Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER
Linear Algebra and its Applications 368 (2003) 71–81
www.elsevier.com/locate/laa

Inequalities for numerical invariants of sets of matrices ☆

Jairo Bochi

IMPA, Estr. D. Castorina 110, 22460-320 Rio de Janeiro, Brazil

Received 22 July 2002; accepted 21 October 2002

Submitted by L. Elsner

Abstract

We prove three inequalities relating some invariants of sets of matrices, such as the joint spectral radius. One of the inequalities, in which proof we use geometric invariant theory, has the generalized spectral radius theorem of Berger and Wang as an immediate corollary.

Keywords: Joint spectral radius; Geometric invariant theory

1. Introduction

Let $M(d)$ be the space of $d \times d$ complex matrices. If $A \in M(d)$, we indicate by $\rho(A)$ the spectral radius of A, that is, the maximum absolute value of an eigenvalue of A. Given a norm $\|\cdot\|$ in $\mathbb{C}^d$, we endow the space $M(d)$ with the operator norm $\|A\| = \sup\{\|Av\|;\ \|v\| = 1\}$.

For every $A \in M(d)$ and every norm $\|\cdot\|$ in $\mathbb{C}^d$, we have $\rho(A) \leqslant \|A\|$. On the other hand, there is also a *lower* bound for $\rho(A)$ in terms of norms:

$$\|A^d\| \leqslant C\rho(A)\|A\|^{d-1}, \quad \text{where } C = 2^d - 1. \tag{1}$$

In particular, if $\rho(A) \ll \|A\|$ then $\|A^d\| \ll \|A\|^d$.

Inequality (1) is a very simple consequence of the Cayley–Hamilton theorem. Indeed, let $p(z) = z^d - \sigma_1 z^{d-1} + \cdots + (-1)^d \sigma_d$ be the characteristic polynomial of A. Since $p(A) = 0$,

☆ Supported by CNPQ-Profix.

E-mail address: bochi@impa.br (J. Bochi).

doi:10.1016/S0024-3795(02)00658-4

$$\|A^d\| \leqslant \sum_{i=1}^{d} |\sigma_i| \, \|A\|^{d-i}.$$

Since the σ_i are the elementary symmetric functions on the eigenvalues of A,

$$|\sigma_i| \leqslant \binom{d}{i} \rho(A)^i \leqslant \binom{d}{i} \rho(A) \|A\|^{i-1}.$$

Therefore (1) follows.

The *spectral radius theorem* (for the finite-dimensional case) asserts that

$$\rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}. \tag{2}$$

The formula above may be deduced from the inequality (1), as we now show. Since $\|A^{n+m}\| \leqslant \|A^n\|\|A^m\|$, the limit in (2) exists (see [11, Problem I.98]); let us call it r. Clearly, $r \geqslant \rho(A)$. Applying (1) to A^n in the place of A, using that $\rho(A^n) = \rho(A)^n$ and taking the $1/dn$-power, we obtain

$$\|A^{dn}\|^{1/dn} \leqslant C^{1/dn} \rho(A)^{1/d} \|A^n\|^{(d-1)/dn}.$$

Taking limits when $n \to \infty$, we get $r \leqslant \rho(A)^{1/d} r^{(d-1)/d}$, that is, $r \leqslant \rho(A)$, proving (2). The author ignores whether this proof has ever appeared in the literature.

Now, let Σ be a non-empty bounded subset of $M(d)$. Define

$$\|\Sigma\| = \sup_{A\in\Sigma} \|A\|, \quad \rho(\Sigma) = \sup_{A\in\Sigma} \rho(A).$$

If $n \in \mathbb{N}$, we denote by Σ^n the set of the products $A_1 \cdots A_n$, with all $A_i \in \Sigma$. Since $\|\Sigma^{n+m}\| \leqslant \|\Sigma^n\|\|\Sigma^m\|$, the limit

$$\mathscr{R}(\Sigma) = \lim_{n\to\infty} \|\Sigma^n\|^{1/n}$$

exists and equals $\inf_n \|\Sigma^n\|^{1/n}$. Besides, it is independent of the chosen norm. The quantity $\mathscr{R}(\Sigma)$ was introduced by Rota and Strang [15] and is called the *joint spectral radius* of the set Σ. For a nice geometrical interpretation of the joint spectral radius, see [13] (or [14]).

