Eduardo OREGÓN-REYES

A new inequality about matrix products and a Berger-Wang formula

A NEW INEQUALITY ABOUT MATRIX PRODUCTS AND
A BERGER-WANG FORMULA

by Eduardo Oregón-Reyes

Abstract. — We prove an inequality relating the norm of a product of matrices $A_n \cdots A_1$ with the spectral radii of subproducts $A_j \cdots A_i$ with $1 \leqslant i \leqslant j \leqslant n$. Among the consequences of this inequality, we obtain the classical Berger-Wang formula as an immediate corollary, and give an easier proof of a characterization of the upper Lyapunov exponent due to I. Morris. As main ingredient for the proof of this result, we prove that for a large enough n, the product $A_n \cdots A_1$ is zero under the hypothesis that $A_j \cdots A_i$ are nilpotent for all i, j such that $1 \leqslant i \leqslant j \leqslant n$.

Résumé (Une nouvelle inégalité sur les produits de matrices et une formule de Berger-Wang)

Nous montrons une inégalité reliant la norme d'un produit $A_n \cdots A_1$ de matrices aux rayons spectraux des sous-produits $A_j \cdots A_i$ avec $1 \leqslant i \leqslant j \leqslant n$. Comme conséquences de cette inégalité, nous obtenons la formule classique de Berger-Wang comme corollaire immédiat, et nous donnons une preuve plus simple de la caractérisation, due à I. Morris, de l'exposant de Liapounov supérieur. Nous montrons, comme ingrédient principal de la preuve de ce résultat, que pour n assez grand, le produit $A_n \cdots A_1$ est nul si les $A_j \cdots A_i$ sont nilpotents pour tout i, j tel que $1 \leqslant i \leqslant j \leqslant n$.

Contents

1. Introduction

Let k be a field, and let $M_d(k)$ be the algebra of $d \times d$ matrices with coefficients in k. If $k = \mathbb{R}$ or $\mathbb{C}$, let $\|.\|$ be any norm on k^d, with the corresponding operator norm on $M_d(k)$ also denoted by $\|.\|$. The spectral radius of a matrix A will be denoted by $\rho(A)$. Given a bounded set $\mathscr{M} \subset M_d(k)$, the *joint spectral radius* of $\mathscr{M}$ is defined by the formula

$$(1) \qquad \mathfrak{R}(\mathscr{M}) = \lim_{n\to\infty} \left(\sup \left\{ \|A_1 \cdots A_n\| : A_i \in \mathscr{M} \right\} \right)^{1/n}.$$

By a submultiplicative argument, this quantity is well defined and finite, and the limit in the right hand side of (1) can be replaced by the infimum over n.

The joint spectral radius was introduced by Rota and Strang [26], and for a set $\mathscr{M} \subset M_d(k)$, represents the maximal exponential growth rate of the partial sequence of products $(A_1 \cdots A_n)_n$ of a sequence of matrices $A_1, A_2, \ldots$ with $A_i \in \mathscr{M}$. For this reason, this quantity has appeared in several mathematical contexts, making it an important object of study (see e.g. [13, 14, 22, 29]). In particular, the question of whether the joint spectral radius may be approximated by periodic sequences plays an important role. The Berger-Wang formula gives a positive answer to this question in the case of bounded sets of matrices [2]:

THEOREM 1.1 (Berger-Wang formula). — *If $\mathscr{M} \subset M_d(\mathbb{C})$ is bounded, then*

$$(2) \qquad \mathfrak{R}(\mathscr{M}) = \limsup_{n\to\infty} \left(\sup \left\{ \rho(A_1 \cdots A_n) : A_i \in \mathscr{M} \right\} \right)^{1/n}.$$

This result has been generalized by Morris, to the context of linear cocycles (including infinite dimensional ones) [21], by using multiplicative ergodic theory. In the finite dimensional case, the problem of finding a formula similar to (2), when there is a Markov-type constraint on the allowed products was presented by Kozyakin [16]. Although the result of Morris already applies to this kind of constraints, the novelty in Kozyakin's proof is that his arguments are purely linear algebraic, and are consequences of Theorem 1.1.

Another tool to obtain results related to joint spectral radius was found by Bochi in [4]. In that work, he proved some inequalities that may be seen as lower bounds for joint spectral radii of sets of matrices in terms of the norms of such matrices. Following that method, the purpose of this article is to present an inequality relating the norm of the product of matrices with the spectral radii of subproducts. We will give an upper bound for the norm of the product of matrices $A_N \cdots A_1$ in terms of the spectral radii of its subproducts $A_\beta A_{\beta-1} \cdots A_{\alpha+1} A_\alpha$. This inequality will allow us to obtain relations similar to (2). It holds in an arbitrary local field where the notions of absolute value, norm, and spectral radius are well defined (see Section 4 for a detailed explanation). Our main result is the following:

THEOREM 1.2. — *Let $d \in \mathbb{N}$, k be a local field, and $\|.\|$ be a submultiplicative norm on $M_d(k)$. There exist constants $N = N(d) \leqslant \prod_{i=1}^{d} \binom{d}{i}$, $r = r(d, N) \leqslant (Nd+1)^{Nd^2+2}$,*

and $C = C(d, \|.\|) > 1$ such that for all $n \geqslant N$ and $A_1, \ldots, A_n \in M_d(k)$:

$$(3) \qquad \|A_n \cdots A_1\| \leqslant C \Big(\prod_{1 \leqslant i \leqslant n} \|A_i\| \Big) \max_{1 \leqslant \alpha \leqslant \beta \leqslant n} \Big(\frac{\rho(A_\beta \cdots A_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|} \Big)^{1/r},$$

where the right hand side is treated as zero if one of the A_i is the zero matrix.

So for large enough n, if the norm of the product $A_n \cdots A_1$ is comparable to (that is, not much smaller than) the product of the norms, then there exists a subproduct $A_\beta \cdots A_\alpha$ whose spectral radius is comparable to (that is, not much smaller than) $\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|$.

Note that inequality (3) is homogeneous in each variable A_i. We will later show that the upper bound $N(d) \leqslant \prod_{i=1}^{d} \binom{d}{i}$ is not sharp, because $N(3) \leqslant 5$ (see Proposition 2.4). In addition, when $k = \mathbb{C}$, the constant C in (3) may be chosen independent of the norm $\|.\|$ and found explicitly, provided that $\|.\|$ is an operator norm (see Proposition 4.2 and Remark 4.5).

The approach of using inequalities to prove results similar to (2) was first used by Elsner [9] in his proof of the Berger-Wang formula (with an inequality of a different nature from Bochi's work). Inequalities like (3) also have been applied by I. Morris to study matrix pressure functions [23] and by the author in the context of isometries in Gromov hyperbolic spaces [24]. The novelty of the inequality presented here is that it respects the order in which the matrices are multiplied. While previous works considered a sum or a maximum over *all* possible subproducts of length N with respect to a given alphabet of matrices, in Theorem 1.2 we consider just one product of length N together with its subproducts, hence our inequality does not follow from previously known Bochi-type results. In addition, our error in the upper bound in terms of spectral radii is multiplicative (the constant C) and not additive as in the case of Elsner's work. These distinctions allow inequality (3) to be used in cases where only some specific kinds of products are allowed (see Theorem 1.4 below), as well to relate asymptotic quantities (like the joint spectral radius) to non-asymptotic expressions, in a uniformly controlled way (see Theorem 4.4).

