Ergod. Th. & Dynam. Sys. (2020), **40**, 2219–2238 © *Cambridge University Press, 2018*
doi:10.1017/etds.2018.129

The weak Bernoulli property for matrix Gibbs states

MARK PIRAINO

Department of Mathematics and Statistics, University of Victoria, Canada
(e-mail: mpiraino@uvic.ca)

(*Received* 13 *June* 2018 *and accepted in revised form* 5 *November* 2018)

Abstract. We study the ergodic properties of a class of measures on $\Sigma^{\mathbb{Z}}$ for which $\mu_{\mathcal{A},t}[x_0 \cdots x_{n-1}] \approx e^{-nP}\|A_{x_0} \cdots A_{x_{n-1}}\|^t$, where $\mathcal{A} = (A_0, \ldots, A_{M-1})$ is a collection of matrices. The measure $\mu_{\mathcal{A},t}$ is called a matrix Gibbs state. In particular, we give a sufficient condition for a matrix Gibbs state to have the weak Bernoulli property. We employ a number of techniques to understand these measures, including a novel approach based on Perron–Frobenius theory. We find that when t is an even integer the ergodic properties of $\mu_{\mathcal{A},t}$ are readily deduced from finite-dimensional Perron–Frobenius theory. We then consider an extension of this method to $t > 0$ using operators on an infinite-dimensional space. Finally, we use a general result of Bradley to prove the main theorem.

Key words: symbolic dynamics, thermodynamic formalism, weak Bernoulli property, matrix equilibrium states
2010 Mathematics Subject Classification: 37A25 (Primary)

1. *Introduction*

We recall the definition of a scalar Gibbs state. Let Σ_A be a shift of finite type and $\varphi : \Sigma_A \to \mathbb{R}$. We say that a shift-invariant measure, μ_φ, is a scalar Gibbs state for φ provided there exist $C > 0$ and P such that

$$C^{-1} \leq \frac{\mu_\varphi([x_0 \cdots x_{n-1}])}{e^{-nP+S_n\varphi(x)}} \leq C$$

for all $x \in \Sigma_A$ and $n > 0$ (where $S_n\varphi(x) = \sum_{k=0}^{n-1} \varphi(\sigma^k x)$). By analogy, if $\mathcal{A} = (A_0, \ldots, A_{M-1}) \in M_d(\mathbb{R})^M$ and $t > 0$ we say that a shift-invariant measure $\mu_{\mathcal{A},t}$ is a matrix Gibbs state for $(\mathcal{A}, t)$ provided there exist a constant $C > 0$ and P such that

$$C^{-1}\mu_{\mathcal{A},t}([x_0 \cdots x_{n-1}]) \leq e^{-nP}\|A_{x_0} \cdots A_{x_{n-1}}\|^t \leq C\mu_{\mathcal{A},t}([x_0 \cdots x_{n-1}]) \qquad (1)$$

for all $x \in \Sigma^{\mathbb{Z}}$ ($\Sigma = \{0, \ldots, M-1\}$) and $n > 0$. Notice we are working with the two-sided shift and not, as has been done in previous literature, the one-sided shift. Thus in a

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

strict sense one may consider that we are working with the invertible extension of matrix Gibbs states, this is important when working on the isomorphism problem and it is also necessary so that we can apply the results in [**6**]. When $t = 1$ we refer to the measure simply as the Gibbs state for $\mathcal{A}$. P is uniquely determined by (1) and is called the pressure, denoted $P(\mathcal{A}, t)$. A computation shows that

$$P(\mathcal{A}, t) = \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{x_0\cdots x_{n-1}} \|A_{x_0} \cdots A_{x_{n-1}}\|^t \right).$$

For the remainder of this article a Gibbs state will always refer to a matrix Gibbs state. Matrix Gibbs states are also equilibrium states for a sub-additive variational principle [**7**]

$$P(\mathcal{A}, t) = \sup_{\mu\in\mathcal{M}(\sigma)} [h(\mu) + t\Lambda(\mathcal{A}, \mu)], \tag{2}$$

where $\Lambda(\mathcal{A}, \mu)$ is the maximal Lyapunov exponent

$$\Lambda(\mathcal{A}, \mu) = \lim_{n\to\infty} \frac{1}{n} \int \log \|A_{x_0} \cdots A_{x_{n-1}}\| \, d\mu(x).$$

Measures which achieve the supremum are called matrix equilibrium states. Such measures always exist by weak* compactness and upper semi-continuity of $h(\mu) + t\Lambda(\mathcal{A}, \mu)$. The connection between Gibbs states and equilibrium states for the variation principle (2) was studied in [**13**]. The study of these measures was originally motivated by their applications to dimension theory [**14**]. Recently, however, interest has been shown in determining their ergodic properties [**20**, **21**]. In the classical case for Hölder continuous functions, scalar Gibbs states are well known to have many nice statistical properties. It is natural to ask to what extent matrix Gibbs states share these properties.

One of the strongest of these properties is that the dynamical system defined by the shift map and a scalar Gibbs state for a Hölder potential is isomorphic to a Bernoulli shift, and this is the problem we will focus on this article. This is a particularly appealing property because Bernoulli shifts are classified up to isomorphism by their entropy [**22**]. In general, it is very difficult to explicitly construct isomorphisms between measure-preserving systems. One of the most common methods for demonstrating that a measure-preserving system is isomorphic to a Bernoulli shift is to show that it is weak Bernoulli and appeal to [**15**]. This is the strategy we will follow in this paper. The same method has been used by Bowen [**2**] for scalar Gibbs states. Recall what it means for a dynamical system to be weak Bernoulli.

Definition 1.1. We say that partitions $\mathcal{Q}$ and $\mathcal{R}$ are ε-independent (written $\mathcal{Q} \perp^{\varepsilon} \mathcal{R}$) if

$$\sum_{q\in\mathcal{Q}, r\in\mathcal{R}} |\mu(q \cap r) - \mu(q)\mu(r)| < \varepsilon.$$

We say that a partition $\mathcal{P}$ is weak Bernoulli if for every $\varepsilon > 0$ there exists N such that $\bigvee_{i=0}^{s-1} \sigma^{-i}\mathcal{P} \perp^{\varepsilon} \bigvee_{i=t}^{t+r-1} \sigma^{-i}\mathcal{P}$ for all $r, s \geq 0$ and $t \geq s + N$. We say that $\mu_{\mathcal{A},t}$ is weak Bernoulli if the standard partition $\mathcal{P} = \{[i] : 0 \leq i \leq M - 1\}$ is weak Bernoulli.

For a word $I = i_0 i_1 \cdots i_{n-1}$ we write

$$A_I := A_{i_0} A_{i_1} \cdots A_{i_{n-1}}$$

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

and we denote the length of the word I by $|I|$. We say that $\mathcal{A} = (A_0, \ldots, A_{M-1}) \in M_d(\mathbb{R})^M$ is *irreducible* if the matrices have no common proper and non-trivial invariant subspace. This implies that there exists a constant $\delta > 0$ such that

$$\sum_{|K| \le d} \|A_I A_K A_J\| \ge \delta \|A_I\| \|A_J\| \tag{3}$$

for all I, J. With this in mind we make the following definition

Definition 1.2. We say that $\mathcal{A} = (A_0, \ldots, A_{M-1})$ is *primitive* if there exists an N and a $\delta > 0$ such that

$$\sum_{|K| = N} \|A_I A_K A_J\| \ge \delta \|A_I\| \|A_J\| \tag{4}$$

for all I, J.

For both irreducible and primitive collections of matrices, matrix Gibbs states are known to exist and be unique [**12**, Theorem 5.5] for all $t > 0$. The terms 'irreducible' and 'primitive' are familiar from Perron–Frobenius theory, and indeed the notions are connected. Let $L_{\mathcal{A}} : M_d(\mathbb{R}) \to M_d(\mathbb{R})$ be defined by $L_{\mathcal{A}} B = \sum_i A_i^* B A_i$; then $L_{\mathcal{A}}$ preserves the cone of positive semi-definite matrices. The operator $L_{\mathcal{A}}$ appears in connection with a class of measures related to fractal geometry called Kusuoka measures [**19**] (see Example 2.3). One can check that if $L_{\mathcal{A}}$ is irreducible (respectively, primitive) in the sense of Perron–Frobenius theory then $\mathcal{A}$ satisfies equation (3) (respectively, equation (4)); for the details see Proposition A.7. Our main theorem is the following.