Our first main result is a generalization of (1) to sets of matrices:

Theorem A. *Given $d \geqslant 1$, there exists $C_1 > 1$ such that, for every bounded set $\Sigma \subset M(d)$ and every norm $\|\cdot\|$ in $\mathbb{C}^d$,*

$$\|\Sigma^d\| \leqslant C_1 \mathscr{R}(\Sigma) \|\Sigma\|^{d-1}.$$

Our next result relates the joint spectral radius of Σ with spectral radii of products of matrices in Σ:

Theorem B. *Given $d \geqslant 1$, there exists $C_2 > 1$ and $k \in \mathbb{N}$ such that, for every bounded set $\Sigma \subset M(d)$,*

$$\mathscr{R}(\Sigma) \leqslant C_2 \max_{1\leqslant j\leqslant k} \rho(\Sigma^j)^{1/j}.$$

Using Theorem B, we can extend the spectral radius theorem (2):

Corollary 1 (Berger–Wang generalized spectral radius theorem). *If $\Sigma \subset M(d)$ is bounded then*

$$\mathscr{R}(\Sigma) = \limsup_{n\to\infty} \rho(\Sigma^n)^{1/n}.$$

Proof. The inequality $\mathscr{R}(\Sigma) \geqslant \limsup \rho(\Sigma^n)^{1/n}$ is trivial. Applying Theorem B to Σ^n and using that $\mathscr{R}(\Sigma^n) = \mathscr{R}(\Sigma)^n$, we obtain

$$\mathscr{R}(\Sigma) \leqslant C_2^{1/n} \max_{1\leqslant j\leqslant k} \rho(\Sigma^{jn})^{1/jn}.$$

Taking lim sup when $n \to \infty$, we get the result. □

The result above was conjectured by Daubechies and Lagarias [2] and proved by Berger and Wang [1]. Other proofs were given in [3,16].

The proof of Theorem A is elementary, while in the proof of Theorem B we shall use some geometric invariant theory. We also give another generalization of (1), Proposition 12, whose proof is elementary.

Remark. For all Σ and $m, n \in \mathbb{N}$, we have $\rho(\Sigma^{mn})^{1/mn} \geqslant \rho(\Sigma^n)^{1/n}$ (because $\Sigma^{nm} \subset (\Sigma^n)^m$). So in Theorem B it is sufficient to take the maximum of $\rho(\Sigma^j)^{1/j}$ over j with $k/2 < j \leqslant k$. Another consequence of the latter fact is that

$$\limsup_{n\to\infty} \rho(\Sigma^n)^{1/n} = \sup_{n\in\mathbb{N}} \rho(\Sigma^n)^{1/n}.$$

2. Proof of Theorem A

We first prove a weaker inequality:

Lemma 2. *Let $\|\cdot\|_{\mathrm{e}}$ be the euclidian norm in $\mathbb{C}^d$. There exists $C_0 = C_0(d)$ such that*

$$\|S\Sigma^d S^{-1}\|_{\mathrm{e}} \leqslant C_0 \|\Sigma\|_{\mathrm{e}} \|S\Sigma S^{-1}\|_{\mathrm{e}}^{d-1}.$$

for every non-empty bounded set $\Sigma \subset M(d)$ and every $S \in \mathrm{GL}(d)$.

Proof. We shall also consider the norm in $M(d)$ defined by

$$\|A\|_0 = \max |a_{ij}|, \qquad \text{where } A = (a_{ij})_{i,j=1,\dots,d}.$$

We first assume S is a diagonal matrix $\mathrm{diag}(\lambda_1, \dots, \lambda_d)$, with $\lambda_1, \dots, \lambda_d > 0$. Take d matrices $A_1, \dots, A_d \in \Sigma$, and write $A_\ell = (a_{ij}^{(\ell)})$. Then

$$\|S\Sigma S^{-1}\|_0 \geqslant \max_{i,j,\ell} \left|\lambda_i a_{ij}^{(\ell)} \lambda_j^{-1}\right|$$

and

$$\left\|SA_1\cdots A_dS^{-1}\right\|_0 \leqslant C_0 \max_{i_0,\ldots,i_d}\left|\lambda_{i_0}a^{(1)}_{i_0i_1}\cdots a^{(d)}_{i_{d-1}i_d}\lambda^{-1}_{i_d}\right|,$$

where $C_0 = d^{d-1}$. Given integers $i_0,\ldots,i_d \in \{1,\ldots,d\}$, by the pigeon-hole principle there exists $1 \leqslant k \leqslant d$ such that $\lambda_{i_{k-1}} \leqslant \lambda_{i_k}$. Therefore

$$\begin{aligned}\left|\lambda_{i_0}a^{(1)}_{i_0i_1}\cdots a^{(d)}_{i_{d-1}i_d}\lambda^{-1}_{i_d}\right| &= \prod_{1\leqslant \ell\leqslant d}\left|\lambda_{i_{\ell-1}}a^{(\ell)}_{i_{\ell-1}i_\ell}\lambda^{-1}_{i_\ell}\right| \\ &\leqslant \left|a^{(k)}_{i_{k-1},i_k}\right|\prod_{\ell\neq k}\left|\lambda_{i_{\ell-1}}a^{(\ell)}_{i_{\ell-1}i_\ell}\lambda^{-1}_{i_\ell}\right| \\ &\leqslant \|\Sigma\|_0\|S\Sigma S^{-1}\|_0^{d-1}.\end{aligned}$$

It follows that $\|S\Sigma^dS^{-1}\|_0 \leq C_0\|\Sigma\|_0\|S\Sigma S^{-1}\|_0^{d-1}$. Up to changing C_0, the same inequality holds for the euclidian norm $\|\cdot\|_{\mathrm{e}}$.

