The Frobenius norm and the commutator

Albrecht Böttcher [a,1] and David Wenzel [a,2]

[a] *Fakultät für Mathematik, TU Chemnitz, 09107 Chemnitz, Germany*

Abstract

In an earlier paper we conjectured an inequality for the Frobenius norm of the commutator of two matrices. This conjecture was recently proved by Seak-Weng Vong and Xiao-Qing Jin. We here give a completely different proof of this inequality, prove some related results, and embark on the corresponding question for unitarily invariant norms.

Keywords: commutator, Frobenius norm, unitarily invariant norm

MSC: Primary: 15A45; Secondary: 15A69

1 Introduction

In [4] we raised the conjecture that the Frobenius norm of the commutator of two real matrices satisfies the inequality

$$\|XY - YX\|_{\mathsf{F}} \le \sqrt{2}\,\|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}}. \tag{1}$$

We there proved this for real 2×2 matrices and also showed that the inequality is true with $\sqrt{2}$ replaced by $\sqrt{3}$. Subsequently László [5] was able to verify (1) for real 3×3 matrices and recently Vong and Jin [6] found a proof of the inequality for real $n\times n$ matrices. Vong and Jin's proof is very clever but based on extensive calculations. We here give a completely new proof of (1). We also extend this inequality to complex matrices, which is not a mere triviality. In Section 3 we provide improvements of (1) and restatements of this inequality, Section 4 is about equality in (1), and Section 5 contains results and questions pertaining to the extension of the inequality to unitarily invariant norms.

[1] Corresponding author. Email: aboettch@mathematik.tu-chemnitz.de
[2] Email: david.wenzel@s2000.tu-chemnitz.de

2 Main result and its proof

We denote by $M_{n,m}(\mathbf{C})$ the linear space of all complex $n \times m$ matrices with the inner product $(Z, W) := \operatorname{tr}(W^* Z)$, where tr denotes the trace and W^* is the Hermitian adjoint of W. The corresponding norm $\|Z\|_{\mathsf{F}} := \sqrt{(Z, Z)}$ is known under the names Frobenius norm, Hilbert-Schmidt norm, or Euclidean norm. Clearly, if $Z = (z_{jk})$ then $\|Z\|_{\mathsf{F}}^2 = \sum_{j,k} |z_{jk}|^2$. We identify $\mathbf{C}^n$ with $M_{n,1}(\mathbf{C})$, which means that we think of vectors in $\mathbf{C}^n$ as columns. Moreover, vectors in $\mathbf{C}^n$ will be denoted by lower-case letters and for $z \in \mathbf{C}^n$, we denote $\|z\|_{\mathsf{F}}$ simply by $\|z\|$. We abbreviate $M_{n,n}(\mathbf{C})$ to $M_n(\mathbf{C})$. Finally, for $Z = (z_{jk}) \in M_{n,m}(\mathbf{C})$ we define $\overline{Z} \in M_{n,m}(\mathbf{C})$ by $\overline{Z} = (\overline{z_{jk}})$. Throughout this paper, $n \geq 2$.

Our main result, Theorem 2.2 below, states that (1) is true for all X, Y in $M_n(\mathbf{C})$. The proof is based on a lemma. For $a, b, u, v \in \mathbf{C}^n$, Cauchy's inequality gives

$$\begin{aligned}
&|(x, a) + (z, b) + (x, u) + (z, v)|^2 \\
&\leq (\|a\|^2 + \|b\|^2 + \|u\|^2 + \|v\|^2)(\|x\|^2 + \|z\|^2 + \|x\|^2 + \|z\|^2). \qquad (2)
\end{aligned}$$

Let $\operatorname{Re} z$ be the real part of a complex number z. Since always

$$2 \frac{\operatorname{Re}[(x, u)(\overline{z}, \overline{v})]}{\|u\| \, \|v\|} \leq \|x\|^2 + \|y\|^2,$$

the following lemma sharpens (2) at the price of a quite exotic hypothesis.

