

Non-commutative standard polynomials applied to matrices

Denis Serre

► **To cite this version:**

Denis Serre. Non-commutative standard polynomials applied to matrices. Linear Algebra and its Applications, Elsevier, 2016, 490, pp.202 - 223. <10.1016/j.laa.2015.11.003>. <ensl-01402364>

HAL Id: ensl-01402364

<https://hal-ens-lyon.archives-ouvertes.fr/ensl-01402364>

Submitted on 24 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-commutative standard polynomials applied to matrices

Denis Serre*

November 3, 2015

Abstract

The Amitsur–Levitski Theorem tells us that the standard polynomial in $2n$ non-commuting indeterminates vanishes identically over the matrix algebra $\mathbf{M}_n(K)$. For $K = \mathbb{R}$ or \mathbb{C} and $2 \leq r \leq 2n - 1$, we investigate how big $\mathcal{S}_r(A_1, \dots, A_r)$ can be when A_1, \dots, A_r belong to the unit ball. We privilege the Frobenius norm, for which the case $r = 2$ was solved recently by several authors. Our main result is a closed formula for the expectation of the square norm. We also describe the image of the unit ball when $r = 2$ or 3 and $n = 2$.

MSC classification : 15A24, 15A27, 15A60

Key words : standard polynomial, random matrices.

1 The problem. First results

Let $r \geq 2$ be an integer. The standard polynomial in r *non-commuting* indeterminates x_1, \dots, x_r is defined as usual by

$$\mathcal{S}_r(x_1, \dots, x_r) := \sum \{ \varepsilon(\sigma) x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(r)} : \sigma \in \mathfrak{S}_r \},$$

where \mathfrak{S}_r is the symmetric group in r symbols and ε is the signature. Each monomial is a word in the letters x_j , affected by a sign ± 1 . Despite its superficial similarity with the determinant of $r \times r$ matrices, \mathcal{S}_r is a completely different object: on the one hand, its arguments are non-commuting indeterminates, on the other hand, there are only r indeterminates instead of the r^2 entries of a matrix. We list here elementary properties of \mathcal{S}_r :

*UMPA, UMR CNRS–ENS Lyon # 5669. École Normale Supérieure de Lyon, 46, allée d’Italie, F–69364 Lyon, cedex 07.

1. \mathcal{S}_r is alternating.
2. $\mathcal{S}_{r+1}(x_1, \dots, x_{r+1}) = \sum_i (-1)^{i+1} x_i \mathcal{S}_r(\hat{x}_i)$, where $\hat{x}_i := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{r+1})$.
3. If r is even, and an x_i commutes with all other x_j 's, then $\mathcal{S}_r(x_1, \dots, x_r) = 0$. Mind that this is false if r is odd.

The first polynomial $x_1x_2 - x_2x_1$ of the list is the commutator. When applied to the elements of an algebra \mathcal{A} , it leads us to distinguish between commutative and non-commutative algebras. More generally, the polynomials \mathcal{S}_r measure somehow the degree of non-commutativity of a given algebra. A classical theorem tells us that for a given matrix $A \in \mathbf{M}_n(\mathbb{C})$, the commutator \mathcal{S}_2 vanishes identically over the algebra $\langle A, A^* \rangle$ (in other words, A is normal) if and only if A is unitarily diagonalizable. It is less known that $\mathcal{S}_{2\ell}$ vanishes identically over the algebra $\langle A, A^* \rangle$ if and only if A is unitarily blockwise diagonalizable, where the diagonal blocks have at most size $\ell \times \ell$; see Exercise 324 in [10].

In addition, we have the theorem of Amitsur and Levitski [2], of which an elegant proof has been given by Rosset [8].

Theorem 1.1 (Amitsur–Levitski.) *Let K be a field (a commutative one, needless to say). The standard polynomial \mathcal{S}_{2n} of degree $2n$ vanishes identically over $\mathbf{M}_n(K)$. However the standard polynomials of degree less than $2n$ do not vanish identically.*

In the sequel, we focus on the algebra $\mathbf{M}_n(K)$ ($K = \mathbb{R}$ or \mathbb{C}) of real or complex matrices. A norm over $\mathbf{M}_n(K)$ is submultiplicative if it satisfies $\|AB\| \leq \|A\| \|B\|$. The main examples are operator norms

$$\|A\| := \sup_{x \in K^n, x \neq 0} \frac{|Ax|}{|x|},$$

where $|\cdot|$ is a given norm over K^n . One often says that $\|\cdot\|$ is *induced* by $|\cdot|$. In particular, $\|\cdot\|_2$ is the norm induced by the standard Euclidian/Hermitian norm. We are also interested in the Frobenius norm

$$\|A\|_F := \sqrt{\sum_{i,j} |a_{ij}|^2},$$

which is not induced, because $\|I_n\|_F = \sqrt{n} > 1$. Nevertheless, it is submultiplicative. A more exotic norm is the *numerical radius*

$$h(A) := \sup \left\{ \frac{|x^*Ax|}{\|x\|_2^2} : x \in \mathbb{C}^n \setminus \{0\} \right\}.$$

This is not a submultiplicative norm but it satisfies $h(A^k) \leq h(A)^k$. The numerical radius is therefore a *super-stable norm*.

The general question that we address is to find precise bounds of

$$\frac{\|\mathcal{S}_r(A_1, \dots, A_r)\|}{\prod_{j=1}^r \|A_j\|}.$$

This can be cast into two sub-problems. On the one hand, we are interested in the best constant $C = C(r, n)$ satisfying

$$\|\mathcal{S}_r(A_1, \dots, A_r)\| \leq C \prod_{j=1}^r \|A_j\|, \quad \forall A_1, \dots, A_r \in \mathbf{M}_n(K).$$

On the other hand, we may ask what is a typical ratio.

For the first task, we look for the smallest ball containing the image when each argument is taken out of the unit ball of $\mathbf{M}_n(K)$. We may even ask for an accurate description of this image. For instance, we shall prove that for 2×2 matrices (i.e. $n = 2$) and the Frobenius norm, the image of the unit ball under \mathcal{S}_2 (the commutator) is the ball of radius $\sqrt{2}$, while the image under \mathcal{S}_3 is an ellipsoid. In order to tackle the second problem, we shall compute the closed form of

$$\frac{\mathbb{E}(\|\mathcal{S}_r(A_1, \dots, A_r)\|_F^2)}{\mathbb{E}(\prod_{j=1}^r \|A_j\|_F^2)},$$

when A_1, \dots, A_r are chosen independently and uniformly in the Frobenius unit ball. Strangely enough, our proof makes use of the Amitsur–Levitski Theorem.