Next consider the general case $S \in \mathrm{GL}(d)$. By the singular value decomposition theorem, there exist unitary matrices U, V and a diagonal matrix $D = \mathrm{diag}(\lambda_1,\ldots,\lambda_d)$, with $\lambda_1,\ldots,\lambda_d > 0$, such that $S = UDV$. Since U and V preserve the euclidian norm,

$$\begin{aligned}\|S\Sigma^dS^{-1}\|_{\mathrm{e}} &= \|D(V\Sigma V^{-1})^dD^{-1}\|_{\mathrm{e}} \\ &\leqslant C_0\|V\Sigma V^{-1}\|_{\mathrm{e}}\|DV\Sigma V^{-1}D^{-1}\|_{\mathrm{e}}^{d-1} \\ &= C_0\|\Sigma\|_{\mathrm{e}}\|S\Sigma S^{-1}\|_{\mathrm{e}}^{d-1}.\end{aligned}$$

This proves the lemma. □

To make the constant in Lemma 2 independent of the norm, we will use:

Lemma 3. *There exists $C = C(d)$ such that, for every two norms $\|\cdot\|_1$, $\|\cdot\|_2$ in $\mathbb{C}^d$, there is $S \in \mathrm{GL}(d)$ such that*:

1. $C^{-1}\|v\|_1 \leqslant \|Sv\|_2 \leqslant \|v\|_1$ *for all* $v \in \mathbb{C}^d$;
2. $C^{-1}\|A\|_1 \leqslant \|SAS^{-1}\|_2 \leqslant C\|A\|_1$ *for all* $A \in M(d)$.

Proof. The second part is an immediate consequence of the first one. To prove the first part, it is enough to show that for every $\|\cdot\|$ in $\mathbb{C}^d$, there is $S \in \mathrm{GL}(d)$ such that

$$C_d^{-1}\|v\| \leqslant \|Sv\|_0 \leqslant \|v\| \quad \forall v \in \mathbb{C}^d, \tag{3}$$

where $\|\cdot\|_0$ is the sup-norm in $\mathbb{C}^d$ and $C_d = 2^d - 1$. The proof is by induction. Let $\|\cdot\|$ be a norm in $\mathbb{C}^{d+1}$. Restrict it to the subspace $\mathbb{C}^d = \mathbb{C}^d \times \{0\} \subset \mathbb{C}^{d+1}$. By induction hypothesis, there is $S \in \mathrm{GL}(d)$ such that (3) holds. If $\pi_j : \mathbb{C}^{d+1} \to \mathbb{C}$ is the projection in the jth coordinate, then $|\pi_j \circ S(v)| \leqslant \|v\|$ for all $v \in \mathbb{C}^d$ and $1 \leqslant j \leqslant d$. By the Hahn–Banach theorem, there are linear functionals $\Lambda_j : \mathbb{C}^{d+1} \to \mathbb{C}$

such that $\Lambda_j|\mathbb{C}^d = \pi_j \circ S$ and $|\Lambda_j(w)| \leqslant \|w\|$ for all $w \in \mathbb{C}^{d+1}$ and $1 \leqslant j \leqslant d$. Let $a = \|\pi_{d+1}\|$ and define a linear map $\bar{S} : \mathbb{C}^{d+1} \to \mathbb{C}^{d+1}$ by

$$\pi_j \circ \bar{S} = \begin{cases} \Lambda_j & \text{if } 1 \leqslant j \leqslant d, \\ a^{-1}\pi_{d+1} & \text{if } j = d. \end{cases}$$

Then $\bar{S} \in \mathrm{GL}(d+1)$ and $\|\bar{S}w\|_0 \leqslant \|w\|$, so $\bar{S}$ satisfies the second inequality in (3). To prove the first one, let $\xi \in \mathbb{C}^{d+1}$ be such that $\pi_{d+1}(\xi) = a$ and $\|\xi\| = 1$. Write $\bar{S}(\xi) = \eta + e_{d+1}$ with $\eta \in \mathbb{C}^d$ and $e_{d+1} = (0, \ldots, 0, 1) \in \mathbb{C}^{d+1}$. We have $\|\eta\|_0 \leqslant \|\xi\| = 1$ and so $\|\bar{S}^{-1}(\eta)\| = \|S^{-1}(\eta)\| \leqslant C_d\|\eta\|_0 \leqslant C_d$. Therefore

$$\|\bar{S}^{-1}(e_{d+1})\| \leqslant \|\xi\| + \|\bar{S}^{-1}(\eta)\| \leqslant 1 + C_d.$$