The proof of this inequality is based on the non trivial case of equality, where the right hand side of (3) is zero but the matrices A_i are non-zero. This occurs when $\rho(A_j \cdots A_i) = 0$ for all $1 \leqslant i \leqslant j \leqslant N$, that is, when $A_j \cdots A_i$ are all nilpotent. Denote by $\mathcal{N}_d(k)$ the set of nilpotent elements of $M_d(k)$. Then define, for $n \geqslant 1$, the set $\mathcal{N}_d^n(k)$ of n-tuples $(A_1, \ldots, A_n) \in M_d(k)^n$ such that $A_j \cdots A_i \in \mathcal{N}_d(k)$ for all $1 \leqslant i \leqslant j \leqslant n$. The particular case of (3) that we highlighted can be restated as follows:

THEOREM 1.3. — *For all $d \geqslant 1$ there exists an integer $N = N(d) \geqslant 1$ such that, for every field k, if $(A_1, \ldots, A_N) \in \mathcal{N}_d^N(k)$, then the product $A_N \cdots A_1$ is zero.*

The proof of Theorem 1.3 is purely linear algebraic, exploiting the properties of the n-exterior power functor. This result may be compared with Levitzki's Theorem [25, Th. 2.1.7], that asserts that for an algebraically closed field k, every semigroup

$S \subset M_d(k)$ of nilpotent matrices is simultaneously triangularizable. That is, there is some $B \in \mathrm{GL}_d(k)$ such that BAB^{-1} is upper triangular with zero diagonal for every $A \in S$ (compare also with the Burnside-Schur Theorem for semigroups of matrices [20]). In particular, if $A_1, \ldots, A_d \in S$, then the product $A_1 \cdots A_d$ is zero. As we show in Section 2.1, the optimal $N(d)$ in Theorem 1.3 is in general larger than d, therefore the result presented here does not follow from Levitzki's Theorem nor Burnside-Schur Theorem, and we don't expect to obtain any information about the semigroup generated by $A_1, \ldots, A_N$. In general, the matrices satisfying the hypothesis of Theorem 1.3 admit no normal form as simple as in Levitzki's Theorem.

Applications to Ergodic theory. — Let $(X, \mathscr{F}, \mu)$ be a probability space, and let $T : X \to X$ be a measure preserving map. By a *linear cocycle* over X, we mean a measurable map $A : X \to M_d(k)$ together with the family of maps A^n defined by the formula

$$A^n(x) = A(T^{n-1}x) \cdots A(Tx)A(x), \quad \text{for } n \geqslant 1, \ x \in X.$$

These maps satisfy the multiplicative cocycle relation $A^{m+n}(x) = A^m(T^n x)A^n(x)$ for all $m, n \geqslant 1, x \in X$.

We usually denote a linear cocycle by $\mathscr{A} = (X, T, A)$, and say that $\mathscr{A}$ is *integrable* if $\max(\log \|A\|, 0)$ is integrable. In this case, Kingman's theorem implies that, for μ-almost all $x \in X$, the limit $\lambda(x) = \lim_{n \to \infty} \frac{\log \|A^n(x)\|}{n} \in [-\infty, \infty)$ exists, and moreover, λ is T-invariant. This function is the upper *Lyapunov exponent* of A, and is one of the most important concepts in multiplicative ergodic theory.

As an application of our inequality, we reprove the following theorem due to I. Morris (first tested numerically in [11] and proved by Avila-Bochi for $\mathrm{SL}_2(\mathbb{R})$ in [1, Th. 15]).

Theorem 1.4. — [21, Th. 1.6] *Let T be a measure-preserving transformation of a probability space $(X, \mathscr{F}, \mu)$ and let $A : X \to M_d(k)$ be an integrable linear cocycle. If λ is as before, then for μ-almost all $x \in X$ we have*

$$(4) \qquad \limsup_{n \to \infty} \frac{\log(\rho(A^n(x)))}{n} = \lambda(x).$$

While Morris's proof of this result relies on Oseledets Theorem, we will mainly use Theorem 1.2 and a quantitative version of Poincaré's Recurrence Theorem.

Organization of the paper. — In Section 2 we prove Theorem 1.3 and compute $N(d)$ for $d = 2, 3$. Then in Section 3, via Nullstellensatz we translate this theorem into a polynomial identity, from which we deduce Theorem 1.2 in Section 4. We prove Theorem 1.4 in Section 5, and discuss some geometric consequences and analogies of this result in Section 6.

2. Proof of Theorem 1.3

We begin the proof of Theorem 1.3 with some useful results. For a given vector space V (over an arbitrary field), let $\operatorname{End}(V)$ be the algebra of linear endomorphisms of V. The dimension of the image of a linear transformation $T \in \operatorname{End}(V)$ will be denoted as $\operatorname{rank}(T)$. Also, let $\mathscr{N}^n(V)$ be the set of n-tuples $(T_1, \ldots, T_n) \in \operatorname{End}(V)^n$ such that $T_j \cdots T_i$ is nilpotent for all $1 \leqslant i \leqslant j \leqslant n$. With our previous notation, we have $\mathscr{N}^n(k^d) = \mathscr{N}^n_d(k)$.

Proposition 2.1. — *Let $n \geqslant 1$ and $(T_1, \ldots, T_n) \in \mathscr{N}^n(V)$ be such that $\operatorname{rank}(T_j) \leqslant 1$ for all $1 \leqslant j \leqslant n-1$. If $v \in V$ and $T_n \cdots T_1 v \neq 0$, then $v, T_1 v, T_2 T_1 v, \ldots, T_n \cdots T_1 v$ are all distinct and form a linearly independent set.*

Proof. — We will use induction on n. The case $n = 1$ comes from the nilpotence of T_1. So, assume that the result holds for tuples in $\mathscr{N}^{n-1}(V)$ and let $(T_1, \ldots, T_n) \in \mathscr{N}^n(V)$ and $v \in V$ be as in the hypothesis. Take a linear combination of $v, T_1 v, \ldots, T_n \cdots T_1$ of the form

$$(5) \qquad \lambda_0 v + \lambda_1 T_1 v + \cdots + \lambda_{n-1} T_{n-1} \cdots T_1 v + \lambda_n T_n \cdots T_1 v = 0,$$

and suppose that this linear combination is non trivial. As $(T_1, \ldots, T_{n-1}) \in \mathscr{N}^{n-1}(V)$ also satisfies the hypothesis with respect to v, by our inductive assumption we have $\lambda_n \neq 0$. Now, apply $T_n \cdots T_1$ in (5). The rank condition over the maps T_j and the fact that $(T_1, \ldots, T_n) \in \mathscr{N}^n(V)$ imply that $(T_j \cdots T_1)^2 = 0$, for all $1 \leqslant j \leqslant n$. Hence, the left hand side of (5) becomes $\lambda_0 T_n \cdots T_1 v$, forcing $\lambda_0 = 0$. But in that case, equation (5) would be a non trivial linear combination of $\{w, T_2 w, T_3 T_2 w, \ldots, T_n \cdots T_2 w\}$, with $w = T_1 v$. This is impossible by our inductive assumption, since $(T_2, \ldots, T_n) \in \mathscr{N}^{n-1}(V)$ satisfies the hypothesis of the proposition with respect to w. We conclude that all linear combinations of $v, T_1 v, T_2 T_1 v, \ldots, T_n \cdots T_1 v$ of the form (5) are trivial, and hence this set is linearly independent with exactly $n+1$ elements. $\qquad \square$

Corollary 2.2. — *If $(T_1, \ldots, T_d) \in \mathscr{N}^d(V)$ and $\operatorname{rank}(T_j) \leqslant 1$ for all $1 \leqslant j \leqslant d-1$, then $T_d \cdots T_1 = 0$.*

Proof. — Assume the contrary and let $v \in V$ be such that $T_d \cdots T_1 v \neq 0$. Then by Proposition 2.1, the set $\{v, T_1 v, T_2 T_1 v, \ldots, T_d \cdots T_1 v\}$ would be a linearly independent set of cardinality greater than $\dim V$. A contradiction. $\qquad \square$

For the next steps in our proof we need some fact about exterior powers. Recall that if V is a vector space of dimension d, the *r-fold exterior power* $\Lambda^r V$ is the vector space of alternating r-linear forms on the dual space V^* (see e.g. [18, XIX.1]). Given a basis $\{v_1, \ldots, v_d\}$ of V, the set $\{v_{i_1} \wedge \cdots \wedge v_{i_r} : 1 \leqslant i_1 < \cdots < i_r \leqslant d\}$ is a basis of $\Lambda^r V$. Hence $\dim \Lambda^r V = \binom{d}{r}$.