When using a submultiplicative norm, the trivial bound

$$\|\mathcal{S}_r(A_1, \dots, A_r)\| \leq r! \prod_{j=1}^r \|A_j\|, \quad \forall A_1, \dots, A_r \in \mathbf{M}_n(K),$$

suggests to work with the normalized polynomial

$$\mathcal{T}_r := \frac{1}{r!} \mathcal{S}_r,$$

which now satisfies

$$\|\mathcal{T}_r(A_1, \dots, A_r)\| \leq \prod_{j=1}^r \|A_j\|, \quad \forall A_1, \dots, A_r \in \mathbf{M}_n(K).$$

However this inequality ignores the cancellations that are likely to occur because of the signs $\varepsilon(\sigma)$ in the definition of \mathcal{T}_r . For instance, the left-hand side vanishes whenever $r \geq 2n$. For this reason, we are interested in the *norm* $\tau(r, n)$ of \mathcal{T}_r , defined by

$$\tau(r, n) := \sup \left\{ \frac{\|\mathcal{T}_r(A_1, \dots, A_r)\|}{\prod_{j=1}^r \|A_j\|} : A_1, \dots, A_r \in \mathbf{M}_n(K) \setminus \{0_n\} \right\}.$$

Alternatively,

$$\tau(r, n) := \sup \{ \|\mathcal{T}_r(A_1, \dots, A_r)\| : A_1, \dots, A_r \in \mathbf{M}_n(K), \|A_1\|, \dots, \|A_r\| \leq 1 \} = \frac{C(r, n)}{r!}.$$

Our definition above depends on the chosen norm on $\mathbf{M}_n(K)$, and our notation should indicate this dependence. In this sense, the Frobenius norm and the operator norm $\|\cdot\|_2$ yield the numbers $\tau_F(r, n)$ and $\tau_2(r, n)$, respectively.

As said above, we always have $\tau(r, n) \leq 1$. However this bound is very poor as for instance $\tau(2n, n) = 0$. Besides, one trivially has $\tau(1, n) = 1$. The first non-trivial case comes when $r = 2$, where \mathcal{S}_2 is the commutator. When $n = 1$, clearly $\tau(2, 1) = 0$. But for $n \geq 2$, the result depends on the norm we are considering. For instance, we know that

$$\tau_F(2, n) = \frac{\sqrt{2}}{2},$$

the case $n = 2$ being due to Böttcher and Wenzel [3], and the case $n = 3$ to László [6]. The equality for every n was conjectured in [3] and proved by [9], and independently by [7] for $K = \mathbb{R}$ and [1] for $K = \mathbb{C}$. See also [4]. The situation is significantly different with the standard operator norm, induced by the Hermitian norm. It is known ([4], Example 5.2) that

$$\tau_2(2, n) = 1, \quad \forall n \geq 2.$$

All the subtlety in the upcoming analysis comes from cancellations. We shall prove several inequalities that hold true for every submultiplicative norm. The simplest one, $\tau(r+s, n) \leq \tau(r, n)\tau(s, n)$, is not sharp. But it suggests to extend our study to operators in infinite dimension. An interesting class in this respect is that of Hilbert–Schmidt operators, whose norm generalizes the Frobenius norm. We thus define the upper bound

$$\tau_F(r) := \sup_{n \geq 1} \tau_F(r, n) = \lim_{n \rightarrow +\infty} \tau_F(r, n).$$

Again, one has $\tau_F(r+s) \leq \tau_F(r)\tau_F(s)$. Since Theorem 1.1 is lost as $n \rightarrow +\infty$, we have $\tau_F(r) > 0$ for every r , and it becomes interesting to compute the *rate of cancellation*

$$\rho_F := \lim_{r \rightarrow +\infty} \tau_F(r)^{1/r} = \inf_{r \geq 1} \tau_F(r)^{1/r}.$$

Because of Böttcher–Wenzel’s inequality, we have $\tau_F(2) = \sqrt{2}/2$ and $\rho_F \leq 2^{-1/4}$. One of our results from below is the improved bound

$$\rho_F \leq \tau_F(2k+1)^{1/2k}, \quad \forall k \geq 1.$$

2 The commutator in terms of the numerical radius

For the following result concerning the numerical radius h , we benefited of a fruitful discussion with Piotr Migdal.

Theorem 2.1 *Let $n \geq 2$ be given. Then*

$$(1) \quad h([A, B]) \leq 4h(A)h(B), \quad \forall A, B \in \mathbf{M}_n(\mathbb{C}).$$

The constant 4 is the smallest possible.

Proof If $M \in \mathbf{M}_n(\mathbb{C})$, we define the *real* and *imaginary parts* of M by

$$\Re M = \frac{1}{2}(M + M^*) \in \mathbb{H}_n \quad \Im M = \frac{1}{2i}(M - M^*).$$

Then

$$\begin{aligned} h(M) &= \sup_x \sup_{\varphi} \frac{\Re(e^{i\varphi} x^* M x)}{\|x\|_2^2} = \sup_x \sup_{\varphi} \frac{x^* \Re(e^{i\varphi} M) x}{\|x\|_2^2} \\ &= \sup_{\varphi} \sup_x \frac{x^* \Re(e^{i\varphi} M) x}{\|x\|_2^2} = \sup_{\varphi \in \mathbb{R}/2\pi\mathbb{Z}} \|\Re(e^{-i\varphi} M)\|_2, \end{aligned}$$

where the supremum is actually a maximum, because $\varphi \mapsto \|\Re(e^{-i\varphi} M)\|_2$ is continuous over the compact set $\mathbb{R}/2\pi\mathbb{Z}$.

Let $A, B \in \mathbf{M}_n(\mathbb{C})$ be given. We write $M = [A, B]$ and chose a φ with $h(M) = \|\Re(e^{-i\varphi} M)\|_2$. Let us denote $X = \Re(e^{-i\varphi} A)$, $Y = \Im(e^{-i\varphi} A)$, $Z = \Re B$ and $T = \Im B$. From

$$e^{-i\varphi} M = [X + iY, Z + iT] = [X, Z] + [T, Y] + i([X, T] + [Y, Z])$$

and the fact that the commutator of Hermitian matrices is skew-Hermitian, we derive

$$\Re(e^{-i\varphi} M) = i([X, T] + [Y, Z]).$$

We infer

$$h([A, B]) = \|[X, T] + [Y, Z]\|_2 \leq \|[X, T]\|_2 + \|[Y, Z]\|_2 \leq 2(\|X\|_2 \cdot \|T\|_2 + \|Y\|_2 \cdot \|Z\|_2).$$

This proves immediately (1).

The constant 4 is attained for the choice

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \oplus I_{n-2}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \oplus I_{n-2},$$

for which

$$[A, B] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus I_{n-2}.$$

We have

$$h(A) = h(B) = \frac{1}{2}, \quad h([A, B]) = 1.$$

This completes the proof.

Inequality (1) is of little interest; a similar proof yields the same optimal factor 4 if one replace the commutator by $AB + BA$:

$$h(AB + BA) \leq 4h(A)h(B).$$

Consequently the numerical radius does not detect the possible cancellations, in contrast to the Frobenius norm.

When considering polynomials of higher degree, the same proof provides the inequality

$$h(\mathcal{S}_r(A_1, \dots, A_r)) \leq 2^{r-1} r! \prod_{j=1}^r h(A_j),$$

which seems to be very far from optimal.

3 General inequalities

We assume for the remainder that the norm is submultiplicative:

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

3.1 Two formulæ about \mathcal{S}_r

It is well-known that

$$\mathcal{S}_r(x_1, \dots, x_r) = \sum_{i=1}^r (-1)^{i-1} x_i \mathcal{S}_{r-1}(\hat{x}_i),$$

where \hat{x}_i is obtained from x by deleting x_i . The following formula generalizes the one above.

