A RAPIDLY-CONVERGING LOWER BOUND FOR THE JOINT SPECTRAL RADIUS VIA MULTIPLICATIVE ERGODIC THEORY

IAN D. MORRIS

ABSTRACT. We use ergodic theory to prove a quantitative version of a theorem of M. A. Berger and Y. Wang, which relates the joint spectral radius of a set of matrices to the spectral radii of finite products of those matrices. The proof rests on a structure theorem for continuous matrix cocycles over minimal homeomorphisms having the property that all forward products are uniformly bounded. MSC primary 15A18, 37H15, 65F15, secondary 37M25.

1. INTRODUCTION

Let A be a bounded nonempty set of $d \times d$ complex matrices. The joint spectral radius of A, introduced by G.-C. Rota and G. Strang in [44], is defined to be the quantity

$$\varrho(\mathsf{A}) := \lim_{n\to\infty} \sup \left\{ \|A_n \cdots A_1\|^{1/n} \colon A_i \in \mathsf{A} \right\}, \tag{1}$$

where $\|\cdot\|$ denotes any norm on $\mathbb{C}^d$. This limit exists (by a simple subadditivity argument described below) and yields a finite value which is independent of the choice of norm. The joint spectral radius arises naturally in a range of topics including control and stability [2, 26, 32], coding theory [38], wavelet regularity [16, 17, 37], numerical solutions to ordinary differential equations [25], and combinatorics [18]. The problem of computing the joint spectral radius of a finite set of matrices has therefore attracted substantial research interest [5, 23, 24, 32, 35, 36, 42, 47, 48]. In this article we shall prove a new estimate relevant to the computation of the joint spectral radius.

Let $\mathrm{Mat}_d(\mathbb{C})$ denote the set of all $d \times d$ complex matrices. The following theorem was proved by M. A. Berger and Y. Wang [4], having originally been conjectured by I. Daubechies and J. C. Lagarias [16]:

Theorem 1.1 (Berger-Wang formula). *Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be a bounded nonempty set. Then*

$$\varrho(\mathsf{A}) = \limsup_{n\to\infty} \sup \left\{ \rho(A_n \cdots A_1)^{1/n} \colon A_i \in \mathsf{A} \right\}, \tag{2}$$

where $\rho(A)$ denotes the ordinary spectral radius of a matrix A.

Some alternative proofs are given in [7, 19, 46]. In this article we shall study the rate of convergence in the expression (2). This has potential implications for some approaches to the computation of the joint spectral radius such as the algorithm given by G. Gripenberg [23].

Let $\|\cdot\|$ be any norm on $\mathbb{C}^d$. For each $n \in \mathbb{N}^+$ define

$$\varrho_n^+(\mathsf{A}, \|\cdot\|) = \sup\left\{\|A_n \cdots A_1\|^{1/n} \colon A_i \in \mathsf{A}\right\},$$

$$\varrho_n^-(\mathsf{A}) = \sup\left\{\rho(A_n \cdots A_1)^{1/n} \colon A_i \in \mathsf{A}\right\}.$$

For fixed A it is clear that $\varrho_{n+m}^+(\mathsf{A}, \|\cdot\|)^{n+m} \leq \varrho_n^+(\mathsf{A}, \|\cdot\|)^n \varrho_m^+(\mathsf{A}, \|\cdot\|)^m$ for all $n, m \in \mathbb{N}^+$, which implies via Fekete's subadditivity lemma that the limit in (1) exists and is equal to the infimum over all $n \in \mathbb{N}^+$ of the same quantity. Conversely, since $\rho(A^m)^{1/m} = \rho(A)$ for all $m \in \mathbb{N}^+$ and any matrix A, one may easily show that $\varrho_{nm}^-(\mathsf{A}) \geq \varrho_n^-(\mathsf{A})$ for every $n, m \in \mathbb{N}^+$ and hence the limit superior in (2) is also a supremum. In general this limit superior can fail to be a limit: G. Gripenberg [23] notes the example

$$\mathsf{A} = \left\{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right\}$$

for which it is easily seen that $\varrho_n^-(\mathsf{A})$ is equal to 0 or 1 according as n is odd or even. In this article we shall present a proof of the following theorem, which extends Theorem 1.1 in the case where A is finite:

Theorem 1.2. *Let A be a nonempty finite set of $d \times d$ complex matrices. Then for every positive real number r,*

$$\varrho(\mathsf{A}) - \max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A}) = O\left(\frac{1}{n^r}\right).$$

Theorem 1.2 implies in particular that if we wish to compute $\varrho(\mathsf{A})$ to within accuracy ε by means of brute-force estimation of the values $\varrho_n^-(\mathsf{A})$, then the number of matrix products which must be evaluated increases at a slower-than-stretched-exponential rate as a function of $1/\varepsilon$. However, it should be noted that the arguments used in this paper do not seem to be well-suited to the production of an *effective* estimate for the quantity $\varrho(\mathsf{A})$.

Two estimates related to Theorem 1.2 have been established previously. By a theorem of J. Bochi [7], there exist for each $d \in \mathbb{N}^+$ a constant $C_d \geq 1$ and an integer $m \in \mathbb{N}^+$ such that $\varrho(\mathsf{A}) \leq C_d \max_{1 \leq k \leq m} \varrho_k^-(\mathsf{A})$ for every nonempty bounded set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$. An easy consequence is the estimate

$$0 \leq \varrho(\mathsf{A}) - \max_{1 \leq k \leq mn} \varrho_k^-(\mathsf{A}) \leq \left(1 - C_d^{-1/n}\right) \varrho(\mathsf{A}) = O\left(\frac{1}{n}\right).$$

In the other direction, F. Wirth [48] gives the bound

$$\varrho_n^+(\mathsf{A}, \|\cdot\|) - \varrho(\mathsf{A}) = O\left(\frac{1}{n}\right)$$

for any norm $\|\cdot\|$ on $\mathbb{C}^d$ and nonempty bounded set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$, provided that there does not exist a linear space V such that $\{0\} \subset V \subset \mathbb{C}^d$ and $AV \subseteq V$ for every $A \in \mathsf{A}$. When such a subspace V exists this estimate is weakened to $O(\log n/n)$. Unlike Bochi's estimate, the constant in Wirth's estimate may vary between sets of matrices A. The example

$$\mathsf{A} = \left\{\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right\}$$

shows that Wirth's estimate is sharp at least for certain choices of norm: taking $\|\cdot\|$ to be the Euclidean norm we obtain $\varrho_n^+(\mathsf{A}, \|\cdot\|) = 2^{1+1/2n}$ for each $n \in \mathbb{N}^+$, whereas $\varrho_1^-(\mathsf{A}) = 2$ and hence $\varrho(\mathsf{A}) = 2$. Remarkably, for certain A it is possible to choose a norm $\vert\!\vert\!\vert \cdot \vert\!\vert\!\vert$ such that $\varrho_n^+(\mathsf{A}, \vert\!\vert\!\vert \cdot \vert\!\vert\!\vert) = \varrho(\mathsf{A})^n$ for every $n \in \mathbb{N}^+$, but giving an explicit description of such a norm (when it exists) is a separate computational problem which does not appear to be easier than that of calculating $\varrho(\mathsf{A})$ in the first place. A discussion of this topic may be found in [33].

The proof of Theorem 1.2 has some points of resemblance to the proof of Theorem 1.1 given by L. Elsner [19] in the case in which A is finite, which we now elaborate upon. Elsner's proof runs essentially as follows. If $\varrho(\mathsf{A}) = 0$ then the result is trivially true. Otherwise, by normalising we may take $\varrho(\mathsf{A}) = 1$. We then reduce to the case where a uniform bound exists for products of elements of A, and hence there exists a compact subset of $\mathrm{Mat}_d(\mathbb{C})$ which contains $\{A_n \ldots A_1 \colon A_i \in \mathsf{A}\}$ for every n. By using the pigeonhole principle on open ε-balls in $\mathrm{Mat}_d(\mathbb{C})$ and in $\mathbb{C}^d$, we can then guarantee the existence of a finite sequence $A_1, \ldots, A_n$ and a vector v belonging to the unit sphere of $\mathbb{C}^d$ such that $A_n \cdots A_1 v$ is close to v and therefore the spectral radius of $A_n \cdots A_1$ is close to 1.

In our proof of Theorem 1.2 we make this strategy quantitative, replacing the pigeonhole principle with a more delicate recurrence argument. In order to achieve this we first prove a theorem describing the dynamical structure of matrix sequences (A_i) with the property that $\|A_n \cdots A_1\|$ is large for all n, and additionally we achieve some understanding of the structure of the orbits in $\mathbb{C}^d$ which are induced by the action of such sequences. The bulk of this paper, therefore, is concerned with proving a theorem on the dynamical structure of these 'extremal' sequences. We describe these ideas in detail in the following section.

2. Linear cocycles

At this point it is convenient to establish some notation and definitions. In the remainder of this article the symbol $\|\cdot\|$ shall be used to denote the Euclidean norm on $\mathbb{C}^d$, whereas the symbol $\vert\!\vert\!\vert \cdot \vert\!\vert\!\vert$ shall be used to denote an *extremal norm* on $\mathbb{C}^d$, which will be defined shortly. In either case we shall also use the symbols $\|\cdot\|$ and $\vert\!\vert\!\vert \cdot \vert\!\vert\!\vert$ to denote the corresponding operator norms induced on $\mathrm{Mat}_d(\mathbb{C})$. Throughout this article we adhere to the convention $\log 0 := -\infty$.

Let $T \colon X \to X$ be a continuous transformation of a nonempty compact topological space. A *cocycle* over T with values in the complex matrices is a function $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ such that for each $x \in X$ and $n, m \in \mathbb{N}$

$$\mathcal{A}(x, n+m) = \mathcal{A}(T^n x, m)\mathcal{A}(x, n),$$
$$\mathcal{A}(x, 0) = I.$$

We say that the cocycle $\mathcal{A}$ is continuous if $\mathcal{A}(\cdot, n)$ is a continuous function from X to $\mathrm{Mat}_d(\mathbb{C})$ for each $n \in \mathbb{N}$. Abusing notation somewhat, we shall sometimes denote $\mathcal{A}(x, 1)$ simply by $\mathcal{A}(x)$. Since for each x, n

$$\mathcal{A}(x, n) = \mathcal{A}(T^{n-1}x) \cdots \mathcal{A}(Tx)\mathcal{A}(x)$$

the cocycle $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ is completely determined by the function $\mathcal{A} \colon X \to \mathrm{Mat}_d(\mathbb{C})$. Whilst it will always be the case in this article that the map T is a homeomorphism, we do not assume that the values of the function $\mathcal{A}$ are always invertible matrices, so in general $\mathcal{A}$ will not admit an extension to a cocycle with domain $X \times \mathbb{Z}$.

