:} Large Deviation Principle; countable Markov shift; Gibbs state; minimizer.
a generality. Under additional assumptions on the potential which are familiar in
the thermodynamic formalism of countable Markov shifts [16, 17, 24, 25, 26, 27],
we show the uniqueness of minimizer, and use this uniqueness to deduce from the
LDP several dynamical/thermodynamic properties of the Markov shift.

Let $S$ be a countable set and denote by $\mathbb{N}$ the set of non-negative integers.
Denote by $S^\mathbb{N}$ the set of all one-sided infinite sequences over $S$ endowed with the
product topology of the discrete topology on $S$, namely

$$S^\mathbb{N} = \{ x = (x_0, x_1, \ldots) : x_i \in S \quad \text{for every } i \in \mathbb{N} \}.$$ 

This topology is metrizable with a metric $d(x, y) = \exp (- \inf \{ i \in \mathbb{N} : x_i \neq y_i \} )$
with the convention $\exp (- \infty ) = 0$. The left shift $\sigma$ acts continuously on $S^\mathbb{N}$ by
$(\sigma x)_i = x_{i+1}$ $(i \in \mathbb{N})$. Let $T = (t_{ij})_{S \times S}$ be a matrix of zeros and ones with no
column or row which is all made of zeros. A (one-sided) topological Markov shift
$X$ generated by the transition matrix $T$ is given by

$$X = \{ x \in S^\mathbb{N} : t_{x_i x_{i+1}} = 1 \quad \text{for every } i \in \mathbb{N} \}.$$ 

If $\#S = \infty$ (resp. $\#S < \infty$), we call $X$ a countable (resp. finite) Markov shift. If
all entries of the matrix are 1, $X$ is called a full shift. The restriction of the left
shift to $X$ is denoted by $\sigma|_X$.

For two strings $v = v_0 \cdots v_{m-1}$, $w = w_0 \cdots w_{n-1}$ of elements of $S$, denote by
$vw$ the concatenated string $v_0 \cdots v_{m-1}w_0 \cdots w_{n-1}$. An $n$-string
$w_0w_1 \cdots w_{n-1}$ is admissible if $n = 1$, or else $n \geq 2$ and $t_{w_i w_{i+1}} = 1$ holds for $i = 0, 1, \ldots, n-1$. Denote
by $E^n$ the set of $n$-admissible strings and put $E^* = \bigcup_{n=1}^{\infty} E^n$. By convention, put
$E^0 = \emptyset$ and $vw = v = wv$ for $v \in E^*$, $w \in E^0$. A countable Markov shift $X$ is
finitely irreducible if there exists a finite set $\Lambda \subset E^*$ such that for all $i, j \in E^*$
there exists $\lambda \in \Lambda$ for which $i\lambda j \in E^*$. If $X$ is finitely irreducible and the finite set $\Lambda$
consists of strings of the same length $N$, then $X$ is called finitely primitive. Notice
that the set $\Lambda$ associated either with a finitely irreducible or finitely primitive
matrix can be taken to be empty for the full shift (in which case $N = 0$). The
finite primitiveness implies that the left shift is topologically mixing.

Denote by $\mathcal{M}$ the space of Borel probability measures on $X$ endowed with the
weak*-topology, and by $\mathcal{M}(\sigma|_X)$ the set of shift-invariant elements of $\mathcal{M}$. The
space $\mathcal{M}$ is metrizable with the bounded Lipschitz metric. The Kolmogorov-Sinai
entropy of each $\mu \in \mathcal{M}(\sigma|_X)$ with respect to $\sigma|_X$ is denoted by $h_\mu(\sigma|_X)$. Given a
measurable function $\phi : X \to \mathbb{R}$ with $\sup \phi < \infty$ define

$$\mathcal{M}_\phi(\sigma|_X) = \left\{ \mu \in \mathcal{M}(\sigma|_X) : \int \phi d\mu > -\infty \right\}.$$ 

The condition $\sup \phi < \infty$ ensures the well-definedness of $\int \phi d\mu$, though it can be
$-\infty$. For each $n$-string $w = w_0 \cdots w_{n-1} \in E^n$ define an $n$-cylinder

$$[w] = [w_0, \ldots, w_{n-1}] = \{ x \in X : x_i = w_i \quad \text{for } i = 0, \ldots, n-1 \}.$$ 

Put

$$Z_n(\phi) = \sum_{w \in E^n} \sup [w] \exp S_n \phi,$$
where \( S_n \phi = \sum_{i=0}^{n-1} \phi \circ \sigma^i \). A pressure is defined by

\[
P(\phi) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\phi).
\]

Since \( n \mapsto \log Z_n(\phi) \) is sub-additive, the limit in (1.1) exists, and in fact never \(-\infty\). Define

\[
P_0(\phi) = \sup \left\{ h_\mu(\sigma|_X) + \int \phi d\mu : \mu \in \mathcal{M}_\phi(\sigma|_X) \right\}.
\]

In general, \( P(\phi) \geq P_0(\phi) \) holds [16, Theorems 1.4]. We say a variational principle holds if \( P(\phi) < \infty \) and \( P(\phi) = P_0(\phi) \).

If the variational principle holds, a measure in \( \mathcal{M}_\phi(\sigma|_X) \) which attains the supremum in (1.2) is called an equilibrium state for the potential \( \phi \).

A Borel probability measure \( \mu_\phi \) on \( X \) is a Gibbs state (in the sense of Bowen) for the potential \( \phi \) (cf. [2, 16, 23, 24]) if there exist constants \( c \geq 1 \) and \( P \in \mathbb{R} \) such that for every \( n \geq 1 \) and every \( x \in X \),

\[
c^{-1} \leq \frac{\mu_\phi[x_0, \ldots, x_{n-1}]}{\exp (-Pn + S_n \phi(x))} \leq c.
\]

If \( \mu_\phi \) is shift-invariant, then it is called a shift-invariant Gibbs state. If there exists a Gibbs state for the potential \( \phi \), the constant \( P \) in (1.3) is equal to the pressure \( P(\phi) \). It is now classical [2, 23, 28] that for a topologically mixing, finite Markov shift and a Hölder continuous potential there exists a unique shift-invariant Gibbs state, and it coincides with the unique equilibrium state for the potential. The construction of shift-invariant Gibbs states for countable Markov shifts was carried out in [16, 17, 24, 26], see Theorem 1.2 below.

Let \( \phi : X \to \mathbb{R} \) be a measurable function and assume there exists a Gibbs state \( \mu_\phi \) for the potential \( \phi \). We introduce the following three sequences of Borel probability measures on \( \mathcal{M} \):

1. (Empirical means). For each \( x \in X \) and an integer \( n \geq 1 \) define

   \[
   \delta_x^n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\sigma^i x},
   \]

   with \( \delta_{\sigma^i x} \) the unit point mass at \( \sigma^i x \). Denote by \( \xi_n \) the distribution of the \( \mathcal{M} \)-valued random variable \( x \mapsto \delta_x^n \) on the probability space \((X, \mu_\phi)\);

2. (Weighted periodic points). Let \( A \) be a countable subset of \( X \). For each integer \( n \geq 1 \) put

   \[
   Z_n(\phi, A) = \sum_{x \in A} \exp S_n \phi(x).
   \]

   Define

   \[
   \eta_n = \frac{1}{Z_n(\phi, \text{Per}_n(\sigma|_X))} \sum_{x \in \text{Per}_n(\sigma|_X)} \exp S_n \phi(x) \delta_{\delta_x^n},
   \]

   with \( \text{Per}_n(\sigma|_X) = \{ x \in X : \sigma^n x = x \} \) and \( \delta_{\delta_x^n} \) the unit point mass at \( \delta_x^n \);
3. (Weighted iterated pre-images). Fix $y \in X$ and define

$$\zeta_{y,n} = \frac{1}{Z_n(\phi, (\sigma|_X)^{-n}y)} \sum_{x \in (\sigma|_X)^{-n}y} \exp S_n \phi(x) \delta_{\sigma^n x},$$

with $(\sigma|_X)^{-n}y = \{ x \in X : \sigma^n x = y \}$.

We now state the result in [31] we have been leading up to.

**Theorem 1.1.** ([31 Theorem A]). Let $X$ be a finitely irreducible countable Markov shift, $\phi : X \to \mathbb{R}$ a measurable function on $X$ and assume there exists a Gibbs state for the potential $\phi$. Then $\{ \xi_n \}_{n=1}^\infty$ is exponentially tight, i.e., for any $L > 0$ there exists a compact subset $K_L$ of $\mathcal{M}$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \xi_n(K_L^c) \leq -L,$$

and $\{ \xi_n \}_{n=1}^\infty$ satisfies the LDP. The rate function $I : \mathcal{M} \to [0, \infty]$ is given by

$$I(\mu) = -\inf_{\mathcal{G} \ni \mu} \sup_{\mathcal{G}} F,$$

where the infimum is taken over all open subsets $\mathcal{G}$ of $\mathcal{M}$ containing $\mu$, and $F : \mathcal{M} \to [-\infty, 0]$ is defined by

$$F(\nu) = \begin{cases} -P(\phi) + h_\nu(\sigma|_X) + \int \phi d\nu & \text{if } \nu \in \mathcal{M}(\phi, (\sigma|_X)); \\ -\infty & \text{otherwise.} \end{cases}$$

If moreover $X$ is finitely primitive, then $\{ \eta_n \}_{n=1}^\infty$ and $\{ \zeta_{y,n} \}_{n=1}^\infty (y \in X)$ are exponentially tight and satisfy the LDP with the same rate function as that of $\{ \xi_n \}_{n=1}^\infty$.

From the general theory on large deviations [3], the LDP determines the rate function $I$ uniquely, and it is characterized as the Legendre transform of the cumulant generating function (see Lemma A1.1 in Appendix). By the definition (1.4), $I$ is the maximal lower semi-continuous function satisfying $-F \leq I$. If $\phi$ is continuous then $\nu \in \mathcal{M}(\phi, (\sigma|_X)) \mapsto \int \phi d\nu$ is upper semi-continuous (see [12, Lemma 1]), while the entropy is not upper semi-continuous in general (see [12, p.774]), and as a result $-F$ does not coincide with $I$. The affinity character of $F$ implies the convexity of $I$. The exponential tightness and the large deviation bound for all open sets together imply that $I$ is a good rate function [3 Theorem 1.2.18(b)]. Hence, minimizers of $I$ exist, which in turn implies the variational principle. Since $\mathcal{M}(\sigma|_X)$ is a closed subset of $\mathcal{M}$, any minimizer is contained in $\mathcal{M}(\sigma|_X)$. If the Gibbs state in Theorem 1.1 is shift-invariant and ergodic, then it is a minimizer of the rate function $I$, from an elementary consideration using Birkhoff’s ergodic theorem.

For a topologically mixing finite Markov shift $X$ with a Hölder continuous potential $\phi$, the LDP for empirical means was established by Takahashi [29, 30] and Kifer [13] based on different methods, and for weighted periodic points by Kifer [14]. For uniformly hyperbolic systems (Anosov diffeomorphisms) with Hölder continuous derivatives, see Orey-Pelikan [20]. The rate function takes a finite value if and only if $\mu \in \mathcal{M}(\sigma|_X)$ and it is $P(\phi) - h_\mu(\sigma|_X) - \int \phi d\mu$. In particular, the
minimizer coincides with the unique Gibbs-equilibrium state for the potential $\phi$, and the effective domain is the set of shift-invariant measures.