The exterior power also induces a map $\Lambda^r : \operatorname{End}(V) \to \operatorname{End}(\Lambda^r V)$ given by the linear extension of $(\Lambda^r T)(w_1 \wedge \cdots \wedge w_r) = (T w_1 \wedge \cdots \wedge T w_r)$. This map is functorial: The relation $\Lambda^r(ST) = \Lambda^r(S) \Lambda^r(T)$ holds for all $S, T \in \operatorname{End}(V)$. This functor also induces a map $\Lambda^r : \mathscr{N}(V) \to \mathscr{N}(\Lambda^r V)$ that extends to $\mathscr{N}^n(V) \to \mathscr{N}^n(\Lambda^r V)$ for all $n \geqslant 1$.

Another important fact is that, when $T \in \mathscr{N}(V)$ and $\operatorname{rank}(T) = r > 0$, then $\operatorname{rank}(\Lambda^r T) = 1$. This is because the image of $\Lambda^r T$ is generated by any r-form associated to the r-dimensional subspace $T(V)$. This remark is crucial in the end of our proof.

LEMMA 2.3. — *Let $1 \leqslant r \leqslant d$ and $m = \binom{d}{r}$. Given $(T_1, \ldots, T_m) \in \mathscr{N}^m(V)$, with $\operatorname{rank}(T_j) \leqslant r$ for all $1 \leqslant j \leqslant m-1$, we have $\operatorname{rank}(T_m \cdots T_1) < r$.*

Proof. — If that is the case then we will have $\operatorname{rank}(T_j T_{j-1} \cdots T_i) \leqslant r$ for all $1 \leqslant i \leqslant j \leqslant m-1$. Then the tuple $(\Lambda^r T_1, \ldots, \Lambda^r T_m) \in \mathscr{N}^m(\Lambda^r V)$ will satisfy the hypothesis of Corollary 2.2, and hence $\Lambda^r(T_m \cdots T_1) = 0$, which implies that $\operatorname{rank}(T_m \cdots T_1) < r$. ☐

Proof of Theorem 1.3. — Let $1 \leqslant \ell < d$ and $r(\ell) = \binom{d}{1} \cdots \binom{d}{\ell}$. We claim that for all $(T_1, \ldots, T_{r(\ell)}) \in \mathscr{N}^{r(\ell)}(V)$ we have $\operatorname{rank}(T_{r(\ell)} \cdots T_1) < d - \ell$. If so, the result follows with $N = r(d-1) = \binom{d}{1} \cdots \binom{d}{d-1}$.

We will argue by induction. The case $\ell = 1$ is Lemma 2.3 with $r = d - 1$. Now, assume the result for some $\ell < d$, and for $1 \leqslant j \leqslant \binom{d}{\ell+1}$, define $\widehat{T}_j = T_{r(\ell)j} \cdots T_{r(\ell)(j-1)+1}$. Then $(\widehat{T}_1, \ldots, \widehat{T}_{\binom{d}{\ell+1}}) \in \mathscr{N}^{\binom{d}{\ell+1}}(V)$, and by our inductive hypothesis, we obtain $\operatorname{rank}(\widehat{T}_j) \leqslant d - \ell - 1$. So, we are in the assumption of 2.3 with $r = d - \ell - 1$ and we conclude that $\operatorname{rank}(T_{r(\ell+1)} \cdots T_1) = \operatorname{rank}(\widehat{T}_{\binom{d}{\ell+1}} \cdots \widehat{T}_1) < d - \ell - 1$. This proves the claim and concludes the proof of the theorem. ☐

2.1. SOME COMPUTATIONS IN LOW DIMENSION. — Let $N(d)$ be the least value of N for which Theorem 1.3 (and therefore also Theorem 1.2) holds true. From the proof of Theorem 1.3, we can obtain the bound $N(d) \leqslant \binom{d}{1}\binom{d}{2} \cdots \binom{d}{d-1}$ for all d. Also, since for all d we can construct a matrix $A \in \mathscr{N}_d(k)$ of rank $d - 1$, the tuple $(A, \ldots, A) \in \mathscr{N}^{d-1}(k^d)$ satisfies $A^{d-1} \neq 0$ and hence we have the lower bound $N(d) \geqslant d$. In particular, we conclude that $N(2) = 2$, and for higher dimensions we get the bounds $3 \leqslant N(3) \leqslant 9$ and $4 \leqslant N(4) \leqslant 96$. We end this section by finding a better bound for $N(3)$.

PROPOSITION 2.4. — *For any field k, we have $N(3) \leqslant 5$. In addition, if $\operatorname{char} k \neq 2$, then $N(3) = 5$.*

To prove this, we need a lemma:

LEMMA 2.5. — *Let $(C, B, A) \in \mathscr{N}^3(k^3)$. If $\operatorname{rank} B = 1$, then $AB = \lambda B$ or $BC = \lambda B$ for some $\lambda \in k$.*

Proof. — Assume that

$$B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} p & q & r \\ s & t & u \\ v & w & x \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 0 & 0 & b \\ 0 & 0 & e \\ 0 & 0 & h \end{pmatrix} \quad \text{and} \quad BC = \begin{pmatrix} 0 & 0 & 0 \\ v & w & x \\ 0 & 0 & 0 \end{pmatrix}.$$

The nilpotence of AB and BC implies $h = \operatorname{Tr} AB = w = \operatorname{Tr} BC = 0$. Then

$$ABC = \begin{pmatrix} bv & 0 & bx \\ ev & 0 & ex \\ 0 & 0 & 0 \end{pmatrix},$$

and by the nilpotence of ABC, $bv = \operatorname{Tr} ABC = 0$. The case $b = 0$ is $AB = eB$ and the case $v = 0$ is $BC = xB$. □

Corollary 2.6. — *If* $(C, B, A) \in \mathscr{N}^3(k^3)$ *and* $\operatorname{rank}(B) \leqslant 1$, *then* $ABC = 0$.

Proof. — Assume that $\operatorname{rank}(B) = 1$, and $AB, BC \neq 0$. By Lemma 2.5 and after rescaling A or C, we may suppose that $BC = B$ or $AB = B$. In the first case we will have $(C^2, B, A) \in \mathscr{N}^3(k^3)$, and by Corollary 2.2, $ABC = A(BC)C = ABC^2 = 0$. For the case $AB = B$, applying a similar argument to the tuple (A^t, B^t, C^t) of the transposes of A, B, C, we will obtain $(ABC)^t = C^t B^t (A^t)^2 = 0$, and hence $ABC = 0$. □

Proof of Proposition 2.4. — Let $(E, D, C, B, A) \in \mathscr{N}^5(k^3)$. Then (E, BCD, A) belongs to $\mathscr{N}^3(k^3)$, and by Lemma 2.3 with $d = 3, r = 2$, $\operatorname{rank}(BCD) \leqslant 1$. Then, by Corollary 2.6, $ABCDE = 0$ and $N(3) \leqslant 5$. Moreover, when $\operatorname{char} k \neq 2$, it is a straightforward computation to show that $(D, C, B, A) \in \mathscr{N}^4(k^3)$, with

$$A = \begin{pmatrix} -2 & -6 & 1 \\ 3 & 9 & 16 \\ -1 & -3 & -7 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 4 \\ -1 & -1 & -2 \end{pmatrix}, \quad D = \begin{pmatrix} -1 & 3 & 16 \\ 1 & -3 & -16 \\ 1 & 2 & 4 \end{pmatrix}$$

and

$$ABCD = \begin{pmatrix} 4 & 8 & 16 \\ -6 & -12 & -24 \\ 2 & 4 & 8 \end{pmatrix} \neq 0. \qquad \square$$

Remark 2.7. — This last proposition shows that, in general we cannot expect $N(d) = d$. For that reason, the hypothesis of Theorem 1.3 does not imply any kind of simultaneous triangularization or nilpotency of the semigroup generated by sequences in $\mathscr{N}^n(k^d)$. In fact, it is not hard to prove that the matrices in the last example we gave in $\mathscr{N}^4(k^3)$ are not simultaneously triangularizable, since A and B do not have a common invariant subspace of dimension 1.