THEOREM 1.3. *Suppose that $\mathcal{A} = (A_0, \ldots, A_{M-1})$ is primitive. Then for any $t > 0$ the unique t-Gibbs state for $\mathcal{A}$ is weak Bernoulli.*

The proof of Theorem 1.3 can be found in §4. The proof relies on a general result of Bradley [**6**], which is somewhat opaque. With this in mind we also present a method for understanding matrix Gibbs states through transfer operators which is interesting in its own right. Understanding the ergodic/statistical properties of Gibbs states in sub-additive thermodynamic formalism has long been a challenge, with most results being achieved using fairly ad-hoc methods. This is in contrast to the case for scalar Gibbs states which has a well-developed methodology for deducing ergodic/statistical properties relying on the transfer operator. In this article we adapt the classical doctrine of transfer operators for scalar Gibbs states to matrix Gibbs states.

In §2 we show that in the case when t is an even integer the ergodic properties of $\mu_{A,t}$ can be readily understood by studying the convergence properties of a matrix. As a consequence we can obtain an exponential mixing result which includes an explicit rate determined by the spectral gap of a finite-dimensional matrix. This naturally leads to the problem of generalizing this approach to $t > 0$. In §3 we generalize §2 using operators on a suitable infinite- dimensional vector space. A major advantage of the approach in §§2 and 3 is that we can give an explicit construction of certain Gibbs states, including a formula for the measure of a cylinder set. Previous methods have relied on abstract compactness arguments, realizing the Gibbs state as a weak* limit point of a sequence of

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

measures. As many properties are not preserved under weak* limits, this makes an analysis of the Gibbs state difficult. Our transfer operator approach allows us to give direct proofs of ergodic properties. It also provides a strong intuition for understanding how properties of the collection $\mathcal{A}$ are reflected in the ergodic properties of $\mu_{\mathcal{A},t}$.

2. *Matrices which preserve a common cone*

One particular class of matrix Gibbs states has appeared extensively in applications. Consider the following examples.

Example 2.1. Bernoulli measures; take $d = 1$.

Example 2.2. Factors of Markov measures. The 1-Gibbs states for collections of non-negative matrices are precisely factors of Markov measures; for details, see [**4**] or [**8**, **27**]. In fact, allowing the operators in $\mathcal{A}$ to act on an infinite-dimensional space, factors of Gibbs states for Hölder potentials can be viewed as Gibbs states for a suitable collection of operators; see [**24**].

Example 2.3. The Kusuoka measure [**19**] was originally studied because of its connections to fractal geometry. We briefly recall the construction. Let $L_i B = A_i^* B A_i$ and $L_{\mathcal{A}} = \sum_i L_i$. When $\mathcal{A}$ is irreducible there exist U, V positive definite matrices such that $L_{\mathcal{A}} U = \rho(L_{\mathcal{A}})U$, $L_{\mathcal{A}}^* V = \rho(L_{\mathcal{A}})V$ (notice that $L_{\mathcal{A}}^* B = \sum_i A_i B A_i^*$) and $\langle U, V\rangle_{\mathrm{HS}} = 1$ (where $\langle A, B\rangle_{\mathrm{HS}} = \mathrm{tr}(A^* B)$). The Kusuoka measure is then obtained by extending

$$\mu[x_0 \cdots x_{n-1}] = \rho(L_{\mathcal{A}})^{-n} \langle L_{x_0} L_{x_1} \cdots L_{x_{n-1}} U, V\rangle_{\mathrm{HS}}$$

to a measure using Carathéodory's extension theorem. It was shown in [**20**] that the Kusuoka measure is a 2-Gibbs state. We will generalize this result to k-Gibbs states for k even in Example 2.7. Observe that, thinking of the linear maps L_i as matrices, the Kusuoka measure is the 1-Gibbs state for the collection $\widehat{\mathcal{A}} = (L_0, \ldots, L_{M-1})$ each of which preserves the cone of positive semi-definite matrices.

The property shared by all of these matrix equilibrium states is that all of the matrices preserve a common cone. Our goal for this section is, then, to treat these measures in an abstract manner. As one of the applications of this section is the Kusuoka measure, we work with matrices preserving an abstract cone K. For the most part, the reader will lose no intuition by simply thinking of K as being the positive quadrant of $\mathbb{R}^d$. For the reader's convenience we have collected some definitions and facts about abstract cones in finite-dimensional vector spaces in the Appendix. Recall that

$$\mathrm{var}_n\ f = \sup\{|f(x) - f(y)| : x_i = y_i \text{ for all } |i| \le n-1\},$$

and for $\theta \in (0, 1)$ define

$$\mathcal{H}_\theta = \{f \in C(\Sigma^{\mathbb{Z}}) : \text{there exists a constant } K > 0 \text{ for which } \mathrm{var}_n\ f \le K\theta^n\}.$$

We denote the least such constant by $|f|_\theta$, and $\mathcal{H}_\theta$ becomes a Banach space with norm $\|f\|_\theta = \|f\|_\infty + |f|_\theta$. The goal of this section is to prove the following theorem.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

THEOREM 2.4. *Let $\mathcal{A} = (A_0, \ldots, A_{M-1}) \in M_d(\mathbb{R})^M$. Suppose that each A_i is non-negative with respect to a cone K and $A := \sum_i A_i$ is such that $\sum_{k=0}^{d-1} A^k$ maps $K\backslash\{0\}$ into the interior of K (that is, A is K-irreducible). Then there exists a 1-Gibbs state for $\mathcal{A}$, denoted $\mu_{\mathcal{A}}$. Moreover:*

(1) *$\mu_{\mathcal{A}}$ is ergodic and thus unique, and $P(\mathcal{A}, 1) = \log \rho(A)$.*
(2) *If there exists an N such that A^N maps $K\backslash\{0\}$ into the interior of K (that is, A is K-primitive) then:*
   (a) *$\mu_{\mathcal{A}}$ is weak Bernoulli.*
   (b) *$\mu_{\mathcal{A}}$ has exponential decay of correlations for Hölder continuous functions. That is, for a fixed $\theta \in (0, 1)$ there are constants D and $\gamma \in (0, 1)$ such that*

$$\left| \int f \cdot g \circ \sigma^n \, d\mu_{\mathcal{A}} - \int f d\mu_{\mathcal{A}} \int g \, d\mu_{\mathcal{A}} \right| \leq D \| f \|_\theta \| g \|_\theta \gamma^n$$

   for all $f, g \in \mathcal{H}_\theta$, $n \geq 0$. In addition, the rate γ is determined by θ and the eigenvalues of A.

For the Kusuoka measure, part (2)(b) is known [**18**]; however, our proof is fundamentally different and significantly more elementary. In particular, the method in [**18**] uses the g-function for the Kusuoka measure and transfer operator techniques. This is technically challenging largely due to the fact that the g-function can fail to be continuous.

We can explicitly construct the measure $\mu_{\mathcal{A}}$. As A is irreducible we may take u, v to be right and left eigenvectors respectively corresponding to the spectral radius $\rho(A)$ with $\langle u, v \rangle = 1$. On cylinder sets we define

$$\mu_{\mathcal{A}}[x_0 x_1 \cdots x_{n-1}] = \rho(A)^{-n} \langle A_{x_0} A_{x_1} \cdots A_{x_{n-1}} u, v \rangle. \tag{5}$$

Using the fact that u, v are eigenvectors for A, it is readily checked that

$$\sum_i \mu_{\mathcal{A}}[i x_0 \cdots x_{n-1}] = \mu_{\mathcal{A}}[x_0 \cdots x_{n-1}] = \sum_i \mu_{\mathcal{A}}[x_0 \cdots x_{n-1} i].$$

As cylinder sets form a semi-algebra, Carathéodory's extension theorem implies that this extends to a shift-invariant measure on $\Sigma^{\mathbb{Z}}$. Next our goal is to show that this is a 1-Gibbs state for $\mathcal{A}$ and that it is unique. To do so, we prove the following proposition.