For $0 \leq p \leq d$ we let $\mathrm{Gr}(p, d)$ denote the set of all p-dimensional subspaces of $\mathbb{C}^d$. For $V \in \mathrm{Gr}(p, d)$ we let $P_V^\perp$ denote the linear map given by orthogonal projection onto V, and for $u \in \mathbb{C}^d$ we shall write $\mathfrak{d}(u,V) := \inf\{\|u-v\| : v \in V\}$. We equip $\mathrm{Gr}(p,d)$ with the metric given by

$$d_{\mathrm{Gr}}(V,W) := \|P_V^\perp - P_W^\perp\| = \max \left[\sup_{\substack{v\in V\\ \|v\|=1}} \mathfrak{d}(v,W), \sup_{\substack{w\in W\\ \|w\|=1}} \mathfrak{d}(w,V) \right] \tag{3}$$

with respect to which $\mathrm{Gr}(p.d)$ is a compact metric space. A proof of the equality of the two expressions for $d_{\mathrm{Gr}}(V,W)$ given above may be found in [1]. We shall say that a function $\mathcal{V} \colon X \to \mathrm{Gr}(p,d)$ is *invariant* under a cocycle $\mathcal{A}$ if $\mathcal{A}(x)\mathcal{V}(x) \subseteq \mathcal{V}(Tx)$ for all $x \in X$, in which case $\mathcal{A}(x,n)\mathcal{V}(x) \subseteq \mathcal{V}(T^n x)$ for every $x \in X$ and $n \in \mathbb{N}$.

We begin by establishing the following general theorem which will later be applied to study matrix cocycles associated to nonempty compact sets $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$.

Theorem 2.1. *Let $T \colon X \to X$ be a minimal homeomorphism of a nonempty compact topological space, and let $\mathcal{A} \colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ be a continuous linear cocycle. Suppose that there exists $M \geq 1$ such that $\|\mathcal{A}(x,n)\| \leq M$ for all $x \in X$ and all $n \in \mathbb{N}$. Then there exist an integer $0 \leq p \leq d$ and continuous invariant functions $\mathcal{V} \colon X \to \mathrm{Gr}(p,d)$, $\mathcal{W} \colon X \to \mathrm{Gr}(d-p,d)$ such that $\mathcal{V}(x) \oplus \mathcal{W}(x) = \mathbb{C}^d$ for all $x \in X$, and the following additional properties hold. There exist constants $C > 1$ and $\delta, \xi \in (0,1)$ such that for all $x \in X$ and $n \in \mathbb{N}$, $\|\mathcal{A}(x,n)v\| \geq \delta\|v\|$ for every $v \in \mathcal{V}(x)$ and $\|\mathcal{A}(x,n)w\| \leq C\xi^n\|w\|$ for every $w \in \mathcal{W}(x)$. The continuity of the functions $\mathcal{V}$ and $\mathcal{W}$ admits the following quantitative description: there exists a constant $K_1 > 1$ such that for any $x,y \in X$ and $n \in \mathbb{N}$,*

$$d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) \leq K_1 \left(\xi^n + \|\mathcal{A}(T^{-n}x,n) - \mathcal{A}(T^{-n}y,n)\|\right)$$

and

$$d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y)) \leq K_1 \left(\xi^n + \|\mathcal{A}(x,n) - \mathcal{A}(y,n)\|\right).$$

For each $x \in X$ let $P(x) \in \mathrm{Mat}_d(\mathbb{C})$ denote the unique projection with image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$. Then $P(x)$ depends continuously on x, and in particular there exists $K_2 > 1$ such that

$$\|P(x) - P(y)\| \leq K_2\left[d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y))\right]$$

for all $x,y \in X$.

While Theorem 2.1 has a number of features in common with the classical multiplicative ergodic theorem of V. I. Oseledec (see e.g. [34]) our proof is direct and does not make use of any prior multiplicative ergodic theorems. Indeed, since in general we wish to work with non-invertible matrices, the standard statement of Oseledec's theorem does not give the existence even of a measurable splitting of the type given above, giving only an invariant flag (though see [20]). The proof of Theorem 2.1 does however incorporate ideas used in the proofs of Oseledec's theorem given by M. S. Raghunathan [43] and D. Ruelle [45]. For some previous results on the continuity of invariant splittings see [3, Appendix A] and [50].

In order to apply Theorem 2.1 to the study of the joint spectral radius we require some further definitions. We shall say that $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ is *product bounded* if there exists $M > 1$ such that for every $n \in \mathbb{N}^+$ we have $\|A_n \cdots A_1\| \leq M$ for every finite sequence $(A_n, \ldots, A_1) \in \mathsf{A}^n$. Note if such a uniform bound holds for A with respect to some norm on $\mathbb{C}^d$ then it holds with respect to all such norms, subject to variation in the constant M. We shall say that a norm $|\!|\!|\cdot|\!|\!|$ on $\mathbb{C}^d$ is an *extremal norm* for A

if $\|\!|A|\!\| \leq \varrho(\mathsf{A})$ for all $A \in \mathsf{A}$. If $\varrho(\mathsf{A}) > 0$ then an extremal norm exists for A if and only if $\varrho(\mathsf{A})^{-1}\mathsf{A}$ is product bounded [32, 44].

Given a nonempty compact set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$, let us define a metric on $\mathsf{A}^{\mathbb{Z}}$ by

$$d\left[(A_i)_{i\in\mathbb{Z}}, (B_i)_{i\in\mathbb{Z}}\right] := \sum_{i\in\mathbb{Z}} \frac{\|A_i - B_i\|}{2^{|i|}}.$$

If A is compact then $(\mathsf{A}^{\mathbb{Z}}, d)$ is compact. We define the *shift map* $T\colon \mathsf{A}^{\mathbb{Z}} \to \mathsf{A}^{\mathbb{Z}}$ by $T[(A_i)_{i\in\mathbb{Z}}] = (A_{i+1})_{i\in\mathbb{Z}}$. The shift map is a Lipschitz homeomorphism of $\mathsf{A}^{\mathbb{Z}}$. Let $\mathcal{A}\colon \mathsf{A}^{\mathbb{Z}} \to \mathrm{Mat}_d(\mathbb{C})$ be given by $\mathcal{A}[(A_i)_{i\in\mathbb{Z}}] = A_0$, and let $\mathcal{A}(x,n) = \mathcal{A}(T^{n-1}x)\cdots\mathcal{A}(x)$ for all $(x,n) \in \mathsf{A}^{\mathbb{Z}} \times \mathbb{N}$ so that $\mathcal{A}\colon \mathsf{A}^{\mathbb{Z}} \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ is a continuous cocycle. The cocycle $\mathcal{A}$ gives us a framework with which to study Theorem 1.2 using the tools of multiplicative ergodic theory: for each $n \in \mathbb{N}^+$ we have

$$\varrho_n^+(\mathsf{A}, \|\cdot\|) = \sup\left\{\|\mathcal{A}(x,n)\|^{1/n} \colon x \in \mathsf{A}^{\mathbb{Z}}\right\}$$

and

$$\varrho_n^-(\mathsf{A}) = \sup\left\{\rho(\mathcal{A}(x,n))^{1/n} \colon x \in \mathsf{A}^{\mathbb{Z}}\right\}.$$

It follows in particular that the joint spectral radius of A admits the description

$$\log \varrho(\mathsf{A}) = \lim_{n\to\infty} \sup_{x\in\mathsf{A}^{\mathbb{Z}}} \frac{1}{n} \log \|\mathcal{A}(x,n)\|. \tag{4}$$

By applying Theorem 2.1 in this context we will derive the following:

Theorem 2.2. *Let $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ be a nonempty compact set such that $\varrho(\mathsf{A}) = 1$, and suppose that A is product bounded. Let $\|\!|\cdot|\!\|$ be any extremal norm for A and define*

$$Y := \left\{x \in \mathsf{A}^{\mathbb{Z}} \colon \|\!|\mathcal{A}(x,n)|\!\| = 1\ \forall\ n \in \mathbb{N}\right\}.$$

Then the set Y is a compact, nonempty subset of $\mathsf{A}^{\mathbb{Z}}$ such that $TY \subseteq Y$.

Let $Z \subseteq Y$ be any nonempty T-invariant subset such that $T\colon Z \to Z$ is minimal. Then there exists an integer $1 \leq p \leq d$ such that the following properties hold. There exist Hölder continuous invariant functions $\mathcal{V}\colon Z \to \mathrm{Gr}(p,d)$, $\mathcal{W}\colon Z \to \mathrm{Gr}(d-p,d)$ such that $\mathcal{V}(x) \oplus \mathcal{W}(x) = \mathbb{C}^d$ for each $x \in Z$. There exist constants $C > 1$, $\xi \in (0,1)$ such that for all $x \in Z$ and $n \in \mathbb{N}$, $\|\!|\mathcal{A}(x,n)v|\!\| = \|\!|v|\!\|$ for all $v \in \mathcal{V}(x)$ and $\|\!|\mathcal{A}(x,n)w|\!\| \leq C\xi^n \|\!|w|\!\|$ for all $w \in \mathcal{W}(x)$. If for each $x \in Z$ we let $P(x)$ denote the projection with image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$ then $P\colon Z \to \mathrm{Mat}_d(\mathbb{C})$ is Hölder continuous.

To obtain Theorem 1.2 we combine this result with an estimate due to X. Bressaud and A. Quas on the approximation via periodic orbits of closed invariant subsets of shift transformations over finite alphabets (cf. [12]).

We remark that the recently-developed research topic of *ergodic optimisation* has been quite influential on the development of the present article (although the only result in that field which we use directly is that of Bressaud and Quas). Ergodic optimisation is concerned with the following problem: given a continuous dynamical system $T\colon X \to X$ defined on a compact metric space, and a continuous (or only upper semi-continuous) function $f\colon X \to \mathbb{R}$, one studies the greatest possible linear growth rate of the sequence $\sum_{j=0}^{n-1} f(T^j x)$ as x varies over X, which is equal to the supremum of all possible values of the integral of f with respect to a T-invariant probability measure on X. Problems which are considered include the identification and approximation of this maximal growth rate and of those orbits which attain it. Some recent research articles in this area include [8, 9, 11, 12, 14, 29, 40, 51]. The expression (4) shows that the joint spectral radius of a set of matrices A can be interpreted as the maximal linear growth rate of a sequence of observations of a dynamical system in a somewhat related manner. A characterisation of $\varrho(\mathsf{A})$ in terms of integrals with respect to the ergodic measures of the system $(\mathsf{A}^{\mathbb{Z}}, T)$ is described in the article [41], although we do not make use of it here.

An ergodic optimisation approach to the study of the joint spectral radius was previously explored by T. Bousch and J. Mairesse in the article [10]. Indeed, the possibility of applying ergodic optimisation to the joint spectral radius was suggested at an early stage in the unpublished manuscript [15]. However, to the best of the author's knowledge, the article of Bousch and Mairesse is the only article prior to the present work in which ergodic methods have been applied to study the joint spectral radius.

The remainder of this article is structured as follows. In sections 3, 4 and 5 we respectively give the proofs of Theorem 2.1, Theorem 2.2 and Theorem 1.2. In section 6 we describe the obstructions to improving the error term in Theorem 1.2 and to extending that theorem to the case of infinite compact sets A.

3. Proof of Theorem 2.1

Throughout this section we assume that X, T, $\mathcal{A}$, M etc. are as given in the statement of Theorem 2.1. Before commencing the proof proper we require two preliminary results. The first is of a dynamical nature, whereas the second is concerned with the metric structure of the Grassmannian spaces $\mathrm{Gr}(p,d)$.