3. A polynomial identity

For the proof of Theorem 1.2 we need some notation. Let k be a field with algebraic closure $\overline{k}$. For $d, N \in \mathbb{N}$, consider Nd^2 variables $x_{i,j}$ with $1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant d^2$ and let $R_{d,N}$ be the polynomial ring $k[x_{i,j}]$. If $A_1, \ldots, A_N \in M_d(\overline{k})$ and $f \in R_{d,N}$,

by $f(A_1, \ldots, A_N)$ we mean the element $f((a_{i,j})_{i,j})$ where $(a_{i,j})_j$ are the coefficients of A_i in some fixed order.

Recall that a polynomial $f \in k[y_1, \ldots, y_m]$ is *homogeneous* of degree $\lambda \geqslant 0$ if it is of the form $\sum_{i_1+\cdots+i_m=\lambda} c_{i_1\ldots i_m} y_1^{i_1} \ldots y_m^{i_m}$ for some $c_{i_1\ldots i_m} \in k$, $i_1, \ldots, i_m \geqslant 0$. We say that a monomial $f \in R_{d,N}$ is *multihomogeneous* of *multidegree* $\operatorname{mdeg} f = (\lambda_1, \ldots, \lambda_N) \in \mathbb{N}^N$ if it is of the form $f((x_{i,j})_{i,j}) = c \prod_{i,j} x_{i,j}^{u_{i,j}}$, where $c \in k$, $u_{i,j} \geqslant 0$ and $\sum_j u_{i,j} = \lambda_i$ for all $1 \leqslant i \leqslant N$, and that a polynomial $p \in R_{d,N}$ is multihomogeneous of multidegree $\operatorname{mdeg} p$ if it is a finite sum of multihomogeneous monomials of multidegree $\operatorname{mdeg} p$. This is equivalent to say that, for each $1 \leqslant i \leqslant N$, p is homogeneous of degree λ_i in the variables $x_{i,1} \ldots x_{i,d^2}$. Note there is a direct sum decomposition

$$(6) \qquad\qquad R_{d,N} = \bigoplus_{\lambda \in \mathbb{N}^N} R_{d,N,\lambda},$$

where $R_{d,N,\lambda}$ denotes the vector space of multihomogeneous polynomials of multidegree λ.

For $1 \leqslant j \leqslant d^2$ denote by f_j the polynomial in $R_{d,N}$ representing the map that sends the N-tuple $(A_1, \ldots, A_N) \in \overline{k}^{Nd^2}$ to the j-th entry of $A_N \cdots A_1$. Also, for $1 \leqslant \ell \leqslant d$ and $1 \leqslant \alpha \leqslant \beta \leqslant N$, let $T_{\alpha,\beta}^\ell \in R_{d,N}$ be the polynomial that represents the map $(A_1, \ldots, A_N) \mapsto \operatorname{Tr} \Lambda^\ell(A_\beta \cdots A_\alpha)$.

It is not hard to see that f_j are multihomogeneous of multidegree $(1, 1, \ldots, 1, 1)$ and that $T_{\alpha,\beta}^\ell$ are multihomogeneous of multidegree $(0, \ldots, 0, \ell, \ldots, \ell, 0, \ldots, 0)$, with the ℓ's in positions $\alpha, \alpha+1, \ldots, \beta$.

Our purpose is to prove the following:

THEOREM 3.1. — *If $N = N(d)$ is given by Theorem 1.3, there is some $r \leqslant (Nd+1)^{Nd^2+2}$ such that for all $1 \leqslant j \leqslant d^2$ there exist multihomogeneous polynomials $p_{j,\ell}^{\alpha,\beta} \in R_{d,N}$ of multidegree $r \operatorname{mdeg} f_j - \operatorname{mdeg} T_{\alpha,\beta}^\ell \in \mathbb{N}^N$ such that*

$$(7) \qquad\qquad (f_j)^r = \sum_{\alpha,\beta,\ell} p_{j,\ell}^{\alpha,\beta} T_{\alpha,\beta}^\ell.$$

The natural tool to prove this result is Hilbert's Nullstellensatz. If $I \subset k[y_1, \ldots, y_m]$ is an ideal, let $Z(I)$ be its zero locus in $\overline{k}^m$. Also, for $Z \subset \overline{k}^m$, let $I(Z) \in k[y_1, \ldots, y_m]$ be the ideal of polynomials f that vanish on Z. The effective version of Nullstellensatz that we will use comes from applying Rabinowitsch's proof of Nullstellensatz by assuming weak Nullstellensatz (see e.g. [10, §1.7]) and an effective version of weak Nullstellensatz [28]:

THEOREM 3.2 (Effective Nullstellensatz). — *If $I = \subset k[y_1, \ldots, y_m]$ is an ideal, then $I(Z(I))$ is equal to $\sqrt{I}$, the radical ideal of I. Moreover, if $g \in \sqrt{I}$ and I is generated by polynomial $f_1, \ldots, f_s$ satisfying $\max(\deg f_1, \ldots, \deg f_s, \deg g) = k$, then there is some $r \leqslant (k+1)^{m+2}$ and $p_1, \ldots, p_s \in k[y_1, \ldots, y_m]$ such that*

$$g^r = p_1 f_1 + p_2 f_2 + \cdots + p_s f_s.$$

Proof of Theorem 3.1. — Let $I \subset R_{d,N}$ be the ideal generated by the polynomials $T_{\alpha,\beta}^\ell$ and let $W = Z(I)$. Note that for a matrix A of order $d \times d$, the non leading coefficients of its characteristic polynomial are precisely $(-1)^\ell \operatorname{Tr} \Lambda^\ell(A)$, with $1 \leqslant \ell \leqslant d$. By this observation, the set W is precisely the set of N-tuples $(A_1, \ldots, A_N) \in \mathcal{N}_d^N(\overline{k})$. Hence, by our choice of N, Theorem 1.3 guarantees us that $f_j(P) = 0$ for all $P \in W$. Then Nullstellensatz applies and $f_j \in I(Z(W)) = \sqrt{I}$.

Since $\max(\deg f_j, \deg T_{\alpha,\beta}^\ell) = Nd$, by Theorem 3.2 there is some $r \leqslant (Nd+1)^{Nd^2+2}$ and polynomials $p_{j,\ell}^{\alpha,\beta,\gamma} \in R_{d,N}$ satisfying (7) for all j. Finally, by the direct sum decomposition (6) and by comparing multidegrees, we may assume that $p_{j,\ell}^{\alpha,\beta,\gamma}$ are multihomogeneous of multidegree $r \operatorname{mdeg} f_j - \operatorname{mdeg} T_{\alpha,\beta,\gamma}^\ell$. $\qquad\square$

4. Proof of Theorem 1.2

Theorem 3.1 is the fundamental relation that we will need to prove inequality (3). For the next we will assume that k is a *local field.* That is, a field together with an absolute value $|.| : k \to \mathbb{R}^+$ that induces a non-discrete locally compact topology on k via the induced metric. Examples of these include $\mathbb{R}, \mathbb{C}$ with the standard absolute values and fields of p-adic numbers $\mathbb{Q}_p$ for a prime p. For more information about local fields, see [19].