Lemma 2.1 *If $a, b \in \mathbf{C}^n$ and $u, v \in \mathbf{C}^n \setminus \{0\}$ and*

$$\left(a, \frac{u}{\|u\|} \right) + \|u\| = \left(b, \frac{v}{\|v\|} \right) + \|v\|, \qquad (3)$$

then

$$\begin{aligned}
&|(x, a) + (z, b) + (x, u) + (z, v)|^2 \\
&\leq (\|a\|^2 + \|b\|^2 + \|u\|^2 + \|v\|^2) \left(\|x\|^2 + \|z\|^2 + 2 \frac{\operatorname{Re}[(x, u)(\overline{z}, \overline{v})]}{\|u\| \, \|v\|} \right) \qquad (4)
\end{aligned}$$

for every $x, z \in \mathbf{C}^n$.

Proof. Let $\|u\| = \varrho$, $\|v\| = \tau$, $u = \varrho u_0$, $v = \tau v_0$, put $\xi = \begin{pmatrix} x \\ z \end{pmatrix}$, and think of ξ as a column in $\mathbf{C}^{2n}$. We have

$$
\begin{aligned}
&|(x,a)+(z,b)+(x,u)+(z,v)|^2 \\
&= \left(\begin{pmatrix} a+u \\ b+v \end{pmatrix} \begin{pmatrix} a+u \\ b+v \end{pmatrix}^* \begin{pmatrix} x \\ z \end{pmatrix}, \begin{pmatrix} x \\ z \end{pmatrix} \right) =: (M_1\xi, \xi)
\end{aligned}
$$

and

$$
2\,\mathrm{Re}\,[(x,u_0)(\overline{z},\overline{v_0})] = \left(\begin{pmatrix} 0 & u_0 \otimes v_0^* \\ v_0 \otimes u_0^* & 0 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}, \begin{pmatrix} x \\ z \end{pmatrix} \right) =: (M_2\xi, \xi).
$$

We may assume that one of the vectors $a+u$ and $b+v$ is nonzero. The matrix M_1 is Hermitian and of rank 1. The nonzero eigenvalue is

$$
\begin{pmatrix} a+u \\ b+v \end{pmatrix}^* \begin{pmatrix} a+u \\ b+v \end{pmatrix}
$$

and a corresponding eigenvector is

$$
w_0 = \begin{pmatrix} a+u \\ b+v \end{pmatrix} = \begin{pmatrix} a+\varrho u_0 \\ b+\tau v_0 \end{pmatrix}.
$$

It follows that $M_1\zeta = 0$ whenever ζ is orthogonal to w_0. The Hermitian matrix M_2 has rank 2 and its two nonzero eigenvalues are 1 and -1 with the eigenvectors

$$
w_+ = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}, \quad w_- = \begin{pmatrix} u_0 \\ -v_0 \end{pmatrix}.
$$

Again we have $M_2\zeta = 0$ if ζ is orthogonal to both w_+ and w_-. Let $W = \mathrm{span}\,\{w_0, w_+, w_-\}$. Every $\xi \in \mathbf{C}^{2n}$ is of the form $\xi = \eta + \zeta$ with $\eta \in W$ and $\zeta \perp W$. We want to prove that

$$
(M_1\xi, \xi) \le c\|\xi\|^2 + c(M_2\xi, \xi) \tag{5}
$$

where $c = \|a\|^2 + \|b\|^2 + \|u\|^2 + \|v\|^2$. Since $M_1\zeta = M_2\zeta = 0$, we get $(M_1\xi, \xi) = (M_1\eta, \eta)$ and $(M_2\xi, \xi) = (M_2\eta, \eta)$. As $\|\eta\|^2 \le \|\xi\|^2$, inequality (5) will therefore follow once we have shown that

$$
(M_1\eta, \eta) \le c\|\eta\|^2 + c(M_2\eta, \eta).
$$

Thus, we are left to prove (4) for

$$
\xi = \begin{pmatrix} x \\ z \end{pmatrix} \in W = \mathrm{span}\,\{w_0, w_+, w_-\}.
$$

Put

$$w_\perp = w_0 - (w_0, w_+)\frac{w_+}{2} - (w_0, w_-)\frac{w_-}{2}.$$

If $w_\perp \neq 0$, the vectors $w_+/\sqrt{2}$, $w_-/\sqrt{2}$, $w_\perp/\|w_\perp\|$ form an orthonormal basis in W. Otherwise put $\gamma = 0$ in what follows. Take

$$\xi = \begin{pmatrix} x \\ z \end{pmatrix} = \delta\frac{w_+}{\sqrt{2}} + \varepsilon\frac{w_-}{\sqrt{2}} + \gamma\frac{w_\perp}{\|w_\perp\|}.$$