Proposition 3.1 *Let $r = r_1 + \dots + r_\ell$ be a partition into positive integers ($k \mapsto r_k$ need not to be monotonous). Denote by $\mathcal{P}(r_1, \dots, r_\ell)$ the set of ordered partitions of $\llbracket 1, r \rrbracket = \{1, \dots, r\}$ into subsets I_1, \dots, I_ℓ such that $|I_k| = r_k$. If $I = (I_1, \dots, I_\ell) \in \mathcal{P}(r_1, \dots, r_\ell)$, and $I_k = \{i_{k,1} < \dots < i_{k,r_k}\}$, let us denote $\alpha(I)$ the signature of the permutation ρ_I defined by*

$$\rho_I(j + r_1 + \dots + r_{k-1}) = i_{k,j}, \quad 1 \leq j \leq r_k, 1 \leq k \leq \ell.$$

Then

$$(2) \quad \mathcal{S}_r(x_1, \dots, x_r) = \sum_{I \in \mathcal{P}(r_1, \dots, r_\ell)} \alpha(I) \mathcal{S}_{r_1}(X_{I_1}) \cdots \mathcal{S}_{r_\ell}(X_{I_\ell}),$$

where $X_{I_k} = (x_{k,1}, \dots, x_{k,r(k)})$.

Proof Every $\sigma \in \mathfrak{S}_r$ can be factorized as

$$\sigma = (\sigma_1 \times \cdots \times \sigma_\ell) \circ \rho_I$$

with $I \in \mathcal{P}(r_1, \dots, r_\ell)$ and $\sigma_k \in \mathfrak{S}_{r_k}$. For this, take $I_1 = \sigma(\{1, \dots, r_1\})$, etc... This decomposition is unique; actually, one has

$$|\mathfrak{S}_r| = r! = r_1! \cdots r_\ell! \binom{r}{r_1} \binom{r-r_1}{r_2} \cdots \binom{r_\ell}{r_\ell} = |\mathfrak{S}_{r_1} \times \cdots \times \mathfrak{S}_{r_\ell}| \times \text{card } \mathcal{P}(r_1, \dots, r_\ell).$$

The signature of σ is obviously the product of all the signatures of $\sigma_1, \dots, \sigma_\ell, \rho_I$. Therefore the right-hand side of (2) contains exactly once the monomial $x_{\sigma(1)} \cdots x_{\sigma(r)}$, with the sign

$$\alpha(I) \varepsilon(\sigma_1) \cdots \varepsilon(\sigma_\ell) = \varepsilon(\sigma).$$

Hence both sides are equal to each other. This proves the proposition.

Let $1 \leq k < m \leq \ell$ be given, and denote ζ the transposition (k, m) . The ordered partition I^ζ is defined by $I_k^\zeta = I_m$, $I_m^\zeta = I_k$ and $I_j^\zeta = I_j$ otherwise; mind that r_k and r_m have been flipped. We have $\alpha(I^\zeta) = (-1)^{r_k r_m} \alpha(I)$. Hence we derive from (2) the following identities.

Proposition 3.2 *Let $r = r_1 + \cdots + r_\ell$. Then,*

$$(3) \quad \sum_{I \in \mathcal{P}(r_1, \dots, r_\ell)} \alpha(I) \mathcal{S}_\ell(\mathcal{S}_{r_1}(X_{I_1}), \dots, \mathcal{S}_{r_\ell}(X_{I_\ell})) = \begin{cases} \ell! \mathcal{S}_r(X), & \text{if every } r_k \text{ is odd,} \\ 0, & \text{otherwise.} \end{cases}$$

The case $r = 3 = 2 + 1$ (hence $\ell = 2$) of (3) is the Jacobi identity for the commutator.

Proof Let $F(X)$ denote the left-hand side. The transposition $i \leftrightarrow j$ induces an involution S over $\mathcal{P}(r_1, \dots, r_\ell)$ and we have $\alpha(SI) = -\alpha(I)$. Therefore

$$F(x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_{j-1}, x_i, x_{j+1}, \dots, x_r) = -F(X).$$

Since F is a homogenous polynomial, linear in each of the indeterminates, we deduce that F equals a multiple of \mathcal{S}_r .

In order to determine the constant factor, we may focus on the monomial $X^\uparrow = x_1 x_2 \cdots x_r$. This monomial occurs in the sum each time the interval $\llbracket 1, r \rrbracket$ is split into consecutive intervals $I_{\mathbf{v}(1)}, \dots$ of respective lengths $r_{\mathbf{v}(1)}$, where $\mathbf{v} \in \mathfrak{S}_\ell$ is arbitrary. It is accompanied by the sign $\alpha(I_{\mathbf{v}(1)}, \dots) \varepsilon(\mathbf{v})$. If all the r_k are odd, then $\alpha(I_{\mathbf{v}(1)}, \dots) \varepsilon(\mathbf{v}) = +1$ for every \mathbf{v} and all the terms have a positive sign, whence the factor $\ell!$.

Suppose on the contrary that some r_k is even, say r_1 , and let us denote ζ the transposition $1 \leftrightarrow 2$. Then \mathfrak{S}_ℓ is the disjoint union of \mathfrak{A}_ℓ and $\zeta \mathfrak{A}_\ell$. If $\mathbf{v} \in \mathfrak{A}_\ell$, then

$$\alpha(I_{\zeta \mathbf{v}(1)}, \dots) = \alpha(I_{\mathbf{v}(1)}, \dots),$$

while $\varepsilon(\zeta \mathbf{v}) = -\varepsilon(\mathbf{v})$. Therefore half of the occurrences of X^\uparrow have a positive sign, and half of them have a negative sign, whence the second formula and the proposition.

3.2 Submultiplicativity of $\tau(\cdot, n)$

The identity (2) allows us to derive a bound for $\tau(r, n)$ in terms of $\tau(s, n)$ and $\tau(r-s, n)$. If A_1, \dots, A_r are in the unit ball, then each product in the sum above is bounded by

$$s!(r-s)!\tau(s, n)\tau(r-s, n) \prod_{j=1}^r \|A_j\|.$$

Since there are $r!/(s!(r-s)!)$ terms in the sum, we immediately obtain

$$\|\mathcal{T}_r(A_1, \dots, A_r)\| \leq \tau(s, n)\tau(r-s, n) \prod_{j=1}^r \|A_j\|,$$

from which we derive the following estimate.

Proposition 3.3 *If the norm is submultiplicative, then*

$$\tau(r, n) \leq \tau(s, n)\tau(r-s, n), \quad \forall 1 \leq s < r.$$

In particular, $\tau_F(r) \leq \tau_F(s)\tau_F(r-s)$. It is well known that for such a submultiplicative sequence, the sequence $u_r := \tau_F(r)^{1/r}$ converges to its lower bound. We call this limit the *rate of cancellation* and denote it by

$$\rho_F = \lim_{r \rightarrow +\infty} \tau_F(r)^{1/r} = \inf_r \tau_F(r)^{1/r}.$$

Because of $\tau(r, n) \leq 1$, we infer the next statement.