We will work on the finite dimensional vector space k^d, where k is a local field with absolute value $|.|$. In this situation, we consider the norm on $M_d(k)$ given by $\|A\|_0 = \max_{1 \leqslant j \leqslant d^2} |a_j|$, where a_j are the entries of A. Since the absolute value on k extends in a unique way to an absolute value on $\overline{k}$ (see Lang's Algebra [18, XII.2, Prop. 2.5]), the spectral radius of a matrix $A \in M_d(k)$ is then defined in the usual way. The *height* $h(f)$ of a polynomial $f \in k[y_1, \ldots, y_m]$ is defined as the logarithm of the maximum modulus of its coefficients.

We begin with a lemma.

LEMMA 4.1. — *If $f \in R_{d,N}$ is a multihomogeneous polynomial of multidegree $(\lambda_1, \ldots, \lambda_N)$ and height $h(f) \leqslant H$, then*

$$|f(A_1, \ldots, A_N)| \leqslant e^H \prod_{i=1}^N \binom{d^2 - 1 + \lambda_i}{\lambda_i} \|A_1\|_0^{\lambda_1} \cdots \|A_N\|_0^{\lambda_N}$$

for all $A_1, \ldots, A_N \in M_d(k)$.

Proof. — If f is a multihomogeneous monomial with $h(f) \leqslant H$, then $f(X_1, \ldots, X_N) = c \prod_{i=1}^N \prod_{j=1}^{\lambda_i} X_{i,\ell_{i,j}}$, for some $1 \leqslant \ell_{i,j} \leqslant d^2$ and $c \in k$ with $|c| \leqslant e^H$. So, given $A_1, \ldots, A_N \in M_d(k)$,

$$|f(A_1, \ldots, A_N)| = |c| \prod_{i=1}^N \prod_{j=1}^{\lambda_i} |A_{i,\ell_{i,j}}| \leqslant e^H \prod_{i=1}^N \|A_i\|_0^{\lambda_i}.$$

The lemma then follows by noting that a multihomogeneous polynomial of multidegree $(\lambda_1, \ldots, \lambda_N)$ is sum of at most $\prod_{i=1}^N \binom{d^2-1+\lambda_i}{\lambda_i}$ multihomogeneous monomials of the same multidegree. $\qquad\square$

Proof of Theorem 1.2. — Let $N = N(d)$ and $r > 1$ be given by Theorems 1.3 and 3.1 respectively, and consider first $n = N$ and the norm $\|.\|_0$. Let $A_1, \ldots, A_N \in M_d(k)$. First, note that for $1 \leqslant \alpha \leqslant \beta \leqslant N$ and $1 \leqslant \ell \leqslant d$, $T^\ell_{\alpha,\beta}(A_1, \ldots, A_N)$ is the ℓ-th symmetric polynomial evaluated at the eigenvalues of $A_\beta \cdots A_\alpha$. Hence we have $|T^\ell_{\alpha,\beta}(A_1, \ldots, A_N)| \leqslant \binom{d}{\ell} \rho(A_\beta \cdots A_\alpha)^\ell$. Also, as the polynomials $p^{\alpha,\beta}_{j,\ell}$ in the statement of Theorem 3.1 have multidegree $(r, \ldots, r, r - \ell, \ldots, r - \ell, r, \ldots, r)$, by Lemma 4.1 we have

$$|p^{\alpha,\beta}_{j,\ell}(A_1, \ldots, A_N)| \leqslant e^H \binom{d^2 - 1 + r}{r}^{N - \beta + \alpha - 1} \binom{d^2 - 1 + r - \ell}{r - \ell}^{\beta - \alpha + 1}$$
$$\times \left(\prod_{s=1}^N \|A_s\|_0 \right)^r \left(\prod_{t=\alpha}^\beta \|A_t\|_0 \right)^{-\ell}$$

for all j, α, β, ℓ. Thus, from (7) we obtain the following:

$$\|A_N \cdots A_1\|_0^r = \max_j |f_j(A_1, \ldots, A_N)|^r$$
$$\leqslant \max_j \sum_{\alpha, \beta, \ell} |p^{\alpha,\beta}_{j,\ell}(A_1, \ldots, A_N)||T^\ell_{\alpha,\beta}(A_1, \ldots, A_N)|$$
$$\leqslant C_1 \left(\prod_{i=1}^N \|A_i\|_0 \right)^r \max_{\alpha, \beta, \ell} \left(\frac{\rho(A_\beta \cdots A_\alpha)}{\prod_{t=\alpha}^\beta \|A_t\|_0} \right)^\ell,$$

for

$$C_1 = e^H \sum_{\alpha, \beta, \ell} \binom{d^2 - 1 + r}{r}^{N - \beta + \alpha - 1} \binom{d^2 - 1 + r - \ell}{r - \ell}^{\beta - \alpha + 1} \binom{d}{\ell}.$$

Now, let

$$\Lambda = \max_{\alpha, \beta} \left(\frac{\rho(A_\beta \cdots A_\alpha)}{\prod_{t=\alpha}^\beta \|A_t\|_0} \right).$$

An easy computation shows that $\|AB\|_0 \leqslant d\|A\|_0 \|B\|_0$ for all $A, B \in M_d(k)$. Moreover, by the Gelfand's formula $\rho(A) = \lim_{n \to \infty} \|A^n\|_0^{1/n}$ we obtain $\rho(A) \leqslant d\|A\|_0$ for all $A \in M_d(k)$. These facts together imply that $\Lambda \leqslant d^{N+1}$, and hence $\Lambda^d \leqslant d^{(N+1)(d-1)} \Lambda$. Also, depending on whether Λ is greater than 1 or not, we have $\Lambda^\ell \leqslant \max(\Lambda, \Lambda^d) \leqslant d^{(N+1)(d-1)} \Lambda$ for all $1 \leqslant \ell \leqslant d$. Thus we conclude

$$\|A_N \cdots A_1\|_0^r \leqslant d^{(N+1)(d-1)} C_1 \left(\prod_{i=1}^N \|A_i\|_0 \right)^r \cdot \Lambda.$$

Applying r-th root to the last inequality, we obtain (3) with

$$C = C(\|.\|_0) := C_1^{1/r} d^{(N+1)(d-1)/r}.$$

Now, consider an arbitrary submultiplicative norm $\|.\|$ on $M_d(k)$. Since in a finite dimensional vector space over a local field all norms are equivalent [18, XII.2, Prop. 2.2], there is some $D \geqslant 1$ such that $D^{-1}\|A\| \leqslant \|A\|_0 \leqslant D\|A\|$ for all $A \in M_d(k)$.

This implies

$$\|A_N \cdots A_1\| \leqslant D\|A_N \cdots A_1\|_0$$

$$\leqslant DC\bigg(\prod_{1 \leqslant i \leqslant N} \|A_i\|_0\bigg) \max_{1 \leqslant \alpha \leqslant \beta \leqslant N} \bigg(\frac{\rho(A_\beta \cdots A_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|_0}\bigg)^{1/r}$$

$$\leqslant D^{N+1}C\bigg(\prod_{1 \leqslant i \leqslant N} \|A_i\|\bigg) \max_{1 \leqslant \alpha \leqslant \beta \leqslant N} \bigg(\frac{\rho(A_\beta \cdots A_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|}\bigg)^{1/r}$$

and proves the statement for $n = N$. For a general $n \geqslant N$ the result follows by applying (3) to the sequence $A_1, \ldots, A_{N-1}, A_N A_{N+1} \cdots A_n$, and then using the submultiplicativity of $\|.\|$. □

4.1. THE CASE OF THE COMPLEX NUMBERS. — When the base field is $k = \mathbb{C}$ we can say a little more. Recall that for a norm $\|.\|$ on $\mathbb{C}^d$, the *operator norm* on $M_d(\mathbb{C})$ (also denoted by $\|.\|$) is defined by $\|A\| = \sup_{v \in \mathbb{C}^d \setminus \{0\}} \|Av\|/\|v\|$.