For countable Markov shifts, the identification of the set of minimizers is a non-trivial problem. As $-F \geq I$ from the definition (1.4), equilibrium states for the potential $\phi$ are necessarily minimizers. The following example inspired by [1, 17, 26] shows that the converse is not true in general. Consider the potential $\phi$:

$$\phi(x) = \log p_{x_0}$$

where $\{p_k\}_{k \in \mathbb{N}}$ is a sequence with $p_k \in (0, 1)$, $\sum_{k \in \mathbb{N}} p_k = 1$ and $\sum_{k \in \mathbb{N}} p_k \log p_k = -\infty$, e.g., $p_k \propto 1/(k(\log k)^2)$. The Gibbs state $\mu_\phi$ for the potential $\phi$ is the Bernoulli measure associated with the infinite probability vector $(p_k)_{k \in \mathbb{N}}$. It is ergodic and hence a minimizer. As $\int \phi d\mu_\phi = -\infty$, it is not an equilibrium state for the potential $\phi$.

To identify minimizers, we need additional assumptions on the potential. A measurable function $\phi: X \to \mathbb{R}$ is summable if

$$Z_1(\phi) < \infty.$$ 

Notice that the summability of $\phi$ implies $\sup \phi < \infty$, $\inf \phi = -\infty$, $P(\phi) < \infty$, and the existence of a Gibbs state for the potential $\phi$ implies the summability of $\phi$.

For a summable function $\phi$, define a threshold inverse temperature

$$\beta_\infty(\phi) = \inf\{\beta \in \mathbb{R}: \beta \phi \text{ is summable}\}.$$ 

The summability of $\beta_0 \phi$ for some $\beta_0 > 0$ implies that of $\beta \phi$ for every $\beta > \beta_0$. Hence, $0 \leq \beta_\infty(\phi) \leq 1$ holds. Our first result identifies minimizers and the effective domain of the rate function in Theorem 1.1.

**Theorem A.** Let $X$ be a finitely irreducible countable Markov shift, $\phi: X \to \mathbb{R}$ a uniformly continuous summable function satisfying $\beta_\infty(\phi) < 1$, and assume there exists a Gibbs state for the potential $\phi$. Any minimizer of the rate function $I$ in Theorem 1.1 is an equilibrium state for the potential $\phi$. In addition,

$$\{\mu \in \mathcal{M}: I(\mu) < \infty\} = \mathcal{M}_\phi(\sigma|X).$$

To ensure the uniqueness of minimizer, we need to further impose an assumption on the potential. A function $\phi: X \to \mathbb{R}$ has summable variations if

$$\sum_{n=1}^\infty \sup_{w \in E^n} \sup_{x,y \in [w]} |\phi(x) - \phi(y)| < \infty.$$ 

The summability of variations implies the uniform continuity, and does not preclude unbounded functions, since only values of functions within the same cylinders are compared.

**Theorem 1.2.** (Mauldin-Urbański [16, 17], Sarig [24, 26]). Let $X$ be a finitely irreducible countable Markov shift and $\phi: X \to \mathbb{R}$ a summable function with summable variations. Then the variational principle holds, and there exists a unique shift-invariant Gibbs state $\mu_\phi$ for the potential $\phi$. If $\int \phi d\mu_\phi > -\infty$, then $\mu_\phi$ is the unique equilibrium state for the potential $\phi$.

As we show later, the last integrability condition in Theorem 1.2 follows from $\beta_\infty(\phi) < 1$. From Theorem A and Theorem 1.2 we obtain
**Theorem B.** Let $X$ be a finitely irreducible countable Markov shift and $\phi: X \to \mathbb{R}$ a summable function with summable variations. If $\beta_\infty(\phi) < 1$, then the shift-invariant Gibbs state for the potential $\phi$ in Theorem 1.2 is the unique equilibrium state for the potential $\phi$, and the unique minimizer of the rate function $I$ in Theorem 1.1.

At this point it is worthwhile to recall Dobrushin-Lanford-Ruelle’s variational principle in statistical mechanics [6, 7, 15], which states that the set of shift-invariant Gibbs states defined by the DLR-equation, that of equilibrium states (measures minimizing the free energy) and that of minimizers (of the relative entropy density) coincide for shift-invariant absolutely summable interactions. Theorems 1.2 and Theorem B together establish a version of this principle in the context of the thermodynamic formalism for countable Markov shifts.

Donsker and Varadhan have identified three levels of LDPs (see [8] for details). Using the contraction principle, from the level-2 LDP in Theorem 1.1 one obtains the following level-1 LDP. Denote by $C(X)$ the space of real-valued bounded continuous functions on $X$ endowed with the supremum norm. Each $\varphi \in C(X)$ defines a continuous functional on $\mathcal{M}$ by $\varphi(\mu) = \int \varphi d\mu$ ($\mu \in \mathcal{M}$). By the contraction principle (see Lemma A2.1), the sequences $\{\xi_n \circ \varphi^{-1}\}$, $\{\eta_n \circ \varphi^{-1}\}$, $\{\zeta_{y,n} \circ \varphi^{-1}\}$ ($y \in X$) of Borel probability measures on $\mathbb{R}$ satisfy the LDP with the rate function

$$I_\varphi: \alpha \in \mathbb{R} \mapsto \inf \left\{ I(\mu): \mu \in \mathcal{M}, \int \varphi d\mu = \alpha \right\} \in [0, \infty].$$

The contraction principle preserves good rate functions, and therefore preserves the uniqueness of minimizer (see Lemmas A2.1 and A2.2). Since the level-2 rate function $I$ in Theorem B is a good rate function by Theorem 1.1, the uniqueness of minimizer in Theorem B implies the uniqueness of minimizer of the level-1 rate function $I_\varphi$. This leads to the following statement.

**Corollary 1.3.** (Level-1 LDP). Let $X$ be a finitely irreducible countable Markov shift, $\phi: X \to \mathbb{R}$ a summable function with summable variations satisfying $\beta_\infty(\phi) < 1$, and $\mu_\phi$ the shift-invariant Gibbs state for the potential $\phi$ in Theorem 1.2. Let $\varphi \in C(X)$ be such that $c_\varphi < d_\varphi$, where

$$c_\varphi = \inf_{\mu \in \mathcal{M}_\phi(\sigma|X)} \int \varphi d\mu \quad \text{and} \quad d_\varphi = \sup_{\mu \in \mathcal{M}_\phi(\sigma|X)} \int \varphi d\mu.$$

Then $I_\varphi(\alpha) < \infty$ holds if and only if $\alpha \in [c_\varphi, d_\varphi]$, and $I_\varphi(\alpha) = 0$ holds if and only if $\alpha = \int \varphi d\mu_\phi$. For every interval $K$ intersecting $(c_\varphi, d_\varphi)$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_\phi \left\{ x \in X: \frac{1}{n} S_n \varphi(x) \in K \right\} = - \inf_K I_\varphi.$$

This number is negative if $\int \varphi d\mu_\phi \notin K$. If moreover $X$ is finitely primitive, then

$$\lim_{n \to \infty} \frac{1}{n} \log \frac{Z_n(\phi, \{x \in \text{Per}_n(\sigma|X): (1/n)S_n \varphi(x) \in K\})}{Z_n(\phi, \text{Per}_n(\sigma|X))} = - \inf_K I_\varphi,$$
and
\[ \lim_{n \to \infty} \frac{1}{n} \log \frac{Z_n(\phi, \{x \in (\sigma|_X)^{-n}y : (1/n)S_n\phi(x) \in K\})}{Z_n(\phi, (\sigma|_X)^{-n}y)} = -\inf_K \mathcal{I}_\phi. \]

Under additional assumptions on \((X, \phi)\) and a bounded function \(\varphi : X \to \mathbb{R}\), Results of the following type are already known: for any \(\epsilon > 0\) there exists \(\kappa(\epsilon) > 0\) such that
\[ \mu_{\phi} \left\{ x \in X : \frac{1}{n}S_n\varphi(x) - \int \varphi d\mu_{\phi} > \epsilon \right\} \leq \exp (-\kappa(\epsilon)n), \]
for all sufficiently large \(n\) (see [27, Theorem 7.4], [34, Theorem 3.5]). This type of “local large deviations results”, fairly abundant in the literature, indeed provide precise exponential bounds on small fluctuations near the mean \(\int \varphi d\mu_{\phi}\) (for small \(\epsilon\)). However, the bound is far from optimal for large \(\epsilon\), and do not imply the LDP. Corollary 1.3 contains information far from the mean with the definite exponential rate.

In the case \(X\) is the full shift modeling the so-called Gibbs-Markov maps and \(\varphi\) depends only on the first finite number of symbols, the uniqueness of minimizer of the rate function \(\mathcal{I}_\phi\) was shown by Denker-Kabluchko [4, Theorem 3.3]. They further deduced Erdös-Rényi’s law [4, Theorem 4.1], giving the maximal average of a time series over a time window of logarithmic length \([9]\). As a corollary to our main results, we obtain Erdös-Rényi’s law for a much larger class of \((X, \varphi)\) than in [4, Theorem 4.1].

**Corollary 1.4.** (Erdös-Rényi’s law). Let \(X\) be a finitely irreducible countable Markov shift, \(\phi : X \to \mathbb{R}\) a summable function with summable variations satisfying \(\beta_{\infty}(\phi) < 1\), and \(\mu_{\phi}\) the shift-invariant Gibbs state for the potential \(\phi\) in Theorem 1.2. Let \(\varphi : X \to \mathbb{R}\) be a bounded Lipschitz continuous function such that \(c_{\varphi} < d_{\varphi}\). For every \(\alpha \in (c_{\varphi}, d_{\varphi}) \setminus \{\int \varphi d\mu_{\phi}\}, \)
\[ \lim_{n \to \infty} \max_{m = 0, \ldots, n - l_n} \left\{ \frac{S_{l_n}\varphi(\sigma^m x)}{l_n} \right\} = \alpha \quad \text{for } \mu_{\phi}\text{-a.e. } x \in X, \]
where
\[ l_n = \left\lfloor \frac{\log n}{\mathcal{I}_\phi(\alpha)} \right\rfloor. \]

**Proof.** The upper law
\[ \limsup_{n \to \infty} \max_{m = 0, \ldots, n - l_n} \left\{ \frac{S_{l_n}\varphi(\sigma^m x)}{l_n} \right\} \leq \alpha \quad \text{for } \mu_{\phi}\text{-a.e. } x \in X, \]
is a consequence of Corollary 1.3, see [5, Proposition 2.2]. The lower law follows from an adaptation of the proof of [5, Theorem 3.1] based on the exponential decay of correlations [27, Theorem 6.3] for the shift-invariant Gibbs state \(\mu_{\phi}\). \(\square\)

In particular, one can compute the value \(\mathcal{I}_\phi(\alpha)\) by observing a typical time series. It should be noted that the range of \(\alpha\) is not limited to a small neighborhood of the mean \(\int \varphi d\mu_{\phi}\).

The rest of this paper consists of two sections and one appendix. We always denote by \(X\) countable Markov shifts, and take the countable set \(S\) to be \(\mathbb{N}\). A proof
of Theorem A is in §2 that is briefly outlined as follows. Suppose $\mu_{\min} \in \mathcal{M}(\sigma|_X)$ is a minimizer. By definition there exists a sequence $\{\mu_k\}_{k=1}^\infty$ in $\mathcal{M}_\phi(\sigma|_X)$ which converges to $\mu_{\min}$ in the weak*-topology and $F(\mu_k) \to 0$ as $k \to \infty$. There are two possibilities: (a) $\inf_k \int \phi d\mu_k > -\infty$; (b) $h_{\mu_k}(\sigma|_X) \to \infty$ and $\int \phi d\mu_k \to -\infty$ as $k \to \infty$. The condition $\beta_\infty(\phi) < 1$ readily rules out (b) (see Lemma 2.2). As $\phi$ is continuous bounded from above, it is integrated upper semi-continuously (see [12, Lemma 1]):

$$\limsup_{k \to \infty} \int \phi d\mu_k \leq \int \phi d\mu_{\min}.$$ 

In particular, $\mu_{\min} \in \mathcal{M}_\phi(\sigma|_X)$ holds. If the entropy were upper semi-continuous along the sequence $\{\mu_k\}$, it would follow that $\mu_{\min}$ is an equilibrium state. However, this upper semi-continuity is not expected. To overcome this difficulty, we prove a Main Technical Theorem (Theorem 2.4) which states that the mapping

$$\mu \mapsto \frac{h_{\mu}(\sigma|_X)}{P(\phi) - \int \phi d\mu}$$

has a limited form of upper semi-continuity at minimizers. This property was first shown by Fan et al. [10] for a different purpose, for the full shift and geometric potentials arising from Markov interval maps with countably many branches. We develop their argument further for adaptations to our setting. Theorem B readily follows from Theorem A and Theorem 1.2. In §3 we state and prove two conclusions (Theorems C and D) of the uniqueness of minimizer.