Recall that a sequence $(a_n)_{n=1}^{\infty}$ of elements of $\mathbb{R}\cup\{-\infty\}$ is called *subadditive* if $a_{n+m} \leq a_n + a_m$ for every $n,m \in \mathbb{N}^+$. If (a_n) is subadditive then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}^+} \frac{a_n}{n} \in \mathbb{R}\cup\{-\infty\}.$$

For the purposes of this section, we define a *subadditive function sequence* to be a sequence $(f_n)_{n=1}^{\infty}$ of upper semi-continuous functions from X to $\mathbb{R}\cup\{-\infty\}$ such that the relation

$$f_{n+m}(x) \leq f_m(T^n x) + f_n(x)$$

is satisfied for every $x \in X$ and $n,m \in \mathbb{N}^+$. Proposition 3.1, which deals with subadditive function sequences, is key to our application of the hypothesis $\|\mathcal{A}(x,n)\| \leq M$. Some features of this proposition are similar to [39, Theorem 1].

Proposition 3.1. *Let (f_n) be a subadditive function sequence, and suppose that $C := \sup_{x,n} f_n(x)$ is finite. Then one of the following two cases holds: either $|f_n(x)| \leq C$ for every $x \in X$ and $n \in \mathbb{N}^+$, or $\lim_{n\to\infty} \frac{1}{n}\sup_{x\in X} f_n(x) < 0$.*

Proof. The sequence $(a_n)_{n=1}^{\infty}$ defined by $a_n = \sup_x f_n(x)$ is subadditive, and therefore $\lim_{n\to\infty} a_n/n = \inf_{n\in\mathbb{N}^+} a_n/n \in \mathbb{R}\cup\{-\infty\}$. Let us suppose that the first of the two cases described above does not hold; to prove that the second case holds it suffices to exhibit $N \in \mathbb{N}^+$ such that $a_N < 0$.

Choose $z \in X$ and $m, \varepsilon > 0$ such that $f_m(z) < -(C+\varepsilon)$. By semi-continuity there exists an open set U containing z such that $f_m(x) < -(C+\varepsilon)$ for every $x \in U$. Since T is minimal, there exists $p \in \mathbb{N}^+$ such that $X = \bigcup_{i=0}^{p-1} T^{-i}U$ (see for example [21, p.28]). It follows that for every $x \in X$ we may find an integer $r(x)$ such that $f_{r(x)}(x) < -\varepsilon$ and $m \leq r(x) < m+p$, since necessarily $T^i x \in U$ for some $i \in \{0,\ldots,p-1\}$ and therefore $f_{m+i}(x) \leq f_m(T^i x) + f_i(x) < -(C+\varepsilon) + C =$

$-\varepsilon$. Let us define $R(x,1) = r(x)$ for each $x \in X$, and define $R(\cdot,k)\colon X \to \mathbb{N}^+$ inductively by $R(x,k+1) := R(x,k) + r(T^{R(x,k)})$ for each $k \geq 1$. Clearly for each $x \in X$ and $k \geq 1$ we have $m \leq R(x,k+1) - R(x,k) < m+p$ and therefore $R(x,k) < k(m+p)$, and a simple induction argument shows that $f_{R(x,k)}(x) < -k\varepsilon$. Let $N := (m+p)\lceil C\varepsilon^{-1} + 2\rceil$. For each $x \in X$ we may choose an integer $k(x)$ with the property that $0 < N - R(x,k(x)) \leq m+p$, which in particular implies that $k(x)(m+p) > R(x,k(x)) \geq N - m - p$ and therefore $k(x) \geq N(m+p)^{-1} - 1$. Define also $q(x) := N - R(x,k(x))$. For each $x \in X$ we may now estimate

$$f_N(x) \leq f_{q(x)}(T^{R(x,k(x))}x) + f_{R(x,k(x))}(x) < C - k(x)\varepsilon \leq C + \varepsilon - N\varepsilon(m+p)^{-1} \leq -\varepsilon.$$

In particular $a_N < 0$ and the second case of the proposition has been established. □

The following useful result concerning the metric d_{Gr} does not appear to be widely known:

Lemma 3.2. *Let $V, W \in \mathrm{Gr}(p,d)$ where $1 \leq p \leq d$. Then,*

$$d_{\mathrm{Gr}}(V,W) = \sup_{\substack{v \in V \\ \|v\|=1}} \mathfrak{d}(v,W) = \sup_{\substack{w \in W \\ \|w\|=1}} \mathfrak{d}(w,V).$$

Proof. For conciseness let us define $\partial(V,W) := \sup\{\mathfrak{d}(v,W)\colon v \in V; \|v\| = 1\}$, and define $\partial(W,V)$ analogously by interchanging V and W. For every $V, W \in \mathrm{Gr}(p,d)$ we have $\max\{\partial(V,W), \partial(W,V)\} = d_{\mathrm{Gr}}(V,W) \leq 1$, see for example [1, §34] or [31, p.56]. If $\min\{\partial(V,W), \partial(W,V)\} = 1$ then we are done. If this is not the case then by symmetry we may suppose without loss of generality that $\partial(V,W) < 1$. Since

$$\|(I - P_W^\perp)P_V^\perp\| = \sup_{\substack{v \in \mathbb{C}^d \\ \|v\|=1}} \mathfrak{d}(Pv,W) = \sup_{\substack{v \in V \\ \|v\|=1}} \mathfrak{d}(v,W) = \partial(V,W)$$

and V is by hypothesis isomorphic to W, Theorem I.6.34 in Kato's book [31] shows that $\partial(V,W) = \partial(W,V)$ as desired. □

Remark. The reader may note that the hypothesis in Lemma 3.2 that V and W have equal, finite dimension plays a crucial rôle: for example, if V and W are subspaces of $\mathbb{C}^d$ of unequal dimension such that V is properly contained in W, then it is not difficult to show that $\mathfrak{d}(v,W) = 0$ for all $v \in V$ whilst $\max\{\mathfrak{d}(w,V)\colon w \in W; \|w\| = 1\} = 1$ (see e.g. [1, p.70]). Similarly, it is not difficult to see that if V and W are closed subspaces of an infinite-dimensional Hilbert space, then the same outcome may arise even if V and W are isomorphic.

Before proceeding further with the proof of Theorem 2.1 we require some notation and definitions from linear algebra. For each $B \in \mathrm{Mat}_d(\mathbb{C})$ let us write $|B| := \sqrt{B^*B}$, and for $1 \leq i \leq d$ let $\sigma_1(B) \geq \ldots \geq \sigma_d(B)$ denote the singular values of B, which are defined to be the eigenvalues of $|B|$ listed in decreasing order, allowing repetitions if multiplicities occur. Clearly $0 \leq \sigma_i(B) \leq \|B\|$ for every i. The values $\sigma_i(B)$ depend continuously on $B \in \mathrm{Mat}_d(\mathbb{C})$, and if $A, B \in \mathrm{Mat}_d(\mathbb{C})$ then for $1 \leq \ell \leq d$,

$$\prod_{i=1}^{\ell} \sigma_i(AB) \leq \left(\prod_{i=1}^{\ell} \sigma_i(A)\right)\left(\prod_{i=1}^{\ell} \sigma_i(B)\right). \tag{5}$$

This property follows naturally from the fact that $\sigma_1(B)\sigma_2(B)\cdots\sigma_\ell(B)$ is the operator norm of $\wedge^\ell B$ relative to the Hermitian norm on $\wedge^\ell(\mathbb{C}^d)$ induced by the standard

Hermitian form on $\mathbb{C}^d$; see Ruelle [45, p.43]. An alternative proof which does not make use of multilinear algebra may be found in [22].

For each $x \in X$, $n \in \mathbb{N}^+$ and $1 \leq \ell \leq d$ let us define $f_n^\ell(x) = \sum_{i=1}^{\ell} \log \sigma_i(\mathcal{A}(x,n))$. It follows from (5) that for each ℓ, (f_n^ℓ) is a subadditive function sequence. Moreover, since for all $x \in X$, $n \in \mathbb{N}^+$ and $1 < \ell \leq d$ we have

$$\sum_{i=1}^{\ell} \log \sigma_i(\mathcal{A}(x,n)) \leq \sum_{i=1}^{\ell-1} \log \sigma_i(\mathcal{A}(x,n)) + \log M \leq \ell \log M, \tag{6}$$

Proposition 3.1 implies that each of the limits

$$\theta_\ell := \lim_{n\to\infty} \sup_{x \in X} \frac{1}{n} \sum_{i=1}^{\ell} \log \sigma_i(\mathcal{A}(x,n))$$

exists. Clearly (6) also implies that $\theta_{\ell+1} \leq \theta_\ell \leq 0$ for $1 \leq \ell < d$. If $\theta_1 < 0$ then the conclusions of Theorem 2.1 hold with $p = 0$, $\mathcal{V}(x) \equiv \{0\}$, $\mathcal{W}(x) \equiv \mathbb{C}^d$ and $\xi := e^{\frac{1}{2}\theta_1}$, so for the remainder of the proof we shall assume that $\theta_1 = 0$. Take $p \in \mathbb{N}^+$ such that $\theta_\ell = 0$ for $1 \leq \ell \leq p$ and $\theta_\ell < 0$ for $p < \ell \leq d$. By Proposition 3.1, for ℓ in the range $1 \leq \ell \leq p$ we have

$$\sup_{n\in\mathbb{N}^+} \sup_{x \in X} \left| \sum_{i=1}^{\ell} \log \sigma_i(\mathcal{A}(x,n)) \right| \leq \ell \log M.$$

We deduce from this that there is $\delta_0 \in (0,1)$ such that

$$\min_{1\leq i\leq p} \inf_{x\in X} \inf_{n\geq 1} \sigma_i(\mathcal{A}(x,n)) \geq \delta_0. \tag{7}$$

Since $\theta_i < 0$ for $p < i \leq d$ we similarly deduce that there exist $C_0 > 1$, $\xi \in (0,1)$ such that for each $n \in \mathbb{N}$

$$\max_{p<i\leq d} \sup_{x\in X} \sigma_i(\mathcal{A}(x,n)) \leq C_0 \xi^n. \tag{8}$$

Given $x \in X$ and $n \in \mathbb{N}^+$, let $U_n^+(x) \in \mathrm{Gr}(p,d)$ be the vector space spanned by those eigenvectors of $|\mathcal{A}(x,n)|$ which correspond to the eigenvalues $\sigma_1(\mathcal{A}(x,n))$ up to $\sigma_p(\mathcal{A}(x,n))$ and let $U_n^-(x) \in \mathrm{Gr}(d-p,d)$ be the space spanned by those eigenvectors associated to the remaining eigenspaces. If v is an eigenvector of $|\mathcal{A}(x,n)|$ with eigenvalue $\sigma_i(A(x,n))$ then

$$\|\mathcal{A}(x,n)v\|^2 = \langle \mathcal{A}(x,n)v, \mathcal{A}(x,n)v\rangle = \langle \mathcal{A}(x,n)^*\mathcal{A}(x,n)v, v\rangle = \sigma_i(\mathcal{A}(x,n))^2\|v\|^2. \tag{9}$$

Since $|\mathcal{A}(x,n)|$ is a normal matrix there exists an orthonormal basis for $\mathbb{C}^d$ consisting of its eigenvectors. In particular $U_n^+(x)$ is orthogonal to $U_n^-(x)$, and using (9) we may derive

$$\inf \left\{ \|\mathcal{A}(x,n)v\| \colon v \in U_n^+(x) \text{ and } \|v\| = 1 \right\} \geq \delta_0,$$
$$\sup \left\{ \|\mathcal{A}(x,n)v\| \colon v \in U_n^-(x) \text{ and } \|v\| = 1 \right\} \leq C_0 \xi^n$$

for all $x \in X$ and $n \in \mathbb{N}^+$.