Corollary 3.1 *As a function of its first argument r , $\tau(r, n)$ is non-increasing.*

3.3 Improved submultiplicativity

The formula (2) has the drawback that it still involves the product of matrices, for which we have no gain in norm, since we cannot improve $\|AB\| \leq \|A\| \|B\|$ in general. To go further, we use (3), which involves operators \mathcal{T}_ℓ but no single matrix product. It can be recast as

$$(4) \quad \sum_{I \in \mathcal{P}(r_1, \dots, r_\ell)} \alpha(I) \mathcal{T}_\ell(\mathcal{T}_{r_1}(X_{I_1}), \dots, \mathcal{T}_{r_\ell}(X_{I_\ell})) = \begin{cases} \text{card} \mathcal{P}(r_1, \dots, r_\ell) \cdot \mathcal{T}_r(X), & \text{if every } r_k \text{ is odd,} \\ 0, & \text{otherwise.} \end{cases}$$

For instance, we have

$$\begin{aligned} 8\mathcal{T}_4(A, B, C, D) &= [A, \mathcal{T}_3(B, C, D)] + [B, \mathcal{T}_3(A, D, C)] \\ &\quad + [C, \mathcal{T}_3(A, B, D)] + [D, \mathcal{T}_3(A, C, B)], \end{aligned}$$

an identity which immediately gives

$$\tau(4, n) \leq \tau(3, n)\tau(2, n).$$

The latter inequality is tighter than $\tau(4, n) \leq \tau(2, n)^2$ given by Proposition 3.3, since $r \mapsto \tau(r, n)$ is non-increasing.

More generally, we have

$$(5) \quad 4r\mathcal{T}_{2r}(A_1, \dots, A_{2r}) = \sum_{i=1}^{2r} (-1)^{i+1} [A_i, \mathcal{T}_{2r-1}(\hat{A}_i)],$$

with the classical notation \hat{A}_i for the list $A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_r$ in which A_i has been omitted. We deduce immediately

$$\tau(2r, n) \leq \tau(2r-1, n)\tau(2, n).$$

Using again submultiplicativity, this yields

$$\tau(2r, n) \leq \tau(2, n)\tau(3, n)\tau(2r-4, n).$$

By induction, we infer

$$\tau(4k, n) \leq \tau(2, n)^k \tau(3, n)^k, \quad \tau(4k+2, n) \leq \tau(2, n)^{k+1} \tau(3, n)^k,$$

where submultiplicativity alone only grants $\tau(2\ell, n) \leq \tau(2, n)^\ell$.

More generally, (4) yields the inequality

$$(6) \quad \tau(r, n) \leq \tau(\ell, n)\tau(r_1, n) \cdots \tau(r_\ell, n).$$

This improves the submultiplicativity in the following way: let us define a new sequence θ by shifting the first argument

$$\theta(s, n) := \tau(s+1, n).$$

Then (6) becomes

$$(7) \quad \theta(s, n) \leq \theta(s_0, n)\theta(s_1, n) \cdots \theta(s_\ell, n),$$

for $s = s_0 + \cdots + s_\ell$, whenever s_1, \dots, s_ℓ are even (s_0 may be odd). This is exactly submultiplicativity, up to the restriction on parity. In particular, the sequence $\mu(k, n) := \theta(2k, n)$ is submultiplicative.

Likewise, let us denote $\theta(s) = \tau_F(s+1)$. Clearly, $\tau_F(r)^{1/r}$ and $\theta(s)^{1/s}$ have the same limit, which must be equal to the infimum of the $\theta(2k)^{1/2k}$. Combined, this delivers the following result.

Proposition 3.4 *We have*

$$\rho_F \leq \tau_F(2k+1)^{1/2k}, \quad \forall k \in \mathbb{N}.$$

The above bound improves, for odd arguments, the one we had before, namely $\tau(r)^{1/r}$.

4 The Frobenius norm

The Frobenius norm, which it is the one studied in [3], has several advantages for our study. First of all, it enjoys better bounds than either the operator norm, or the numerical radius: the limit $n \rightarrow +\infty$ seems non trivial. Next, it is a smooth, regular norm, and it is possible to use differential calculus when studying r -uplets which realize the norm of \mathcal{T}_r . At last, we have a duality principle, based on the inner product of two matrices

$$\|M\|_F = \sup\{\Re\text{Tr}(M^*N) : \|N\|_F \leq 1\},$$

whence

$$\|[A, B]\|_F = \sup\{\Re\text{Tr}([A, B]C) : C \in B_F\},$$

where B_F denotes the unit ball for the Frobenius norm. Since $3\text{Tr}([A, B]C) = \text{Tr}\mathcal{S}_3(A, B, C)$, we also have (say that $K = \mathbb{R}$)

$$\tau_F(2, n) = \sup\{\text{Tr}\mathcal{T}_3(A, B, C) : A, B, C \in B_F\}.$$

We warn the reader that this identity extends only to the even numbers r :

$$\tau_F(2s, n) = \sup\{\text{Tr}\mathcal{T}_{2s+1}(A_1, \dots, A_{2s+1}) : A_1, \dots, A_{2s+1} \in B_F\},$$

while

$$\text{Tr}\mathcal{T}_{2s}(A_1, \dots, A_{2s}) \equiv 0.$$

The latter identity expresses the fact that $\mathcal{T}_{2s}(A_1, \dots, A_{2s})$ can be written as a sum of commutators, an idea developed in Section 3.3. The vanishing of these traces is used in the proof by Rosset [8] of the Amitsur–Levitski Theorem.

4.1 Average estimate

Let us endow $\mathbf{M}_n(\mathbb{R})$ with the usual probability measure, where the entries a_{ij} are Gaussian independent variables:

$$d\mu(A) = \frac{1}{V_n} e^{-\|A\|^2} da_{11} \cdots da_{nn}, \quad \|A\|^2 = \text{Tr}(A^T A).$$

Here V_n is a normalizing factor. For instance, we have

$$\mathbb{E}(\|A\|^2) = n^2 \frac{\int x^2 e^{-x^2} dx}{\int e^{-x^2} dx} =: n^2 m_2,$$

where $m_2 = \frac{1}{2}$ is the second moment of the Gaussian.

We wish to calculate the expectation of $\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2$ when A^1, \dots, A^ℓ are independent matrices. This amounts to calculating the average of $\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2$ when A^1, \dots, A^ℓ all have unit norm.

Lemma 4.1 *As a function of the size n of the matrices, the expression*

$$n^{2\ell} \mathbb{E}[\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2]$$

is a polynomial.