2. Identifying minimizers and establishing uniqueness

This section is organized as follows. After stating and proving in §2.1 a few general lemmas, in §2.2 we introduce Theorem 2.4 a main technical theorem of this paper, and use it to finish the proof of Theorem A. As the proof of Theorem 2.4 is lengthy, we defer it to §2.3. In §2.4 we prove Theorem B.

2.1. Preliminary lemmas. Throughout this and the next subsections, $X$ is a general countable Markov shift.

Lemma 2.1. Let $\phi: X \to \mathbb{R}$ be a summable function. For any $\delta > 0$ there exists a constant $K(\delta) \in \mathbb{R}$ such that if $\mu \in \mathcal{M}_\phi(\sigma|_X)$ satisfies $-h_{\mu}(\sigma|_X)/\int \phi d\mu > \beta_\infty(\phi) + \delta$ then $\int \phi d\mu \geq K(\delta)$.

Proof. Let $\delta > 0$ and put $\beta_0 = \beta_\infty(\phi) + \delta/2$. Then

$$h_{\mu}(\sigma|_X) + \beta_0 \int \phi d\mu \leq P_0(\beta_0 \phi) \leq P(\beta_0 \phi) < \infty.$$ 

Put $K(\delta) = \frac{P(\beta_0 \phi)}{\beta_0 - \beta_\infty(\phi) - \delta}$. From the assumption on $\mu$,

$$(\beta_0 - \beta_\infty(\phi) - \delta) \int \phi d\mu \leq P(\beta_0 \phi),$$

and therefore $\int \phi d\mu \geq K(\delta)$ as required. \qed
If \( \inf \phi = -\infty \), then \( M_\phi(\sigma|X) \) is not always a closed set, namely, the weak*-convergence \( \mu_k \to \mu_\infty \) of a sequence \( \{\mu_k\}_{k=1}^\infty \) in \( M_\phi(\sigma|X) \) does not always imply \( \int \phi d\mu_\infty > -\infty \). The next lemma provides conditions which rule out the possibility \( \int \phi d\mu_\infty = -\infty \).

**Lemma 2.2.** Let \( \phi: X \to \mathbb{R} \) be a summable function satisfying \( \beta_\infty(\phi) < 1 \). Let \( \{\mu_k\}_{k=1}^\infty \) be a sequence in \( M_\phi(\sigma|X) \) such that \( \{F(\mu_k)\}_{k=1}^\infty \) converges to a finite number as \( k \to \infty \). Then
\[
\inf_k \int \phi d\mu_k > -\infty.
\]

In particular, if \( \{\mu_k\}_{k=1}^\infty \) converges in the weak*-topology to \( \mu_\infty \) as \( k \to \infty \), then
\[
\int \phi d\mu_\infty > -\infty.
\]

**Proof.** If the infimum is \( -\infty \), then it is possible to take a subsequence \( \{\mu_{k_i}\}_{i=1}^\infty \) of \( \{\mu_k\} \) such that \( \int \phi d\mu_{k_i} \to -\infty \) and \( -h_{\mu_{k_i}}(\sigma|X)/\int \phi d\mu_{k_i} \to 1 \) as \( i \to \infty \). Fix \( \delta \in (0, 1 - \beta_\infty(\phi)) \). Since \( -h_{\mu_{k_i}}(\sigma|X)/\int \phi d\mu_{k_i} > \beta_\infty(\phi) + \delta \) holds for sufficiently large \( i \), we obtain a contradiction to Lemma 2.1.

For a function \( \varphi: X \to \mathbb{R} \) and an integer \( n \geq 1 \) define
\[
D_n(\varphi) = \sup_{w \in E^n} \sup_{x,y \in [w]} S_n \varphi(x) - S_n \varphi(y).
\]

Notice that \( D_n(\varphi) \leq nD_1(\varphi) \), and \( D_1(\varphi) \) can be \( \infty \). The regularity of functions needed in most places is \( D_n(\varphi) = o(n) \) \( (n \to \infty) \).

**2.2. Identifying minimizers as equilibrium states.** We now introduce a key ingredient in the proof of Theorem \( \Lambda \).

**Theorem 2.4.** (Main Technical Theorem). Let \( X \) be a countable Markov shift and \( \phi: X \to \mathbb{R} \) a uniformly continuous summable function satisfying \( \beta_\infty(\phi) < 1 \). Let \( \{\mu_j\}_{j=1}^\infty \) be a sequence in \( M_\phi(\sigma|X) \) which converges to \( \mu_\infty \in M_\phi(\sigma|X) \) in the weak*-topology as \( j \to \infty \). Assume \( P_0(\phi) - \int \phi d\mu_\infty > 0 \). If
\[
\liminf_{j \to \infty} \frac{h_{\mu_j}(\sigma|X)}{P_0(\phi) - \int \phi d\mu_j} > \beta_\infty(\phi),
\]
then
\[
\frac{h_{\mu_\infty}(\sigma|X)}{P_0(\phi) - \int \phi d\mu_\infty} \geq \limsup_{j \to \infty} \frac{h_{\mu_j}(\sigma|X)}{P_0(\phi) - \int \phi d\mu_j}.
\]

For finite Markov shifts, the inequality (2.2) is a consequence of the upper semi-continuity of entropy. A couple of remarks on Theorem 2.4 follows.

**Remark 2.5.** The definition of \( P_0(\phi) \) implies that the assumption (2.1) is vacuous for \( \beta_\infty(\phi) = 1 \). As continuous functions on \( X \) bounded from above are integrated upper semi-continuously (see [12] Lemma 1), the assumption \( P_0(\phi) - \int \phi d\mu_\infty > 0 \) in Theorem 2.4 implies \( \liminf_j (P_0(\phi) - \int \phi d\mu_j) > 0 \).
Remark 2.6. The following example inspired by [12, p.774] indicates that [22] does not imply the upper semi-continuity of entropy at $\mu_\infty$ along the sequence $\{\mu_j\}$. Consider the full shift $\mathbb{N}^\mathbb{N}$ and for each integer $j \geq 1$ let $\nu_j$ be the Bernoulli measure given by the 1-cylinders $[j], [j+1], \ldots, [2^j-1]$. Put $\mu_j = (1 - 1/j)\mu_\infty + (1/j)\nu_j$ where $\mu_\infty$ denotes the unit point mass at 0$^\infty$. Then $\mu_j$ converges in the weak*-topology to $\mu_\infty$ as $j \to \infty$, $h_{\mu_j}(\sigma) = \log 2$ and $h_{\mu_\infty}(\sigma) = 0$. Let $\phi : \mathbb{N}^\mathbb{N} \to \mathbb{R}$ be a summable function which is constant on each 1-cylinder such that $(1/j)2^{-j} \sum_{k=j}^{j+2^j-1} \phi(k) \to -\infty$ as $j \to \infty$, where $\phi(k)$ denotes the value of $\phi$ on $[k]$. Then $\int \phi d\mu_j = (1 - 1/j)\phi(0) + (1/j)2^{-j} \sum_{k=j}^{j+2^j-1} \phi(k) \to -\infty$ as $j \to \infty$, and $P_0(\phi) - \int \phi d\mu_\infty = \log(\sum_{k \in \mathbb{N}} e^{\phi(k)}) - \phi(0) > 0$. All the assumptions in Theorem 2.4 are satisfied and (2.2) holds trivially: $0 \geq 0$.

An idea of the proof of Theorem 2.4, inspired by [10], is to view measures in $\mathcal{M}(\sigma|_X)$ as measures on the full shift $\mathbb{N}^\mathbb{N}$, project them to the canonical finite subsystems $\Sigma_p$ ($p \geq 0$) of $\mathbb{N}^\mathbb{N}$, show for each $p \geq 0$ a $p$-th approximation of the inequality (2.2), and finally let $p \to \infty$ to obtain (2.2). The condition $\beta_\infty(\phi) < 1$ is used to control tails arising in approximations of entropy and integrals of the potential.

We defer a proof of Theorem 2.4 to [22] and below finish the proof of Theorem A assuming the conclusions of Theorem 2.4.

Proof of Theorem A. Let $X$ be a finitely irreducible countable Markov shift and $\phi : X \to \mathbb{R}$ a uniformly continuous summable function. Assume there exists a Gibbs state for the potential $\phi$. The existence of minimizer implies the variational principle holds: $P(\phi) = P_0(\phi)$. Assume $\beta_\infty(\phi) < 1$.

Let $\mu_{\min} \in \mathcal{M}(\sigma|_X)$ be a minimizer of the rate function $I$ in Theorem 1.1. By definition, there exists a sequence $\{\mu_k\}_{k=1}^\infty$ in $\mathcal{M}_\phi(\sigma|_X)$ which converges in the weak*-topology to $\mu_{\min}$ with $F(\mu_k) \to 0$ as $k \to \infty$. Lemma 2.2 implies $\inf_k \int \phi d\mu_k > -\infty$. The upper semi-continuity of the mapping $\mu \in \mathcal{M} \mapsto \int \phi d\mu$ as in Remark 2.3 gives $\int \phi d\mu_{\min} > -\infty$. It is convenient to split the rest of the proof into two cases.

Case 1: $\lim \inf_k h_{\mu_k}(\sigma|_X) = 0$. Take a subsequence $\{\mu_{kj}\}_{j=1}^\infty$ with $\lim_j h_{\mu_{kj}}(\sigma|_X) = 0$. We have

$$0 = \lim_{j \to \infty} F(\mu_{kj}) \leq -P(\phi) + \lim_{j \to \infty} h_{\mu_{kj}}(\sigma|_X) + \limsup_{j \to \infty} \int \phi d\mu_{kj}$$

$$\leq -P(\phi) + h_{\mu_{\min}}(\sigma|_X) + \int \phi d\mu_{\min}.$$  

It follows that $\mu_{\min}$ is an equilibrium state for the potential $\phi$.

Case 2: $\lim \inf_k h_{\mu_k}(\sigma|_X) > 0$. Then $\lim \inf_k (P(\phi) - \int \phi d\mu_k) \geq \lim \inf_k h_{\mu_k}(\sigma|_X) > 0$, and

$$0 = \lim_{k \to \infty} F(\mu_k) = \lim_{k \to \infty} \left( P(\phi) - \int \phi d\mu_k \right) \left( \frac{h_{\mu_k}(\sigma|_X)}{P(\phi) - \int \phi d\mu_k} - 1 \right).$$
Proof of Theorem 2.4.

Let $\beta$ as the first remark after Theorem 2.4 and $\mu_{\text{min}}$ be a projection satisfying $\mu_{\text{min}}(2.3) \lim \inf \frac{h_{\mu_{\text{min}}}(\sigma|X)}{P(\phi) - \int \phi d\mu_{\text{min}}} - 1 \geq 0$, which implies that $\mu_{\text{min}}$ is an equilibrium state for the potential $\phi$. If $P(\phi) - \int \phi d\mu_{\text{min}} > 0$, then Theorem 2.4 gives

$$\frac{h_{\mu_{\text{min}}}(\sigma|X)}{P(\phi) - \int \phi d\mu_{\text{min}}} - 1 \geq 0,$$

It follows that

$$\lim_{k \to \infty} \left( \frac{h_{\mu_k}(\sigma|X)}{P(\phi) - \int \phi d\mu_k} - 1 \right) = 0.$$

We have $P(\phi) - \int \phi d\mu_{\text{min}} \geq 0$. If $P(\phi) - \int \phi d\mu_{\text{min}} = 0$, then obviously $\mu_{\text{min}}$ is an equilibrium state for the potential $\phi$. If $P(\phi) - \int \phi d\mu_{\text{min}} > 0$, then Theorem 2.4 gives

$$\frac{h_{\mu_{\text{min}}}(\sigma|X)}{P(\phi) - \int \phi d\mu_{\text{min}}} - 1 \geq 0,$$

which implies that $\mu_{\text{min}}$ is an equilibrium state for the potential $\phi$.