PROPOSITION 2.5. *Suppose that $\mathcal{A} = (A_0, \ldots, A_{M-1}) \in M_d(\mathbb{R})^M$ is such that each A_i is non-negative with respect to a cone K and $A := \sum_i A_i$ is K-irreducible. Then*

(1) *$\mu_{\mathcal{A}}$ is ergodic.*
(2) *$\mu_{\mathcal{A}}$ satisfies the Gibbs inequality (1) with $P = \log \rho(A)$.*

Proof. (1) Observe that

$$A^n = \left(\sum_i A_i \right)^n = \sum_{|K|=n} A_K. \tag{6}$$

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Let I, J be words. Then

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n}\mu_{\mathcal{A}}([I]\cap\sigma^{-k}[J])-\mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])\right|\\
&\quad\le\left|\frac{1}{n}\sum_{k=1}^{|I|}\mu_{\mathcal{A}}([I]\cap\sigma^{-k}[J])\right|\\
&\qquad+\rho(A)^{-|I|-|J|}\left|\left\langle A_I\left(\frac{1}{n}\sum_{k=|I|+1}^{n}\rho(A)^{|I|-k}A^{k-|I|}\right)A_Ju,\,v\right\rangle\right.\\
&\qquad\left.-\langle A_Iu,\,v\rangle\langle A_Ju,\,v\rangle\right|\\
&\quad\xrightarrow{n\to\infty}\;0+\rho(A)^{-|I|-|J|}|\langle A_I\langle A_Ju,\,v\rangle u,\,v\rangle-\langle A_Iu,\,v\rangle\langle A_Ju,\,v\rangle|=0
\end{aligned}$$

by the Perron–Frobenius Theorem A.5(2)(b). As cylinder sets are a generating semi-algebra, this implies that $\mu_{\mathcal{A}}$ is ergodic.

(2) From the Perron–Frobenius theorem we have that $u\in\operatorname{int}(K)$, $v\in\operatorname{int}(K^*)$. Thus the Gibbs inequality follows directly from an application of Lemma A.6. □

As ergodic measures are mutually singular this implies that $\mu_{\mathcal{A}}$ is the unique 1-Gibbs state for $\mathcal{A}$. The proof of the previous lemma shows that mixing properties of $\mu_{\mathcal{A}}$ are related to the convergence of A^n. It is this fact that we will exploit to prove the remaining assertions in Theorem 2.4.

PROPOSITION 2.6. *Suppose that $\mathcal{A}=(A_0,\dots,A_{M-1})\in M_d(\mathbb{R})^M$ is such that each A_i is non-negative with respect to a cone K and $A:=\sum_i A_i$. If A is K-primitive then the measure $\mu_{\mathcal{A}}$ is weak Bernoulli.*

Proof. Let $r,s\ge 1$, $t\ge s$ and take $[I]\in\bigvee_{i=0}^{s-1}\sigma^{-i}\mathcal{P}$ and $[_tJ]\in\bigvee_{i=t}^{t+r-1}\sigma^{-i}\mathcal{P}$. Notice that

$$\begin{aligned}
&|\mu_{\mathcal{A}}([I]\cap[_tJ])-\mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])|\\
&\quad=\left|\sum_{|K|=t-s}\mu_{\mathcal{A}}([IKJ])-\mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])\right|\\
&\quad=\left|\sum_{|K|=t-s}\rho(A)^{-(s+r+(t-s))}\langle A_IA_KA_Ju,\,v\rangle-\rho(A)^{-(s+r)}\langle A_Iu,\,v\rangle\langle A_Ju,\,v\rangle\right|\\
&\quad=\rho(A)^{-(s+r)}\left|\left\langle A_I\left(\rho(A)^{-(t-s)}\sum_{|K|=t-s}A_K\right)A_Ju,\,v\right\rangle-\langle A_Iu,\,v\rangle\langle A_Ju,\,v\rangle\right|.
\end{aligned}$$

Notice that

$$\rho(A)^{-(t-s)}\sum_{|K|=t-s}A_K=\rho(A)^{-(t-s)}A^{t-s}=uv^T+(\rho(A)^{-(t-s)}A^{t-s}-uv^T).$$

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Thus

$$\begin{aligned}
&|\mu_{\mathcal{A}}([I] \cap [_t J]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])| \\
&\quad = \rho(A)^{-(s+r)} |\langle A_I(\rho(A)^{-(t-s)} A^{t-s} - uv^T) A_J u, v\rangle| \\
&\quad \leq \rho(A)^{-(s+r)} \|A_I^* v\| \|A_J u\| \|\rho(A)^{-(t-s)} A^{t-s} - uv^T\| \\
&\quad \leq C\beta^{t-s} \rho(A)^{-s} \|A_I\| \|\rho(A)^{-r} \|A_J\| \\
&\quad \leq C'\beta^{t-s} \mu_{\mathcal{A}}(I)\mu_{\mathcal{A}}(J) \quad \text{by Proposition 2.5,}
\end{aligned}$$

where $\beta = (|\lambda_2| + \varepsilon)/\rho(A) < 1$ for a small $\varepsilon > 0$ as in Perron–Frobenius Theorem A.5. Then we have

$$\sum_{I,J} |\mu_{\mathcal{A}}([I] \cap [_t J]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])| \leq K\beta^{t-s} \sum_{I,J} \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J]) = K\beta^{t-s}.$$

Hence $\mu_{\mathcal{A}}$ is weak Bernoulli. □

Thus we have proven Theorem 2.4(2)(a). Part (2)(b) follows by an approximation argument; see Bowen's book [**3**, Theorem 1.26]. We conclude this section with an example which shows that k-Gibbs states can be understood in terms of matrices preserving a common cone, for k an even integer.

Example 2.7. This example generalizes the Kusuoka measure (the Kusuoka measure is the case of $k = 2$). Let k be an even integer and define

$$S = \text{span}\{v^{\otimes k} : v \in \mathbb{R}^d\}.$$

We consider the following cone in S^*:

$$K = \{w \in S^* : \langle v^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}} \geq 0 \text{ for all } v \in \mathbb{R}^d\}.$$

Note that when k is odd this set is $\{0\}$. When k is even, K is a cone with non-void interior (see Proposition A.8). The cone K is sometimes referred to as the positive semi-definite tensor cone: in the case of $k = 2$ this cone can be identified with positive semi-definite matrices. Suppose that $\mathcal{A} = (A_0, \ldots, A_{M-1})$ is a collection of matrices with no common proper, non-trivial invariant subspace. Consider the collection $\mathcal{A}' = ((A_0^{\otimes k})^*, \ldots, (A_{M-1}^{\otimes k})^*)$. The collection $\mathcal{A}'$ preserves the cone K. We claim that in fact $A = \sum_i (A_i^{\otimes k})^*$ is irreducible with respect to K. To prove this it is enough to show that no eigenvector of A lies on the boundary of K [**25**, Theorem 4.1]. Suppose that $w \in K$, $w \neq 0$, and that $Aw = \lambda w$ and define

$$W = \text{span}\{u : \langle u^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}} = 0\}.$$

We claim that W is invariant under $\mathcal{A}$. If $\langle u^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}} = 0$ then

$$0 = \langle u^{\otimes k}, Aw\rangle_{(\mathbb{R}^d)^{\otimes k}} = \sum_i \langle (A_i u)^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}}.$$

As $w \in K$ this implies that $\langle (A_i u)^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}} = 0$ for each i. Thus W is $\mathcal{A}$ invariant, so it is either $\mathbb{R}^d$ or $\{0\}$. As $w \neq 0$ we must have that $W = \{0\}$. Therefore $w \in \text{int}(K)$ by

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Lemma A.3 and A is irreducible. Constructing the 1-Gibbs state for $\mathcal{A}'$, we see that it satisfies the Gibbs inequality: there exist constants $C > 0$ and P such that

$$C^{-1}\mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]) \leq e^{-nP}\|(A_{x_0}^{\otimes k})^*(A_{x_1}^{\otimes k})^* \cdots (A_{x_{n-1}}^{\otimes k})^*\| \leq C\mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]).$$

As $A_{x_{n-1}}^{\otimes k} A_{x_{n-2}}^{\otimes k} \cdots A_{x_0}^{\otimes k} = (A_{x_{n-1}} A_{x_{n-2}} \cdots A_{x_0})^{\otimes k}$ we have that

$$C^{-1}\mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]) \leq e^{-nP}\|A_{x_{n-1}} A_{x_{n-2}} \cdots A_{x_0}\|^k \leq C\mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]).$$

Strictly speaking, the order of the product of matrices is backwards from the Gibbs inequality in equation (1). By taking $\mathcal{A} = (A_0^*, \ldots, A_{M-1}^*)$ this can be changed (see Proposition A.9). Thus we have found an elementary way of constructing k-Gibbs states for all even integers.

Therefore we have a completely explicit description of Gibbs states when t is an even integer.