Now let $w \in \mathbb{C}^{d+1}$ be given. Write $w = v + te_{d+1}$ with $w \in \mathbb{C}^d$ and $t \in \mathbb{C}$. Then

$$\begin{aligned} \|\bar{S}^{-1}(w)\| &\leqslant \|\bar{S}^{-1}(v)\| + |t|\|\bar{S}^{-1}(e_{d+1})\| \\ &\leqslant C_d\|v\|_0 + (C_d + 1)|t| \\ &\leqslant (2C_d + 1)\max\{\|v\|_0, |t|\} = C_{d+1}\|w\|_0. \end{aligned}$$

This proves that (3) holds with $d + 1$ and $\bar{S}$ in the place of d and S. $\square$

The result below gives another characterization of the joint spectral radius. For a proof, see [3] or [15].

Proposition 4. *For all bounded $\Sigma \subset M(d)$,*

$$\mathscr{R}(\Sigma) = \inf_{\|\cdot\|} \|\Sigma\|,$$

where the infimum is taken over all norms in $\mathbb{C}^d$.

Proof of Theorem A. Let C_0 and C be as in Lemmas 2 and 3. Let $\|\cdot\|_{\mathrm{e}}$ be the euclidian norm, and let $\|\cdot\|_1$, $\|\cdot\|_2$ be any two norms in $\mathbb{C}^d$. Let $S_1, S_2 \in \mathrm{GL}(d)$ be given by Lemma 3 such that

$$C^{-1}\left\|S_i A S_i^{-1}\right\|_{\mathrm{e}} \leqslant \|A\|_i \leqslant C\left\|S_i A S_i^{-1}\right\|_{\mathrm{e}} \quad \forall A \in M(d), i = 1, 2.$$

Take $\Sigma \subset M(d)$. Then (applying Lemma 2 with $S = S_1S_2^{-1}$ and $S_2\Sigma S_2^{-1}$ in the place of Σ)

$$\begin{aligned} \|\Sigma^d\|_1 &\leqslant C\left\|S_1\Sigma^d S_1^{-1}\right\|_{\mathrm{e}} \\ &\leqslant CC_0\left\|S_2\Sigma S_2^{-1}\right\|_{\mathrm{e}}\left\|S_1\Sigma S_1^{-1}\right\|_{\mathrm{e}}^{d-1} \leqslant C^d C_0\|\Sigma\|_2\|\Sigma\|_1^{d-1}. \end{aligned}$$

Taking the infimum over $\|\cdot\|_2$ in the left-hand side, we obtain, by Proposition 4, $\|\Sigma^d\|_1 \leqslant C_1\mathscr{R}(\Sigma)\|\Sigma\|_1^{d-1}$, where $C_1 = C^d C_0$. $\square$

Let us reread Theorem A in terms of another invariant. Given a non-empty bounded $\Sigma \subset M(d)$, we define

$$\mathscr{S}(\Sigma) = \sup_{\|\cdot\|} \frac{\|\Sigma^d\|}{\|\Sigma\|^{d-1}} \quad \text{if } \Sigma \neq \{0\}, \tag{4}$$

and $\mathscr{S}(\{0\}) = 0$. The functions $\mathscr{R}(\cdot)$ and $\mathscr{S}(\cdot)$ are comparable:

Proposition 5. $\mathscr{R}(\Sigma) \leqslant \mathscr{S}(\Sigma) \leqslant C_1 \mathscr{R}(\Sigma)$.

Proof. The second inequality is Theorem A. For any $\|\cdot\|$ we have $\|\Sigma^d\| \geqslant \mathscr{R}(\Sigma)^d$ and so, using Proposition 4,

$$\mathscr{S}(\Sigma) \geqslant \sup_{\|\cdot\|} \frac{\mathscr{R}(\Sigma)^d}{\|\Sigma\|^{d-1}} = \mathscr{R}(\Sigma). \qquad \square$$

3. Proof of Theorem B

We shall need the following general result:

Proposition 6. *Fix $d, \ell \in \mathbb{N}$. Let $f : M(d)^\ell \to [0, \infty)$ be a locally bounded function such that, for every $A_1, \ldots, A_\ell \in M(d)$,*

- $f(SA_1S^{-1}, \ldots, SA_\ell S^{-1}) = f(A_1, \ldots, A_\ell) \forall S \in \mathrm{GL}(d)$;
- $f(tA_1, \ldots, tA_\ell) = |t| f(A_1, \ldots, A_\ell) \forall t \in \mathbb{C}$.