It remains to show the last assertion on the effective domain of the rate function. It is clear from the definition of $I$ that $I(\mu) < \infty$ for any $\mu \in M_\phi(\sigma|X)$. Conversely, let $\mu \in M$ satisfy $I(\mu) < \infty$. By the definition of $I$, there exists a sequence $\{\mu_k\}_{k=1}^\infty$ in $M_\phi(\sigma|X)$ such that $\mu_k$ converges in the weak*-topology to $\mu$ and $F(\mu_k) \to -I(\mu)$ as $k \to \infty$. Lemma 2.2 gives $\inf_k \int \phi d\mu_k < -\infty$, and therefore $\int \phi d\mu > -\infty$. □

2.3. Limited form of upper semi-continuity.

Proof of Theorem 2.4. Let $\phi: X \to \mathbb{R}$ be a uniformly continuous summable function satisfying $\beta(\phi) < 1$. Let $\{\mu_j\}_{j=1}^\infty$ be a sequence in $M_\phi(\sigma|X)$ which converges to $\mu_\infty \in M_\phi(\sigma|X)$ with $P_0(\phi) - \int \phi d\mu_\infty > 0$. By $\lim \inf_j (P_0(\phi) - \int \phi d\mu_j) > 0$ in the first remark after Theorem 2.4 and $\beta(\phi) \geq 0$, if the strict inequality holds then

$$\lim_{j \to \infty} \inf h_{\mu_j}(\sigma|X) > 0. \tag{2.3}$$

It is convenient to split the rest of the proof of Theorem 2.4 into two cases.

Case 1: $\phi$ is constant on each 1-cylinder of $X$. Recall that $\sigma$ denotes the left shift on the full shift $\mathbb{N}^\mathbb{N}$. It is convenient to view each $\mu \in M(\sigma|X)$ as an element of $M(\sigma)$ by setting $\mu(A) = \mu(A \cap X)$ for any Borel subset $A$ of $\mathbb{N}^\mathbb{N}$. Notice that this extension preserves entropy. For each $k \in \mathbb{N}$, denote by $\langle k \rangle$ the corresponding 1-cylinder of $\mathbb{N}^\mathbb{N}$.

For each integer $p \geq 0$ we consider the shift-invariant subspace

$$\Sigma_p = \{ x \in \mathbb{N}^\mathbb{N} : x_i \leq p \text{ for every } i \in \mathbb{N} \}.$$

Define a projection $\pi_p: \mathbb{N}^\mathbb{N} \to \Sigma_p$ as follows: for each $x = x_0x_1x_2 \cdots \in \mathbb{N}^\mathbb{N}$ define $\pi_p(x) \in \Sigma_p$ by replacing in the sequence $x_0x_1x_2 \cdots$ all symbols greater than or equal to $p + 1$ by the symbol $p$. For each $\mu \in M(\sigma)$, write $\mu|_p$ for $\mu \circ \pi_p^{-1}$. Since $\pi_p$ commutes with the shift, $\mu|_p$ is a $\sigma|_{\Sigma_p}$-invariant measure. Put

$$c_p(\mu) = \sum_{k=p+1}^\infty \mu[k] \text{ and } K_p(\mu) = -\sum_{k=p+1}^\infty \phi(k)\mu[k].$$

Notice that $\int \phi d\mu > -\infty$ if and only if $K_p(\mu) \to 0$ as $p \to \infty$. 

UNIQUENESS OF MINIMIZER FOR COUNTABLE MARKOV SHIFTS 11
Proposition 2.7. Assume $\phi : X \to \mathbb{R}$ is constant on each 1-cylinder of $X$. For any $\delta > 0$ there exists $p_0 \geq 0$ such that for every $\mu \in \mathcal{M}_\phi(\sigma|X)$ and every integer $p \geq p_0$,

$$h_\mu(\sigma|X) - h_{\mu|\sigma}(\sigma|\Sigma_p) \leq - (1 - c_p(\mu)) \log(1 - c_p(\mu)) \quad - c_p(\mu) \log c_p(\mu) + (\beta_\infty(\phi) + \delta) K_p(\mu).$$

Proof. Before proceeding let us summarize basic facts on entropy. Let $\mathcal{A} = \{A_k\}_{k \in \mathbb{N}}$ be a countable partition of $X$ into Borel sets and let $\mu \in \mathcal{M}(\sigma|X)$. The entropy of $\mathcal{A}$ with respect to $\mu$ is the number

$$H_\mu(\mathcal{A}) = - \sum_{k \in \mathbb{N}} \mu(A_k) \log \mu(A_k),$$

with the convention $0 \log 0 = 0$. If $H_\mu(\mathcal{A}) < \infty$ then define

$$h_\mu(\mathcal{A}) = \lim_{n \to \infty} \frac{1}{n} H_\mu\left( \bigvee_{i=0}^{n-1} \sigma^{-i} \mathcal{A} \right).$$

where the symbol $\bigvee$ denotes the join of the partitions $\sigma^{-i} \mathcal{A}$ ($i = 0, \ldots, n - 1$). Since $n \mapsto H_\mu(\bigvee_{i=0}^{n-1} \sigma^{-i} \mathcal{A})$ is sub-additive, this limit exists, is finite and $h_\mu(\mathcal{A}) \leq H_\mu(\mathcal{A})$ holds. We have $h_\mu(\sigma|X) = \sup_\mathcal{A} h_\mu(\mathcal{A})$ where the supremum is taken over all countable partitions $\mathcal{A}$ of $X$ with $H_\mu(\mathcal{A}) < \infty$. If $\mathcal{A}$ is a generator of the Borel sigma-field of $X$ and $H_\mu(\mathcal{A}) < \infty$ then $h_\mu(\sigma|X) = h_\mu(\mathcal{A})$ holds.

Now, consider two partitions of $X$:

$$\mathcal{A}_p = \{X \cap \pi_p^{-1}(0), \ldots, X \cap \pi_p^{-1}(p)\}, \quad \mathcal{B}_p = \left\{ \bigcup_{k=0}^{p} [k], [p+1], [p+2], \ldots \right\}.$$

Lemma 2.8. The partition $\mathcal{A}_p \cup \mathcal{B}_p$ is a generator of the Borel sigma-field of $X$.

Proof. Since $[k] = \pi_p^{-1}(k) \cap X = (\pi_p^{-1}(k) \cap X) \cap \bigcup_{j=0}^{p} [j]$ holds for every $k \in \mathbb{N}$, we have $[k] \in \mathcal{A}_p \cup \mathcal{B}_p$ for $k = 0, \ldots, p$. It follows that the smallest sigma-algebra containing $\mathcal{A}_p \cup \mathcal{B}_p$ contains the collection $\{[k]\}_{k \in \mathbb{N}}$ of 1-cylinders, and so the claim holds.

Lemma 2.9. For every $\mu \in \mathcal{M}_\phi(\sigma|X)$, $H_\mu(\mathcal{B}_p) < \infty$ and

$$h_\mu(\mathcal{A}_p \cup \mathcal{B}_p) \leq h_\mu(\mathcal{A}_p) + h_\mu(\mathcal{B}_p).$$

Proof. The assumption $\int \phi d\mu > -\infty$ implies $h_\mu(\sigma|X) < \infty$. From [32, Lemma 2.1], $H_\mu(\mathcal{B}_p) < \infty$ holds. For each integer $n \geq 1$ we introduce a finite partition

$$\mathcal{B}_{p,n} = \left\{ \bigcup_{k=0}^{p} [k], [p+1], [p+2], \ldots, [p+n], \bigcup_{k=p+n+1}^{\infty} [k] \right\}.$$

Then $h_\mu(\mathcal{A}_p \cup \mathcal{B}_{p,n}) \leq h_\mu(\mathcal{A}_p) + h_\mu(\mathcal{B}_{p,n})$ holds. The $\{\mathcal{B}_{p,n}\}_{n=1}^\infty$ defines an increasing sequence of sub Borel sigma-fields of $X$ satisfying $\bigvee_{n=1}^{\infty} \mathcal{B}_{p,n} = \mathcal{B}_p$. By [33, Theorem 4.22], $h_\mu(\mathcal{B}_{p,n}) \to h_\mu(\mathcal{B}_p)$ as $n \to \infty$, and similarly $h_\mu(\mathcal{A}_p \cup \mathcal{B}_{p,n}) \to h_\mu(\mathcal{A}_p \cup \mathcal{B}_p)$.
Lemma 2.10. For every \( \mu \in \mathcal{M}_\phi(\sigma|_X) \),
\[
h_\mu(\mathcal{A}_p) = h_{\mu|_p}(\sigma|_{\Sigma_p}).
\]

Proof. Consider the partitions \( \mathcal{C}_p = \{(0) \cap \Sigma_p, \ldots, (p) \cap \Sigma_p\} \) of \( \Sigma_p \), and \( \pi_p^{-1}\mathcal{C}_p = \{\pi_p^{-1}(0) \cap \Sigma_p, \ldots, \pi_p^{-1}(p) \cap \Sigma_p\} \) of \( \mathbb{N}^N \). Then \( h_\mu(\mathcal{A}_p) = h_\mu(\pi_p^{-1}\mathcal{C}_p) \) holds (the \( \mu \) on the l.h.s. belongs to \( \mathcal{M}(\sigma|_X) \) and the \( \mu \) on the r.h.s. to \( \mathcal{M}(\sigma) \)). Since the projection \( \pi_p \) commutes with the left shift, \( h_\mu(\pi_p^{-1}\mathcal{C}_p) = h_{\mu|_p}(\mathcal{C}_p) \) holds. Since the partition \( \mathcal{C}_p \) is a generator of the Borel sigma-field of \( \Sigma_p \),
\[
h_{\mu|_p}(\mathcal{C}_p) = h_{\mu|_p}(\sigma|_{\Sigma_p}) \)
holds.