We now proceed to construct the function $\mathcal{W}$. Let $x \in X$, $m \in \mathbb{N}^+$ and $u \in \mathbb{C}^d$. Writing $u = u_m^+ + u_m^-$ with $u_m^+ \in U_m^+(x)$ and $u_m^- \in U_m^-(x)$, we have

$$\begin{aligned} \|\mathcal{A}(x,m)u\| &\geq \|\mathcal{A}(x,m)u_m^+\| - \|\mathcal{A}(x,m)u_m^-\| \\ &\geq \delta_0\|u_m^+\| - C_0\xi^m\|u_m^-\| \geq \delta_0.\mathfrak{d}(u, U_m^-(x)) - C_0\xi^m\|u\| \end{aligned} \tag{10}$$

since $\|u_m^+\| = \mathfrak{d}(u, U_m^-(x))$ and $\|u_m^-\| \leq \|u\|$. Now, if $m \geq n \geq 1$ then we have

$$\|\mathcal{A}(x,m)v\| = \|\mathcal{A}(T^n x, m-n)\mathcal{A}(x,n)\| \leq M\|\mathcal{A}(x,n)\|$$

and so we obtain

$$\delta_0.\mathfrak{d}(u, U_m^-(x)) \leq C_0\xi^m + M\|\mathcal{A}(x,n)u\| \tag{11}$$

whenever $x \in X$, $u \in \mathbb{C}^d$ and $m \geq n \geq 1$. Using Lemma 3.2 we may choose $u \in U_n^-(x)$ such that $\mathfrak{d}(u, U_m^-(x)) = d_{\mathrm{Gr}}(U_n^-(x), U_m^-(x))$ and in this case (11) yields:

$$\delta_0 d_{\mathrm{Gr}}(U_n^-(x), U_m^-(x)) = \delta_0.\mathfrak{d}(u, U_m^-(x)) \leq C_0\xi^m + C_0 M\xi^n \leq C_0(M+1)\xi^n. \tag{12}$$

We deduce that for each $x \in X$ the sequence of subspaces $U_n^-(x)$ is a Cauchy sequence in $\mathrm{Gr}(d-p,d)$. Let us define $\mathcal{W}\colon X \to \mathrm{Gr}(d-p,d)$ by setting $\mathcal{W}(x) := \lim_{n\to\infty} U_n^-(x)$. Since this convergence is uniform and the spaces $U_n^\pm(x)$ depend continuously on x, the continuity of $\mathcal{W}$ follows.

We next derive two key inequalities describing the action of $\mathcal{A}(x,n)$ on $\mathcal{W}(x)$ and on $\mathbb{C}^d$. Firstly we assert that given any $x \in X$, $n \in \mathbb{N}$ and $w \in \mathcal{W}(x)$,

$$\|\mathcal{A}(x,n)w\| \leq C\xi^n\|w\| \tag{13}$$

for some constant $C > 1$ as described in the statement of the theorem. For $n = 0$ this is clear. Given $x \in X$ and $n \in \mathbb{N}^+$, note that $d_{\mathrm{Gr}}(\mathcal{W}(x), U_n^-(x)) \leq \delta_0^{-1}C_0M\xi^n$ by taking the limit $m \to \infty$ in (12). Writing $w \in \mathcal{W}(x)$ in the form $w = w_n^+ + w_n^-$ with $w_n^\pm \in U_n^\pm(x)$ and applying Lemma 3.2, we obtain

$$\|w_n^+\| = \mathfrak{d}(w, U_n^-(x)) \leq \|w\|.d_{\mathrm{Gr}}(\mathcal{W}(x), U_n^-(x)) \leq \delta_0^{-1}C_0M\xi^n\|w\|$$

and therefore

$$\begin{aligned}\|\mathcal{A}(x,n)w\| &\leq \|\mathcal{A}(x,n)w_n^+\| + \|\mathcal{A}(x,n)w_n^-\| \\ &\leq \delta_0^{-1}C_0M^2\xi^n\|w\| + C_0\xi^n\|w\| \leq C\xi^n\|w\|\end{aligned}$$

as required, where $C := C_0(\delta_0^{-1}M^2 + 1) > 1$. Secondly we make the following observation: if $x \in X$, $n \in \mathbb{N}$ and $u \in \mathbb{C}^d$, then

$$\|\mathcal{A}(x,n)u\| \geq \delta_1.\mathfrak{d}(u, \mathcal{W}(x)) \tag{14}$$

where $\delta_1 \in (0,1)$ is constant. Indeed, the case $n = 0$ being trivial, this follows simply by taking the limit $m \to \infty$ in (11) and defining $\delta_1 := M^{-1}\delta_0 \in (0,1)$.

The invariance of $\mathcal{W}$ under $\mathcal{A}$ now follows easily. Given $w \in \mathcal{W}(x)$ we must show that $\mathfrak{d}(\mathcal{A}(x)w, \mathcal{W}(Tx)) = 0$. Let $n \in \mathbb{N}^+$ be arbitrary; combining (13) and (14) yields

$$\delta_1.\mathfrak{d}(\mathcal{A}(x)w, \mathcal{W}(Tx)) \leq \|\mathcal{A}(Tx,n)\mathcal{A}(x)w\| = \|\mathcal{A}(x,n+1)w\| \leq C\xi^{n+1}\|w\|,$$

and since n may be taken arbitrarily large this implies the desired result. The continuity estimate for $\mathcal{W}$ is also simple to establish: given $x, y \in X$ and $n \in \mathbb{N}$, applying Lemma 3.2 we may choose $w \in \mathcal{W}(x)$ such that $d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) = \mathfrak{d}(w, \mathcal{W}(y))$ and $\|w\| = 1$. Via (13) and (14) this entails

$$\begin{aligned}\delta_1.d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \leq \|\mathcal{A}(y,n)w\| &\leq \|\mathcal{A}(x,n)w\| + \|(\mathcal{A}(x,n) - \mathcal{A}(y,n))w\| \\ &\leq C\xi^n + \|\mathcal{A}(x,n) - \mathcal{A}(y,n)\|\end{aligned}$$

so that we may take $K_1 := \delta_1^{-1}C > 1$.

We next turn to the construction of $\mathcal{V}$. To achieve this we make the following observations: since $\mathcal{A}$ is a cocycle with respect to the transformation T, it follows directly that the function $\mathcal{A}^*\colon X \times \mathbb{N} \to \mathrm{Mat}_d(\mathbb{C})$ defined by $\mathcal{A}^*(x,n) :=$

$[\mathcal{A}(T^{-n}x,n)]^*$ is a cocycle with respect to the transformation T^{-1}, which of course is also minimal. Moreover, since the singular values of a matrix are equal to the singular values of its conjugate transpose, the inequalities (7) and (8) must also apply to the cocycle $\mathcal{A}^*$ with identical constants C_0, δ_0, ξ and p. By repeating the construction of $\mathcal{W}$ given above for the cocycle $\mathcal{A}^*$, it follows that there exists a continuous function $\mathcal{Y}\colon X \to \mathrm{Gr}(d-p,d)$, invariant under $\mathcal{A}^*$, with the following properties: given $x, y \in X$, $n \in \mathbb{N}$ and $v \in \mathcal{Y}(x)$,

$$\|\mathcal{A}^*(x,n)v\| \leq C\xi^n\|v\| \tag{15}$$

and

$$d_{\mathrm{Gr}}(\mathcal{Y}(x),\mathcal{Y}(y)) \leq K_1(\xi^n + \|\mathcal{A}^*(x,n)-\mathcal{A}^*(y,n)\|). \tag{16}$$

Now, if U, V are linear subspaces of $\mathbb{C}^d$, and $B \in \mathrm{Mat}_d(\mathbb{C})$, then $BU \subseteq V$ if and only if $B^*V^\perp \subseteq U^\perp$, since both properties are equivalent to the statement that $\langle Bu, v\rangle = 0$ whenever $u \in U$ and $v \in V^\perp$. Let us therefore define $\mathcal{V}\colon X \to \mathrm{Gr}(p,d)$ by $\mathcal{V}(x) := \mathcal{Y}(x)^\perp$ for all $x \in X$. The invariance relation $\mathcal{A}^*(x)\mathcal{Y}(x) \subseteq \mathcal{Y}(T^{-1}x)$ now yields $\mathcal{A}(T^{-1}x)\mathcal{V}(T^{-1}x) \subseteq \mathcal{V}(x)$ and hence $\mathcal{V}$ is invariant under the cocycle $\mathcal{A}$. Moreover using (16) it is clear that $\mathcal{V}$ must satisfy the continuity estimate

$$d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) \leq K_1(\xi^n + \|\mathcal{A}(T^{-n}x,n)-\mathcal{A}(T^{-n}y,n)\|)$$

required for Theorem 2.1.

We shall find it useful to have an alternative characterisation of $\mathcal{V}$. For $x \in X$ and $n \in \mathbb{N}$ define $\mathcal{U}_n(x) := \mathcal{A}(T^{-n}x,n)\left[\mathcal{W}(T^{-n}x)^\perp\right]$. We claim that $\mathcal{V}(x) = \lim_{n\to\infty}\mathcal{U}_n(x)$ for each $x \in X$. Indeed, given x and n, applying Lemma 3.2 we may choose $u \in \mathcal{U}_n(x)$ such that $d_{\mathrm{Gr}}(\mathcal{U}_n(x),\mathcal{V}(x)) = \mathfrak{d}(u,\mathcal{V}(x))$ and $\|u\| = 1$. Let us write $u = \mathcal{A}(T^{-n}x,n)v$ where $v \in \mathcal{W}(T^{-n}x)^\perp$. By (14) we have $\delta_1\|v\| \leq \|u\| = 1$, and making use of (15) we obtain

$$\begin{aligned} d_{\mathrm{Gr}}(\mathcal{U}_n(x),\mathcal{V}(x)) &= \mathfrak{d}(u,\mathcal{Y}(x)^\perp) = \sup_{\substack{w\in\mathcal{Y}(x)\\ \|w\|=1}} \langle w,u\rangle = \sup_{\substack{w\in\mathcal{Y}(x)\\ \|w\|=1}} \left\langle w, \mathcal{A}(T^{-n}x,n)v\right\rangle \\ &= \sup_{\substack{w\in\mathcal{Y}(x)\\ \|w\|=1}} \langle \mathcal{A}^*(x,n)w, v\rangle \leq \sup_{\substack{w\in\mathcal{Y}(x)\\ \|w\|=1}} \|\mathcal{A}^*(x,n)w\|.\|v\| \leq \delta_1^{-1}C\xi^n, \end{aligned}$$

which suffices to prove the claim.