Then there exist numbers $k = k(d) \in \mathbb{N}$ and $C = C(d, \ell, f) > 0$ such that

$$f(A_1, \ldots, A_\ell) \leqslant C \max_{1 \leqslant j \leqslant k} \rho(\Sigma^j)^{1/j}, \quad \text{where } \Sigma = \{A_1, \ldots, A_\ell\}. \tag{5}$$

Let us postpone the proof of this proposition and conclude the:

Proof of Theorem B. Let $\mathscr{S}(\cdot)$ be as in (4). Define a function $f : M(d)^d \to [0, \infty)$ by $f(A_1, \ldots, A_d) = \mathscr{S}(\{A_1, \ldots, A_d\})$. By Theorem A, $f(\Sigma) \leqslant C_1 \|\Sigma\|$ (for any norm)—in particular, f is locally bounded. f also satisfies the other hypotheses of Proposition 6, thus there are k and C_2 such that

$$\mathscr{S}(\Sigma) \leqslant C_2 \max_{1 \leqslant j \leqslant k} \rho(\Sigma^j)^{1/j}, \tag{6}$$

for every $\Sigma \subset M(d)$ with at most d elements. But

$$\mathscr{S}(\Sigma) = \sup\{\mathscr{S}(\Sigma'); \Sigma' \subset \Sigma, \#\Sigma' \leqslant d\},$$

hence (6) actually holds for every bounded Σ. Since $\mathscr{R}(\Sigma) \leqslant \mathscr{S}(\Sigma)$ (Proposition 5), Theorem B follows. $\square$

A few preliminaries in geometric invariant theory are necessary to prove Proposition 6. Some references are [6,10].

3.1. Polynomial invariants

Let V be a complex vector space, G be a group and $\iota : G \to \mathrm{GL}(V)$ be a linear representation of G. We shall write $gx = \iota(g)(x)$. The *orbit* of $x \in V$ is the set $\mathcal{O}(x) = \{gx;\, g \in G\}$. Let $\mathbb{C}[V]$ be the ring of polynomial functions $\phi : V \to \mathbb{C}$. A polynomial $\phi \in \mathbb{C}[V]$ is *invariant* if it is constant along each orbit, that is, $\phi(gx) \equiv \phi(x)$. The *ring of invariants*, denoted by $\mathbb{C}[V]^G$, is the set of all invariant polynomials.

For some groups G, called *reductive* groups, a celebrated theorem of Nagata asserts that the ring $\mathbb{C}[V]^G$ is finitely generated. We shall not define a reductive group; but some examples are $\mathrm{GL}(d)$, $\mathrm{SL}(d)$, $\mathrm{PGL}(d)$. We assume from now on that G is reductive. In this case, the theory provides an algebraic quotient of V by G with good properties:

Theorem 7. *Let $\phi_1, \dots, \phi_N$ be a set of generators of $\mathbb{C}[V]^G$. Let $\pi : V \to \mathbb{C}^N$ be the mapping $x \mapsto (\phi_1(x), \dots, \phi_N(x)) \in \mathbb{C}^N$. Then*:

0. *π is G-invariant* (*i.e.*, *constant along orbits*);
1. *$Y = \pi(V)$ is closed*;
2. *$\pi(x_1) = \pi(x_2)$ if and only if the closures $\overline{\mathcal{O}(x_1)}$ and $\overline{\mathcal{O}(x_2)}$ have non-empty intersection*;
3. *for every $y \in Y$, the fiber $\pi^{-1}(y)$ contains an unique closed orbit.*

In the statement above, and in everything that follows, the spaces V and $\mathbb{C}^N$ are endowed with the ordinary (not Zariski) topologies. Notice item 2 says that π separates every pair of orbits that can be separated by a G-invariant continuous function.

Indication of proof. Let $\mathbb{C}(V)^G$ be the field of G-invariant rational functions. It is easy to see that $\mathbb{C}(V)^G$ is the field of quotients of $\mathbb{C}[V]^G$. Let π and Y be as in the statement. Let Z be the Zariski-closure of Y, and consider π as a function $V \to Z$. Then π induces a homomorphism $\pi^* : C(Z) \to \mathbb{C}(V)$ via $f \mapsto f \circ \pi$. One easily shows that π^* is an isomorphism onto $\mathbb{C}(V)^G$, so $\pi : V \to Z$ is an algebraic quotient in the sense of [6, Section II.3.2]. Therefore π is surjective, that is, $Y = Z$, and item 1 follows. Items 2 and 3 are [10, Corollary 3.5.2] and [6, bemerkung 1, Section II.3.2], respectively. In the references above the Zariski topology is used instead. But this makes no difference here, by [6, Section AI.7]. $\square$

Example. Let $G = \mathrm{GL}(d)$ act on $V = M(d)$ by conjugation: $\iota(S)(A) = SAS^{-1}$ for $S \in G$ and $A \in V$. Given $A \in V$, let $\sigma_1(A), \dots, \sigma_d(A)$ be the coefficients of the characteristic polynomial of A. Then $\sigma_1, \dots, \sigma_d \in \mathbb{C}[V]^G$. Moreover, these polynomials generate the ring $\mathbb{C}[V]^G$. Let $\pi = (\sigma_1, \dots, \sigma_d)$. Then π is onto $\mathbb{C}^d$. Every fiber $\pi^{-1}(y)$ consists in finitely many orbits, each one corresponding to a different

Jordan form. The closed orbits are those of diagonalizable matrices. The fiber $\pi^{-1}(0)$ is the set of nilpotent matrices (see [6, Section I.3]).