PROPOSITION 4.2. — *For $d \in \mathbb{N}$ and $N(d)$ given by Theorem 1.2, there is a constant $C = C(d) > 1$ such that if $n \geqslant N$, the inequality (3) holds for all operator norms $\|.\|$ on $M_d(\mathbb{C})$ and $A_1, \ldots, A_n \in M_d(\mathbb{C})$.*

We will need the following lemma which is a consequence of John's ellipsoid theorem [27, Th. 10.12.2] (see also [4, Lem. 3.2]):

LEMMA 4.3. — *For all $d \in \mathbb{N}$ and for every two operator norms $\|.\|$ and $\|.\|_1$ on $M_d(\mathbb{C})$ there exists some $S \in \mathrm{GL}_d(\mathbb{C})$ such that for every $A \in M_d(\mathbb{C})$:*

$$d^{-1}\|A\| \leqslant \|SAS^{-1}\|_1 \leqslant d\|A\|.$$

Proof of Proposition 4.2. — Since operator norms are submultiplicative, we may assume that $n = N$. Fix the ℓ^1-operator norm $\|A\|_1 = \max_j \sum_{i=1}^d |a_{ij}|$ on $M_d(\mathbb{C})$, where a_{ij} are the entries of A. Since $\|A\|_1 \leqslant \|A\|_0 \leqslant d\|A\|_1$ for all A, by the proof of Theorem 1.2 we have $C(\|.\|_1) = d^{N+1}C(\|.\|_0)$ with $C(\|.\|_0)$ as in the end of the proof of Theorem 1.2. If $\|.\|$ is an arbitrary operator norm on $M_d(\mathbb{C})$, let $S \in \mathrm{GL}_d(\mathbb{C})$ be relating $\|.\|$ and $\|.\|_1$ as in Lemma 4.3.

Given $A_1, \ldots, A_N \in M_d(\mathbb{C})$ let $B_i = SA_iS^{-1}$ for all i. We have

$$\|A_N \cdots A_1\| \leqslant d\|B_N \cdots B_1\|_1$$

$$\leqslant d^{N+2}C(\|.\|_0)\bigg(\prod_{1 \leqslant i \leqslant N} \|B_i\|_1\bigg) \max_{1 \leqslant \alpha \leqslant \beta \leqslant N} \bigg(\frac{\rho(B_\beta \cdots B_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|B_i\|_1}\bigg)^{1/r}$$

$$\leqslant d^{2N+2}C(\|.\|_0)\bigg(\prod_{1 \leqslant i \leqslant N} \|A_i\|\bigg) \max_{1 \leqslant \alpha \leqslant \beta \leqslant N} \bigg(\frac{d^{\beta-\alpha+1}\rho(A_\beta \cdots A_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|}\bigg)^{1/r}$$

$$\leqslant d^{3N+2}C(\|.\|_0)\bigg(\prod_{1 \leqslant i \leqslant N} \|A_i\|\bigg) \max_{1 \leqslant \alpha \leqslant \beta \leqslant N} \bigg(\frac{(\rho(A_\beta \cdots A_\alpha)}{\prod_{\alpha \leqslant i \leqslant \beta} \|A_i\|}\bigg)^{1/r}.$$

It is clear that $C(d) := d^{3N+2}C(\|.\|_0)$ does not depend on $\|.\|$. □

Proposition 4.2 allows us to conclude the following inequality:

Theorem 4.4. — *Given $d \in \mathbb{N}$, let $C(d)$ be as in Proposition 4.2. Then the following inequality is valid for all bounded sets $\mathscr{M} \subset M_d(\mathbb{C})$:*

$$\mathfrak{R}(\mathscr{M}) \leqslant C(d) \max_{1 \leqslant j \leqslant N(d)} \left(\sup \{ \rho(A_1 \cdots A_j) : A_i \in \mathscr{M} \} \right)^{1/j}.$$

This inequality was first proved by Bochi in [4] (without giving and effective bound on $N(d)$), and it has Theorem 1.1 as an immediate consequence. In [5, Lem. 2.1], Breuillard gave another proof of this inequality valid for arbitrary local fields: in the Archimedean case the same conclusion of Theorem 4.4 holds with $N(d) = d^2$, while for a non Archimedean field k and for every bounded set $\mathscr{M} \subset M_d(k)$ we have the identity

$$\inf_{S \in \mathrm{GL}_d(k)} \|S \mathscr{M} S^{-1}\|_0 = \max_{1 \leqslant j \leqslant d^2} \left(\sup \{ \rho(A_1 \cdots A_j) : A_i \in \mathscr{M} \} \right)^{1/j}.$$

Breuillard used this result to study semigroups of invertible matrices.

Proof of Theorem 4.4. — For $1 \leqslant j \leqslant N(d)$, define $\rho_j = \sup \{ \rho(A_1 \cdots A_j) : A_i \in \mathscr{M} \}$. For an arbitrary operator norm $\|.\|$ on $M_d(\mathbb{C})$, take supremum for $A_i \in \mathscr{M}$ in both sides of (3). We obtain

$$\begin{aligned}
\mathfrak{R}(\mathscr{M})^N &\leqslant \sup_{A_i \in \mathscr{M}} \|A_N \cdots A_1\| \\
(8) \qquad &\leqslant C(d) \max_{1 \leqslant j \leqslant N} \left(\sup_{A \in \mathscr{M}} \|A\|^{N-j/r} \cdot (\rho_j)^{1/r} \right).
\end{aligned}$$

Now, recall that $\mathfrak{R}(\mathscr{M}) = \inf_{\|.\|} \sup_{A \in \mathscr{M}} \|A\|$, where the infimum is taken over all operator norms on $M_d(\mathbb{C})$ (for a proof, see [26]), and let $\|.\|_n$ be a sequence of operator norms on $M_d(\mathbb{C})$ such that $\sup_{A \in \mathscr{M}} \|A\|_n \to \mathfrak{R}(\mathscr{M})$. Taking a subsequence, we may assume that for all $\|.\|_n$, the maximum in the right hand side of (8) is achieved by the same index $j \in \{1, \ldots, N\}$. Then, taking limit as n tends to infinity in (8) we will have

$$(9) \qquad \mathfrak{R}(\mathscr{M})^N \leqslant C(d)(\rho_j)^{1/r} \cdot \mathfrak{R}(\mathscr{M})^{N-j/r}$$

(here is where we use Proposition 4.2 since $C(d)$ does not depend on n). If $\mathfrak{R}(\mathscr{M}) = 0$ the conclusion is obvious. Otherwise, dividing by $\mathfrak{R}(\mathscr{M})^{N-j/r}$ and taking j/r-th root in (9) we obtain the desired inequality. $\qquad \square$

Remark 4.5. — In the proof of Theorem 1.2 the constant $C(\|.\|_0)$ depends on H, the maximum on the heights of the polynomials $p_{j,\ell}^{\alpha,\beta}$. In the case of the complex numbers we can give an effective upper bound on H by means of the effective arithmetic Nullstellensatz [17, Th. 1]. Applied to our case, we obtain $r \leqslant 4(Nd^2 + 1)(dN)^{Nd^2+1}$ and

$$H \leqslant \log 2(Nd^2 + (N(d^2+1))r) + r(Nd^2+2)(\log(d\binom{N+1}{2}+1) + (Nd^2+8)\log(Nd^2+2)dN).$$