Now, given $x \in X$ suppose that $u \in \mathcal{U}_n(x)$ for some $n \in \mathbb{N}$, and let $u = \mathcal{A}(T^{-n}x,n)v$ with $v \in \mathcal{W}(T^{-n}x)^\perp$. Writing $u = w_1 + w_2$ with $w_1 \in \mathcal{W}(x)$ and $w_2 \in \mathcal{W}(x)^\perp$, and applying (13), we obtain

$$\begin{aligned} \|\mathcal{A}(x,m)u\| &\leq \|\mathcal{A}(x,m)w_1\| + \|\mathcal{A}(x,m)w_2\| \\ &\leq C\xi^m\|w_1\| + M\|w_2\| \leq C\xi^m\|u\| + M.\mathfrak{d}(u,\mathcal{W}(x)) \end{aligned}$$

for all $m \in \mathbb{N}$. On the other hand, from (14) we arrive at the inequality

$$\|\mathcal{A}(x,m)u\| = \|\mathcal{A}(x,m)\mathcal{A}(T^{-n}x,n)v\| = \|\mathcal{A}(T^{-n}x,n+m)v\| \geq \delta_1\|v\| \geq \delta_1 M^{-1}\|u\|$$

since $v \in \mathcal{W}(T^{-n}x)^\perp$ and $\|u\| = \|\mathcal{A}(T^{-n}x,n)v\| \leq M\|v\|$. Combining the above inequalities yields

$$\delta_1 M^{-1}\|u\| \leq C\xi^m\|u\| + M\mathfrak{d}(u,\mathcal{W}(x))$$

whenever $u \in \bigcup_{n=0}^\infty \mathcal{U}_n(x)$ and $m \in \mathbb{N}$. Since $\mathcal{V}(x) = \lim_{n\to\infty}\mathcal{U}_n(x)$ and m may be taken arbitrarily large, we deduce that for every $v \in \mathcal{V}(x)$

$$\mathfrak{d}(v,\mathcal{W}(x)) \geq \delta_1 M^{-2}\|v\|. \tag{17}$$

Combining this inequality with (14) it follows that for $x \in X$, $n \in \mathbb{N}$ and $v \in \mathcal{V}(x)$ we have $\|\mathcal{A}(x,n)v\| \geq \delta\|v\|$ as required for Theorem 2.1, where $\delta := \delta_1^2 M^{-2} \in (0,1)$. The inequality (17) also shows that for every $x \in X$ we have $\mathcal{V}(x) \cap \mathcal{W}(x) = \{0\}$ and therefore $\mathcal{V}(x) \oplus \mathcal{W}(x) = \mathbb{C}^d$ as required.

For each $x \in X$ let $P(x)$ denote the projection having image $\mathcal{V}(x)$ and kernel $\mathcal{W}(x)$. It remains only to prove that $P(x)$ depends continuously on x in the desired manner. Now, for any $x \in X$ and $u \in \mathbb{C}^d$, since $P(x)u \in \mathcal{V}(x)$ and $u - P(x)u \in \mathcal{W}(x)$ we obtain from (17)

$$\|u\| \geq \mathfrak{d}(u, \mathcal{W}(x)) = \mathfrak{d}(P(x)u, \mathcal{W}(x)) \geq \delta_1 M^{-2}\|P(x)u\|$$

so that $\|P(x)\| \leq \delta_1^{-1}M^2$. Since $P(x)$ fixes every element of $\mathcal{V}(x)$ we also have $\|P(x)\| \geq 1$. We will show that if $x, y \in X$ satisfy

$$3\|P(x)\|.[d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y))] < \frac{1}{2} \tag{18}$$

then

$$\|P(x) - P(y)\| \leq 21\|P(x)\|^3.[d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y))]. \tag{19}$$

The result then follows by taking $K_2 := 21\sup\|P\|^3 \leq 21M^6\delta_1^{-3}$.

For notational convenience we write $Q(x) = I - P(x)$ for all $x \in X$. For $x, y \in X$ define $U(x,y) = P^{\perp}_{\mathcal{V}(y)}P(x) + P^{\perp}_{\mathcal{W}(y)}Q(x)$, where $P^{\perp}_Z$ denotes orthogonal projection onto Z. Clearly $\|U(x,y)\| \leq 2\|P(x)\| + 1 \leq 3\|P(x)\|$. Since $I = P(x) + Q(x) = P^{\perp}_{\mathcal{V}(x)}P(x) + P^{\perp}_{\mathcal{W}(x)}Q(x)$ we have

$$\|U(x,y) - I\| \leq (2\|P(x)\| + 1).\left[d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y))\right]. \tag{20}$$

Now suppose that x, y satisfy (18). Then $U(x,y)$ is invertible and

$$\begin{aligned} \|U(x,y)^{-1} - I\| &\leq \sum_{n=1}^{\infty}\|U(x,y) - I\|^n \\ &\leq 6\|P(x)\|\left[d_{\mathrm{Gr}}(\mathcal{V}(x),\mathcal{V}(y)) + d_{\mathrm{Gr}}(\mathcal{W}(x),\mathcal{W}(y))\right]. \end{aligned} \tag{21}$$

Since for each $v \in \mathcal{V}(x)$ and $w \in \mathcal{W}(x)$ we have

$$U(x,y)P(x)(v+w) = U(x,y)v = P(y)U(x,y)(v+w)$$

it follows that $P(y) = U(x,y)P(x)U(x,y)^{-1}$. Using the estimate

$$\|P(x) - P(y)\| \leq \|(I - U(x,y))P(x)\| + \|U(x,y)P(x)(I - U(x,y)^{-1})\|$$

in combination with (20) and (21) yields (19) and the proof is complete.

4. Proof of Theorem 2.2

Let A and $|\!|\!| \cdot |\!|\!|$ be as in the statement of the theorem, and choose $M \geq 1$ such that $|\!|\!| v |\!|\!| \leq M\|v\| \leq M^2 |\!|\!| v |\!|\!|$ for all $v \in \mathbb{C}^d$. As in the introduction we let $\mathcal{A}\colon \mathsf{A}^{\mathbb{Z}} \to \mathrm{Mat}_d(\mathbb{C})$ be given by projection onto the zeroth co-ordinate, let $T\colon \mathsf{A}^{\mathbb{Z}} \to \mathsf{A}^{\mathbb{Z}}$ be the shift map, and let d to be the metric on $\mathsf{A}^{\mathbb{Z}}$ defined previously. Clearly $\mathcal{A}$ and T are Lipschitz continuous. For each $n \in \mathbb{N}^+$ we have $\left\{\mathcal{A}(x,n)\colon x \in \mathsf{A}^{\mathbb{Z}}\right\} = \{A_n \cdots A_1\colon A_i \in \mathsf{A}\}$ and therefore $\sup\left\{|\!|\!|\mathcal{A}(x,n)|\!|\!|\colon x \in \mathsf{A}^{\mathbb{Z}}\right\} = \varrho_n^+(\mathsf{A}, |\!|\!| \cdot |\!|\!|) = 1$. For each $n \in \mathbb{N}$ let us define $Y_n := \{x \in \mathsf{A}^{\mathbb{Z}}\colon |\!|\!|\mathcal{A}(x,n)|\!|\!| = 1\}$. Clearly each Y_n is compact and nonempty. If $x \in Y_{n+1}$ then since

$$1 = |\!|\!|\mathcal{A}(x,n+1)|\!|\!| = |\!|\!|\mathcal{A}(T^n x)\mathcal{A}(x,n)|\!|\!| \leq |\!|\!|\mathcal{A}(T^n x)|\!|\!|.|\!|\!|\mathcal{A}(x,n)|\!|\!| \leq |\!|\!|\mathcal{A}(x,n)|\!|\!| \leq 1$$

we have $x \in Y_n$ also, and it follows that the compact set $Y := \bigcap_{n=0}^{\infty} Y_n$ must be nonempty. Similarly, $x \in Y_{n+1}$ also implies

$$1 = |||\mathcal{A}(x, n+1)||| \leq |||\mathcal{A}(Tx, n)|||.|||\mathcal{A}(x)||| \leq |||\mathcal{A}(Tx, n)||| \leq 1$$

and hence $Tx \in Y_n$. We deduce that $TY \subseteq Y$ as required. We remark that the construction of Y using the extremal norm $|||\cdot|||$ could be compared to the use of sub-actions in [14] or the "Mañé lemma" in [8].

Let $Z = TZ$ be any nonempty minimal set contained in Y. Note that for all $x \in Z$ and $n \in \mathbb{N}$ we have $\|\mathcal{A}(x, n)\| \leq M^2$ since $|||\mathcal{A}(x, n)||| = 1$. We may therefore apply Theorem 2.1 to the minimal set Z and the cocycle $\mathcal{A}$. If we were to have $p = 0$ then we would have $|||\mathcal{A}(x, n)||| < 1$ for some $x \in Z$ and $n \in \mathbb{N}$, so it must be the case that $p \geq 1$. To prove Theorem 2.2, we must show firstly that the functions $\mathcal{V}, \mathcal{W}$ and P provided by Theorem 2.1 are Hölder continuous, and secondly that for all $x \in Z$ and $n \in \mathbb{N}$ one has $|||\mathcal{A}(x, n)v||| = |||v|||$ for every $v \in \mathcal{V}(x)$.

The proof of Hölder continuity is straightforward. Let K_1, ξ be as given by Theorem 2.1, and let $D = \operatorname{diam} Z$ and $\alpha = \log \xi / (\log \xi - \log 2) \in (0, 1)$. If $D = 0$ then the Hölder continuity of $\mathcal{V}$, $\mathcal{W}$ and P holds trivially, so we assume $D > 0$. Given two distinct points $x, y \in Z$, choose $n \in \mathbb{N}^+$ such that $D(\xi/2)^n < d(x, y) \leq D(\xi/2)^{n-1}$. We have

$$\begin{aligned}
\|\mathcal{A}(x, n) - \mathcal{A}(y, n)\| &\leq \sum_{i=0}^{n-1} \left\|\mathcal{A}\left(T^{i+1}x, n-i-1\right)\left(\mathcal{A}\left(T^i x\right) - \mathcal{A}(T^i y)\right)\mathcal{A}\left(y, i\right)\right\| \\
&\leq M^2 \sum_{i=0}^{n-1} \left\|\mathcal{A}\left(T^i x\right) - \mathcal{A}\left(T^i y\right)\right\| \\
&\leq M^2 2^n \sum_{i \in \mathbb{Z}} \frac{\left\|\mathcal{A}\left(T^i x\right) - \mathcal{A}\left(T^i y\right)\right\|}{2^{|i|}} \\
&= M^2 2^n d(x, y) \leq 2M^2 D \xi^{n-1}
\end{aligned}$$

and therefore

$$d_{\mathrm{Gr}}(\mathcal{W}(x), \mathcal{W}(y)) \leq K_1\left(2M^2 D\xi^{-1} + 1\right)\xi^n \leq K_1 D^{-\alpha}\left(2M^2 D\xi^{-1} + 1\right) d(x, y)^\alpha$$

using the continuity estimate provided by Theorem 2.1. An almost identical argument shows that $\mathcal{V}$ is α-Hölder, and by Theorem 2.1 this implies that P is α-Hölder also.

It remains to show that for every $x \in Z$ and $n \in \mathbb{N}$ we have $|||\mathcal{A}(x, n)v||| = |||v|||$ for every $v \in \mathcal{V}(x)$. We shall prove the following stronger result: for each $x \in Z$, there exists an increasing sequence of natural numbers (n_r) such that $\lim_{r\to\infty} \mathcal{A}(x, n_r) = P(x)$. To see that this implies the desired result, note that if $v \in \mathcal{V}(x)$ and $|||\mathcal{A}(x, n)v||| \leq (1-\varepsilon)|||v|||$ then $|||\mathcal{A}(x, n_r)v||| \leq (1-\varepsilon)|||v|||$ for all large enough r, and therefore $|||P(x)v||| \leq (1-\varepsilon)|||v|||$; but since $P(x)$ is a projection and v belongs to its image, this is only possible in the case where v is zero. We conclude that $|||\mathcal{A}(x, n)v||| = |||v|||$ whenever $v \in \mathcal{V}(x)$ and $n \in \mathbb{N}$ as desired.