Returning to the proof of Proposition 2.7, let \( \mu \in \mathcal{M}_\phi(\sigma|_X) \). In the case \( c_p(\mu) = 0 \) we have \( h_\mu(\sigma|_X) - h_{\mu|_p}(\sigma|_{\Sigma_p}) = 0 \). Hence we assume \( c_p(\mu) > 0 \). By Lemmas 2.8 and 2.9 for each \( \mu \in \mathcal{M}_\phi(\sigma|_X) \) we have
\[
h_\mu(\sigma|_X) = h_\mu(\mathcal{A}_p \lor \mathcal{B}_p) \leq h_\mu(\mathcal{A}_p) + h_\mu(\mathcal{B}_p).
\]
By Lemma 2.10
\[
h_\mu(\sigma|_X) - h_{\mu|_p}(\sigma|_{\Sigma_p}) \leq h_\mu(\mathcal{B}_p) \leq H_\mu(\mathcal{B}_p)
\]
(2.4)
\[
- \sum_{k=0}^{p} \mu[k] \log \sum_{k=0}^{p} \mu[k] - \sum_{k=p+1}^{\infty} \mu[k] \log \mu[k]
\]
To treat the last summand in (2.4), define a potential \( \varphi : \mathbb{N}^N \to \mathbb{R} \) which is constant on each 1-cylinder of \( \mathbb{N}^N \) by \( \varphi(k) = \phi(k) \), and denote by \( P(\varphi) \) the pressure as defined in (1.1). The summability of \( \phi \) gives
\[
P(\varphi) \leq \sum_{k \in \mathbb{N}} \sup_{k \in \mathbb{N}} \exp \phi = \sum_{k \in \mathbb{N}} \exp \phi(k) < \infty.
\]
Denote by \( \nu_p \) the Bernoulli measure on \( \mathbb{N}^N \) which assigns to each 1-cylinder \( \langle k \rangle \), \( k \geq p + 1 \) the probability \( \mu[k]/c_p(\mu) \). Notice that
\[
h_{\nu_p}(\sigma) = - \sum_{k=p+1}^{\infty} \frac{\mu[k]}{c_p(\mu)} \log \frac{\mu[k]}{c_p(\mu)}.
\]
(2.5)
\[
\int \varphi d\nu_p = -K_p(\mu)/c_p(\mu) > -\infty, \quad \text{and} \quad \int \varphi d\nu_p \leq \sup_{k \geq p+1} \phi(k).
\]
The summability of \( \phi \) implies \( \sup_{k \geq p+1} \phi(k) \to -\infty \) as \( p \to \infty \), and so \( \int \varphi d\nu_p \to -\infty \). Lemma 2.1 applied to \( (\mathbb{N}^N, \varphi) \) shows that for any \( \delta > 0 \) there exists \( p_0 \geq 1 \) independent of \( \mu \) such that for every \( p \geq p_0 \),
\[
c_p(\mu)h_{\nu_p}(\sigma) \leq K_p(\mu)(\beta_\infty(\phi) + \delta).
\]
(2.6)
Plugging (2.5) into (2.6) and then rearranging the result gives
\[
- \sum_{k=p+1}^{\infty} \mu[k] \log \mu[k] \leq -c_p(\mu) \log c_p(\mu) + K_p(\mu)(\beta_\infty(\phi) + \delta).
\]
Plugging this inequality into (2.4) yields the desired one. \( \square \)
Returning to the proof of Theorem 2.4 in view of (2.1) fix \( \delta > 0 \) such that
\[
\liminf_{j \to \infty} \frac{h_{\mu_j}(\sigma|x)}{P_0(\phi) - \int \phi d\mu_j} > \beta_\infty(\phi) + \delta.
\]

Let \( \epsilon > 0 \) be such that
\[
(2.7) \quad (1 - \epsilon) \liminf_{j \to \infty} \frac{h_{\mu_j}(\sigma|x)}{P_0(\phi) - \int \phi d\mu_j} \geq \beta_\infty(\phi) + \delta.
\]

Since \( \mu_j \) converges weakly to \( \mu_\infty \) as \( j \to \infty \), Portmanteau’s theorem implies
\[
\limsup_j c_p(\mu_j) \to 0 \quad \text{as} \quad p \to \infty.
\]
From this and (2.3), there exists \( p_0 \geq 0 \) such that for each integer \( p \geq p_0 \) the following holds for sufficiently large \( j \):
\[
(2.8) \quad (1 - c_p(\mu_j)) \log(1 - c_p(\mu_j)) \geq -\epsilon h_{\mu_j}(\sigma|x).
\]

Since \( \int \phi d\mu_j > -\infty \) we have
\[
\int \phi d\mu_j = \sum_{k=0}^p \phi(k) \mu_j[k] + \sum_{k=p+1}^\infty \phi(k) \mu_j[k] = \int \phi d\mu_j - K_p(\mu_j).
\]

The equation (2.9) for \( \mu_\infty \) in the place of \( \mu_j \) implies \( \int \phi d\mu_\infty|_p \to \int \phi d\mu_\infty \) as \( p \to \infty \).
In what follows we assume \( p \) is large enough so that \( P_0(\phi) - \int \phi d\mu_\infty|_p > 0 \). Since \( \phi \) is bounded continuous on \( \Sigma_p \), the weak*-convergence of \( \mu_j|_p \) to \( \mu_\infty|_p \) as \( j \to \infty \) gives
\[
\int \phi d\mu_j|_p \to \int \phi d\mu_\infty|_p.
\]
In particular, \( P_0(\phi) - \int \phi d\mu_j|_p > 0 \) holds for sufficiently large \( j \), and therefore
\[
\frac{h_{\mu_j|_p}(\sigma|\Sigma_p)}{P_0(\phi) - \int \phi d\mu_j|_p} \geq \frac{h_{\mu_j}(\sigma|x) - (\beta_\infty(\phi) + \delta)K_p(\mu_j) + (1 - c_p(\mu_j)) \log(1 - c_p(\mu_j))}{P_0(\phi) - \int \phi d\mu_j - K_p(\mu_j)} \geq (1 - \epsilon) \frac{h_{\mu_j}(\sigma|x)}{P_0(\phi) - \int \phi d\mu_j}.
\]

The first inequality is a consequence of Proposition 2.7, the second of (2.8) and (2.9). The last inequality is trivial if \( K_p(\mu_j) = 0 \). Otherwise we appeal to the following: let \( a, b, c, d > 0 \) be such that \( c > d \) and \( a/c \geq b/d \). Then \( (a-b)/(c-d) \geq a/c \). Apply this with \( a = (1 - \epsilon)h_{\mu_j}(\sigma|x) \), \( b = (\beta_\infty(\phi) + \delta)K_p(\mu_j) \), \( c = P_0(\phi) - \int \phi d\mu_j \), \( d = K_p(\mu_j) \). The condition \( a/c \geq b/d \) is satisfied by virtue of (2.7).

The convergence of \( \mu_j|_p \) to \( \mu|_p \) as \( j \to \infty \) takes place in the space of shift-

invariant measures on \( \Sigma_p \) where the entropy is upper semi-continuous and \( \phi \) is bounded continuous. Hence
\[
\frac{h_{\mu_\infty|_p}(\sigma|\Sigma_p)}{P_0(\phi) - \int \phi d\mu_\infty|_p} \geq (1 - \epsilon) \limsup_{j \to \infty} \frac{h_{\mu_j}(\sigma|x)}{P_0(\phi) - \int \phi d\mu_j}.
\]
By Proposition 2.7 and (2.9) for $\mu_\infty$ in the place of $\mu_j$,

$$\lim_{p \to \infty} \frac{h_{\mu_\infty|_p}(\sigma|_{\Sigma_p})}{P_0(\phi) - \int \phi d\mu_\infty|_p} = \frac{h_{\mu_\infty}(\sigma|_X)}{P_0(\phi) - \int \phi d\mu_\infty}.$$  

Since $\epsilon > 0$ is arbitrary the desired inequality holds.

Case 2: $\phi$ is not constant on some 1-cylinder of $X$. The uniform continuity of $\phi$ permits to reduce part of our analysis to Case 1. We apply the result in Case 1 to accelerated (induced) systems and associated induced functions which are constant on each 1-cylinder.

Let $m \geq 1$ an integer. Consider the countable full shift $(E^m)^N$ with symbols in $E^m$ and denote by $\hat{\sigma}$ the left shift on $(E^m)^N$. The map $\theta: \{x_i\}_{i=0}^\infty \in (E^m)^N \mapsto x_0x_1\cdots \in \mathbb{N}^\mathbb{N}$ is a homeomorphism onto its image and commutes with $\hat{\sigma}$ and $(\sigma|_X)^m$. Define a countable Markov shift

$$\hat{X} = \{\{x_i\}_{i=0}^\infty \in (E^m)^N: \theta(x) \in X\}.$$  

The restriction of $\theta$ to $\hat{X}$ is also denoted by $\theta$:

$$\begin{array}{ccc}
\hat{X} & \xrightarrow{\hat{\sigma}} & \hat{X} \\
\theta \downarrow & & \theta \\
X & \xrightarrow{(\sigma|_X)} & X.
\end{array}$$

Denote by $\widehat{E}^n$ the set of admissible $n$-strings of elements of $E^m$. Since $\widehat{E}^1$ is canonically identified with $E^m$, the 1-cylinder in $\hat{X}$ corresponding to each $w \in \widehat{E}^1$ is denoted by $[w]$ with a slight abuse of notation.

Given a measurable function $\psi: X \to \mathbb{R}$ with $\sup \psi < \infty$, define an induced function $\hat{\psi}: \hat{X} \to \mathbb{R}$ which is constant on each 1-cylinder of $\hat{X}$ by

$$\hat{\psi}|_{[w]} = \sup_{[w]} S_m \psi \quad \text{for each } w \in \widehat{E}^1.$$  

Lemma 2.11. The following holds:

(a) for each $\mu \in \mathcal{M}(\sigma|_X)$ with $\int \psi d\mu > -\infty$,

$$\left| m \int \psi d\mu - \int \hat{\psi}d(\mu \circ \theta) \right| \leq D_m(\psi);$$

(b) for each $\mu \in \mathcal{M}(\sigma|_X)$, $\int \psi d\mu > -\infty$ holds if and only if $\int \hat{\psi}d(\mu \circ \theta) > -\infty$.

Proof. Let $\mu \in \mathcal{M}(\sigma|_X)$. If $\int \psi d\mu > -\infty$ then the shift-invariance of $\mu$ implies $\int \psi \circ \theta d(\mu \circ \theta) = (1/m) \int S_m \psi(\theta(x))d(\mu \circ \theta)(x)$, and therefore

$$\left| m \int \psi d\mu - \int \hat{\psi}d(\mu \circ \theta) \right| \leq \sup_{x \in \hat{X}} \left| S_m \psi(\theta(x)) - \hat{\psi}(x) \right| \leq D_m(\psi),$$

as required.
For each $l \in \mathbb{N}$ put $\psi_l = \max(\psi, -l)$. By Lemma 2.11(a), for each $\mu \in \mathcal{M}(\sigma|X)$ we have
\[
\left| m \int \psi_l d\mu - \int \widehat{\psi}_l d(\mu \circ \theta) \right| \leq D_m(\psi_l).
\]
As $l \to \infty$, the monotone convergence theorem gives $\int \psi_l d\mu \to \int \psi d\mu$ and $\int \widehat{\psi}_l d(\mu \circ \theta) \to \int \widehat{\psi} d(\mu \circ \theta)$, and $D_m(\psi_l) \to D_m(\psi)$ holds. This implies Lemma 2.11(b).

To finish the proof of Theorem 2.4, let $\epsilon \in (0, P_0(\phi) - \int \phi d\mu_\infty)$. In view of Lemma 2.11(a) and Lemma 2.3, we assume $m \geq 1$ is a large integer so that $|\int \widehat{\phi} d(\mu \circ \theta) - m \int \phi d\mu| \leq m\epsilon/2$ holds for every $\mu \in \mathcal{M}_\phi(\sigma|X)$. Moreover, $h_{\mu \circ \theta}(\widehat{\sigma}) = mh_\mu(\sigma|X)$ holds. By Lemma 2.11(b), measures which take part in the definition of $P_0(\widehat{\phi})$ and those which take part in the definition of $P_0(\phi)$ are in one-to-one correspondence. It follows that $|P_0(\widehat{\phi}) - mP_0(\phi)| \leq m\epsilon/2$, and
\[
(2.10) \quad \left| P_0(\widehat{\phi}) - \int \widehat{\phi} d(\mu \circ \theta) - \left( mP_0(\phi) - m \int \phi d\mu \right) \right| \leq m\epsilon.
\]
To apply the result in Case 1 to the Markov shift $\widehat{X}$ and the function $\widehat{\phi}: \widehat{X} \to \mathbb{R}$, we need to check the necessary assumptions. Since $\phi$ is uniformly continuous, Lemma 2.3 gives
\[
\sup_{w \in \mathbb{E}^1} \sup_{x,y \in [w]} \widehat{\phi}(x) - \widehat{\phi}(y) = o\left( \frac{1}{m} \right).
\]
For every $\beta > 0$,
\[
\log Z_1(\beta \widehat{\phi}) = o\left( \frac{1}{m} \right) + \log Z_m(\beta \phi),
\]
which implies $\beta_\infty(\widehat{\phi}) \leq \beta_\infty(\phi) < 1$. The inequality (2.10) for $\mu_\infty$ in the place of $\mu$ gives $P_0(\widehat{\phi}) - \int \widehat{\phi} d(\mu_\infty \circ \theta) > 0$. Hence we obtain
\[
\frac{h_{\mu_\infty}(\sigma|X)}{P_0(\widehat{\phi}) - \int \widehat{\phi} d\mu_\infty - \epsilon} \geq \frac{h_{\mu_\infty \circ \theta}(\widehat{\sigma})}{P_0(\widehat{\phi}) - \int \widehat{\phi} d(\mu_\infty \circ \theta)} \geq \limsup_{j \to \infty} \frac{h_{\mu_j}(\sigma|X)}{P_0(\widehat{\phi}) - \int \widehat{\phi} d\mu_j + \epsilon} \geq \frac{\inf_j (P_0(\phi) - \int \phi d\mu_j) + \epsilon}{\inf_j (P_0(\phi) - \int \phi d\mu_j) + \epsilon} \limsup_{j \to \infty} \frac{h_{\mu_j}(\sigma|X)}{P_0(\phi) - \int \phi d\mu_j}.
\]
The first and the third inequalities are from (2.10). The result in Case 1 applied to $(\widehat{X}, \widehat{\phi})$ gives the second inequality. Letting $\epsilon \to 0$ yields (2.2).