3.2. Topological considerations

Fix π and $Y = \pi(V)$ as in Theorem 7. We endow the set $Y \subset \mathbb{C}^N$ with the induced topology. The following theorem was proved independently by Luna [7] and Neeman [8, Corollary 1.6, Remark 1.7] (see also [9]):

Theorem 8. *The topology in Y coincides with the quotient topology induced by $\pi : V \to Y$ (i.e., $U \subset Y$ is open if and only if $\pi^{-1}(U)$ is open in V).*

Corollary 9. *The mapping $\pi : V \to Y$ is* semiproper, *that is, for every compact set $L \subset Y$ there exists a compact set $K \subset V$ such that $\pi(K) \supset L$.*

Proof. Suppose that for some compact $L \subset Y$ there is no compact set $K \subset V$ such that $\pi(K) \supset L$. Take compact sets $K_n \subset V$ such that $K_n \subset \operatorname{int} K_{n+1}$ and $\bigcup_n K_n = V$. Then, for each n, there exists $y_n \in L$ such that $\pi^{-1}(y_n) \cap K_n = \emptyset$. Up to replacing (y_n) with a subsequence, we may assume that $y = \lim y_n$ exists and $y_n \neq y$ for each n. Then the set $F = \{y_n; n \in \mathbb{N}\}$ is not closed in Y, but $\pi^{-1}(F) = \bigcup_n \pi^{-1}(y_n)$ is closed in V, contradicting Theorem 8. $\square$

Let us derive a consequence of the above results:

Lemma 10. *If $f : V \to [0, \infty)$ is a G-invariant locally bounded function then there exists a locally bounded $h : Y \to [0, \infty)$ such that $f \leqslant h \circ \pi$.*

Proof. Given $x \in V$, let $F_x = \pi^{-1}(\pi(x))$ be the fiber containing x. Set

$$\bar{f}(x) = \inf\left\{ \sup f|U;\ U \text{ is a } G\text{-invariant open set containing } F_x \right\}.$$

(Here "U is G-invariant" means $\mathcal{O}(x) \subset U$ for all $x \in U$.) We claim that $\bar{f}(x)$ is finite for all $x \in V$. Indeed, each fiber F_x contains an unique closed orbit $\mathcal{O}(x_0)$, by Theorem 7. Let U_0 be a bounded neighborhood of x_0; so $\sup f|U_0$ is finite. Let $U = \bigcup_{x \in U_0} \mathcal{O}(x)$; then U is a G-invariant open set and $\sup f|U = \sup f|U_0$. Moreover, U contains F_x: for every $\xi \in F_x$, we have, by Theorem 7, $\overline{\mathcal{O}(\xi)} \cap \mathcal{O}(x_0) \neq \emptyset$, hence $\mathcal{O}(\xi) \cap U_0 \neq \emptyset$ and $\xi \in U$. This proves that $\bar{f}(x) \leqslant \sup f|U < \infty$.

The function $\bar{f} : V \to \mathbb{R}$ satisfies $\bar{f} \geqslant f$ and is also locally bounded. Since $\bar{f}$ is constant on fibers, there exist $h : Y \to \mathbb{R}$ such that $\bar{f} = h \circ \pi$. The function h is locally bounded, because if $L \subset Y$ is a compact set then, by Corollary 9, there is some compact $K \subset V$ such that $\pi(K) \supset L$ and, in particular, $h|L \leqslant (h \circ \pi)|K = \bar{f}|K < \infty$. $\square$

3.3. ℓ-uples of matrices and end of the proof

From now on we set $G = \mathrm{GL}(d)$, $V = M(d)^d$ and

$$\iota(S)(A_1, \ldots, A_\ell) = (SA_1S^{-1}, \ldots, SA_\ell S^{-1}).$$

In this case, a finite set of generators for $\mathbb{C}[V]^G$ is known:

Theorem 11 (Procesi [12], Theorem 3.4a). *The ring of invariants is generated by the polynomials* $\mathrm{tr}(A_{i_1} \cdots A_{i_j})$ *with* $1 \leqslant j \leqslant k$, *where* $k = 2^d - 1$.