Let us therefore define

$$\mathcal{S}(x) := \left\{B \colon \liminf_{n\to\infty} \max\left[d(T^n x, x), |||\mathcal{A}(x, n) - B|||\right] = 0\right\}$$

for each $x \in Z$. Note that $\mathcal{S}(x)$ is a subset of the "limit semigroup" $\mathcal{S}_\infty$ of A considered by F. Wirth [48, 49]. We shall show that each $\mathcal{S}(x)$ is a closed subsemigroup of $\mathcal{S}_\infty$.

Fix $x \in Z$. Since T acts minimally on Z, x is recurrent, and since $|||\mathcal{A}(x,n)||| = 1$ for each n we have $\mathcal{S}(x) \neq \emptyset$. If $B = \lim_{k\to\infty} B_k$ with $B_k \in \mathcal{S}(x)$ for each k then we may choose a strictly increasing sequence (n_k) such that $d(T^{n_k}x, x) < 1/k$, $|||B_k - B||| \leq 1/k$ and $|||\mathcal{A}(x, n_k) - B_k||| < 1/k$ for each $k \in \mathbb{N}^+$, which shows that $B \in \mathcal{S}(x)$ and therefore $\mathcal{S}(x)$ is closed. Since clearly $|||B||| = 1$ for all $B \in \mathcal{S}(x)$ it follows that $\mathcal{S}(x)$ is compact.

We now show that $\mathcal{S}(x)$ is a semigroup. Let $B_1, B_2 \in \mathcal{S}(x)$; it suffices to show that for any $N, \varepsilon > 0$ there is $n > N$ such that $d(T^n x, x) < \varepsilon$ and $|||\mathcal{A}(x,n) - B_1B_2||| < \varepsilon$. Since $B_1 \in \mathcal{S}(x)$ we can choose $n_1 > N$ such that $|||\mathcal{A}(x,n_1) - B_1||| < \varepsilon/3$ and $d(T^{n_1}x, x) < \varepsilon/2$. Since $B_2 \in \mathcal{S}(x)$ we may choose $n_2 > N$ such that $|||\mathcal{A}(x,n_2) - B_2||| < \varepsilon/3$ and such that $d(T^{n_2}x, x)$ is so small as to guarantee $|||\mathcal{A}(T^{n_2}x, n_1) - \mathcal{A}(x,n_1)||| < \varepsilon/3$ and $d(T^{n_1+n_2}x, T^{n_1}x) < \varepsilon/2$. We have

$$\begin{aligned} |||\mathcal{A}(x, n_1+n_2) - B_1B_2||| &\leq |||\mathcal{A}(T^{n_2}x, n_1)\mathcal{A}(x,n_2) - \mathcal{A}(x,n_1)\mathcal{A}(x,n_2)||| \\ &\quad + |||\mathcal{A}(x,n_1)\mathcal{A}(x,n_2) - \mathcal{A}(x,n_1)B_2||| \\ &\quad + |||\mathcal{A}(x,n_1)B_2 - B_1B_2||| < \varepsilon \end{aligned}$$

and

$$d(T^{n_1+n_2}x, x) \leq d(T^{n_1+n_2}x, T^{n_1}x) + d(T^{n_1}x, x) < \varepsilon$$

as required to prove the claim.

We now finish the proof. Since $\mathcal{S}(x)$ is a nonempty compact semigroup, it contains an idempotent element P, i.e. a projection (see e.g. [28]). Let $P = \lim_{r\to\infty} \mathcal{A}(x, n_r)$ where (n_r) is an increasing sequence of natural numbers such that $T^{n_r}x \to x$. Clearly $\|Pw\| = \lim_{r\to\infty} \|\mathcal{A}(x,n_r)w\| = 0$ when $w \in \mathcal{W}(x)$. If v is a nonzero element of $\mathcal{V}(x)$ then $\mathcal{A}(x,n_r)v \in \mathcal{V}(T^{n_r}x)$ by the invariance of $\mathcal{V}$, and since $\mathcal{V}$ is continuous it follows that $Pv \in \mathcal{V}(x)$. By Theorem 2.1 it also follows that $\|Pv\| = \lim_{r\to\infty} \|\mathcal{A}(x,n_r)v\| \geq \delta\|v\| > 0$ and therefore $Pv \neq 0$. We have shown that the kernel of P is equal to $\mathcal{W}(x)$ whilst the image of P equals $\mathcal{V}(x)$, and therefore $P = P(x)$ as claimed. This completes the proof of Theorem 2.2.

5. Proof of Theorem 1.2

Let A be as given in the statement of Theorem 1.2. If $\varrho(\mathsf{A}) = 0$ then the conclusion of the theorem follows from the fact that $0 \leq \varrho_n^-(\mathsf{A}) \leq \varrho(\mathsf{A})$ for every $n \in \mathbb{N}^+$, so we shall assume that $\varrho(\mathsf{A}) > 0$. It is clear that the validity of the conclusion of the theorem is unaffected if we rescale A by a positive real number, so without loss of generality we may assume that $\varrho(\mathsf{A}) = 1$.

The following lemma shows that there is also no loss of generality in assuming that A is product bounded. Results of this general type are also used in the proofs of Theorem 1.1 given by Berger-Wang [4], Elsner [19], and Shih *et al.* [46].

Lemma 5.1. *Let* $\mathsf{A} = \{A_1, \ldots, A_r\} \subset \mathrm{Mat}_d(\mathbb{C})$ *be a nonempty finite set such that* $\varrho(\mathsf{A}) = 1$*, and suppose that* A *is not product bounded. Then there exist a positive integer* $\hat{d} < d$ *and a set* $\hat{\mathsf{A}} = \{\hat{A}_1, \ldots, \hat{A}_r\} \subset \mathrm{Mat}_{\hat{d}}(\mathbb{C})$ *such that* $\varrho(\hat{\mathsf{A}}) = 1$*,* $\varrho_n^-(\mathsf{A}) \geq \varrho_n^-(\hat{\mathsf{A}})$ *for every* $n \in \mathbb{N}^+$*, and* A *is product bounded.*

Proof. If A is not product-bounded then necessarily $d \geq 2$. By [19, Lemma 4] there exist an invertible matrix $U \in \mathrm{Mat}_d(\mathbb{C})$ and a positive integer $d' < d$ such that for every $i = 1, \ldots, r$, the matrix $U^{-1}A_iU$ may be written in block upper-triangular form,

$$U^{-1}A_iU = \begin{pmatrix} B_i & * \\ 0 & C_i \end{pmatrix},$$

where the matrices B_i, C_i have dimensions $d' \times d'$ and $(d-d') \times (d-d')$ respectively. Let $\mathsf{B} = \{B_1, \ldots, B_r\}$ and $\mathsf{C} = \{C_1, \ldots, C_r\}$. Clearly we have $\varrho(\mathsf{A}) = \max\{\varrho(\mathsf{B}), \varrho(\mathsf{C})\}$ and $\varrho_n^-(\mathsf{A}) = \max\{\varrho_n^-(\mathsf{B}), \varrho_n^-(\mathsf{C})\}$ for each $n \in \mathbb{N}^+$. Define $\hat{\mathsf{A}} := \mathsf{B}$ and $\hat{d} := d'$ if $\varrho(\mathsf{B}) = 1$, and $\hat{\mathsf{A}} := \mathsf{C}$, $\hat{d} := d - d'$ otherwise. It is clear that the new set $\hat{\mathsf{A}}$ has all of the required features except perhaps for product boundedness. Repeating the above procedure by inductive descent we must eventually either obtain a new product bounded set $\hat{\mathsf{A}}$ with $\hat{d} > 1$, or else reduce to the case $\hat{d} = 1$ in which case product boundedness is satisfied automatically. □

For the remainder of this section we shall assume that A is a nonempty finite set of $d \times d$ matrices such that $\varrho(\mathsf{A}) = 1$ and A is product bounded. Since A is finite, the metric described in the introduction is Lipschitz equivalent to the more easily-used metric given by

$$d\left[(A_i)_{i\in\mathbb{Z}}, (B_i)_{i\in\mathbb{Z}}\right] = 2^{-\sup\{n\in\mathbb{N}\colon A_i = B_i \text{ for } |i|\leq n\}}.$$

The following proposition may be obtained easily by modifying a result of X. Bressaud and A. Quas [12, Theorem 1].

Proposition 5.2. *Let A be finite, let $Z \subseteq \mathsf{A}^{\mathbb{Z}}$ be compact with $TZ = Z$, and let $N \in \mathbb{N}^+$. Then there exist sequences of integers (r_n), (m_n) and a sequence of points $x_n \in \mathsf{A}^{\mathbb{Z}}$ such that $m_n^{-1}\log n \to 0$ and such that for all sufficiently large n each r_n is divisible by N, $r_n \leq n$, $T^{r_n}x_n = x_n$ and*

$$\max_{0\leq k<r_n} d(T^k x_n, Z) \leq 2^{-m_n}.$$

Now let $|\!|\!| \cdot |\!|\!|$ be an extremal norm for A, let Y be as in Theorem 2.2, and let $Z \subseteq Y$ be any nonempty minimal set. Let $\mathcal{V}$, $\mathcal{W}$, P, C, ξ and p be as given by Theorem 2.2, and define $Q(x) = I - P(x)$ for each $x \in Z$. If $p = d$ then for each $x \in Z$ and $n \in \mathbb{N}$ the matrix $\mathcal{A}(x,n)$ is an isometry with respect to $|\!|\!| \cdot |\!|\!|$; we therefore have $\rho(\mathcal{A}(x,n)) = 1$ for every $x \in Z$ and $n \in \mathbb{N}$ so that the result follows trivially. For the remainder of the proof we shall therefore assume that $1 \leq p < d$. Note that for $v \in \mathcal{V}(x)$ and $w \in \mathcal{W}(x)$ we have

$$\mathcal{A}(x,n)P(x)(v+w) = \mathcal{A}(x,n)v = P(T^n x)\mathcal{A}(x,n)(v+w)$$

and therefore $\mathcal{A}(x,n)P(x) = P(T^n x)\mathcal{A}(x,n)$ for all $x \in Z$ and $n \in \mathbb{N}$. Clearly this implies that $\mathcal{A}(x,n)Q(x) = Q(T^n x)\mathcal{A}(x,n)$ for all $x \in Z$ and $n \in \mathbb{N}$.