3. *Transfer operators and exponential mixing*

The goal of this section is to explore a method for constructing matrix Gibbs states and proving ergodic and statistical properties using transfer operators. This approach is interesting for number of reasons. In particular, it is an application of transfer operator methods to a problem in sub-additive ergodic theory. It is also a reasonable generalization of Example 2.7 using operators on infinite-dimensional spaces. We will need the following definitions.

Definition 3.1. We say that a collection of invertible $d \times d$ matrices $(A_0, \ldots, A_{M-1})$ is *strongly irreducible* if they do not preserve a finite union of proper and non-trivial subspaces.

Definition 3.2. An element $B \in M_d(\mathbb{R})$ is called *proximal* if B has a simple eigenvalue of modulus $\rho(B)$ and any other eigenvalue has modulus strictly smaller then $\rho(B)$. The collection $(A_0, \ldots, A_{M-1})$ is called *proximal* if there exists a product $B = A_{x_0} \cdots A_{x_n}$ that is proximal.

We have the following theorem.

THEOREM 3.3. *Suppose that $\mathcal{A} = (A_0, \ldots A_{M-1})$ is a collection of real invertible $d \times d$ matrices which is proximal and strongly irreducible. Then for any $t \geq 0$ there exists a unique Gibbs state for $(\mathcal{A}, t)$, $\mu_{\mathcal{A},t}$. Moreover:*

(1) *$\mu_{\mathcal{A},t}$ is weak Bernoulli;*
(2) *$\mu_{\mathcal{A},t}$ has exponential decay of correlations for Hölder continuous functions. That is, for a fixed $\theta \in (0, 1)$ there are constants D and $\gamma \in (0, 1)$ such that*

$$\left| \int f \cdot g \circ \sigma^n \, d\mu_{\mathcal{A},t} - \int f \, d\mu_{\mathcal{A},t} \int g \, d\mu_{\mathcal{A},t} \right| \leq D\|f\|_\theta \|g\|_\theta \gamma^n$$

for all $f, g \in \mathcal{H}_\theta$, $n \geq 0$.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

In the previous section we have seen that the role of the transfer operator for $t = 2k$ was played by $A = \sum_i A_i^{\otimes 2k}$; we need to find a suitable replacement. By identifying 2-tensors with bilinear forms which are in turn a subspace of the 2-homogeneous functions one is naturally led to consider the action of the matrices on t-homogeneous functions. This is then equivalent to the action of the matrices on the projective space $\mathbb{RP}^{d-1}$ weighted by the functions $\|A_i(u/\|u\|)\|^t$. That is, define a transfer operator by

$$L_t f(\overline{u}) = \sum_{i=0}^{M-1} \left\| A_i \frac{u}{\|u\|} \right\|^t f(\overline{A_i u}), \tag{7}$$

which acts on $C(\mathbb{RP}^{d-1})$. The connection between matrix Gibbs states and this operator is made clear in Proposition 3.4. First we fix some notation. For a function h and a measure ν we write

$$\langle h, \nu \rangle = \int h \, d\nu.$$

Recall that $\mathbb{RP}^{d-1}$ is obtained by taking the quotient of $\mathbb{R}^d \setminus \{0\}$ by the equivalence relation $x \sim y$ if and only if $x = \lambda y$ for some $\lambda \neq 0$. We denote the equivalence class of a vector v by $\overline{v}$. Define a metric on $\mathbb{RP}^{d-1}$ by

$$d(\overline{u}, \overline{w}) = \inf\{\|u' - w'\| : \|u'\| = \|w'\| = 1 \text{ and } \overline{u'} = \overline{u}, \overline{w'} = \overline{w}\}.$$

PROPOSITION 3.4. *Let $t \geq 0$ and $\mathcal{A} = (A_0, \ldots, A_{M-1})$ be a collection of invertible matrices. Suppose that there exist ν_t a Borel probability measure not supported on a projective subspace and h_t a strictly positive continuous function such that $L_t h_t = \rho(L_t) h_t$, $L_t^* \nu_t = \rho(L_t)\nu_t$ and $\langle h_t, \nu_t \rangle = 1$. Define L_i by $L_i f(\overline{u}) = \|A_i(u/\|u\|)\|^t f(\overline{A_i u})$. Then the formula*

$$\mu_{A,t}[x_0 x_1 \cdots x_{n-1}] = \rho(L_t)^{-n} \int_{\mathbb{RP}^{d-1}} L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h_t(\overline{u}) \, d\nu_t(\overline{u}) \tag{8}$$

extends to a shift-invariant measure on $\Sigma^{\mathbb{Z}}$. Moreover, $\mu_{\mathcal{A},t}$ is a Gibbs state for $(\mathcal{A}, t)$.

Proof. The assumption that h_t, ν_t are eigenvectors corresponding to $\rho(L_t)$ implies that the formula in (8) extends to a shift-invariant measure by Carathéodory's extension theorem. All that remains to be shown is that $\mu_{\mathcal{A},t}$ satisfies the Gibbs inequality. To see why the Gibbs inequality holds, notice that

$$A \mapsto \int_{\mathbb{RP}^{d-1}} \left\| A \frac{u}{\|u\|} \right\|^t d\nu_t(\overline{u})$$

is continuous and strictly positive (by the assumption that ν_t is not supported on a projective subspace) from the set of norm-one $d \times d$ matrices to $\mathbb{R}$. Take $C > 0$ such that

$$\int_{\mathbb{RP}^{d-1}} \left\| A \frac{u}{\|u\|} \right\|^t d\nu(\overline{u}) \geq C \|A\|^t$$

for all $A \in M_d(\mathbb{R})$. Thus

$$\rho(L)^{-n} \langle L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h, \nu \rangle \geq (\inf h) C \rho(L)^{-n} \|A_{x_0} A_{x_1} \cdots A_{x_{n-1}}\|^t$$

and

$$\rho(L)^{-n} \langle L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h, \nu \rangle \leq (\sup h) \rho(L)^{-n} \|A_{x_0} A_{x_1} \cdots A_{x_{n-1}}\|^t,$$

which shows that the measure $\mu_{\mathcal{A},t}$ satisfies the Gibbs inequality. □

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

If $I = i_0 i_1 \cdots i_{n-1}$ we will use the notation that

$$L_I := L_{i_{n-1}} \cdots L_{i_1} L_{i_0}.$$

Notice that this is backward from the definition of A_I. To see why consider

$$\begin{aligned} L_{x_1} L_{x_0} f(\overline{u}) &= \left\| A_{x_1} \frac{u}{\|u\|} \right\|^t L_{x_0} f(\overline{A_{x_1} u}) \\ &= \left\| A_{x_1} \frac{u}{\|u\|} \right\|^t \left\| A_{x_0} \frac{A_{x_1} u}{\|A_{x_1} u\|} \right\|^t f(\overline{A_{x_0} A_{x_1} u}) \\ &= \left\| A_{x_0} A_{x_1} \frac{u}{\|u\|} \right\|^t f(\overline{A_{x_0} A_{x_1} u}). \end{aligned}$$

As we can see, pre-composition reverses the order of the products.

Operators like L_t have appeared frequently in the study of random matrix products. This is, however, the first time they have been used to construct a measure on $\Sigma^{\mathbb{Z}}$ and deduce ergodic and statistical properties. To prove Theorem 3.3 all we require is a suitable Perron–Frobenius theorem. For each $\varepsilon > 0$ denote by $C^\varepsilon(\mathbb{RP}^{d-1})$ the space of ε-Hölder continuous functions in the d metric on $\mathbb{RP}^{d-1}$. This becomes a Banach space in the usual way with norm $\|\cdot\|_\varepsilon = \|\cdot\|_\infty + |\cdot|_\varepsilon$ (where $|f|_\varepsilon$ is the least ε-Hölder constant for f). Set $\bar{t} = \min\{1, t\}$. The following theorem is a result of Guivarc'h and Le Page [**16**].