2.4. Uniqueness of minimizer.

Proof of Theorem B. Let $X$ be a finitely irreducible countable Markov shift and $\phi: X \to \mathbb{R}$ a summable function with summable variations. By Theorem 1.2, the variational principle holds and there exists a unique shift-invariant Gibbs state...
for the potential $\phi$, denoted by $\mu_\phi$. Assume $\beta_\infty(\phi) < 1$. Below we show that $\int \phi d\mu_\phi > -\infty$. Then by Theorem 1.2, $\mu_\phi$ is the unique equilibrium state for the potential $\phi$. By Theorem A, the minimizer is unique and it is $\mu_\phi$.

By re-ordering $\mathbb{N}$ if necessary, we may assume $k \in \mathbb{N} \mapsto \sup_{[k]} \exp \phi$ is decreasing with no loss of generality. Since $\mu_\phi$ is a Gibbs state, $D_1(\phi) < \infty$ holds (see [31, Lemma 2.2]) and (1.3) gives $\mu_\phi[k] \leq c k - P(\phi) \sup_{[k]} \exp \phi$ for every $k \in \mathbb{N}$. Fix $\delta \in (0, 1 - \beta_\infty(\phi))$ and a constant $C(\delta) > 0$ such that for every $k \in \mathbb{N}$,

$$\sup_{[k]} |\phi| \leq D_1(\phi) + \inf_{[k]} |\phi| \leq C(\delta) + D_1(\phi) + \exp(-\delta \phi),$$

Then

$$\left| \int \phi d\mu_\phi \right| \leq \sum_{k \in \mathbb{N}} \sup_{[k]} |\phi| \mu_\phi[k] \leq ce^{-P(\phi)} ((C(\delta) + D_1(\phi)) Z_1(\phi) + Z_1((1 - \delta) \phi)) < \infty. \quad \square$$

3. Conclusions from uniqueness of minimizer

Yoichiro Takahashi told that the LDP is not a mere limit theorem, but a unifying principle which encompasses a number of important conclusions. In this section we take up two conclusions: the equidistributions of weighted periodic points and weighted iterated pre-images (§3.1); the differentiation of pressure (§3.2).

3.1. Equidistributions from the uniqueness of minimizer. Let $\phi : X \to \mathbb{R}$ be a measurable function and assume there exists a Gibbs state for the potential $\phi$. Define two sequences in $\mathcal{M}$:

$$p_n = \frac{1}{Z_n(\phi, \text{Per}_n(\sigma|X))} \sum_{x \in \text{Per}_n(\sigma|X)} \exp S_n \phi(x) \delta^n_x,$$

$$\nu_{y,n} = \frac{1}{Z_n(\phi, (\sigma|X)^{-n}y)} \sum_{x \in (\sigma|X)^{-n}y} \exp S_n \phi(x) \delta^n_x,$$

where $y \in X$ is fixed. Of interest is the behavior of these sequences as $n \to \infty$. Since $\mathcal{M}$ is non-compact, even the existence of accumulation points is an issue.

**Theorem C.** (Equidistributions). Let $X$ be a finitely primitive countable Markov shift, $\phi : X \to \mathbb{R}$ a measurable function and assume there exists a Gibbs state for the potential $\phi$. Then $\{p_n\}_{n=1}^\infty$ and $\{(1/n) \sum_{i=0}^{n-1} \nu_{y,n} \circ \sigma^{-i}\}_{n=1}^\infty (y \in X)$ are tight and all their accumulation points are minimizers of the rate function $I$. In particular, the minimizer is unique, denoted by $\mu_{\min}$, then

$$p_n \longrightarrow \mu_{\min} \quad \text{in the weak}\,^*\text{-topology as } n \to \infty,$$

and

$$\frac{1}{n} \sum_{i=0}^{n-1} \nu_{y,n} \circ \sigma^{-i} \longrightarrow \mu_{\min} \quad \text{in the weak}\,^*\text{-topology as } n \to \infty,$$

for every $y \in X$. 

For finite Markov shifts with continuous potentials, Misiurewicz’s proof of the variational principle [19] implies that any accumulation point of the sequences as in Theorem 3.1 is an equilibrium state for the potential. Hence, the uniqueness of equilibrium state implies the equidistribution. Theorem C ties the equidistribution for countable Markov shifts with the uniqueness of minimizer.

We apply our results to the Gauss transformation $T: (0, 1] \to [0, 1)$ given by $Tx = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$. Following the orbits of $T$ over the infinite Markov partition one can model $T$ with the countable full shift. The function $-\log |DT|$ (which is assumed to be the appropriate one-sided derivative at each discontinuity point $\frac{1}{k}$) induces a potential $\phi$ on the symbolic space which is summable with summable variations [11, Section 7]. It is not hard to show $\beta_\infty(\phi) < 1$, and in fact $\beta_\infty(\phi) = \frac{1}{2}$ (see [18]). For each $\beta > \frac{1}{2}$ denote by $\mu_\beta$ the $T$-invariant Borel probability measure which corresponds to the unique Gibbs state for the potential $\beta \phi$. By Theorem A, $\mu_\beta$ corresponds to the unique minimizer of the rate function associated with the potential $\beta \phi$.

For each integer $n \geq 1$ write

$$\text{Per}_n(T) = \{x \in (0, 1): T^n x = x\},$$

and

$$T^{-n} y = \{x \in (0, 1): T^n x = y\},$$

with $y \in (0, 1) \setminus \mathbb{Q}$ fixed. Theorem C reads as follows:

$$\frac{1}{\text{Per}_n(T)} \sum_{x \in \text{Per}_n(T)} |DT^n(x)|^{-\beta} \sum_{x \in \text{Per}_n(T)} |DT^n(x)|^{-\beta} \delta_x \longrightarrow \mu_\beta \quad \text{as } n \to \infty,$$

and

$$\frac{1}{n} \sum_{i=0}^{n-1} \nu_{y,n} \circ T^{-i} \longrightarrow \mu_\beta \quad \text{as } n \to \infty,$$

where

$$\nu_{y,n} = \frac{1}{\text{Per}_n(T)} \sum_{x \in T^{-n} y} |DT^n(x)|^{-\beta} \sum_{x \in T^{-n} y} |DT^n(x)|^{-\beta} \delta_x^n.$$

The convergences are with respect to the weak*-topology on the space of $T$-invariant Borel probability measures on $(0, 1) \setminus \mathbb{Q}$. The weak*-convergence of weighted periodic points for the case $\beta = 1$ was first proved by Fiebig-Fiebig-Yuri [11] by directly showing the tightness of the sequence of measures on the full shift $\mathbb{N}^\mathbb{N}$. The $\mu_1$ is the Gauss measure: $d\mu_1 = \frac{1}{\log 2} \frac{dx}{1+x}$. Their method certainly works for all $\beta > 1/2$. The Cesàro mean convergence associated with $\nu_{y,n}$ is not treated in [11].

**Proof of Theorem C.** Let $X$ be finitely primitive, $\phi: X \to \mathbb{R}$ a measurable function and $\mu_\phi$ a Gibbs state for the potential $\phi$. By Theorem [11] the sequences $\{\eta_n\}_n$ and $\{\zeta_{y,n}\}_n (y \in X)$ of measures on $\mathcal{M}$ are exponentially tight, and so they are tight and satisfy the LDP with the good rate function $I$ in Theorem [11]. Define

$$I^{-1}(0) = \{\mu \in \mathcal{M}(\sigma|_X): I(\mu) = 0\},$$

and denote by $\mathcal{N}$ the set of Borel probability measures on $\mathcal{M}$ whose supports are contained in $I^{-1}(0)$. By Lemma A2.3, all accumulation points of $\{\eta_n\}$ and $\{\zeta_{y,n}\}$
are contained in \( \mathcal{N} \). Since \( I \) is a good rate function, \( I^{-1}(0) \) is compact and so \( \mathcal{N} \) is compact. It is a convex set, and the set of its extremal points is the set of unit point masses. As a consequence of Choquet’s representation theorem \([21]\) and the approximation of integrals of bounded continuous functions by those of simple functions, every measure in \( \mathcal{N} \) is weak*-approximated by a finite convex combination of unit point masses.

Define a mapping \( \Pi: \mathcal{N} \to I^{-1}(0) \) as follows. For \( \tau \in \mathcal{N} \) of the form \( \tau = \sum_{i=1}^{k} \alpha_i \delta_{\mu_i} \) with \( 0 < \alpha_i < 1 \), \( \sum_{i=1}^{k} \alpha_i = 1 \) and \( \mu_i \in I^{-1}(0) \) for \( i = 1, \ldots, k \) with \( \delta_{\mu_i} \), the unit point mass at \( \mu_i \), define \( \Pi(\tau) = \sum_{i=1}^{k} \alpha_i \mu_i \). Otherwise, choose a sequence \( \{\tau_m\}_{m=1}^{\infty} \) of finite convex combinations of unit point masses at points in \( I^{-1}(0) \) which converges in the weak* topology to \( \tau \) as \( m \to \infty \), and define \( \Pi(\tau) \) to be the weak*-limit point of \( \{\Pi(\tau_m)\}_{m=1}^{\infty} \) as \( m \to \infty \). In the next paragraph we show that this limit exists.

Write \( \tau_m = \sum_{i=1}^{m} \alpha_{i,m} \delta_{\mu_{i,m}} \in \mathcal{N} \). Then \( \Pi(\tau_m) = \sum_{i=1}^{m} \alpha_{i,m} \mu_{i,m} \) holds. Since \( I^{-1}(0) \) is compact, \( \{\Pi(\tau_m)\}_{m} \) has a convergent subsequence. If there were two subsequences \( \{\sum_{j=1}^{k_{p(j)}} \alpha_{i,p(j)} \mu_{i,p(j)}\}_{j=1}^{\infty} \) and \( \{\sum_{j=1}^{k_{q(j)}} \alpha_{i,q(j)} \mu_{i,q(j)}\}_{j=1}^{\infty} \) converging to different limits as \( j \to \infty \), there would exist \( \varphi \in C(X) \) satisfying

\[
\lim_{j \to \infty} \sum_{i=1}^{k_{p(j)}} \alpha_{i,p(j)} \int \varphi d\mu_{i,p(j)} \neq \lim_{j \to \infty} \sum_{i=1}^{k_{q(j)}} \alpha_{i,q(j)} \int \varphi d\mu_{i,q(j)}.
\]

This would imply

\[
\lim_{j \to \infty} \int \varphi(\mu) d\tau_{p(j)}(\mu) \neq \lim_{j \to \infty} \int \varphi(\mu) d\tau_{q(j)}(\mu),
\]

a contradiction to the choice of \( \{\tau_m\}_m \). Since \( I \) is lower semi-continuous, \( \Pi(\tau) \in I^{-1}(0) \) holds. Therefore, the mapping \( \Pi: \mathcal{N} \to I^{-1}(0) \) is well-defined.