Proof Denoting $A^\pi = A^{\pi(1)} \dots A^{\pi(\ell)}$, we have

$$\begin{aligned} \mathbb{E}[\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2] &= \mathbb{E}[\text{Tr} \mathcal{S}_\ell(A^1, \dots, A^\ell)^T \mathcal{S}_\ell(A^1, \dots, A^\ell)] \\ &= \sum_{\pi \in \mathfrak{S}_\ell} \mathbb{E}[\varepsilon(\pi) \text{Tr} (A^\pi)^T \mathcal{S}_\ell(A^1, \dots, A^\ell)] \\ &= \sum_{\pi \in \mathfrak{S}_\ell} \mathbb{E}[\varepsilon(\pi) \text{Tr} (B^{\text{id}})^T \mathcal{S}_\ell(B^{\pi^{-1}(1)}, \dots, B^{\pi^{-1}(\ell)})] \\ &= \ell! \mathbb{E}[\text{Tr} (B^{\text{id}})^T \mathcal{S}_\ell(B^1, \dots, B^\ell)] \\ &= \ell! \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) \mathbb{E}[\text{Tr} (B^{\text{id}})^T B^\pi]. \end{aligned}$$

Given $\pi \in \mathfrak{S}_\ell$, we have

$$\text{Tr} (A^{\text{id}})^T A^\pi = a_{\alpha_{\ell-1}\alpha_\ell}^\ell \cdots a_{\alpha_1\alpha_2}^2 a_{\beta_1\alpha_1}^1 a_{\beta_1\beta_2}^{\pi(1)} \cdots a_{\beta_\ell\alpha_\ell}^{\pi(\ell)},$$

with Einstein's convention of summation over repeated indices. We stress that β_1 plays the role of an α_0 , and α_ℓ plays the role of a $\beta_{\ell+1}$. When taking the expectation of a given monomial, we obtain either $(m_2)^\ell$ or 0, according to whether every entry shows up with a square, or not. A non-zero contribution happens when

$$(\beta_k, \beta_{k+1}) = (\alpha_{\pi(k)-1}, \alpha_{\pi(k)}),$$

or equivalently if

$$(8) \quad \alpha_{\pi(k)-1} = \alpha_{\pi(k-1)} = \beta_k$$

for every $k = 1, \dots, \ell + 1$. Hereabove we have to extend π by

$$(9) \quad \pi(0) = 0, \quad \pi(\ell + 1) = \ell + 1.$$

Let G^π be the graph whose vertices are the indices $0 \leq j \leq \ell$ for the α 's, and the indices $1 \leq k \leq \ell + 1$ for the β 's. Thus G^π has $2(\ell + 1)$ vertices. The edges correspond to every equality of the form either $j = \pi(k - 1)$ or $j = \pi(k) - 1$. This includes the edge between the vertices $j = 0$ and $k = 1$, and the edge between the vertices $j = \ell$ and $k = \ell + 1$. Notice that j and k may be connected by two edges, in case $\pi(k) - 1 = \pi(k - 1) = j$.

Given the permutation π , many among the monomials

$$a_{\alpha_{\ell-1}\alpha_\ell}^\ell \cdots a_{\alpha_1\alpha_2}^2 a_{\beta_1\alpha_1}^1 a_{\beta_1\beta_2}^{\pi(1)} \cdots a_{\beta_\ell\alpha_\ell}^{\pi(\ell)}$$

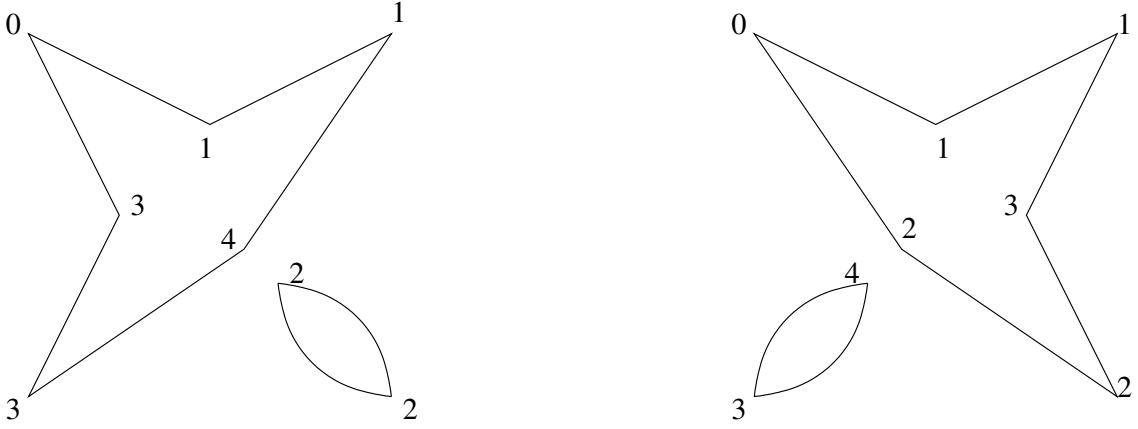


Figure 1: The graph G^π when $\ell = 3$. The α -indices, from 0 to 3, are outer; the β -indices, from 1 to 4, are inner. Left: the cycle (123). Right: the transposition (12). In both cases, the graph has two connected components. For the identity, it should have four of them.

have zero expectation. The remaining ones have expectation m_2^ℓ ; they are parametrized by the maps

$$G^\pi \xrightarrow{(\alpha, \beta)} [[1, n]]$$

that are constant on each connected component. The number of such maps (α, β) is $n^{N(\pi)}$, where $N(\pi)$ is the number of connected components of G^π . Therefore, we obtain

$$\mathbb{E}[\text{Tr}(A^{\text{id}})^T A^\pi] = n^{N(\pi)} m_2^\ell$$

and

$$\mathbb{E}[\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2] = \ell! m_2^\ell \sum_{\pi \in \mathfrak{S}_\ell} \varepsilon(\pi) n^{N(\pi)} =: \ell! P_\ell(n) m_2^\ell,$$

for some $P_\ell \in \mathbb{Z}[X]$. We infer

$$(10) \quad \mathbb{E}[\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2] = \frac{\ell! P_\ell(n)}{n^{2\ell}} \mathbb{E}[\|A^1\|^2 \dots \|A^\ell\|^2].$$

We are now going to express the polynomial P_ℓ in closed form. To begin with, we note that always $N(\pi) \geq 1$, and therefore $P_\ell(0) = 0$. Actually, the quantity $N(\pi)$ can further be restricted.

Proposition 4.1 *For every $\pi \in \mathfrak{S}_\ell$, we have $1 \leq N(\pi) \leq \ell + 1$ and*

$$N(\pi) \equiv \ell + 1 \pmod{2}.$$

In addition,

$$(N(\pi) = \ell + 1) \iff (\pi = \text{id}).$$

Proof The edges of π always link an α -index j to a β -index k . In addition, every vertex has valence 2. Therefore a connected component is an even cycle in which the j 's and the k 's alternate. We may therefore construct a graph J^π , whose vertices are the indices $j \in \llbracket 0, \ell \rrbracket$ and there is an edge between j and j' if $\{j, j'\} = \{S\pi k, \pi S k\}$ for some k , where S is the shift $k \mapsto k - 1$ over $\llbracket 1, \ell + 1 \rrbracket$. For completeness, we define $S(0) = \ell + 1$, so that S is a permutation of $\llbracket 0, \ell + 1 \rrbracket$. This amounts to saying that J^π is the graph associated with the permutation $\rho := \pi S \pi^{-1} S^{-1}$ in $\llbracket 0, \ell \rrbracket$ (because of (9), $\ell + 1$ is a fixed point of ρ). For instance, the examples displayed in Figure 1 yield $\rho = (031)$ (left) or $\rho = (021)$ (right); in both cases, the resulting J^π consists of a three-node cycle and an isolated vertex.