THEOREM 3.5. (Guivarc'h and Le Page [**16**]) *Let $t \geq 0$. Suppose that $(A_0, \ldots, A_{M-1})$ are real, invertible, strongly irreducible and proximal. Then there exists an ε with $0 < \varepsilon \leq \bar{t}$ such that the following hold.*

(1) *$L_t : C^\varepsilon(\mathbb{RP}^{d-1}) \to C^\varepsilon(\mathbb{RP}^{d-1})$, that is, L_t preserves the space of ε-Hölder functions.*

(2) *The spectral radius of $L_t : C^\varepsilon(\mathbb{RP}^{d-1}) \to C^\varepsilon(\mathbb{RP}^{d-1})$ is equal to $e^{P(\mathcal{A},t)}$. That is,*

$$\log \rho(L_t) = \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{|I|=n} \|A_I\|^t \right) = P(\mathcal{A}, t).$$

(3) *There exists a unique Borel probability measure ν_t on $\mathbb{RP}^{d-1}$, not supported on a projective subspace, such that $L_t^* \nu_t = \rho(L_t)\nu_t$.*

(4) *There exists a unique $\bar{t}$-Hölder function $h_t : \mathbb{RP}^{d-1} \to (0, \infty)$ such that $L_t h_t = \rho(L_t) h_t$ and $\langle h_t, \nu_t \rangle = 1$.*

(5) *The operator L_t has a spectral gap on $C^\varepsilon(\mathbb{RP}^{d-1})$. That is to say, there exists decomposition of L_t as $L_t = \rho(L_t)(P_t + R_t)$ where $\rho(R_t) < 1$, $P_t R_t = R_t P_t = 0$ and*

$$P_t f = \langle f, \nu_t \rangle h_t \quad \text{for all } f \in C^\varepsilon(\mathbb{RP}^{d-1}).$$

Proof. If we take the measure on $GL_d(\mathbb{R})$ to be $\mu = (1/M) \sum_{i=0}^{M-1} \delta_{A_i}$ then the operator called P^t in [**16**] is a scalar multiple of L_t and the result follows from [**16**, Theorem 8.8]. That h_t is $\bar{t}$-Hölder is [**16**, Lemma 4.8]. □

COROLLARY 3.6. *Under the assumptions of Theorem 3.5 there exist a constant $C > 0$ and β with $0 < \beta < 1$ such that for any $f \in C^\varepsilon(\mathbb{RP}^{d-1})$ we have*

$$\|\rho(L_t)^{-n} L_t^n f - \langle f, \nu_t \rangle h_t\|_\varepsilon \leq C \|f\|_\varepsilon \beta^n$$

for all $n \geq 0$.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Proof. Notice that $\rho(L_t)^{-n}L_t^{-n} = P_t + R_t^n$. Thus

$$\|\rho(L_t)^{-n}L_t^n f - \langle f, \nu_t\rangle h_t\|_\varepsilon = \|R_t^n f\|_\varepsilon \le \|R_t^n\|_{\varepsilon,\mathrm{op}}\|f\|_\varepsilon.$$

Taking $\beta = \rho(R_t) + \eta < 1$ for a small $\eta > 0$, we have the result. □

In order to obtain decay of correlation results we are thus forced into controlling the regularity of $L_J h_t$. This is the content of the next lemma.

LEMMA 3.7.

(1) *For any $A \in GL_d(\mathbb{R})$ we have that*

$$d(\overline{Au}, \overline{Aw}) \le \frac{2\|A\|}{\|A(u/\|u\|)\|} d(\overline{u}, \overline{w})$$

for all $u, w \in \mathbb{R}^d$.

(2) *For any $A \in GL_d(\mathbb{R})$ and $t \ge 0$ we have that*

$$\left| \left\| A\frac{u}{\|u\|} \right\|^t - \left\| A\frac{w}{\|w\|} \right\|^t \right| \le (t+1)\|A\|^t d(\overline{u}, \overline{w})^{\bar{t}}$$

for all $u, w \in \mathbb{R}^d$.

(3) *For any $0 < \varepsilon \le \bar{t}$ there exists a constant K such that $\|L_J h_t\|_\varepsilon \le K\|A_J\|^t$ for all J.*

Proof. (1) This is essentially [**16**, Lemma 4.6]. We provide the details for the sake of completeness. Notice that, for any u, w,

$$\begin{aligned}\|Au\|\|Aw\|\left(\frac{Au}{\|Au\|} - \frac{Aw}{\|Aw\|}\right) &= \|Aw\|Au - \|Au\|Aw \\ &= \|Aw\|Au - \|Aw\|Aw + \|Aw\|Aw - \|Au\|Aw \\ &= \|Aw\|(Au - Aw) + (\|Aw\| - \|Au\|)Aw.\end{aligned}$$

By taking the norm of both sides we have that

$$\|Au\|\|Aw\|\left\|\frac{Au}{\|Au\|} - \frac{Aw}{\|Aw\|}\right\| \le 2\|Aw\|\|A(u-w)\|.$$

Thus

$$\begin{aligned}d(\overline{Au}, \overline{Aw}) &\le \left\| \frac{A(u/\|u\|)}{\|A(u/\|u\|)\|} - \frac{A(w/\|w\|)}{\|A(w/\|w\|)\|} \right\| \\ &\le \frac{2}{\|A(u/\|u\|)\|}\left\| A\left(\frac{u}{\|u\|} - \frac{w}{\|w\|}\right)\right\| \\ &\le \frac{2\|A\|}{\|A(u/\|u\|)\|}\left\| \frac{u}{\|u\|} - \frac{w}{\|w\|} \right\|.\end{aligned}$$

The same argument holds for $\|-u/\|u\| - w/\|w\|\|$. Hence the result.

(2) This is [**16**, Lemma 4.6].

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

(3) Notice that

$$\begin{aligned}
&|L_J h(\overline{u}) - L_J h(\overline{w})| \\
&\quad = \left| \left\| A_J \frac{u}{\|u\|} \right\|^t h_t(\overline{A_J u}) - \left\| A_J \frac{w}{\|w\|} \right\|^t h_t(\overline{A_J w}) \right| \\
&\quad \le \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t(\overline{A_J u}) - h_t(\overline{A_J w})| + \|h_t\|_\infty \left| \left\| A_J \frac{u}{\|u\|} \right\|^t - \left\| A_J \frac{w}{\|w\|} \right\|^t \right| \\
&\quad \le \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t|_{\bar{t}} d(\overline{A_J u}, \overline{A_J w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\overline{u}, \overline{w})^{\bar{t}} \\
&\quad \le \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t|_{\bar{t}} \left(\frac{2\|A_J\|}{\|A_J(u/\|u\|)\|} \right)^{\bar{t}} d(\overline{u}, \overline{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\overline{u}, \overline{w})^{\bar{t}} \\
&\quad = \left\| A_J \frac{u}{\|u\|} \right\|^{t-\bar{t}} \|A_J\|^{\bar{t}} |h_t|_{\bar{t}} 2^{\bar{t}} d(\overline{u}, \overline{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\overline{u}, \overline{w})^{\bar{t}} \\
&\quad \le \|A_J\|^t |h_t|_{\bar{t}} 2^{\bar{t}} d(\overline{u}, \overline{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\overline{u}, \overline{w})^{\bar{t}} \\
&\quad = [|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1)] \|A_J\|^t d(\overline{u}, \overline{w})^{\bar{t}}.
\end{aligned}$$

Thus for $0 < \varepsilon \le \bar{t}$ we have

$$|L_J h_t|_\varepsilon \le 2^{\bar{t}-\varepsilon} |L_J h_t|_{\bar{t}} \le \|A_J\|^t 2^{\bar{t}-\varepsilon} [|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1)].$$

Therefore

$$\begin{aligned}
\|L_J h_t\|_\varepsilon &= \|L_J h_t\|_\infty + |L_J h_t|_\varepsilon \\
&\le \|A_J\|^t (\|h_t\|_\infty + 2^{\bar{t}-\varepsilon} (|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1))).
\end{aligned}$$

□

The proof of Theorem 3.3 then follows in exactly the same way as Theorem 2.4.