The following two lemmas, and the general strategy of their application, are suggested by [30]. For each $x \in Z$ and $\theta > 0$ let us define

$$\mathfrak{C}(x,\theta) = \left\{v \in \mathbb{C}^d \colon \theta |\!|\!| P(x)v |\!|\!| \geq |\!|\!| Q(x)v |\!|\!|\right\}.$$

Lemma 5.3. *Let $x, y \in Z$ and suppose that $|\!|\!| P(x) - P(y) |\!|\!| \leq \theta < 1/5$. Then $\mathfrak{C}(x,\theta) \subseteq \mathfrak{C}(y, 3\theta)$.*

Proof. If $v \notin \mathfrak{C}(y,3\theta)$ then $|\!|\!| Q(y)v |\!|\!| > 3\theta |\!|\!| P(y)v |\!|\!|$ and therefore

$$\begin{aligned} 3\theta |\!|\!| P(x)v |\!|\!| \leq 3\theta |\!|\!| P(y) |\!|\!| + 3\theta^2 |\!|\!| v |\!|\!| < |\!|\!| Q(y)v |\!|\!| + 3\theta^2 |\!|\!| v |\!|\!| \leq |\!|\!| Q(x)v |\!|\!| + (\theta + 3\theta^2) |\!|\!| v |\!|\!| \\ \leq \left(1 + \theta + 3\theta^2\right) |\!|\!| Q(x)v |\!|\!| + \left(\theta + 3\theta^2\right) |\!|\!| P(x)v |\!|\!| \end{aligned}$$

and therefore

$$\theta\|P(x)v\| \leq \frac{2\theta - 3\theta^2}{1+\theta+3\theta^2}\|P(x)v\| < \|Q(x)v\|$$

so that $v \notin \mathfrak{C}(x,\theta)$. □

Lemma 5.4. *Let $x \in Z$ and $n \in \mathbb{N}$, and suppose that $v \in \mathfrak{C}(x,\theta)$ for some $\theta \in (0,1]$. Then $\mathcal{A}(x,n)v \in \mathfrak{C}(T^n x, L_1\xi^n\theta)$ and $\|\mathcal{A}(x,n)v\| \geq (1-\theta-L_1\xi^n\theta)\|v\|$, where $L_1 > 1$ does not depend on x, n, θ or v.*

Proof. Let $M = \sup_{z\in Z}\|Q(z)\| \geq 1$ and $L_1 = 2CM > 1$. If $v \in \mathfrak{C}(x,\theta)$ then clearly

$$\|v\| \leq \|P(x)v\| + \|Q(x)v\| \leq (1+\theta)\|P(x)v\|.$$

Using Theorem 2.2 it follows that

$$\|P(T^n x)\mathcal{A}(x,n)v\| = \|\mathcal{A}(x,n)P(x)v\| = \|P(x)v\| \geq (1+\theta)^{-1}\|v\|$$

and

$$\|Q(T^n x)\mathcal{A}(x,n)v\| = \|\mathcal{A}(x,n)Q(x)v\| \leq C_1\xi^n\|Q(x)v\| \leq CM\theta\xi^n\|v\|.$$

Consequently

$$\|\mathcal{A}(x,n)v\| \geq \|P(T^n x)\mathcal{A}(x,n)v\| - \|Q(T^n x)\mathcal{A}(x,n)v\| \geq (1-\theta-L_1\theta\xi^n)\|v\|$$

and

$$\|Q(T^n x)\mathcal{A}(x,n)v\| \leq L_1\xi^n\theta\|P(T^n x)\mathcal{A}(x,n)v\|$$

as required. □

We now prove Theorem 1.2. Let $L_2 > 1$ and $\alpha \in (0,1)$ such that $\|P(x)-P(y)\| \leq L_2 d(x,y)^\alpha$ for all $x, y \in Z$, let $N \geq 1$ be large enough that $L_1\xi^N < 1/3$, and let $(x_n), (m_n), (r_n)$ be as given by Proposition 5.2. Suppose that n is large enough that $L_2 2^{\alpha(N-m_n)} < 1/5$, $m_n \geq N$, and all of the properties listed in Proposition 5.2 are satisfied. Let $q = r_n/N$ and choose $z_1, \ldots, z_q \in Z$ such that $d(z_i, T^{(i-1)N}x_n) \leq 2^{-m_n}$ for each i. We then have

$$d(T^N z_i, z_{i+1}) = \max\{d(T^N z_1, T^{iN}x_n), d(T^{iN}x_n, z_{i+1})\} \leq 2^{N-m_n}$$

for $1 \leq i < q$, and similarly $d(T^N z_q, z_1) \leq 2^{N-m_n}$. If $v \in \mathfrak{C}(z_i, L_2 2^{\alpha(N-m_n)})$ for $1 \leq i < q$ then we may apply Lemmas 5.3 and 5.4 to deduce that $\mathcal{A}(z_i,N)v \in \mathfrak{C}(z_{i+1}, L_2 2^{\alpha(N-m_n)})$ and $\|\mathcal{A}(z_i,N)v\| \geq (1 - L_2 2^{1+\alpha(N-m_n)})\|v\|$, and similarly if $v \in \mathfrak{C}(z_q, L_2 2^{\alpha(N-m_n)})$ then $\mathcal{A}(z_q,N)v \in \mathfrak{C}(z_1, L_2 2^{\alpha(N-m_n)})$ and $\|\mathcal{A}(z_q,N)v\| \geq (1 - L_2 2^{1+\alpha(N-m_n)})\|v\|$. It follows that if $v \in \mathfrak{C}(z_1, L_2 2^{\alpha(N-m_n)})$ then

$$\mathcal{A}(x_n, r_n)v = \mathcal{A}(z_q,N)\cdots\mathcal{A}(z_1,N)v \in \mathfrak{C}\left(z_1, L_2 2^{\alpha(N-m_n)}\right)$$

(where we have used $m_n \geq N$) and

$$\|\mathcal{A}(x_n,r_n)v\| = \|\mathcal{A}(z_q,N)\cdots\mathcal{A}(z_1,N)v\| \geq (1 - L_2 2^{1+\alpha(N-m_n)})^{r_n/N}\|v\|.$$

(Note that $1 - L_2 2^{1+\alpha(N-m_n)} > 3/5 > 0$ by our assumption on n.) If we choose $v \in \mathfrak{C}(z_1, L_2 2^{\alpha(N-m_n)})$ with $\|v\| = 1$, then since $r_n \leq n$ we deduce

$$\begin{aligned}\max_{1\leq k\leq n} \varrho_k^-(\mathsf{A}) \geq \rho(\mathcal{A}(x_n, r_n))^{1/r_n} &= \left(\lim_{k\to\infty} \|\mathcal{A}(x_n, r_n)^k\|^{1/k}\right)^{1/r_n} \\ &\geq \left(\liminf_{k\to\infty} \|\mathcal{A}(x_n, r_n)^k v\|^{1/k}\right)^{1/r_n} \\ &\geq (1 - L_2 2^{1+\alpha(N-m_n)})^{1/N} \geq 1 - L_2 2^{1+\alpha(N-m_n)}.\end{aligned}$$

It follows that for all large enough n

$$0 \leq \varrho(\mathsf{A}) - \max_{1\leq k\leq n} \varrho_k^-(\mathsf{A}) \leq \left(L_2 2^{1+\alpha N}\right) 2^{-\alpha m_n}.$$

To complete the proof we have only to observe that the condition $m_n^{-1} \log n \to 0$ is equivalent to the assertion that $e^{-\varepsilon m_n} = O(1/n^r)$ for every $r, \varepsilon > 0$.

6. Discussion on possible extensions of Theorem 1.2

We shall now briefly discuss some of the limitations of the method of proof of Theorem 1.2 and the prospects for an extension of that theorem using the approach of the present article.

Fix some nonempty compact set $\Omega \subset \mathbb{C}^d$, and consider the metric space $\Omega^{\mathbb{Z}}$ equipped with the metric $d[(x_i), (y_i)] = \sum_{i\in\mathbb{Z}} 2^{-|i|}\|x_i - y_i\|$ together with the shift map $T\colon \Omega^{\mathbb{Z}} \to \Omega^{\mathbb{Z}}$, which is Lipschitz continuous with Lipschitz constant 2. Given a nonempty compact T-invariant set $Z \subseteq \Omega^{\mathbb{Z}}$, let us define

$$\varepsilon(Z, n) = \min_{1\leq k\leq n} \inf_{T^k x = x} \max_{0\leq i<k} \operatorname{dist}(T^i x, Z),$$

where $\operatorname{dist}(y, Z) := \inf\{d(y, z) \colon z \in Z\}$. The magnitude of the error term in the proof of Theorem 1.2 is determined by the result of X. Bressaud and A. Quas in [12] which asserts that if Ω is a finite set, then $\varepsilon(Z, n) = O(1/n^r)$ for every $r > 0$. (To simplify our proof we in fact considered only approximations using periodic orbits whose period is divisible by N, but this requirement could be dispensed with without difficulty.) Bressaud and Quas' result is essentially sharp: see [12] and related work in [13]. In the case where Ω is compact but not finite, the rate of decrease of $\varepsilon(Z, n)$ can be much slower, and this is the principal obstacle in extending Theorem 1.2 to the case in which A is compact but infinite. The following simple example illustrates the problem.

Suppose that $\Omega = S^1 \subset \mathbb{C}$. Let $\gamma = (1 - \sqrt{5})/2$ and define

$$Z = \left\{\left(e^{2\pi i m\gamma}\omega\right)_{m\in\mathbb{Z}} : \omega \in S^1\right\},$$

which is clearly compact and T-invariant. Let $n \in \mathbb{N}^+$ and $1 \leq k \leq n$, and suppose that $x \in \Omega^{\mathbb{Z}}$ has $T^k x = x$ and $\max_{0\leq j<k} \operatorname{dist}(T^j x, Z) = \varepsilon(Z, n)$. For $j = 0, \ldots, k-1$ choose $z_j = (e^{2\pi i m\gamma}\omega_j)_{m\in\mathbb{Z}} \in Z$ such that $d(T^j x, z) \leq \varepsilon(Z, n)$, and define also $z_k = z_0$ and $\omega_k = \omega_0$. For $0 \leq j < k$ we have

$$\left|e^{2\pi i\gamma}\omega_j - \omega_{j+1}\right| \leq d(Tz_j, z_{j+1}) \leq d(Tz_j, T^{j+1}x) + d(T^{j+1}x, z_{j+1}) \leq 3\varepsilon(Z, n),$$

and it follows that

$$\left|e^{2\pi i k\gamma}\omega_0 - \omega_0\right| \leq \sum_{j=0}^{k-1} \left|e^{2\pi i j\gamma}\omega_j - e^{2\pi i(j+1)\gamma}\omega_{j+1}\right| \leq 3k\varepsilon(Z, n).$$

However, it is well-known [27] that there exists $\delta > 0$ such that $|e^{2\pi i m\gamma} - 1| \geq \delta/m$ for every $m \in \mathbb{N}^+$, and we deduce that $\varepsilon(Z,n) \geq \delta/3k^2 \geq \delta/3n^2$.

We conclude that if $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ is some compact set of matrices which is isometric to S^1, then there exists a minimal invariant set $Z \subset \mathsf{A}^{\mathbb{Z}}$ such that $\varepsilon(Z,n)$ is not $o(n^{-2})$. In particular, the method of Theorem 1.2 is in this case not strong enough even to show that

$$\varrho(\mathsf{A}) - \max_{1 \leq k \leq n} \varrho_k^-(\mathsf{A}) = O\left(\frac{1}{n^{2\alpha}}\right),$$

where $\alpha \in (0,1)$ is the Hölder exponent of the function P given by Theorem 2.2. Since α could be arbitrarily small this estimate would anyway be inferior to the estimate of J. Bochi described in the introduction. If we wish to achieve further progress using the methods of the present article, therefore, the key step must be to show that for a given set $\mathsf{A} \subset \mathrm{Mat}_d(\mathbb{C})$ there is an extremal norm $|\!|\!| \cdot |\!|\!|$ for which the set

$$Y = \left\{x \in \mathsf{A}^{\mathbb{Z}} \colon \varrho(\mathsf{A})^{-n} |\!|\!| \mathcal{A}(x,n) |\!|\!| = 1 \ \forall\ n \in \mathbb{N}^+\right\} \tag{22}$$

contains a minimal set Z such that the quantity $\varepsilon(Z,n)$ decreases with some specified rapidity as a function of n.