Every connected component $(j_1, k_1, j_2, k_2, \dots)$ of G^π corresponds to a connected component (j_1, j_2, \dots) of J^π . Thus $N(\pi)$ is nothing but the number of orbits of ρ .

Because ρ is a commutator, it is an even permutation of $\llbracket 0, \ell \rrbracket$. The number N_{even} of its cycles of even length (these are odd permutations !) is thus even. Counting the elements *modulo* 2, we deduce that $N_{\text{odd}} \equiv \ell + 1$. Hence $N(\pi) = N_{\text{even}} + N_{\text{odd}} \equiv \ell + 1$.

If π is the identity, then ρ is the identity too and $N(\pi) = \ell + 1$. Conversely, if $N(\pi) = \ell + 1$, we have $\rho = \text{id}$, that is $\pi(k - 1) = \pi(k) - 1$ for every $k \in \llbracket 1, \ell + 1 \rrbracket$. With (9), we deduce that π is the identity.

Corollary 4.1 *The monic polynomial P_ℓ is odd if ℓ is even, and it is even if ℓ is odd. It has the form $P_\ell(X) = X^{\ell+1} + \text{l.o.t.}$, without constant term.*

Finally, we invoke the theorem of Amitsur and Levistki. If $2n \leq \ell$, we have $P_\ell(n) = 0$, thus $X - n$ divides $P_\ell(X)$. Because of the parity, we infer that $X^2 - n^2$ divides P_ℓ . Hence

$$P_\ell(X) = Q_\ell(X)(X^2 - 1)(X^2 - 4) \cdots (X^2 - r^2), \quad r = \left\lfloor \frac{\ell}{2} \right\rfloor.$$

By the corollary, we know that Q_ℓ is a monic polynomial of degree $\ell + 1 - 2r$. If ℓ is even, then Q_ℓ is odd, of degree 1, hence equals X . If ℓ is odd, Q_ℓ is even of degree 2, vanishes at 0 and is therefore equal to X^2 . Summarizing these thoughts, we obtain the desired relation.

Theorem 4.1 *Let A^j be i.i.d. Gaussians. Then,*

$$\mathbb{E}[\|\mathcal{S}_\ell(A^1, \dots, A^\ell)\|^2] = \frac{\ell! P_\ell(n)}{n^{2\ell}} \mathbb{E}[\|A^1\|^2 \cdots \|A^\ell\|^2]$$

with

$$\begin{aligned} P_{2k}(X) &= X(X^2 - 1^2) \cdots (X^2 - k^2), \\ P_{2k+1}(X) &= X^2(X^2 - 1^2) \cdots (X^2 - k^2). \end{aligned}$$

4.2 Asymptotic properties

We focus on the even case, $\ell = 2k$ which behaves a little nicer than the odd one, even if their asymptotics are quite similar.

Let us denote the ratio given by Theorem 4.1 by

$$\omega(k; n) = \frac{(2k)! P_{2k}(n)}{n^{4k}}.$$

First fix n and vary k from 1 to $n-1$. We have

$$\omega(1; n) = \frac{2}{n} \left(1 - \frac{1}{n^2}\right),$$

a formula already known to Böttcher & Wenzel [3]. Then

$$\frac{\omega(k; n)}{\omega(k-1; n)} = \frac{2k(2k-1)}{n^4} (n^2 - k^2).$$

Denoting $t = \frac{k}{n}$, this reads

$$\frac{\omega(k; n)}{\omega(k-1; n)} = 4t \left(t - \frac{1}{2n}\right) (1 - t^2) < 4t^2(1 - t^2) \leq 1.$$

Therefore the sequence $k \mapsto \omega(k; n)$ is strictly decreasing. It has a critical point for $t_n \sim \frac{\sqrt{2}}{2}$ (that is $k_n \sim \frac{n\sqrt{2}}{2}$). The decay is faster for small values of k , and also for k approaching n .

We next investigate the behaviour of $\omega(k; n)$ as $n \rightarrow +\infty$, while $k/n \rightarrow t \in (0, 1)$. We start from

$$\omega(k; n) = \frac{(2k)!(n+k)!}{n^{4k}(n-k-1)!}.$$

The Stirling formula yields

$$\begin{aligned} \omega(k; n) &\sim 2n^{-4k} \left(\frac{2k}{e}\right)^{2k} \left(\frac{n+k}{e}\right)^{n+k} \left(\frac{e}{n-k-1}\right)^{n-k-1} \sqrt{\pi k \frac{n+k}{n-k-1}} \\ &= 2 \left(\frac{4k^2(n+k)(n-k-1)}{e^4 n^4}\right)^k \left(\frac{n+k}{n-k-1}\right)^n \frac{n-k-1}{e} \sqrt{\pi k \frac{n+k}{n-k-1}}. \end{aligned}$$

Hence

$$\log \omega(k; n) \sim dk,$$

where

$$d = d(t) := \log(4t^2(1-t^2)) - 4 + \frac{1}{t} \log \frac{1+t}{1-t}.$$

A simple calculation gives that τ is increasing over $(0, x)$ and decreasing over $(x, 1)$, where $x \approx 0.95 \in (0, 1)$ is the unique root of

$$(11) \quad 4t = \log \frac{1+t}{1-t}.$$

Then

$$\max d(t) = d(x) = \log 4x^2(1-x^2) < 0.$$

In conclusion, $n \mapsto \omega(\lfloor tn \rfloor; n)$ decays exponentially fast, with a rate not larger than $4x^2(1-x^2)$, a number strictly less than one.

5 The law of distribution of the commutator ($n = 2$)

We continue our investigation for the Frobenius norm. We already know that

$$\|[A, M]\|_F \leq \sqrt{2} \|A\|_F \|M\|_F,$$

where the constant $\sqrt{2}$ is optimal [3]. As noted by several authors, we may always restrict to the hyperplane H of zero-trace matrices, because on the one hand $[A + tI_n, M] = [A, M] = [A, M + sI_n]$ for every $s, t \in \mathbb{R}$, and on the other hand the projection $A \mapsto A - \frac{1}{n}(\text{Tr } A)I_n$ diminishes the Frobenius norm. We denote by $B = B_H$ the unit ball in H and ask two questions:

- What is the range of the map $(A, M) \mapsto \frac{\sqrt{2}}{2} [A, M]$ over $B_H \times B_H$? We already know that it is contained in B_H .
- What is the distribution law of $[A, M]$ when A and M are chosen uniformly and independently in B_H ?

We solve these questions for the case $n = 2$. The following assertion is immediate

Proposition 5.1 *The linear map*

$$\begin{pmatrix} a & b \\ c & -a \end{pmatrix} \mapsto \begin{pmatrix} a\sqrt{2} \\ b \\ c \end{pmatrix}$$

is an isometry between H and the standard Euclidean space \mathbb{R}^3 .