We are now able to give the:

Proof of Proposition 6. Let $k = 2^d - 1$, $N = \ell + \ell^2 + \cdots + \ell^k$, and let $\alpha_1, \ldots, \alpha_N$ be all the sequences $\alpha = (i_1, \ldots, i_j) \in \{1, \ldots, \ell\}^j$ of length $|\alpha| = j$, $1 \leqslant j \leqslant k$. Let $\pi = (\phi_1, \ldots, \phi_N) : V \to \mathbb{C}^N$ be given by

$$\phi_i(A_1, \ldots, A_\ell) = \mathrm{tr}(A_{i_1} \cdots A_{i_j}), \quad \text{where } \alpha_i = (i_1, \ldots, i_j).$$

Let $Y = \pi(V)$. Define another function $\tau : \mathbb{C}^N \to \mathbb{R}$ by

$$\tau(z_1, \ldots, z_N) = \max\left\{|z_\ell|^{1/|\alpha_i|};\ 1 \leqslant i \leqslant N\right\}.$$

So if $x = (A_1, \ldots, A_\ell)$ then

$$\tau(\pi(x)) = \max\left\{|\mathrm{tr}(A_{i_1} \cdots A_{i_j})|^{1/j};\ 1 \leqslant j \leqslant k,\ 1 \leqslant i_1, \ldots, i_j \leqslant \ell\right\}.$$

Since $|\mathrm{tr}(A)| \leqslant d\rho(A)$ for every $A \in M(d)$, we have

$$\tau(\pi(x)) \leqslant d \max_{1 \leqslant j \leqslant k} \rho(\Sigma^j)^{1/j}, \quad \text{where } \Sigma = \{A_1, \ldots, A_\ell\}.$$

Notice $\tau(\pi(tx)) = |t|\tau(\pi(x))$ for all $t \in \mathbb{C}$. Let $h : Y \to [0, \infty)$ be given by Lemma 10. Since $K = \tau^{-1}(1)$ is compact and Y is closed, $C_0 = \sup h|(Y \cap K)$ is finite. Given $x \in V$, let $t = \tau(\pi(x))$. If $t \neq 0$ then

$$\begin{aligned} f(x) &= tf(t^{-1}x) \leqslant th(\pi(t^{-1}x)) \leqslant C_0 t \\ &= C_0\tau(\pi(x)) \leqslant dC_0 \max_{1 \leqslant j \leqslant k} \rho(\Sigma^j)^{1/j}. \end{aligned}$$

Let $C = dC_0$. Then (5) holds. If $t = 0$, that is, $\pi(x) = 0$, we argue differently. By Theorem 7, the orbit of x accumulates at 0. It follows from the hypotheses on f that $f(0) = 0$ and f is continuous at 0. Therefore $f(x) = 0$ and (5) holds. This completes the proof of Proposition 6 and so of Theorem B. □

4. Another inequality and some questions

We shall prove another inequality, Proposition 12, which generalizes (1) and is also an elementary consequence of the Cayley–Hamilton theorem.

We need some notation. For $s \geqslant 1$, let S_s be the set of permutations of $\{1, 2, \dots, s\}$. Given $\sigma \in S_s$, decompose σ in disjoint cycles, including the ones of length 1:

$$\sigma = (i_1 \cdots i_k)(j_1 \cdots j_h) \cdots (t_1 \cdots t_e).$$

Then, given matrices $A_1, \dots, A_s$, we set

$$\Phi_\sigma(A_1, \dots, A_s) = \operatorname{tr}(A_{i_1} \cdots A_{i_k})\operatorname{tr}(A_{j_1} \cdots A_{j_h}) \cdots \operatorname{tr}(A_{t_1} \cdots A_{t_e}).$$

Letting $\varepsilon(\sigma)$ be the sign of σ, we define

$$F(A_1, \dots, A_s) = \sum_{\sigma \in S_s} \varepsilon(\sigma)\Phi_\sigma(A_1, \dots, A_s). \tag{7}$$

Define also $P(A_1, \dots, A_s) = \sum_{\sigma \in S_s} A_{\sigma(1)} \cdots A_{\sigma(s)}$. The *trace identity* from [12, Corollary 4.4] (which follows from the Cayley–Hamilton theorem by an elementary process, see also [4, Section 4]) is

$$\sum_{s=0}^{d} \sum (-1)^s F(A_{i_1}, \dots, A_{i_s}) P(A_{j_1}, \dots, A_{j_{d-s}}) = 0, \tag{8}$$

where the second sum runs over all partitions of $\{1, \dots, d\}$ into two disjoint subsets $\{i_1 < \cdots < i_s\}$ and $\{j_1 < \cdots < j_{d-s}\}$; it is understood that $F(\emptyset) = 1$ and $P(\emptyset) = I$.

Proposition 12. *Given $d \geqslant 1$, there exists $C > 1$ such that for every operator norm $\|\cdot\|$ and every d matrices $A_1, \dots, A_d \in M(d)$, we have*

$$\|P(A_1, \dots, A_d)\| \leqslant C\|\Sigma\|^{d-1} \max_{1 \leqslant j \leqslant d} \rho(\Sigma^j)^{1/j},$$

where $\Sigma = \{A_1, \dots, A_d\}$.