It should be remarked that the explicit structure of the set Y defined in (22) is for the most part unknown, and so the range of minimal sets Z which may be contained in such a set Y could in principle be quite limited, potentially leading to improved estimates in Theorem 1.2. Indeed, the the *finiteness conjecture* of J. Lagarias and Y. Wang, proposed in [35], was equivalent to the statement that Y must always contain a periodic orbit. The existence of counterexamples to the finiteness conjecture was established by T. Bousch and J. Mairesse [10], with a simpler argument subsequently being given in [6]. At present, the only well-understood examples of sets A in which Y does not contain a periodic orbit have the property that the orbits in Y are "Sturmian" or "balanced" [10]. When Z consists of Sturmian orbits one may show that $\varepsilon(Z,n)$ decreases exponentially as a function of n, and in particular the arguments used in this article could be applied to obtain an exponential estimate in Theorem 1.2 in this special case.

7. Acknowledgments

The author would like to thank an anonymous referee for a number of helpful suggestions which have led to the substantial streamlining of several proofs. The author would also like to thank M. Pollicott for suggesting the reference [20].

This research was supported by EPSRC grant EP/E020801/1.

References

1. N. I. Akhiezer and I. M. Glazman, *Theory of linear operators in Hilbert space*, Dover Publications Inc., New York, 1993, Translated from the Russian and with a preface by Merlynd Nestell, Reprint of the 1961 and 1963 translations, Two volumes bound as one. MR MR1255973 (94i:47001)
2. N. E. Barabanov, *On the Lyapunov exponent of discrete inclusions. I*, Avtomat. i Telemekh. (1988), no. 2, 40–46. MR MR940263 (89e:34025)
3. L. Barreira and Ya. Pesin, *Lectures on Lyapunov exponents and smooth ergodic theory*, Smooth ergodic theory and its applications (Seattle, WA, 1999), Proc. Sympos. Pure Math., vol. 69, Amer. Math. Soc., Providence, RI, 2001, Appendix A by M. Brin and Appendix B by D. Dolgopyat, H. Hu and Pesin, pp. 3–106. MR MR1858534 (2003a:37039)

4. Marc A. Berger and Yang Wang, *Bounded semigroups of matrices*, Linear Algebra Appl. **166** (1992), 21–27. MR MR1152485 (92m:15012)
5. Vincent D. Blondel and Yurii Nesterov, *Computationally efficient approximations of the joint spectral radius*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 1, 256–272 (electronic). MR MR2176820 (2006k:15027)
6. Vincent D. Blondel, Jacques Theys, and Alexander A. Vladimirov, *An elementary counterexample to the finiteness conjecture*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 4, 963–970 (electronic). MR MR2003315 (2004g:15010)
7. Jairo Bochi, *Inequalities for numerical invariants of sets of matrices*, Linear Algebra Appl. **368** (2003), 71–81. MR MR1983195 (2004f:15034)
8. Thierry Bousch, *Le poisson n'a pas d'arêtes*, Ann. Inst. H. Poincaré Probab. Statist. **36** (2000), no. 4, 489–508. MR MR1785392 (2001i:37005)
9. Thierry Bousch and Oliver Jenkinson, *Cohomology classes of dynamically non-negative C^k functions*, Invent. Math. **148** (2002), no. 1, 207–217. MR MR1892849 (2003c:37048)
10. Thierry Bousch and Jean Mairesse, *Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture*, J. Amer. Math. Soc. **15** (2002), no. 1, 77–111 (electronic). MR MR1862798 (2002j:49008)
11. Julien Brémont, *Entropy and maximizing measures of generic continuous functions*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 3-4, 199–201. MR MR2393641
12. Xavier Bressaud and Anthony Quas, *Rate of approximation of minimizing measures*, Nonlinearity **20** (2007), no. 4, 845–853. MR MR2307883 (2008f:37003)
13. D. Collier and I. D. Morris, *Approximating the maximum ergodic average via periodic orbits*, Ergodic Theory Dynam. Systems **28** (2008), no. 4, 1081–1090. MR MR2437220
14. G. Contreras, A. O. Lopes, and Ph. Thieullen, *Lyapunov minimizing measures for expanding maps of the circle*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1379–1409. MR MR1855838 (2002i:37038)
15. J.-P. Conze and Y. Guivarc'h, *Croissance des sommes ergodiques*, unpublished manuscript, circa 1990.
16. Ingrid Daubechies and Jeffrey C. Lagarias, *Sets of matrices all infinite products of which converge*, Linear Algebra Appl. **161** (1992), 227–263. MR MR1142737 (93f:15006)
17. ———, *Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal. **23** (1992), no. 4, 1031–1079. MR MR1166574 (93g:39001)
18. Jean Marie Dumont, Nikita Sidorov, and Alain Thomas, *Number of representations related to a linear recurrent basis*, Acta Arith. **88** (1999), no. 4, 371–396. MR MR1690457 (2000c:11030)
19. L. Elsner, *The generalized spectral-radius theorem: an analytic-geometric proof*, Proceedings of the Workshop "Nonnegative Matrices, Applications and Generalizations" and the Eighth Haifa Matrix Theory Conference (Haifa, 1993), vol. 220, 1995, pp. 151–159. MR MR1334574 (96e:15010)
20. Gary Froyland, Simon Lloyd, and Anthony Quas, *Coherent structures and isolated spectrum for Perron-Frobenius cocycles*, preprint, 2008.
21. H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981, M. B. Porter Lectures. MR MR603625 (82j:28010)
22. I. C. Gohberg and M. G. Kreĭn, *Introduction to the theory of linear nonselfadjoint operators*, Translated from the Russian by A. Feinstein. Translations of Mathematical Monographs, Vol. 18, American Mathematical Society, Providence, R.I., 1969. MR MR0246142 (39 #7447)
23. Gustaf Gripenberg, *Computing the joint spectral radius*, Linear Algebra Appl. **234** (1996), 43–60. MR MR1368770 (97c:15043)
24. N. Guglielmi, F. Wirth, and M. Zennaro, *Complex polytope extremality results for families of matrices*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 3, 721–743 (electronic). MR MR2208331 (2007b:93106)
25. N. Guglielmi and M. Zennaro, *On the zero-stability of variable stepsize multistep methods: the spectral radius approach*, Numer. Math. **88** (2001), no. 3, 445–458. MR MR1835466 (2002d:65066)
26. Leonid Gurvits, *Stability of discrete linear inclusion*, Linear Algebra Appl. **231** (1995), 47–85. MR MR1361100 (96i:93056)
27. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, sixth ed., Oxford University Press, Oxford, 2008, Revised by D. R. Heath-Brown and J. H. Silverman. MR MR2445243

28. Karl Heinrich Hofmann and Paul S. Mostert, *Elements of compact semigroups*, Charles E. Merr ll Books, Inc., Columbus, Ohio, 1966. MR MR0209387 (35 #285)
29. Oliver Jenkinson, *Ergodic optimization*, Discrete Contin. Dyn. Syst. **15** (2006), no. 1, 197–224. MR MR2191393
30. Boris Kalinin, *Livsic theorem for matrix cocycles*, preprint, 2008.
31. Tosio Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1980 edition. MR MR1335452 (96a:47025)
32. V. S. Kozyakin, *Algebraic unsolvability of a problem on the absolute stability of desynchronized systems*, Avtomat. i Telemekh. (1990), no. 6, 41–47. MR MR1071607 (91e:93065)
33. ______, *Iterative building of Barabanov norms and computation of the joint spectral radius for matrix sets*, arXiv preprint 0810.2154v2, 2008.
34. Ulrich Krengel, *Ergodic theorems*, de Gruyter Studies in Mathematics, vol. 6, Walter de Gruyter & Co., Berlin, 1985, With a supplement by Antoine Brunel. MR MR797411 (87i:28001)
35. Jeffrey C. Lagarias and Yang Wang, *The finiteness conjecture for the generalized spectral radius of a set of matrices*, Linear Algebra Appl. **214** (1995), 17–42. MR MR1311628 (95k:15038)
36. Mohsen Maesumi, *An efficient lower bound for the generalized spectral radius of a set of matrices*, Linear Algebra Appl. **240** (1996), 1–7. MR MR1387282 (97b:15027)
37. ______, *Calculating joint spectral radius of matrices and Hölder exponent of wavelets*, Approximation theory IX, Vol. 2 (Nashville, TN, 1998), Innov. Appl. Math., Vanderbilt Univ. Press, Nashville, TN, 1998, pp. 205–212. MR MR1744409
38. Bruce E. Moision, Alon Orlitsky, and Paul H. Siegel, *On codes that avoid specified differences*, IEEE Trans. Inform. Theory **47** (2001), no. 1, 433–442. MR MR1820392 (2001k:94084)
39. I. D. Morris, *A sufficient condition for the subordination principle in ergodic optimization*, Bull. Lond. Math. Soc. **39** (2007), no. 2, 214–220. MR MR2323451 (2008c:37052)
40. Ian D. Morris, *Maximizing measures of generic Hölder functions have zero entropy*, Nonlinearity **21** (2008), no. 5, 993–1000. MR MR2412323 (2009b:37055)
41. Ian David Morris, *The generalized Berger-Wang formula and the spectral radius of linear cocycles*, arXiv preprint 0906.2915, 2009.
42. Pablo A. Parrilo and Ali Jadbabaie, *Approximation of the joint spectral radius of a set of matrices using sum of squares*, Hybrid systems: computation and control, Lecture Notes in Comput. Sci., vol. 4416, Springer, Berlin, 2007, pp. 444–458. MR MR2363632 (2008m:15027)
43. M. S. Raghunathan, *A proof of Oseledec's multiplicative ergodic theorem*, Israel J. Math. **32** (1979), no. 4, 356–362. MR MR571089 (81f:60016)
44. Gian-Carlo Rota and Gilbert Strang, *A note on the joint spectral radius*, Nederl. Akad. Wetensch. Proc. Ser. A 63 = Indag. Math. **22** (1960), 379–381. MR MR0147922 (26 #5434)
45. David Ruelle, *Ergodic theory of differentiable dynamical systems*, Inst. Hautes Études Sci. Publ. Math. (1979), no. 50, 27–58. MR MR556581 (81f:58031)
46. Mau-Hsiang Shih, Jinn-Wen Wu, and Chin-Tzong Pang, *Asymptotic stability and generalized Gelfand spectral radius formula*, Linear Algebra Appl. **252** (1997), 61–70. MR MR1428628 (97k:15028)
47. John N. Tsitsiklis and Vincent D. Blondel, *The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate*, Math. Control Signals Systems **10** (1997), no. 1, 31–40. MR MR1462278 (99h:65238a)
48. Fabian Wirth, *The generalized spectral radius and extremal norms*, Linear Algebra Appl. **342** (2002), 17–40. MR MR1873424 (2003g:15025)
49. ______, *On the structure of the set of extremal norms of a linear inclusion*, Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005 (Seville, Spain), December 2005, pp. 3019–3024.
50. Jean-Christophe Yoccoz, *Some questions and remarks about* $\mathrm{SL}(2, \mathbf{R})$ *cocycles*, Modern dynamical systems and applications, Cambridge Univ. Press, Cambridge, 2004, pp. 447–458. MR MR2093316 (2005h:37068)
51. Guocheng Yuan and Brian R. Hunt, *Optimal orbits of hyperbolic systems*, Nonlinearity **12** (1999), no. 4, 1207–1224. MR MR1709845 (2000g:37029)