Its inverse $L : \mathbb{R}^3 \rightarrow H$ satisfies

$$L(x \times y) = \frac{\sqrt{2}}{2} [Lx, Ly]^T.$$

Therefore we have the commutative diagram

$$\begin{array}{ccc} (A, M) & \longrightarrow & \frac{\sqrt{2}}{2} [A, M] \\ L \uparrow & & L \uparrow \\ (x, y) & \longrightarrow & F(x \times y) \end{array}$$

where L is an isometry, and

$$F : \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mapsto \begin{pmatrix} a \\ c \\ b \end{pmatrix}$$

is an isometry in \mathbb{R}^3 . Because the diagram's bottom line is onto B_3 , we already see that

$$(A, M) \in B_H \times B_H \mapsto \frac{\sqrt{2}}{2} [A, M] \in B_H$$

is surjective, too.

Let us denote $\mathrm{d}\nu$ the pushforward of $\mathrm{d}A \otimes \mathrm{d}M$, and $\mathrm{d}\mu$ that of $\mathrm{d}x \otimes \mathrm{d}y$. Because of the equivariance $R(x \times y) = (Rx) \times (Ry)$ under every rotation R , we know that $\mathrm{d}\mu$ is rotationally invariant. Therefore

$$\begin{aligned} \int_{B_H} f(Z) \mathrm{d}\nu(Z) &:= \iint_{B_H \times B_H} f\left(\frac{\sqrt{2}}{2} [A, M]\right) \mathrm{d}A \mathrm{d}M \\ &= \iint_{B_3 \times B_3} f\left(\frac{\sqrt{2}}{2} [Lx, Ly]\right) \mathrm{d}x \mathrm{d}y \\ &= \iint_{B_3 \times B_3} f(LF(x \times y)) \mathrm{d}x \mathrm{d}y = \int_{B_3} (f \circ L \circ F)(z) \mathrm{d}\mu(z) \\ &= \int_{B_3} (f \circ L)(Fz) \mathrm{d}\mu(z) = \int_{B_3} (f \circ L)(z) \mathrm{d}\mu(z). \end{aligned}$$

This shows that the distribution of $(A, M) \mapsto \frac{\sqrt{2}}{2} [A, M]$ over $B_H \times B_H$ is exactly the same as the distribution of $(x, y) \mapsto x \times y$ over the unit ball B_3 , up to the identification provided by the isometry L .

Proposition 5.2 *The map $(x, y) \mapsto x \times y$ from $B_3 \times B_3$ is onto B_3 . The pushforward $\mathrm{d}\mu$ of $\mathrm{d}x \otimes \mathrm{d}y$ is radial, with density*

$$h(\rho) \mathrm{d}\rho \mathrm{d}\omega$$

where $\mathrm{d}\omega$ is the normalized area over S^2 , and

$$h(\rho) = 9\rho^2 \left(\frac{\sqrt{1-\rho^2}}{\rho} + 2 \arctan \sqrt{\frac{1+\rho}{1-\rho}} - \pi \right).$$

This instantly transfers to the commutator setting.

Corollary 5.1 *The image of $B_H \times B_H$ under the commutator is $\sqrt{2}B_H$. The pushforward $\mathrm{d}\nu$ of $\mathrm{d}A \otimes \mathrm{d}M$ under the map $(A, M) \mapsto \frac{\sqrt{2}}{2} [A, M]$ is radial, with distribution determined by*

$$\int_{B_H} \phi(\|Z\|_F) \mathrm{d}\nu(Z) = \int_0^1 h(\rho) \phi(\rho) \mathrm{d}\rho, \quad \forall \phi \in C([0, 1]),$$

with h as in Proposition 5.2.

Proof We only have to calculate the density h . Because it is radial, we only need to consider radial functions $f(z) = \phi(|z|)$; we have

$$\int_0^1 \phi(\rho) h(\rho) \mathrm{d}\rho = \iint \phi(|x \times y|) \mathrm{d}x \mathrm{d}y,$$

where f denotes the normalized integral over B_3 . Using again the rotational invariance, we have

$$\begin{aligned}
\int_0^1 \phi(\rho)h(\rho)d\rho &= 3 \int_0^1 r^2 dr \int \phi(r|\vec{e}_1 \times y|) dy = 9 \int_0^1 \int_0^1 r^2 s^2 dr ds \int_{S^2} \phi(rs|\vec{e}_1 \times e|)d\omega(e) \\
&= \frac{9}{2} \int_0^1 \int_0^1 r^2 s^2 dr ds \int_0^\pi \phi(rs \sin \theta) \sin \theta d\theta \\
&= \frac{9}{2} \int_0^1 \int_0^1 r^2 s^2 dr ds \int_{-1}^1 \phi(rs\sqrt{1-c^2}) dc \\
&= 9 \int_0^1 \int_0^1 r^2 s^2 dr ds \int_0^1 \phi(rs\sqrt{1-c^2}) dc.
\end{aligned}$$

Denoting $\rho := rs\sqrt{1-c^2}$, we have

$$\rho \leq rs, \quad |drdsd\rho| = \frac{rs}{\rho} \sqrt{r^2 s^2 - \rho^2} |drdsdc|,$$

hence

$$\int_0^1 \phi(\rho)h(\rho)d\rho = 9 \int_0^1 \int_0^1 rs dr ds \int_0^{rs} \phi(\rho) \frac{\rho d\rho}{\sqrt{r^2 s^2 - \rho^2}}.$$

We infer

$$h(\rho) = 9\rho \iint_{0 < r, s < 1, rs > \rho} \frac{rs dr ds}{\sqrt{r^2 s^2 - \rho^2}}.$$

We integrate first with respect to s , which varies from ρ/r to 1. From

$$d\sqrt{r^2 s^2 - \rho^2} = \frac{r^2 s ds}{\sqrt{r^2 s^2 - \rho^2}},$$

we deduce

$$h(\rho) = 9\rho \int_\rho^1 \frac{dr}{r} \left[\sqrt{r^2 s^2 - \rho^2} \right]_{\rho/r}^1 = 9\rho \int_\rho^1 \sqrt{r^2 - \rho^2} \frac{dr}{r}.$$

Let us parametrize

$$r = \rho \frac{a^2 + 1}{a^2 - 1}, \quad \sqrt{r^2 - \rho^2} = \rho \frac{2a}{a^2 - 1}, \quad a > \sqrt{\frac{1+\rho}{1-\rho}}.$$

We obtain

$$h(\rho) = 9\rho^2 \int_{\sqrt{\frac{1+\rho}{1-\rho}}}^{+\infty} \frac{4a^2}{a^2 - 1} \left(\frac{da}{a^2 - 1} - \frac{da}{a^2 + 1} \right) = 9\rho^2 \left[-\frac{2a}{a^2 - 1} - 2 \arctan a \right]_{\sqrt{\frac{1+\rho}{1-\rho}}}^{+\infty},$$

which is the required formula.