A straightforward computation shows that

$$w_\perp = \begin{pmatrix} a - (a, u_0)u_0 \\ b - (b, v_0)v_0 \end{pmatrix}.$$

Hence

$$\begin{pmatrix} x \\ z \end{pmatrix} = \frac{\delta}{\sqrt{2}}\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \frac{\varepsilon}{\sqrt{2}}\begin{pmatrix} u_0 \\ -v_0 \end{pmatrix} + \frac{\gamma}{\|w_\perp\|}\begin{pmatrix} a - (a, u_0)u_0 \\ b - (b, v_0)v_0 \end{pmatrix}$$

and consequently,

$$\begin{aligned}
(x, a) &= \frac{\delta}{\sqrt{2}}(u_0, a) + \frac{\varepsilon}{\sqrt{2}}(u_0, a) + \frac{\gamma}{\|w_\perp\|}\Big(\|a\|^2 - |(a, u_0)|^2\Big), \\
(z, b) &= \frac{\delta}{\sqrt{2}}(v_0, b) - \frac{\varepsilon}{\sqrt{2}}(v_0, b) + \frac{\gamma}{\|w_\perp\|}\Big(\|b\|^2 - |(b, v_0)|^2\Big), \\
(x, u) &= \frac{\delta}{\sqrt{2}}\varrho + \frac{\varepsilon}{\sqrt{2}}\varrho, \qquad (z, v) = \frac{\delta}{\sqrt{2}}\tau - \frac{\varepsilon}{\sqrt{2}}\tau.
\end{aligned}$$

Adding these equalities, taking into account assumption (3), which is equivalent to $(u_0, a) + \varrho = (v_0, b) + \tau$, and using the obvious equality

$$\|w_\perp\|^2 = \|a\|^2 + \|b\|^2 - |(a, u_0)|^2 - |(b, v_0)|^2$$

we get

$$|(x, a) + (z, b) + (x, u) + (z, v)|^2 = \left|\delta\frac{(u_0, a) + \varrho}{\sqrt{2}} + \delta\frac{(v_0, b) + \tau}{\sqrt{2}} + \gamma\|w_\perp\|\right|^2.$$

By Cauchy's inequality, this is at most

$$\begin{aligned}
&(2|\delta|^2 + |\gamma|^2)\left(\frac{1}{2}|(u_0, a) + \varrho|^2 + \frac{1}{2}|(v_0, a) + \tau|^2 + \|w_\perp\|^2\right) \\
&\leq (2|\delta|^2 + |\gamma|^2)(|(u_0, a)|^2 + \varrho^2 + |(v_0, b)|^2 + \tau^2 + \|w_\perp\|^2) \\
&= (2|\delta|^2 + |\gamma|^2)(\|u\|^2 + \|v\|^2 + \|a\|^2 + \|b\|^2).
\end{aligned}$$

Finally, as

$$
\begin{aligned}
&\|x\|^2 + \|z\|^2 + 2\,\mathrm{Re}\,[(x,u_0)(\overline{z},\overline{v_0})] \\
&= |\delta|^2 + |\varepsilon|^2 + |\gamma|^2 + 2\,\mathrm{Re}\left[\left(\frac{\delta}{\sqrt{2}} + \frac{\varepsilon}{\sqrt{2}}\right)\left(\frac{\overline{\delta}}{\sqrt{2}} - \frac{\overline{\varepsilon}}{\sqrt{2}}\right)\right] \\
&= |\delta|^2 + |\varepsilon|^2 + |\gamma|^2 + |\delta|^2 - |\varepsilon|^2 = 2|\delta|^2 + |\gamma|^2,
\end{aligned}
$$

we arrive at (4). $\square$

Here is our main result.

Theorem 2.2 *If $X, Y \in M_n(\mathbf{C})$, then*

$$\|XY - YX\|_{\mathsf{F}} \le \sqrt{2}\,\|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}}.$$