Proof of Theorem 3.3. Notice that

$$\begin{aligned}
&|\mu_{\mathcal{A},t}([J] \cap \sigma^{-n-|J|}[I]) - \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J])| \\
&\quad = \left| \sum_{|K|=n} \mu_{\mathcal{A},t}([JKI]) - \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J]) \right| \\
&\quad = \left| \sum_{|K|=n} \rho(L)^{-(n+|I|+|J|)} \langle L_I L_K L_J h_t, \nu_t \rangle - \rho(L)^{-(|I|+|J|)} \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle \right| \\
&\quad = \rho(L)^{-(|I|+|J|)} \left| \left\langle L_I \left(\rho(L)^{-n} \sum_{|K|=n} L_K \right) L_J h_t, \nu_t \right\rangle - \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle \right| \\
&\quad = \rho(L)^{-(|I|+|J|)} |\langle L_I \rho(L)^{-n} L^n L_J h_t, \nu_t \rangle - \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle| \quad \text{by (6)} \\
&\quad = \rho(L)^{-(|I|+|J|)} |\langle L_I (\rho(L)^{-n} L^n L_J h_t - \langle L_J h_t, \nu_t \rangle h_t), \nu_t \rangle| \\
&\quad \le \rho(L)^{-(|I|+|J|)} \|L_I\|_{\infty,\mathrm{op}} \|\rho(L)^{-n} L^n L_J h_t - \langle L_J h_t, \nu_t \rangle h_t\|_\infty \\
&\quad \le \rho(L)^{-(|I|+|J|)} \|L_I\|_{\infty,\mathrm{op}} \|L_J h_t\|_\varepsilon \beta^n \quad \text{by Corollary 3.6} \\
&\quad \le K \rho(L)^{-(|I|+|J|)} \|A_I\|^t \|A_J\|^t \beta^n \quad \text{by Lemma 3.7} \\
&\quad \le C^2 K \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J]) \beta^n \quad \text{by Proposition 3.4.}
\end{aligned}$$

This proves Theorem 3.3(1), and (2) follows by an approximation argument as in Bowen's book [**3**, Theorem 1.26]. □

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Recently, in addition to the interest in Gibbs states associated with the norms of matrices, there has been significant interest in the so-called singular value potential [**1**, **10**]. One can associate a suitable transfer operator to this potential. It seems likely that the method presented in this chapter could be extended to give decay of correlations results for Gibbs states of the singular value potential (in particular, taking advantage of [**16**, Theorem 8.10]). In addition, it seems likely this method could be particularly well suited to studying Gibbs states when $t < 0$. From the perspective of thermodynamic formalism it is likely that these measures for $t < 0$ are significantly more interesting; for example, it is known that the pressure function can fail to be analytic [**11**] and thus one expects that these systems can exhibit phase transitions. We leave this for future work.

4. *The weak Bernoulli property*

The purpose of this section is to prove Theorem 1.3. The proof is similar to [**26**] where scalar potentials satisfying the Bowen property are considered. The key tool is a result of Bradley on ψ-mixing sequences of random variables [**6**] which implies the following lemma.

LEMMA 4.1. *Let μ be a shift-invariant measure on $\Sigma^{\mathbb{Z}}$. Suppose that for some $N > 0$ there exists a constant $C > 0$ such that*

$$C^{-1}\mu([I])\mu([J]) \le \mu([I] \cap \sigma^{-N-|J|}[J]) \le C\mu([I])\mu([J]) \tag{9}$$

for all words I, J. Then μ is weak Bernoulli.

Proof sketch. Notice that for $n \ge N$ we have that

$$\begin{aligned}\mu([I] \cap \sigma^{-n-|J|}[J]) &= \sum_{|K|=n-N} \mu([I] \cap \sigma^{-N-|K|-|J|}[KJ]) \\ &\ge C^{-1} \sum_{|K|=n-N} \mu([I])\mu([KJ]) \\ &= C^{-1}\mu([I]) \sum_{|K|=n-N} \mu([KJ]) \\ &= C^{-1}\mu([I])\mu([J]).\end{aligned}$$

A similar argument for the other inequality shows that in fact (9) holds with the same constant C for all $n \ge N$. Thus we have by an approximation argument that

$$\limsup_{n\to\infty} \mu(X \cap \sigma^{-n}Y) \le C\mu(X)\mu(Y)$$

and

$$\liminf_{n\to\infty} \mu(X \cap \sigma^{-n}Y) \ge C^{-1}\mu(X)\mu(Y)$$

for all X, Y Borel measurable. The second inequality gives that μ is totally ergodic and the first then implies that μ is mixing by a theorem of Ornstein [**23**, Theorem 2.1]. By an approximation argument we have that

$$\psi_n^* = \sup\left\{\tfrac{\mu(A\cap B)}{\mu(A)\mu(B)} : A \in \textstyle\bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P},\ B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P},\ \mu(A)\mu(B) > 0\right\} \le C,$$

$$\psi_n' = \inf\left\{\tfrac{\mu(A\cap B)}{\mu(A)\mu(B)} : A \in \textstyle\bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P},\ B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P},\ \mu(A)\mu(B) > 0\right\} \ge C^{-1}$$

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

for all $n \geq N$. A result of Bradley [**6**, Theorem 1] implies that μ is ψ-mixing; that ψ-mixing implies weak Bernoulli is trivial. □

The lemma is essentially a rephrasing of [**5**, Theorem 4.1(2)]. With this lemma in hand the proof of Theorem 1.3 is merely an application of the Gibbs inequality.

Proof of Theorem 1.3. Let N be as in the definition of 'primitive'. Let $t > 1$ and take q such that $1/t + 1/q = 1$. Then for any I, J,

$$\begin{aligned}
\mu_{\mathcal{A},t}([I] \cap \sigma^{-N-|J|}[J]) &= \sum_{|K|=N} \mu_{\mathcal{A},t}([IKJ]) \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \sum_{|K|=N} \|A_I A_K A_J\|^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} M^{-Nt/q} \left(\sum_{|K|=N} \|A_I A_K A_J\| \right)^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} M^{-Nt/q} \delta^t \|A_I\|^t \|A_J\|^t \\
&\geq C^{-2} e^{-NP(\mathcal{A},t)} M^{-Nt/q} \delta^t \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J])
\end{aligned}$$

where $M = |\Sigma|$. For $0 < t \leq 1$ we have that

$$\begin{aligned}
\mu_{\mathcal{A},t}([I] \cap \sigma^{-N-|J|}[J]) &= \sum_{|K|=N} \mu_{\mathcal{A},t}([IKJ]) \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \sum_{|K|=N} \|A_I A_K A_J\|^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \left(\sum_{|K|=N} \|A_I A_K A_J\| \right)^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \delta^t \|A_I\|^t \|A_J\|^t \\
&\geq C^{-2} e^{-NP(\mathcal{A},t)} \delta^t \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J]).
\end{aligned}$$

For matrix Gibbs states the right-hand inequality in (9) always holds, as was noticed in [**20**]. The right-hand inequality is a simple consequence of the Gibbs inequality and the fact that the norm is sub-multiplicative. The result then follows from Lemma 4.1. □

Acknowledgements. I am grateful to Anthony Quas for suggesting this problem as well as many helpful discussions and comments on drafts of this manuscript. I am also grateful to Chris Bose for helpful discussions and comments on drafts of this manuscript. I am grateful to the referee for a careful reading of the paper, providing helpful suggestions and for pointing me to [**16**] which significantly improved the contents of §3.

A. *Appendix.*

Here we collect some facts as well as the basic definitions and properties of cones in finite-dimensional vector spaces. Most of material on cones will be familiar from the classical Perron–Frobenius theory for non-negative matrices. In addition, we collect some elementary propositions and lemmas which we use in the article.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Definition A.1. A subset $K \subseteq \mathbb{R}^d$ is called a cone if:
(1) $K \cap (-K) = \{0\}$;
(2) $\lambda K = K$ for all $\lambda > 0$;
(3) K is convex.

If K is a cone then we define the dual of K by

$$K^* := \{w : \langle u, w\rangle \geq 0 \text{ for all } u \in K\}.$$

Definition A.2. Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be a linear map and K be a cone.
- We say that A is K-non-negative provided $AK \subseteq K$ and we write $A \geq^K 0$.
- We say that A is K-positive if $A(K\backslash\{0\}) \subseteq \text{int}(K)$ and we write $A >^K 0$.
- We say $A \geq^K 0$ is K-primitive if there exists an N such that A^N is K-positive.
- We say $A \geq^K 0$ is K-irreducible if $\sum_{k=0}^{d-1} A^k$ is K-positive.

Often when K is understood we suppress the notation.

There are various definitions of K-irreducibility in the literature. It is known that these conditions are all equivalent, though finding a complete proof in the literature is surprising difficult. Thus for the sake of completeness we include Proposition A.4. In order to show that part (3) of the proposition implies part (4), we need the following lemma.

LEMMA A.3. *Suppose that $K \subseteq \mathbb{R}^d$ is a cone. Then*

$$\text{int}(K) = \{u : \langle u, v\rangle > 0 \textit{ for all } v \in K^*\backslash\{0\}\}.$$

Proof. First we recall that

$$K = \{u : \langle u, v\rangle \geq 0 \text{ for all } v \in K^* \text{ with } \|v\| = 1\}.$$

This is a very general fact for closed cones in Banach spaces which follows from a suitable Hahn–Banach theorem; see [**9**] for a nice discussion. Let

$$u \in \{u : \langle u, v\rangle > 0 \text{ for all } v \in K^*\backslash\{0\}\}.$$

Now take $\delta > 0$ such that $\langle u, v\rangle > \delta$ for all $v \in K^*$ with $\|v\| = 1$. Suppose that $\|w - u\| < \delta/2$. Then for any $v \in K^*$ with $\|v\| = 1$ we have

$$|\langle u, v\rangle - \langle w, v\rangle| = |\langle u - w, v\rangle| \leq \|w - u\| < \delta/2.$$

This implies that $\langle w, v\rangle \geq \delta/2 > 0$ and thus $w \in K$. As $B(u, \delta/2) \subseteq K$, we have that $u \in \text{int}(K)$.