Hence

$$C(d) = d^{3N+2}C(\|.\|_0)$$
$$= e^{H/r}d^{3N+2+(N+1)(d-1)/r}\left(\sum_{\alpha,\beta,\ell}\binom{d^2-1+r}{r}^{N-\beta+\alpha-1}\binom{d^2-1+r-\ell}{r-\ell}^{\beta-\alpha+1}\binom{d}{\ell}\right)^{1/r}.$$

5. ERGODIC-THEORETICAL CONSEQUENCES

For the proof of Theorem 1.4, we will need the following result which may be seen as a quantitative version of Poincaré's Recurrence theorem for measure preserving transformations. It is a consequence of Birkhoff Ergodic Theorem, and the fact that for a measurable set U of positive measure, for almost all points x in U, the frequency of points of the sequence $x, Tx, T^2x, \ldots$ that belong to U is positive (compare with the subbaditive ergodic theorem of Karlsson-Gouëzel [12, Th. 1.1]). For a detailed proof, see [3, Lem. 3.12].

LEMMA 5.1. — *Let $T : X \to X$ be a measure preserving map over the probability space $(X, \mathscr{F}, \mu)$, and let $U \in \mathscr{F}$ have positive measure. Given $\gamma > 0$, there exists a measurable map $N_0 : U \to \mathbb{N}$ such that, for μ-a.e. $x \in U$ and for every $n \geqslant N_0(x)$ and $t \in [0,1]$ there is some $\ell \in \{1, \ldots, n\}$ with $T^\ell(x) \in U$ and $|(\ell/n) - t| < \gamma$.*

Proof of Theorem 1.4. — Fix an operator norm $\|.\|$ on $M_d(k)$, and let

$$Y = \{x \in X : \lambda(x) \in \mathbb{R}\}.$$

This is a measurable T-invariant set, and since $\rho(A) \leqslant \|A\|$ for all $A \in M_d(k)$, we have that both sides of (4) equal $-\infty$ for μ-almost all $x \in X \backslash Y$. So we only have to check the result μ-a.e. in Y.

Assume the contrary. That is, assume the existence of some $\varepsilon > 0, K \in \mathbb{N}$ and a measurable set $U \subset Y$ of positive measure such that, for all $x \in U$, if $n \geqslant K$, then $\log \rho(A^n(x))/n + \varepsilon \leqslant \lambda(x)$. By Egorov's theorem, and restricting to a smaller subset if necessary, we may assume that on U, $\log \|A^n(x)\|/n$ converges uniformly to $\lambda(x)$.

Let N, r and C be as in the statement of Theorem 1.2 and let $\varepsilon' = \varepsilon/(2 + 6Nr)$. By the uniform convergence assumption, there is some $M \geqslant 1$ such that, $n \geqslant M$ implies

$$(10) \qquad |\log \|A^n(x)\| - n\lambda(x)| < n\varepsilon' \text{ for all } x \in U.$$

Take $x \in U$ and $N_0(x) \in \mathbb{N}$ such that Lemma 5.1 holds with $\gamma = 1/3N$, and let $n \geqslant \max(3NM, 3NK, 3Nr \log C/\varepsilon', N_0(x))$. Let $m_0 = 0$, and given $1 \leqslant i \leqslant N$ let m_i be such that $1 \leqslant m_i \leqslant n$ and

$$(11) \qquad \left|\frac{m_i}{n} - \frac{i}{N}\right| < \frac{1}{3N}$$

and $T^{m_i}x \in U$. We have

$$m_i - m_{i-1} > (in/N - n/3N) - ((i-1)/N + n/3N) = n/3N \geqslant \max(M, K, r \log C/\varepsilon')$$

for all $1 \leqslant i \leqslant N$.

Now apply Theorem 1.2 to $A_i = A^{m_i - m_{i-1}}(T^{m_{i-1}} x)$. By the cocycle relation, we obtain $A_N \cdots A_1 = A^{m_N}(x)$, and hence

$$(12) \quad \log \|A^{m_N}(x)\| \leqslant \log C + \sum_{i=1}^{N} \log \|A^{m_i - m_{i-1}}(T^{m_{i-1}} x)\|$$

$$+ \frac{1}{r} \left(\log \rho(A^{m_\beta - m_{\alpha-1}}(T^{m_{\alpha-1}} x)) - \sum_{i=\alpha}^{\beta} \log \|A^{m_i - m_{i-1}}(T^{m_{i-1}} x)\| \right)$$

for some $1 \leqslant \alpha \leqslant \beta \leqslant N$. But, by definition, $T^{m_i} x \in U$ for all i, and as $m_i - m_{i-1} \geqslant M$, (10) applies. Combining it with (12) we have

$$\log \rho(A^m{}_{()}\beta - m_{\alpha-1} T^{m_{\alpha-1}} x) \geqslant \sum_{i=\alpha}^{\beta} \log \|A^{m_i - m_{i-1}}(T^{m_{i-1}} x)\|$$

$$+ r \left(\log \|A^{m_N}(x)\| - \sum_{i=1}^{N} \log \|A^{m_i - m_{i-1}}(T^{m_{i-1}} x)\| - \log C \right)$$

$$> (m_\beta - m_{\alpha-1})(\lambda(x) - \varepsilon') - r(\log C + 2\varepsilon' m_N)$$

$$= (m_\beta - m_{\alpha-1})\lambda(x) - (\varepsilon'((m_\beta - m_{\alpha-1}) + 2rm_N)) + r \log C).$$

On the other hand, by (11) we have

$$\frac{m_N}{m_\beta - m_{\alpha-1}} < \frac{n}{n(\beta - \alpha + 1)/N - 2n/3N} \leqslant 3N.$$

But, since $T^{m_{\alpha-1}} x \in U$, and $(m_\beta - m_{\alpha-1}) \geqslant \max(K, r \log C / \varepsilon')$ we conclude

$$\frac{\varepsilon'((m_\beta - m_{\alpha-1}) + 2r(m_N)) + r \log C}{m_\beta - m_{\alpha-1}} = \varepsilon' + \frac{\varepsilon' 2r(m_N)}{(m_\beta - m_{\alpha-1})} + \frac{r \log C}{(m_\beta - m_{\alpha-1})}$$

$$\leqslant \varepsilon' + \varepsilon' 6Nr + \frac{r \log C}{(m_\alpha - m_{\alpha-1})}$$

$$\leqslant (2 + 6Nr)\varepsilon' = \varepsilon.$$

This is the desired contradiction and the proof is complete. $\qquad \square$

6. Geometric remarks

We can observe that the main ingredients of the proof of Theorem 1.4 are Theorem 1.2 and Poincaré's recurrence Theorem. Therefore, if we have another situation where an analogue of inequality (3) holds, then we should obtain a result similar to Theorem 1.4. This is the case of cocycles of isometries of Gromov hyperbolic spaces. For definition and further properties of Gromov hyperbolicity see [6, 7, 8].

As it was proved in [24, Th. 1.2], if M is a Gromov hyperbolic space with distance d, then there is a constant $C > 0$ such that, for all $o \in M$ and f, g isometries of M we have

$$d(fgo, o) \leqslant C + \max\left(d(fo, o) + d^\infty(g), d^\infty(f) + d(go, o), \frac{d(fo, o) + d(go, o) + d^\infty(fg)}{2} \right),$$

where $d^\infty(h) = \lim_{n \to \infty} d(h^n o, o)/n$ is the stable length.