Let us compare the distribution dv with the uniform distribution $dM = 3\rho^2 d\rho d\omega$ over B_H . We have $dv = g(\rho)dM$, where

$$g(\rho) = 3 \left(\frac{\sqrt{1-\rho^2}}{\rho} + 2 \arctan \sqrt{\frac{1+\rho}{1-\rho}} - \pi \right).$$

Clearly, $g(0) = +\infty$, $g(1) = 0$ and $g'(\rho) = -3\rho^{-2}\sqrt{1-\rho^2} < 0$. Therefore g is monotonous decreasing. Large commutators (with norms $\approx \sqrt{2}$) are rare while small ones are likely. This phenomenon becomes stronger for larger matrix sizes n . On the one hand, the average ratio for $\|[A, B]\|_F^2 / \|A\|_F \|B\|_F$ is $\frac{2}{n} \left(1 - \frac{1}{n^2}\right)$ (see [3] or Paragraph 4.1). On the other hand, when $n \geq 3$, very few trace-less matrices of norms $\sqrt{2}$ can be written as $[A, B]$ with $\|A\|_F = \|B\|_F = 1$; see Section 4 in [4], or [5].

6 The law of distribution of \mathcal{S}_3 over $\mathbf{M}_2(\mathbb{R})$

The situation changes significantly when passing from $r = 2$ (the commutator) to $r = 3$. On the one hand, the addition of tI_n becomes harmful:

$$\mathcal{S}_3(A + tI_n, B, C) = \mathcal{S}_3(A, B, C) + t[B, C].$$

Therefore we may not restrict to zero-trace matrices; incidentally, the trace of $\mathcal{S}_3(A, B, C)$ itself does not vanish in general.

By direct inspection, one may verify the formula

$$(12) \quad \mathcal{S}_3(A, B, C) = \begin{pmatrix} W & X \\ Y & Z \end{pmatrix}$$

with

$$W = \begin{vmatrix} 2a_{11} - a_{22} & 2b_{11} - b_{22} & 2c_{11} - c_{22} \\ a_{12} & b_{12} & c_{12} \\ a_{21} & b_{21} & c_{21} \end{vmatrix}, \quad X = - \begin{vmatrix} a_{11} & b_{11} & c_{11} \\ a_{12} & b_{12} & c_{12} \\ a_{22} & b_{22} & c_{22} \end{vmatrix},$$

$$Y = \begin{vmatrix} a_{22} & b_{22} & c_{22} \\ a_{21} & b_{21} & c_{21} \\ a_{11} & b_{11} & c_{11} \end{vmatrix}, \quad Z = \begin{vmatrix} a_{21} & b_{21} & c_{21} \\ a_{12} & b_{12} & c_{12} \\ 2a_{22} - a_{11} & 2b_{22} - b_{11} & 2c_{22} - c_{11} \end{vmatrix}.$$

We interpret these formulæ in terms of the mixed product in \mathbb{R}^4 . If $a, b, c \in \mathbb{R}^4$, we denote $a \times b \times c$ the vector defined by

$$(a \times b \times c) \cdot x = \det(a, b, c, x), \quad \forall x \in \mathbb{R}^4.$$

Then, identifying

$$A \sim a = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{pmatrix}, \quad B \sim b = \begin{pmatrix} b_{11} \\ b_{12} \\ b_{21} \\ b_{22} \end{pmatrix}, \quad C \sim c = \begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix},$$

we have

$$W = 2f_4 + f_1, \quad X = -f_3, \quad Y = -f_2, \quad Z = 2f_1 + f_4, \quad f := a \times b \times c.$$

Observe that the image of $B_4 \times B_4 \times B_4$ by the mixed product is B_4 itself. Therefore the image of $B_F \times B_F$ is given by the inequality

$$\left(\frac{2Z-W}{3}\right)^2 + \left(\frac{2W-Z}{3}\right)^2 + X^2 + Y^2 \leq 1.$$

This is an ellipsoid centered at the origin, whose main semi-axis have lengths 1, 1, 3 and 3. We deduce the optimal inequality

$$\|\mathcal{S}_3(A, B, C)\|_F^2 \leq 9\|A\|_F^2\|B\|_F^2\|C\|_F^2, \quad \forall A, B, C \in \mathbf{M}_2(\mathbb{R}).$$

The equality holds precisely when a, b, c form an orthogonal basis of the orthogonal complement of the vector

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

in other words when A, B and C are mutually orthogonal and trace-less. The factor 9 above must be compared with the average ratio given by Theorem 4.1,

$$\frac{3!P_3(2)}{2^6} = \frac{9}{8}.$$

We remark that this average ratio is larger than the inner radius of the ellipsoid.

When A, B, C are drawn uniformly and independently from B_F , the law of $\mathcal{S}_3(A, B, C)$ is the push-forward of the law of $a \times b \times c$ (a, b, c chosen uniformly and independently in B_4) under the linear map $f \mapsto (W, X, Y, Z)$ declared previously. If $R \in \mathbf{SO}_4$, we have $(Ra) \times (Rb) \times (Rc) = R(a \times b \times c)$, while the distribution of (Ra, Rb, Rc) is the same as that of (a, b, c) . Therefore the law of $a \times b \times c$ is radial.

Another consequence of (12) is a statement of the annihilating case.

Corollary 6.1 *For $A, B, C \in \mathbf{M}_2(\mathbb{R})$, the following statements are equivalent:*

- We have $\mathcal{S}_3(A, B, C) = 0_2$.
- There exists a $D \in \mathbf{M}_2(\mathbb{R}) \setminus \{0_2\}$ such that

$$\text{Tr}(AD) = \text{Tr}(BD) = \text{Tr}(CD) = 0.$$

Acknowledgement. I am grateful to the referee, who did a remarkable and detailed work. The report helped to improve significantly the presentation of the paper.

References

- [1] K. Audenaert. Variance bounds, with an application to norm bounds for commutators. *Linear Algebra Appl.*, **432** (2010), pp 1126–1143.
- [2] S. A. Amitsur, J. Levitski. Minimal identities for algebras. *Proc. Amer. Math. Soc.*, **1** (1950), pp 449–463.
- [3] A. Böttcher, D. Wenzel. How big can the commutator be and how big is it typically? *Linear Algebra Appl.*, **403** (2005), pp 216–228.
- [4] A. Böttcher, D. Wenzel. The Frobenius norm and the commutator. *Linear Algebra Appl.*, **429** (2008), pp 1864–1885.
- [5] C.-M. Cheng, S.-W. Vong, D. Wenzel. Commutators with maximal Frobenius norm. *Linear Algebra Appl.*, **432** (2010), pp 292–306.
- [6] L. László. Proof of Böttcher and Wenzel’s conjecture on commutator norms for 3-by-3 matrices. *Linear Algebra Appl.*, **422** (2007), pp 659–663.
- [7] Zhiqin Lu. Proof of the normal scalar curvature conjecture. Available from [arxiv:0711.3510v1](https://arxiv.org/abs/0711.3510v1) [math.DG] 22, November 2007.
- [8] S. Rosset. A new proof of the Amitsur-Levitski identity. *Israël J. Math.*, **23** (1976), pp 187–188.
- [9] Seak-Weng Vong, Xiao-Qing Jin. Proof of Böttcher and Wenzel’s conjecture. *Oper. Matrices*, **2** (2008), pp 435–442.
- [10] D. Serre. *Matrices : Theory & applications* (Springer-Verlag GTM **216**), *Additional exercises*, <http://www.umpa.ens-lyon.fr/~serre/DPF/exobis.pdf>.