Now suppose that $u \in \text{int}(K)$. Take $\delta > 0$ such that $B(u, \delta) \subseteq K$. If $\|w\| = \delta/2$ then

$$\|(u + w) - u\| = \|w\| < \delta \quad \text{and} \quad \|(u - w) - u\| = \|w\| < \delta$$

implies that $u - w, u + w \in K$. Thus for any $v \in K^*$,

$$0 \leq \langle u + w, v\rangle = \langle u, v\rangle + \langle w, v\rangle \quad \text{and} \quad 0 \leq \langle u - w, v\rangle = \langle u, v\rangle - \langle w, v\rangle$$

which implies

$$-\langle w, v\rangle \leq \langle u, v\rangle \leq \langle w, v\rangle.$$

Then we have

$$\|v\| = 2/\delta \sup_{\|w\|=\delta/2} |\langle w, v\rangle| \leq 2/\delta \langle u, v\rangle.$$

In particular, if $v \neq 0$ then $\langle u, v\rangle > 0$. □

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

PROPOSITION A.4. *Suppose that A preserves a non-void cone $K \subseteq \mathbb{R}^d$. The following are equivalent.*

(1) *A has no eigenvector contained in ∂K.*
(2) *A has no invariant faces.*
(3) *$(I + A)^{d-1}$ is K-positive.*
(4) *$\sum_{k=0}^{d-1} A^k$ is K-positive.*

Proof. (1 $\iff$ 2) is [**25**, Theorem 4.2].

(2 $\implies$ 3) is [**25**, Lemma 4.2].

(4 $\implies$ 1) is clear, for if A has an eigenvector contained in ∂K then so does $\sum_{k=0}^{d-1} A^k$.

(3 $\implies$ 4) Suppose that $u \in K\backslash\{0\}$. By the assumption that $(I + A)^{d-1} >^K 0$ we have that for any $v \in K^*\backslash\{0\}$,

$$0 < \langle (I + A)^{d-1} u, v\rangle = \sum_{k=0}^{d-1} \binom{d-1}{k} \langle A^k u, v\rangle$$

by Lemma A.3. This implies that

$$0 < \sum_{k=0}^{d-1} \langle A^k u, v\rangle = \left\langle \left(\sum_{k=0}^{d-1} A^k \right) u, v \right\rangle,$$

and hence $\sum_{k=0}^{d-1} A^k u \in \operatorname{int}(K)$ by Lemma A.3 and $\sum_{k=0}^{d-1} A^k$ is K-positive. □

The Perron–Frobenius theorem holds for abstract finite-dimensional cones just as it does for the positive quadrant.

THEOREM A.5. *Suppose that K is a closed cone with non-void interior.*

(1) *If A is K-non-negative then:*
   (a) *$\rho(A)$ is an eigenvalue.*
   (b) *K contains an eigenvector corresponding to $\rho(A)$.*
(2) *If A is K-irreducible then:*
   (a) *$\rho(A)$ is a simple eigenvalue, and any other eigenvalue with the same modulus is simple.*
   (b) *Suppose that u is an eigenvector for A corresponding to $\rho(A)$ and v is an eigenvector of A^T corresponding to $\rho(A)$ normalized so that $\langle u, v\rangle = 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \rho(A)^{-k} A^k = P$$

   where $Pw = \langle w, v\rangle u$.
(3) *If A is K-primitive then:*
   (a) *$\rho(A)$ is a simple eigenvalue, which is greater in modulus then any other eigenvalue.*
   (b) *Suppose that u is an eigenvector for A corresponding to $\rho(A)$ and v is an eigenvector of A^T corresponding to $\rho(A)$ normalized so that $\langle u, v\rangle = 1$. Then for all small $\varepsilon > 0$ there exists $C > 0$ such that for all $n \geq 0$,*

$$\|\rho(A)^{-n} A^n - P\| \leq C \left(\frac{|\lambda_2| + \varepsilon}{\rho(A)} \right)^n$$

   where $Pw = \langle w, v\rangle u$.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Proof.
(1) This is [**25**, Theorem 3.1].
(2) Part (a) can be found in [**25**, Theorem 4.3]. Part (b) follows from (a) using the same proof as for non-negative matrices, which can be found in [**17**, Theorem 8.6.1].
(3) This result is well known. It can be proved, for example, using Hilbert's projective metric and holds in significant generality; see, for example, [**9**, Theorem 2.3] (although the result in [**9**] is significantly more powerful than needed here). The article [**25**] contains a proof this result when A is assumed K-positive. □

It is clear from the definition of 'irreducible' and 'primitive' that the eigenvector corresponding to $\rho(A)$ is contained in the interior of the cone K. This agrees with the fact from classical Perron–Frobenius theory that the eigenvector has all positive entries. Notice also that for K-irreducible/primitive matrices there always exist vectors u and v with u an eigenvector for A corresponding to $\rho(A)$ and v an eigenvector for A^* corresponding to $\rho(A)$ such that $\langle u, v\rangle = 1$ (by (1)(b) and the observation that $u \in \operatorname{int}(K)$ by irreducibility/primitivity). This ensures that (2)(b) and (3)(b) are never vacuous. We need the following lemma to produce the Gibbs inequality.

LEMMA A.6. *Suppose that K is a cone and that $D \subset \operatorname{int}(K)$ and $D^* \subset \operatorname{int}(K^*)$ are non-empty and compact. Then there exists a constant $C > 0$ such that*

$$C^{-1}\|A\| \le \langle Au, v\rangle \le C\|A\|$$

for all $u \in D$, $v \in D^*$ and $A \ge^K 0$.

Proof. Suppose $A \ge^K 0$, and that for some $u \in \operatorname{int}(K)$ and $v \in \operatorname{int}(K^*)$ we have $\langle Au, v\rangle = 0$. Then $Au = 0$ by an arguement similar to Lemma A.3. Thus for any $w \in K^*$ we have that $\langle u, A^*w\rangle = 0$, which implies that $A^*w = 0$ by Lemma A.3. Therefore $A^* = 0$ and of course $A = 0$. Thus the function

$$(A, u, v) \mapsto \langle Au, v\rangle$$

is continuous and $\langle Au, v\rangle > 0$ for all $A \ge^K 0$, $A \ne 0$, and $u \in D$, $v \in D^*$. As the set of norm-one K-non-negative matrices cross $D \times D^*$ is compact, we can find a $C > 0$ such that

$$C^{-1} \le \left\langle \frac{Au}{\|A\|}, v \right\rangle \le C$$

for all $A \ge^K 0$, $A \ne 0$, and $u \in D$, $v \in D^*$. Clearly the inequality holds for $A = 0$, hence we have the result. □

With the preceding lemma the proof of the following proposition which relates the definition of irreducibility and primitivity from the introduction to that for operators is straightforward.