**Lemma 3.1.** If \( \tau \in \mathcal{N} \) then \( \int \varphi(\mu) d\tau(\mu) = \int \varphi d\Pi(\tau) \) holds for every \( \varphi \in C(X) \).

**Proof.** The equality in the last assertion of Lemma 3.1 holds for every element of \( \mathcal{N} \) which is a finite convex combination of unit point masses at points in \( I^{-1}(0) \). Hence, it holds for any element of \( \mathcal{N} \). \( \square \)

Returning to the proof of Theorem C, Let \( \{p_{n_j}\}_j \) be a subsequence of \( \{p_n\}_n \). Since \( \{\eta_n\}_n \) is tight and \( \mathcal{M} \) is a Polish space, by Prohorov’s theorem it is possible to choose a subsequence \( \{p_{n_{j(k)}}\}_k \) of \( \{p_{n_j}\}_j \) such that the corresponding subsequence \( \{\eta_{n_{j(k)}}\}_k \) of \( \{\eta_n\}_n \) converges weakly, say to \( \tau \in \mathcal{N} \) as \( k \to \infty \). Let \( \varphi \in C(X) \).
Since the functional \( \mu \in \mathcal{M} \mapsto \int \varphi d\mu = \varphi(\mu) \) is bounded continuous,
\[
\int \varphi dp_{n_j(k)} = \frac{1}{n_j(k)} \sum_{i=0}^{n_j(k)-1} \int \varphi d\left(p_{n_j(k)} \circ (|x|^{-1}) \right)
= \frac{1}{n_j(k)} \sum_{i=0}^{n_j(k)-1} \frac{1}{Z_{n_j(k)}(\phi, \text{Per}_{n_j(k)}(|x|))} \sum_{x \in \text{Per}_{n_j(k)}(|x|)} \exp S_{n_j(k)}(\phi(x)) \varphi(\sigma^i x)
= \frac{1}{Z_{n_j(k)}(\phi, \text{Per}_{n_j(k)}(|x|))} \sum_{x \in \text{Per}_{n_j(k)}(|x|)} \exp S_{n_j(k)}(\phi(x)) \sum_{i=0}^{n_j(k)-1} \varphi(\sigma^i x)
= \int \varphi(\mu) d\eta_{n_j(k)}(\mu).
\]
Hence
\[
\lim_{k \to \infty} \int \varphi dp_{n_j(k)} = \lim_{k \to \infty} \int \varphi(\mu) d\eta_{n_j(k)}(\mu) = \int \varphi(\mu) d\tau(\mu) = \int \varphi \Pi(\tau),
\]
the last equality by Lemma 3.1. Since \( \varphi \in C(X) \) is arbitrary, it follows that \( p_{n_j(k)} \to \Pi(\tau) \) in the weak\(^*\)-topology as \( k \to \infty \). Since any subsequence has a convergent subsequence, \( \{p_n\}_n \) is tight. The tightness of the Cesàro mean sequence can be proved in the same way.

3.2. Differentiation of pressure from the uniqueness of minimizer. In thermodynamics, important macroscopic parameters are derived from Helmholtz’s free energy. A counterpart of Helmholtz’s free energy in the thermodynamic formalism is the pressure. For a topologically mixing finite Markov shift with a Hölder continuous potential, the uniqueness of equilibrium state and the general differentiability result in large deviations [3] Theorem II. 6.3] yield the Gâteaux differentiability of the pressure at \( \phi \). The next theorem extends this statement to countable Markov shifts.

**Theorem D.** Let \( X \) be a finitely irreducible countable Markov shift, \( \phi: X \to \mathbb{R} \) a uniformly continuous function and assume there exists a Gibbs state for the potential \( \phi \). If the minimizer is unique, denoted by \( \mu_{\min} \), then for any \( \varphi \in C(X) \) which is uniformly continuous and satisfies \( c_\varphi < d_\varphi \), the function \( \beta \in \mathbb{R} \mapsto P(\phi + \beta \varphi) \) is differentiable at \( \beta = 0 \) and satisfies
\[
\left. \frac{d}{d\beta} \right|_{\beta=0} P(\phi + \beta \varphi) = \int \varphi d\mu_{\min}.
\]

Assuming \( \phi \) and \( \varphi \) have summable variations, Sarig [27, Theorem 6.5] showed the analyticity of the function \( \beta \in \mathbb{R} \mapsto P(\phi + \beta \varphi) \) at \( \beta = 0 \), as well as formulas for the first and the second-order derivatives. The uniqueness of minimizer implies the existence of the first-order derivative for a broader class of functions.
Without the assumption of the uniqueness of minimizer, using Lemma A1.1 one can show the following: a measure \( \mu \in \mathcal{M}(\sigma|_X) \) is a minimizer if and only if
\[
P(\phi + \varphi) - P(\phi) \geq \int \varphi d\mu \quad \text{for every uniformly continuous } \varphi \in C(X).
\]

**Proof of Theorem D.** Let \( X \) be finitely irreducible, \( \phi: X \to \mathbb{R} \) uniformly continuous and \( \mu_\phi \) a Gibbs state for the potential \( \phi \). Then \( \sup \phi < \infty \), \( \inf \phi = -\infty \) and \( D_1(\phi) < \infty \) hold (see [31, Lemma 2.2]). By Theorem [14], the LDP holds. By Varadhan’s integral lemma [3, Theorem 4.3.1], for each \( \varphi \in C(X) \) the limit
\[
Q(\varphi) = \lim_{n \to \infty} \frac{1}{n} \log \int \exp S_n \varphi d\mu \phi
\]
eexists and satisfies
\[
Q(\varphi) = \sup_{\mu \in \mathcal{M}} \left( \int \varphi d\mu - I(\mu) \right).
\]

**Lemma 3.2.** For any \( \varphi \in C(X) \),
\[
Q(\varphi) = \sup_{\mu \in \mathcal{M}} \left( \int \varphi d\mu + F(\mu) \right).
\]

**Proof.** The relation \(-I \geq F\) immediately gives \( Q(\varphi) = \sup_{\mu \in \mathcal{M}} \left( \int \varphi d\mu + F(\mu) \right) \). If this inequality were strict, there would exist \( \mu_\infty \in \mathcal{M} \) such that \( \int \varphi d\mu_\infty - I(\mu_\infty) > \sup_{\mu \in \mathcal{M}} \left( \int \varphi d\mu + F(\mu) \right) \). Since \( I(\mu_\infty) < \infty \), there would exist a sequence \( \{\mu_k\}_{k=1}^\infty \in \mathcal{M}_\phi(\sigma|_X) \) which converges in the weak*-topology to \( \mu_\infty \) and satisfies \( F(\mu_k) \to -I(\mu_\infty) \) as \( k \to \infty \). Since \( \varphi \in C(X) \), \( \int \varphi d\mu_k + F(\mu_k) \to \int \varphi d\mu_\infty - I(\mu_\infty) \) as \( k \to \infty \) and arises a contradiction. \( \square \)

Now, let \( \varphi \in C(X) \) be uniformly continuous and let \( \beta \in \mathbb{R} \). By Lemma 3.2 and the definition of \( F \),
\[
Q(\beta \varphi) = P_0(\phi + \beta \varphi) - P(\phi).
\]
Since \( X \) is finitely irreducible, \( \varphi \) is bounded uniformly continuous, \( \phi + \beta \varphi \) is uniformly continuous and satisfies \( D_1(\phi + \beta \varphi) \leq D_1(\phi) + \beta D_1(\varphi) < \infty \), by [16, Theorems 1.3 and 1.4] the variational principle holds for the potential \( \phi + \beta \varphi \): \( P_0(\phi + \beta \varphi) = P(\phi + \beta \varphi) \). If the minimizer of the rate function in (1.1) is unique, then from the exponential convergence in Corollary 1.3 and [8, Theorem II. 6.3] we obtain the desired conclusion. \( \square \)

**APPENDIX**

**A1. Representation of rate function.** The rate function in Theorem [14] has a canonical representation well-known in the theory of large deviations.

**Lemma A1.1.** Let \( X \) be a finitely irreducible countable Markov shift, \( \phi: X \to \mathbb{R} \) a measurable function and \( \mu_\phi \) a Gibbs state for the potential \( \phi \). The rate function \( I \) in Theorem [7,1] satisfies
\[
I(\mu) = \sup_{\varphi \in C(X)} \left( \int \varphi d\mu - Q(\varphi) \right) \quad \text{for every } \mu \in \mathcal{M}.
\]
Proof. We extend the rate function $I$ to a lower semi-continuous, convex function on the locally convex topological vector space of signed measures on $X$: for each signed measure $\mu$ with the Jordan decomposition $\mu = \mu^+ - \mu^-$, define $I(\mu) = \mu^+(X) \cdot I(\mu^+ / \mu^+(X))$ if $\mu^+ \neq 0$ and $I(\mu) = \infty$ otherwise. Since the space $C(X)$ determining the weak*-topology of the space of signed measures is a separating vector space in the algebraic dual of the space of signed measures, the topological dual of this space is $C(X)$ (see [22, Theorem 3.10]). By the convex duality [3, Lemma 4.5.8], the desired equality holds.  

A2. General results on the LDP. This subsection collects rather general results in the theory of large deviations. The following, known as the contraction principle, states that the LDP is preserved under continuous mappings. Together with Lemma A2.2, this is used in the deduction of Corollary 1.3.

Lemma A2.1. (e.g., [3, Theorem 4.2.1]). Let $\mathcal{X}$, $\mathcal{Y}$ be Hausdorff spaces and $f: \mathcal{X} \to \mathcal{Y}$ a continuous map. Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of Borel probability measures on $\mathcal{X}$ which satisfies the LDP with a rate function $I: \mathcal{X} \to [0, \infty]$. Then the sequence $\{\mu_n \circ f^{-1}\}_{n=1}^\infty$ of Borel probability measures on $\mathcal{Y}$ satisfies the LDP with a rate function $I': \mathcal{Y} \to [0, \infty]$ defined by

$$I'(y) = \inf \{I(x) : x \in \mathcal{X}, y = f(x)\}.$$ 

If $I$ is a good rate function, then so is $I'$.

Lemma A2.2. Under the hypotheses and notation in Theorem A.1, assume $\mathcal{X}$ is a metric space and $I$ is a good rate function. If $y \in \mathcal{Y}$ is a minimizer of $I'$ then there exists a minimizer $x \in \mathcal{X}$ of $I$ such that $y = f(x)$.

Proof. Let $y \in \mathcal{Y}$ be a minimizer of $I'$. Take a sequence $\{x_k\}_{k=1}^\infty$ in $\mathcal{X}$ such that $y = \lim_k f(x_k)$ and $\lim_k I(x_k) = 0$. Since $I$ is a good rate function, $\{x_k\}$ is contained in a compact set, and so has an accumulation point, say $x$. The lower semi-continuity of $I$ implies $I(x) = 0$. Since $f$ is continuous, $y = f(x)$ holds.  

The next lemma is used in the proof of Theorem C.

Lemma A2.3. Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of Borel probability measures on a normal topological space $\mathcal{X}$ which satisfies the LDP with a good rate function $I$. The support of any accumulation point of $\{\mu_n\}_{n=1}^\infty$ in the weak*-topology is contained in the set $I^{-1}(0)$ of minimizers of $I$.

Proof. Taking a subsequence if necessary we may assume $\{\mu_n\}_{n=1}^\infty$ converges in the weak*-topology to a measure $\mu$. As $I$ is lower semi-continuous, for each $\alpha \in (0, \infty)$ the set $I^\alpha = \{x \in \mathcal{X} : I(x) > \alpha\}$ is open and $\overline{I^\alpha} = \{x \in \mathcal{X} : I(x) \geq \alpha\}$. By Portmanteau’s theorem and the large deviations bound for closed sets, we have

$$\mu(I^\alpha) \leq \liminf_{n \to \infty} \mu_n(I^\alpha) \leq \limsup_{n \to \infty} \mu_n(\overline{I^\alpha}) \leq \limsup_{n \to \infty} \exp(-\alpha n) = 0. \tag{3.1}$$

Since the set $I^{-1}(\infty) = \{x \in \mathcal{X} : I(x) = \infty\}$ is contained in $I^\alpha$, we obtain $\mu(I^{-1}(\infty)) = 0$.