Proof. We estimate terms in (8) for $1 \leqslant s \leqslant d$. If σ is a permutation of $\{i_1 < \cdots < i_s\}$ with cycles of lengths $k_1, \dots, k_h$ then

$$|\Phi_\sigma(A_{i_1}, \dots, A_{i_s})| \leqslant C_0 \rho(\Sigma^{k_1}) \cdots \rho(\Sigma^{k_h}),$$

where C_0 is a constant. The right-hand side is $\leqslant C_0 \rho(\Sigma^{k_i})^{1/k_i} \|\Sigma\|^{s-1}$ for any k_i. Plugging this estimate in (7), we get

$$|F(A_{i_1}, \dots, A_{i_s})| \leqslant C_0 \|\Sigma\|^{s-1} \max_{1 \leqslant j \leqslant s} \rho(\Sigma^j)^{1/j}.$$

Using the inequality above and the obvious bound $\|P(A_{j_1}, \dots, A_{j_{d-s}})\| \leqslant (d-s)!\|\Sigma\|^{d-s}$, the result follows from (8). $\square$

We do not know whether the methods of the proof of Proposition 12 can be improved to give an elementary proof of Theorem B. Notice that if k in Theorem B were equal to d then Proposition 12 would follow from Theorems A and B.

Question. What is the minimum k such that Theorem B holds? Can one take $k = d$?

The answer is yes when $d = 2$. The ring of invariants of two 2×2 matrices A_1 and A_2 is generated by $\operatorname{tr} A_1$, $\det A_1$, $\operatorname{tr} A_2$, $\det A_2$, $\operatorname{tr} A_1 A_2$, see [4, Section 7]. Since $\det A$ can be expressed as a polynomial in $\operatorname{tr} A$ and $\operatorname{tr} A^2$, one can take $k = 2$ in Theorem 11, and so also in Theorem B, when $d = 2$. Moreover, since $\rho(\Sigma) \leqslant \rho(\Sigma^2)^{1/2}$, Theorem B assumes the form

$$\mathscr{R}(\Sigma) \leqslant C_2 \rho(\Sigma^2)^{1/2}.$$

Using this inequality, it is easy to show that the sequence $\rho(\Sigma^{2n})^{1/2n}$ converges. However, the sequence $\rho(\Sigma^n)^{1/n}$ itself does not necessarily converge. We reproduce an example from [5]:

$$\Sigma = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right\} \quad \Rightarrow \quad \rho(\Sigma^n) = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ 1 & \text{if } n \text{ is even.} \end{cases}$$

Acknowledgement

I would like to thank J.V. Pereira for helpful discussions.

References

[1] M.A. Berger, Y. Wang, Bounded semigroups of matrices, Linear Algebra Appl. 166 (1992) 21–27.
[2] I. Daubechies, J.C. Lagarias, Sets of matrices all infinite products of which converge, Linear Algebra Appl. 161 (1992) 227–263; Corrigendum/Addendum 327 (2001) 69–83.
[3] L. Elsner, The generalized spectral-radius theorem: an analytic geometric proof, Linear Algebra Appl. 220 (1995) 151–159.
[4] E. Formanek, The polynomial identities of matrices, in: S.A. Amitsur, D.J. Saltman, G.B. Seligman (Eds.), Algebraists' Homage, Contemporary Mathematics 13 (1982) 41–79.
[5] G. Gripenberg, Computing the joint spectral radius, Linear Algebra Appl. 234 (1996) 43–60.
[6] H. Kraft, Geometrische Methoden in der Invariantentheorie, Fried, Vieweg & Sohn, 1985.
[7] D. Luna, Adhérences d'orbites et invariants, Invent. Math. 29 (1973) 231–238.
[8] A. Neeman, The topology of quotient varieties, Ann. Math. 122 (1985) 419–459.
[9] A. Neeman, Analytic questions in geometric invariant theory, in: R. Fossum, W. Haboush, M. Hochster, V. Lakshmibai (Eds.), Invariant theory, Contemporary Mathematics 88 (1989) 11–22.
[10] P.E. Newstead, Introduction to moduli problems and orbit spaces, Tata Institute of Fundamental Research, Springer-Verlag, 1978.
[11] G. Pólya, G. Szegö, Problems and Theorems in Analysis I, Springer-Verlag, 1978.
[12] C. Procesi, The invariant theory of $n \times n$ matrices, Adv. Math. 19 (1976) 306–381.
[13] V.Yu. Protasov, Joint spectral radius and invariant sets of several linear operators, Fund. Prikl. Mat. 2 (1996) 205–231 (in Russian).
[14] V. Yu. Protasov, The generalized joint spectral radius. A geometric approach, Izvestiya Math. 61 (1997) 995–1030 (English Transl.).
[15] G.C. Rota, W.G. Strang, A note on the joint spectral radius, Indag. Math. 22 (1960) 379–381.
[16] M.H. Shih, J.W. Wu, C.T. Pang, Asymptotic stability and generalized Gelfand spectral radius formula, Linear Algebra Appl. 252 (1997) 61–70.