PROPOSITION A.7. *Let $\mathcal{A} = (A_0, \ldots, A_{M-1}) \in M_d(\mathbb{R})$ and define $L_{\mathcal{A}}$ as in Example 2.3.*
(1) *If $L_{\mathcal{A}}$ is irreducible then $\mathcal{A}$ satisfies equation (3).*
(2) *If $L_{\mathcal{A}}$ is primitive then $\mathcal{A}$ satisfies equation (4).*

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Proof. We will prove (2); then (1) will be similar. Let L_i be as in Example 2.3. Take N such that $L_{\mathcal{A}}^N >^K 0$ and U, V positive definite matrices. Set $D = \{U\}$ and $D^* = (L_{\mathcal{A}}^*)^N(\{W \in K^* : \langle U, W\rangle_{\mathrm{HS}} = 1\})$. Notice that $\{W \in K^* : \langle U, W\rangle_{\mathrm{HS}} = 1\}$ is closed and bounded (by Lemma A.6), hence compact. Take $C > 0$ as in Lemma A.6. Then

$$\begin{aligned}\sum_{|K|=N} \|A_I A_K A_J\|^2 &\geq C^{-1}\langle L_I L_{\mathcal{A}}^N L_J U, V\rangle_{\mathrm{HS}} \\ &= C^{-1}\langle L_I U, V\rangle_{\mathrm{HS}}\left\langle L_J U, (L_{\mathcal{A}}^*)^N \frac{L_I^* V}{\langle L_I U, V\rangle}\right\rangle_{\mathrm{HS}} \\ &\geq C^{-3}\|A_I\|^2\|A_J\|^2,\end{aligned}$$

where we have used the fact that $\|L_I\| = \|A_I\|^2$. Therefore

$$\sum_{|K|=N} \|A_I A_K A_J\| \geq \left(\sum_{|K|=N} \|A_I A_K A_J\|^2\right)^{1/2} \geq C^{-3/2}\|A_I\|\|A_J\|. \qquad \square$$

PROPOSITION A.8. *Let k be an even number and define*

$$S = \operatorname{span}\{v^{\otimes k} : v \in \mathbb{R}^d\}$$

and

$$K = \{w \in S^* : \langle v^{\otimes k}, w\rangle_{(\mathbb{R}^d)^{\otimes k}} \geq 0 \text{ for all } v \in \mathbb{R}^d\}.$$

Then K is a closed cone with non-void interior.

Proof. That K is a closed cone is trivial. Thus we turn our attention to showing that K has a non-void interior. First we note that there exist elements $w \in K$ such that $\langle v^{\otimes k}, w\rangle > 0$ for all $v \in \mathbb{R}^d\backslash\{0\}$. For example, define a multi-linear map $f : (\mathbb{R}^d)^k \to \mathbb{R}$ by

$$f(v^1, v^2, \ldots, v^k) = \sum_{i=1}^d v_i^1 v_i^2 \cdots v_i^k.$$

This gives a linear map $f : (\mathbb{R}^d)^{\otimes k} \to \mathbb{R}$ such that

$$f(v^{\otimes k}) = \sum_{i=1}^d v_i^k > 0.$$

Now if $v^n \xrightarrow{n\to\infty} w$ then

$$\begin{aligned}(v^n)^{\otimes k} &= \sum_{i_1\cdots i_k} v_{i_1}^n \cdots v_{i_k}^n e_{i_1} \otimes \cdots \otimes e_{i_k} \\ &\xrightarrow{n\to\infty} \sum_{i_1\cdots i_k} w_{i_1} \cdots w_{i_k} e_{i_1} \otimes \cdots \otimes e_{i_k} \\ &= w^{\otimes k}.\end{aligned}$$

Thus $\{v^{\otimes k} : \|v^{\otimes k}\| = 1\}$ is compact. Take $\delta > 0$ such that $f(v^{\otimes k}) > \delta$ for all v for which $\|v^{\otimes k}\| = 1$. Now if $g \in S^*$ is such that $\|f - g\| < \delta/2$ then $g \in K$. Hence $\operatorname{int}(K) \neq \emptyset$. $\square$

PROPOSITION A.9. *Suppose that $(A_0, \ldots, A_{M-1})$ is irreducible. Then $(A_0^*, \ldots, A_{M-1}^*)$ is also irreducible.*

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

Proof. Notice that if $A_i^* W \subseteq W$ then $A_i W^\perp \subseteq W^\perp$. To see this, consider that for any $u \in W^\perp$ and $w \in W$ we have

$$0 = \langle A_i^* w, u \rangle = \langle w, A_i u \rangle,$$

which implies that $A_i u \in W^\perp$. If $A_i W \subseteq W$ for all $0 \le i \le M-1$ then $W^\perp = \{0\}$ or $\mathbb{R}^d$, hence $W = \{0\}$ or $\mathbb{R}^d$. □

References

[1] J. Bochi and I. D. Morris. Equilibrium states of generalised singular value potentials and applications to affine iterated function systems. *Geom. Funct. Anal.* **28**(4) (2018), 995–1028.
[2] R. Bowen. Bernoulli equilibrium states for Axiom A diffeomorphisms. *Math. Syst. Theory* **8**(4) (1974/75), 289–294.
[3] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms (Lecture Notes in Mathematics, 470)*, Revised edn. Ed. J.-R. Chazottes. Springer, Berlin, 2008, with a preface by David Ruelle.
[4] M. Boyle and K. Petersen. Hidden Markov processes in the context of symbolic dynamics. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems (London Mathematical Society Lecture Note Series, 385)*. Cambridge University Press, Cambridge, 2011, pp. 5–71.
[5] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** (2005), 107–144. Update of, and a supplement to, the 1986 original.
[6] R. C. Bradley Jr. On the ψ-mixing condition for stationary random sequences. *Trans. Amer. Math. Soc.* **276**(1) (1983), 55–66.
[7] Y.-L. Cao, D.-J. Feng and W. Huang. The thermodynamic formalism for sub-additive potentials. *Discrete Contin. Dyn. Syst.* **20**(3) (2008), 639–657.
[8] J.-R. Chazottes and E. Ugalde. Projection of Markov measures may be Gibbsian. *J. Stat. Phys.* **111**(5–6) (2003), 1245–1272.
[9] S. P. Eveson and R. D. Nussbaum. Applications of the Birkhoff–Hopf theorem to the spectral theory of positive linear operators. *Math. Proc. Cambridge Philos. Soc.* **117**(3) (1995), 491–512.
[10] K. J. Falconer. The Hausdorff dimension of self-affine fractals. *Math. Proc. Cambridge Philos. Soc.* **103**(2) (1988), 339–350.
[11] D.-J. Feng. The limited Rademacher functions and Bernoulli convolutions associated with Pisot numbers. *Adv. Math.* **195**(1) (2005), 24–101.
[12] D.-J. Feng. Equilibrium states for factor maps between subshifts. *Adv. Math.* **226**(3) (2011), 2470–2502.
[13] D.-J. Feng and A. Käenmäki. Equilibrium states of the pressure function for products of matrices. *Discrete Contin. Dyn. Syst.* **30**(3) (2011), 699–708.
[14] D.-J. Feng and K.-S. Lau. The pressure function for products of non-negative matrices. *Math. Res. Lett.* **9**(2–3) (2002), 363–378.
[15] N. A. Friedman and D. S. Ornstein. On isomorphism of weak Bernoulli transformations. *Adv. Math.* **5**(1970) (1970), 365–394.
[16] Y. Guivarc'h and E. Le Page. Simplicité de spectres de Lyapounov et propriété d'isolation spectrale pour une famille d'opérateurs de transfert sur l'espace projectif. *Random Walks and Geometry*. Walter de Gruyter, Berlin, 2004, pp. 181–259.
[17] R. A. Horn and C. R. Johnson. *Matrix Analysis*, 2nd edn. Cambridge University Press, Cambridge, 2013.
[18] A. Johansson, A. Öberg and M. Pollicott. Ergodic theory of Kusuoka measures. *J. Fract. Geom.* **4**(2) (2017), 185–214.
[19] S. Kusuoka. Dirichlet forms on fractals and products of random matrices. *Publ. Res. Inst. Math. Sci.* **25**(4) (1989), 659–680.
[20] I. D. Morris. Ergodic properties of matrix equilibrium states. *Ergod. Th. & Dynam. Sys.* **38**(6) (2018), 2295–2320.
[21] I. D. Morris. A necessary and sufficient condition for a matrix equilibrium state to be mixing. *Ergod. Th. & Dynam. Sys.*, to appear.
[22] D. S. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.* **4** (1970), 337–352.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press

[**23**] D. S. Ornstein. On the root problem in ergodic theory. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability Theory*. Eds. L. M. LeCam, J. Neyman and E. L. Scott. University of California Press, Berkeley, CA, 1972, pp. 347–356.
[**24**] M. Piraino. Projections of Gibbs states for Hölder potentials. *J. Stat. Phys.* **170**(5) (2018), 952–961.
[**25**] J. S. Vandergraft. Spectral properties of matrices which have invariant cones. *SIAM J. Appl. Math.* **16** (1968), 1208–1222.
[**26**] P. Walters. Regularity conditions and Bernoulli properties of equilibrium states and g-measures. *J. Lond. Math. Soc. (2)* **71**(2) (2005), 379–396.
[**27**] J. Yoo. On factor maps that send Markov measures to Gibbs measures. *J. Stat. Phys.* **141**(6) (2010), 1055–1070.

https://doi.org/10.1017/etds.2018.129 Published online by Cambridge University Press