In this context, given a probability space $(X, \mathscr{F}, \mu)$ and a measure preserving map $T : X \to X$, a *cocycle of isometries* of M is a measurable map $A : X \to Isom(M)$, where $Isom(M)$ is the group of isometries of M, endowed with the Borel $\sigma-$algebra induced by the compact-open topology. We say that the cocycle A is *integrable* if the map $x \mapsto d(A(x)o, o)$ is integrable for some (and hence all) $o \in M$. In the same way as for linear cocycles, we define the family of maps $A^n : X \to Isom(M)$. For references about cocycles of isometries, see e.g. [12, 15].

Following the same steps of the proof of Theorem 1.4, we can obtain the following:

PROPOSITION 6.1. — *Let M be a Gromov hyperbolic space, $o \in M$, and let T be a measure-preserving transformation of a probability space $(X, \mathscr{F}, \mu)$. Also, let $A : X \to Isom(M)$ be an integrable cocycle of isometries of M. Then for μ-almost all $x \in X$ and we have the following limits exist in $\mathbb{R}_0^+$ and are equal:*

$$\limsup_{n \to \infty} \frac{d^\infty(A^n(x))}{n} = \lim_{n \to \infty} \frac{d(A^n(x)o, o)}{n}.$$

A result similar to Proposition 6.1 is far from being true if we do not assume a negative curvature condition on M.

EXAMPLE 6.2. — Let $X = \mathbb{S}^1$ and μ be the Lebesgue measure on X. If $T(z) = z^2$ is the doubling map on X, which preserves μ, and $R_a(p) = p + a$ is the translation by $a \neq 0$ on $\mathbb{R}^2$, define a cocycle $A : \mathbb{S}^1 \to Isom(\mathbb{R}^2)$ by $A(z)p = T(z)R_a(z^{-1}p)$ for all $p \in \mathbb{R}^2$. Note that $A^n(z)p = T^n(z)R_a^n(z^{-1}p)$ and hence the limit $\lim_{n \to \infty} d(A^n(z)p, p)/n$ exists and equals $|a| > 0$ for all $z \in \mathbb{S}^1$ and $p \in \mathbb{R}^2$. On the other hand, if z is not a periodic point for T, then $A^n(z)$ is not a translation and hence has a fixed point. Thus we have that $d^\infty(A^n(z)) = 0$ for all $n \in \mathbb{N}$ and all z in the set of non periodic point of T, which is a full measure set with respect to μ.

References

[1] A. Avila & J. Bochi – "A formula with some applications to the theory of Lyapunov exponents", *Israel J. Math.* **131** (2002), p. 125–137.

[2] M. A. Berger & Y. Wang – "Bounded semigroups of matrices", *Linear Algebra Appl.* **166** (1992), p. 21–27.

[3] J. Bochi – "Genericity of zero Lyapunov exponents", *Ergodic Theory Dynam. Systems* **22** (2002), no. 6, p. 1667–1696.

[4] ———, "Inequalities for numerical invariants of sets of matrices", *Linear Algebra Appl.* **368** (2003), p. 71–81.

[5] E. Breuillard – "A height gap theorem for finite subsets of $\mathrm{GL}_d(\overline{\mathbb{Q}})$ and nonamenable subgroups", *Ann. of Math. (2)* **174** (2011), no. 2, p. 1057–1110.

[6] M. R. Bridson & A. Haefliger – *Metric spaces of non-positive curvature*, Grundlehren Math. Wiss., vol. 319, Springer-Verlag, Berlin, 1999.

[7] M. Coornaert, T. Delzant & A. Papadopoulos – *Géométrie et théorie des groupes. Les groupes hyperboliques de Gromov*, Lect. Notes in Math., vol. 1441, Springer-Verlag, Berlin, 1990.

[8] T. Das, D. Simmons & M. Urbański – *Geometry and dynamics in Gromov hyperbolic metric spaces. With an emphasis on non-proper settings*, Math. Surveys and Monographs, vol. 218, American Mathematical Society, Providence, RI, 2017.

[9] L. Elsner – "The generalized spectral-radius theorem: an analytic-geometric proof", *Linear Algebra Appl.* **220** (1995), p. 151–159, Proceedings of the Workshop "Nonnegative Matrices, Applications and Generalizations" and the Eighth Haifa Matrix Theory Conference (Haifa, 1993).

[10] W. Fulton – *Algebraic curves. An introduction to algebraic geometry*, Advanced Book Classics, Addison-Wesley Publishing Company, Redwood City, CA, 1989.

[11] I. Goldhirsch, P.-L. Sulem & S. A. Orszag – "Stability and Lyapunov stability of dynamical systems: a differential approach and a numerical method", *Phys. D* **27** (1987), no. 3, p. 311–337.

[12] S. Gouëzel & A. Karlsson – "Subadditive and multiplicative ergodic theorems", *J. Eur. Math. Soc. (JEMS)* (to appear).

[13] L. Gurvits – "Stability of discrete linear inclusion", *Linear Algebra Appl.* **231** (1995), p. 47–85.

[14] R. Jungers – *The joint spectral radius. Theory and applications*, Lect. Notes in Control and Information Sci., vol. 385, Springer-Verlag, Berlin, 2009.

[15] A. Karlsson & G. A. Margulis – "A multiplicative ergodic theorem and nonpositively curved spaces", *Comm. Math. Phys.* **208** (1999), no. 1, p. 107–123.

[16] V. Kozyakin – "The Berger-Wang formula for the Markovian joint spectral radius", *Linear Algebra Appl.* **448** (2014), p. 315–328.

[17] T. Krick, L. M. Pardo & M. Sombra – "Sharp estimates for the arithmetic Nullstellensatz", *Duke Math. J.* **109** (2001), no. 3, p. 521–598.

[18] S. Lang – *Algebra*, third ed., Graduate Texts in Math., vol. 211, Springer-Verlag, New York, 2002.

[19] F. Lorenz – *Algebra. Vol. II: Fields with structure, algebras and advanced topics*, Universitext, Springer, New York, 2008.

[20] R. McNaughton & Y. Zalcstein – "The Burnside problem for semigroups", *J. Algebra* **34** (1975), p. 292–299.

[21] I. D. Morris – "The generalised Berger-Wang formula and the spectral radius of linear cocycles", *J. Funct. Anal.* **262** (2012), no. 3, p. 811–824.

[22] ———, "Mather sets for sequences of matrices and applications to the study of joint spectral radii", *Proc. London Math. Soc. (3)* **107** (2013), no. 1, p. 121–150.

[23] ———, "An inequality for the matrix pressure function and applications", *Adv. Math.* **302** (2016), p. 280–308.

[24] E. Oregón-Reyes – "Properties of sets of isometries of Gromov hyperbolic spaces", *Groups Geom. Dyn.* **12** (2018), no. 3, p. 889–910.

[25] H. Radjavi & P. Rosenthal – *Simultaneous triangularization*, Universitext, Springer-Verlag, New York, 2000.

[26] G.-C. Rota & G. Strang – "A note on the joint spectral radius", *Indag. Math.* **22** (1960), p. 379–381.

[27] R. Schneider – *Convex bodies: the Brunn-Minkowski theory*, Encyclopedia of Math. and its Appl., vol. 151, Cambridge University Press, Cambridge, 2014.

[28] M. Sombra – "A sparse effective Nullstellensatz", *Adv. in Appl. Math.* **22** (1999), no. 2, p. 271–295.

[29] J. N. Tsitsiklis & V. 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, p. 31–40, Correction: *Ibid.*, no. 4, p. 381.

Eduardo Oregón-Reyes, Department of Mathematics, University of California at Berkeley
848 Evans Hall, Berkeley, CA 94720-3860, U.S.A.
E-mail : eoregon@berkeley.edu
Url : https://math.berkeley.edu/~eoregon/