Let $\mathcal{G} \subset \mathcal{X}$ be an arbitrary open set containing $I^{-1}(0)$. Since $I$ is a good rate function, for each $\alpha \in (0, \infty)$ the level set $I_\alpha = \{x \in \mathcal{X} : I(x) \leq \alpha\}$ is compact.
By the normality of $\mathcal{X}$, one can choose a finite number of open sets $\mathcal{G}_1, \ldots, \mathcal{G}_m$ such that for each $k = 1, \ldots, m$ there exists $\alpha_k \in (0, \infty)$ such that $\mathcal{G}_k \subset \mathcal{G}^\alpha$, and $\mathcal{G}^\alpha \cap I_\alpha \subset \bigcup_{k=1}^m \mathcal{G}_k$. From (3.1), $\mu(\mathcal{G}_k) = 0$ holds. Hence $\mu(\mathcal{G}^\alpha \cap I_\alpha) = 0$ and letting $\alpha \to \infty$ yields $\mu(\mathcal{G}^\alpha \setminus I^{-1}(\infty)) = 0$. Together with $\mu(I^{-1}(\infty)) = 0$ we obtain $\mu(\mathcal{G}^\alpha) = 0$, namely, the support of $\mu$ is contained in $\mathcal{G}$. As $\mathcal{G}$ is an arbitrary open set containing $I^{-1}(0)$ and $\mathcal{X}$ is regular, it follows that the support of $\mu$ is contained in $I^{-1}(0)$.

A3. Existence of natural equilibrium states. Let $X$ be a countable Markov shift and $\phi: X \to \mathbb{R}$ a measurable function with $\sup \phi < \infty$. Measures in $\mathcal{M}_\phi(\sigma|_X)$ which attain the supremum in (1.2) are called potential natural equilibrium states. It makes sense to treat the existence of natural equilibrium states.

As a by-product of Theorem 2.4 and a tightness argument, we obtain the existence of natural equilibrium states with mild assumptions on the potential, with no assumption on the connectivity of the shift space.

**Theorem A3.1.** Let $X$ be a countable Markov shift and $\phi: X \to \mathbb{R}$ a uniformly continuous summable function satisfying $\beta_\infty(\phi) < 1$. There exists a natural equilibrium state for the potential $\phi$.

**Proof.** Take a sequence $\{\mu_k\}_{k=1}^\infty$ in $\mathcal{M}_\phi(\sigma|_X)$ such that $h_{\mu_k}(\sigma|_X) + \int \phi d\mu_k \to P_0(\phi)$ as $k \to \infty$. Below we show the tightness of $\{\mu_k\}_{k=1}^\infty$. As $X$ is a Polish space, by Prohorov’s theorem $\{\mu_k\}_{k=1}^\infty$ has an accumulation point. We pick one and denote it by $\mu_\infty$. A slight modification of the proof of Theorem A in §2 shows that $\mu_\infty$ is a natural equilibrium state for the potential $\phi$.

By the assumption $\beta_\infty(\phi) < 1$ and Lemma 2.2, $\inf_k \int \phi d\mu_k > -\infty$ holds. The uniform continuity of $\phi$ implies $D_m(\phi) < \infty$ for some $m \geq 1$. Taking a constant $c < 0$ and considering $\phi + c$ instead of $\phi$ if necessary, we may assume $\sup \phi < 0$ with no loss of generality. We introduce an order $\prec$ on the set $E^m$ of admissible $m$-strings of elements of $\mathbb{N}$ so that $v \in E^m \mapsto \sup [v] \phi$ decreases as $v$ increases. Then, for every $v \in E^m$ and every $k \geq 1$,

$$-\sup_{[v]} \phi \cdot \sum_{w \in E^m : w \succ v} \mu_k[w] \leq - \sum_{w \in E^m : w \succ v} \int [w] \phi d\mu_k \leq - \inf_k \int \phi d\mu_k.$$

Therefore

$$\sup_{k \geq 1} \sum_{w \in E^m : w \succ v} \mu_k[w] \leq \frac{\inf_{k \geq 1} \int \phi d\mu_k}{\sup [v] \phi} \quad \text{for every } v \in E^m. \quad (3.2)$$

From $\inf \phi = -\infty$ and $D_m(\phi) < \infty$, $\sup [v] \phi$ diverges to $-\infty$ as $v$ increases. Let $\epsilon > 0$. Choose an increasing sequence $\{w_i\} \subset E^m$ of elements of $E^m$ such that

$$\sup_{k \geq 1} \sum_{w \in E^m : w \succ w_i} \mu_k[w] \leq \frac{\epsilon}{2^{i+1}} \quad \text{for every } i \in \mathbb{N}.$$

This choice is feasible by (3.2). Define a compact subset $K$ of $X$ by

$$K = \{x \in X : x_{m_i} x_{m_i+1} \cdots x_{m_i+m-1} < w_i \quad \text{for every } i \in \mathbb{N}\}.$$
For every $k \geq 1$ we have
\[
\mu_k(K) = \mu_k \left( \left( \bigcup_{i \in \mathbb{N}} \left\{ x \in X : x_{mi}x_{mi+1} \cdots x_{mi+m-1} \succ w_i \right\} \right)^c \right)
\geq 1 - \sum_{i \in \mathbb{N}} \mu_k \left\{ x \in X : x_{mi}x_{mi+1} \cdots x_{mi+m-1} \succ w_i \right\}
= 1 - \sum_{i \in \mathbb{N}} \sum_{w \in E^m \, w \succ w_i} \mu_k[w]
\geq 1 - \sum_{i \in \mathbb{N}} \sup_{k \geq 1} \sum_{w \in E^m \, w \succ w_i} \mu_k[w]
\geq 1 - \epsilon,
\]
the second equality from the shift-invariance of $\mu_k$. As $\epsilon > 0$ is arbitrary, this shows the tightness of $\{\mu_k\}_{k=1}^\infty$. □

Acknowledgments. This research was partially supported by the Grant-in-Aid for Young Scientists (A) of the JSPS 15H05435 and the Grant-in-Aid for Scientific Research (B) of the JSPS 16KT0021.

References

[1] Arnol’d, V. I., Avez, A.: Ergodic problems of classical mechanics, Benjamin, New York (1982)
[2] Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Second revised edition. Lecture Notes in Mathematics, 470 Springer-Verlag, Berlin 2008.
[3] Dembo, A., Zeitouni, O.: Large deviations techniques and applications, Applications of Mathematics 38, Springer, second edition (1998)
[4] Denker, M., Kabluchko, Z.: An Erdős-Rényi law for mixing processes. Probab. Math. Statist. 27, 139–149 (2007)
[5] Denker, M., Nicol, M.: Erdős-Rényi laws for dynamical systems. J. London Math. Soc. 87, 497–508 (2013)
[6] Dobrushin, R. L.: The description of a random field by means of conditional probabilities and conditions of its regularity. Theor. Prob. Appl. 13, 197–224. (1968)
[7] Dobrushin, R. L.: Gibbsian random fields. The general case. Funct. Anal. Appl. 3, 22–28. (1969)
[8] Ellis, R.S.: Entropy, large deviations, and statistical mechanics, Grundlehren der Mathematischen Wissenschaften 271, Springer (1985)
[9] Erdős, P., Rényi, A.: On a new law of large numbers. J. Anal. Math. 23, 103–111 (1970)
[10] Fan, A.-H., Jordan, T., Liao, L., Rams, M.: Multifractal analysis for expanding interval maps with infinitely many branches. Trans. Amer. Math. Soc. 367, 1847–1870 (2015)
[11] Fiebig, D., Fiebig, U.-R., Yuri, M.: Pressure and equilibrium states for countable state Markov shifts. Israel J. Math. 131, 221–257 (2002)
[12] Jenkinson, O., Mauldin, R. D., Urbański, M.: Zero temperature limits of Gibbs-equilibrium states for countable alphabet subshifts of finite type. Journal of Statistical Physics. 119, 765–776 (2005)
[13] Kifer, Y.: Large deviations in dynamical systems and stochastic processes, Trans. Amer. Math. Soc. 321, 505–524 (1990)
[14] Kifer, Y.: Large deviations, averaging and periodic orbits of dynamical systems, Commun. Math. Phys. 162, 33–46 (1994)
[15] Lanford, O. E., Ruelle, D.: Observables at infinity and states with short range correlations in statistical mechanics. Commun. Math. Phys. 13, (1969) 194–215.

[16] Mauldin, R. D., Urbański, M.: Gibbs states on the symbolic space over infinite alphabet. Israel J. Math. 125, 93–130 (2001)

[17] Mauldin, R. D., Urbański, M.: *Graph directed Markov systems: Geometry and Dynamics of Limit Sets*. Cambridge Tracts in Mathematics 148 Cambridge University Press (2003)

[18] Mayer, D. H.: On the thermodynamic formalism for the Gauss transformation. Commun. Math. Phys. 130, 311–333 (1990)

[19] Misiurewicz, M.: A short proof of the variational principle for a \( \mathbb{Z}^N \) action on a compact space. Asterisque 40, 147–187 (1976)

[20] Orey, S., Pelikan, S.: Deviations of trajectory averages and the defect in Pesin’s formula for Anosov diffeomorphisms. Trans. Amer. Math. Soc. 315, 741–753 (1989)

[21] Phelps, R.: *Lectures on Choquet’s Theorem*. Van Nostrand, Princeton, N.J., 1966.

[22] Rudin, W.: *Functional Analysis*, Second Edition. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991

[23] Ruelle, D.: *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics*. Second edition. Cambridge University Press (2004)

[24] Sarig, O.: Thermodynamic formalism for countable Markov shifts. Ergodic Theory and Dynamical Systems 19, 1565–1593 (1999)

[25] Sarig, O.: Phase transitions for countable topological Markov shifts. Commun. Math. Phys. 217, 555–577.

[26] Sarig, O.: Characterization of the existence of Gibbs states for countable Markov shifts, Proc. Amer. Math. Soc. 131, 1751–1758 (2003)

[27] Sarig, O.: Thermodynamic formalism for countable Markov shifts. Proceedings of Symposia in Pure Mathematics. 89, 81–117 (2015)

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

[29] Takahashi, Y.: Entropy functional (free energy) for dynamical systems and their random perturbations. In *Stochastic analysis (Katata/Kyoto, 1982)*, North-Holland Math. Library, 32, 437–467. North-Holland, Amsterdam, 1984.

[30] Takahashi, Y.: Asymptotic behaviours of measures of small tubes: entropy, Liapunov’s exponent and large deviation. In *Dynamical systems and applications (Kyoto, 1987)*, World Sci. Adv. Ser. Dynam. Systems 5, 1–21 (1987)

[31] Takahasi, H.: Large deviation principles for countable Markov shifts. Trans. Amer. Math. Soc. in press

[32] Takahasi, H.: Entropy-approachability for transitive Markov shifts over infinite alphabet. Proc. Amer. Math. Soc. in press

[33] Walters, P.: *Introduction to ergodic theory*, Graduate Texts in Mathematics 79, Springer (1982)

[34] Yuri, M.: Large deviations for countable to one Markov systems. Commun. Math. Phys. 258, 455–474 (2005)

KEIO INSTITUTE OF PURE AND APPLIED SCIENCES (KiPAS), DEPARTMENT OF MATHEMATICS, KEIO UNIVERSITY, YOKOHAMA, 223-8522, JAPAN

E-mail address: hiroki@math.keio.ac.jp

URL: http://www.math.keio.ac.jp/~hiroki/