Proof. Let $X = USV$ be the singular value decomposition with the diagonal matrix $S = \mathrm{diag}\,(s_1, \ldots, s_n)$. Put $C = VYV^*$ and $D = U^*YU$. Then

$$
\begin{aligned}
&\|XY - YX\|_{\mathsf{F}}^2 = \|USVY - YUSV\|_{\mathsf{F}}^2 = \|SVYV^* - U^*YUS\|_{\mathsf{F}}^2 \\
&= \|SC - DS\|_{\mathsf{F}}^2 = \sum_{j,k=1}^{n} |s_j c_{jk} - s_k d_{jk}|^2 \\
&= \sum_{j \ne k} \left(s_j^2|c_{jk}|^2 - 2\,\mathrm{Re}\,(s_j c_{jk} s_k \overline{d_{jk}}) + s_k^2|d_{jk}|^2\right) + \sum_{j=1}^{n} s_j^2 |c_{jj} - d_{jj}|^2 \\
&\le \sum_{j \ne k} \left(s_j^2|c_{jk}|^2 + s_k^2|c_{jk}|^2 + s_j^2|d_{jk}|^2 + s_k^2|d_{jk}|^2\right) + \sum_{j=1}^{n} s_j^2 |c_{jj} - d_{jj}|^2 \\
&= \sum_{j=1}^{n} s_j^2 \Delta_j
\end{aligned}
$$

with

$$\Delta_j = |c_{jj} - d_{jj}|^2 + \sum_{k \ne j} (|c_{jk}|^2 + |c_{kj}|^2 + |d_{jk}|^2 + |d_{kj}|^2).$$

Thus, it remains to prove that $\Delta_j \le 2\|Y\|_{\mathsf{F}}^2$ for all j. Obviously, we may restrict ourselves to the case $j = 1$. Put $A = U^*YV^*$ and $Q = VU$. Then $C = QA$ and $D = AQ$ and we are left with proving that $\Delta_1 \le 2\|A\|_{\mathsf{F}}^2$. We write

$$A = e^{i\varphi}\begin{pmatrix} \sigma & y^* \\ x & B \end{pmatrix}, \quad Q = e^{i\psi}\begin{pmatrix} \omega & \sqrt{1-\omega^2}\,q^* \\ \sqrt{1-\omega^2}\,p & R \end{pmatrix} \tag{6}$$

with numbers $\varphi, \psi \in [0, 2\pi)$, $\sigma \in [0, \infty)$, $\omega \in [0, 1]$, columns $x, y, p, q \in \mathbf{C}^{n-1}$, and matrices $B, R \in M_{n-1}(\mathbf{C})$. Since Q is unitary, we have $\|p\| = \|q\| = 1$. Clearly,

$$\Delta_1 = |(QA)_{11}|^2 - 2\,\mathrm{Re}\,[(QA)_{11}\overline{(AQ)_{11}}] + |(AQ)_{11}|^2 + \sum_{k=2}^{n}\Big(|(QA)_{1k}|^2 + |(QA)_{k1}|^2 + |(AQ)_{1k}|^2 + |(AQ)_{k1}|^2\Big).$$

Let $e_1 = (\,1\,0\,\ldots\,0\,)^\top$. Taking into account that Q is unitary we get

$$\sum_{k=1}^{n} |(QA)_{k1}|^2 = \|QAe_1\|^2 = \|Ae_1\|^2 = \sigma^2 + \|x\|^2.$$

Furthermore,

$$\sum_{k=2}^{n} |(QA)_{1k}|^2 = \|\omega y^* + \sqrt{1-\omega^2}\, q^* B\|^2 = \omega^2\|y\|^2 + 2\omega\sqrt{1-\omega^2}\ \mathrm{Re}\,(\overline{y}, \overline{B^* q}) + (1-\omega^2)\|B^* q\|^2.$$

Analogously,

$$\sum_{k=1}^{n} |(AQ)_{1k}|^2 = \sigma^2 + \|y\|^2,$$
$$\sum_{k=2}^{n} |(AQ)_{k1}|^2 = \omega^2\|x\|^2 + 2\omega\sqrt{1-\omega^2}\ \mathrm{Re}\,(x, Bp) + (1-\omega^2)\|Bp\|^2.$$

Finally, using that $\mathrm{Re}\,\overline{y^* p} = \mathrm{Re}\,(\overline{y}, \overline{p})$ we see that

$$-2\,\mathrm{Re}\,[(QA)_{11}\overline{(AQ)_{11}}] = -2\,\mathrm{Re}\,[(\omega\sigma + \sqrt{1-\omega^2}\, q^* x)(\omega\sigma + \sqrt{1-\omega^2}\,\overline{y^* p})]$$
$$= -2\omega^2\sigma^2 - 2\omega\sigma\sqrt{1-\omega^2}\ \mathrm{Re}\,[(x,q) + (\overline{y},\overline{p})] - 2(1-\omega^2)\,\mathrm{Re}\,[(x,q)(y,p)].$$

Summing up we obtain

$$\Delta_1 = \alpha\omega^2 + \beta\omega\sqrt{1-\omega^2} + \gamma$$

with

$$\alpha = \|x\|^2 + \|y\|^2 - \|Bp\|^2 - \|B^* q\|^2 - 2\sigma^2 + 2\,\mathrm{Re}\,[(x,q)(y,p)],$$
$$\beta = 2\,\mathrm{Re}\,[(x, Bp) + (\overline{y}, \overline{B^* q}) - \sigma(x,q) - \sigma(\overline{y},\overline{p})],$$
$$\gamma = 2\sigma^2 + \|x\|^2 + \|y\|^2 + \|Bp\|^2 + \|B^* q\|^2 - 2\,\mathrm{Re}\,[(x,q)(y,p)].$$

Writing $\omega = \cos\frac{t}{2}$ with $t \in [0,\pi]$ we get

$$\Delta_1 = \alpha\left(\cos\frac{t}{2}\right)^2 + \beta\cos\frac{t}{2}\sin\frac{t}{2} + \gamma = \frac{\alpha}{2} + \frac{1}{2}\,(\alpha\cos t + \beta\sin t) + \gamma \le \frac{\alpha}{2} + \frac{1}{2}\sqrt{\alpha^2 + \beta^2} + \gamma =: \widetilde{\Delta}.$$

We prove that

$$\widetilde{\Delta} \le 2\sigma^2 + 2\|x\|^2 + 2\|y\|^2 + \|Bp\|^2 + \|B^*q\|^2. \tag{7}$$

This will imply the assertion, because (7) gives

$$\Delta_1 \le \widetilde{\Delta} \le 2\sigma^2 + 2\|x\|^2 + 2\|y\|^2 + \|B\|_{\mathsf{F}}^2 + \|B^*\|_{\mathsf{F}}^2 = 2\|A\|_{\mathsf{F}}^2. \tag{8}$$

Inequality (7) is equivalent to the inequality

$$\sqrt{\alpha^2 + \beta^2} \le 2\sigma^2 + \|x\|^2 + \|y\|^2 + \|Bp\|^2 + \|B^*q\|^2 + 2\,\mathrm{Re}\,[(x,q)(y,p)],$$

which with

$$d := \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\,[(x,q)(y,p)], \quad c := 2\sigma^2 + \|Bp\|^2 + \|B^*q\|^2$$

is the inequality $\sqrt{(d-c)^2 + \beta^2} \le d + c$ and hence the inequality $(\beta/2)^2 \le cd$. But the last inequality follows from the inequality

$$\begin{aligned} &|(x, Bp) + (\overline{y}, \overline{B^*q}) - \sigma(x,q) - \sigma(\overline{y}, \overline{p})|^2 \\ &\le (2\sigma^2 + \|Bp\|^2 + \|B^*q\|^2)(\|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\,[(x,q)(y,p)]), \end{aligned}$$

which in turn is Lemma 2.1 with $z = \overline{y}$, $a = Bp$, $b = \overline{B^*q}$, $u = -\sigma q$, $v = -\sigma\overline{p}$.
□

3 Equivalent statements and improvements

Clearly, Theorem 2.2 is equivalent to saying that

$$\|XY - YX\|_{\mathsf{F}} \le \sqrt{2} \quad \text{for} \quad X, Y \in M_n(\mathbf{C}) \quad \text{with} \quad \|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1. \tag{9}$$

The following theorem strengthens this inequality to a chain of inequalities.

Theorem 3.1 *If* $X, Y \in M_n(\mathbf{C})$ *and* $\|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1$ *then*

$$\begin{aligned} \|XY - YX\|_{\mathsf{F}} &\le \|X \otimes Y - Y \otimes X\|_{\mathsf{F}} = \sqrt{2\,(1 - |\mathrm{tr}(Y^*X)|^2)} \\ &\le \sqrt{\|X+Y\|_{\mathsf{F}}\|X-Y\|_{\mathsf{F}}} \le \sqrt{2}. \end{aligned}$$

Proof. First of all,

$$
\begin{aligned}
&\|X \otimes Y - Y \otimes X\|_{\mathsf{F}}^2 = \operatorname{tr}(X \otimes Y - Y \otimes X)^*(X \otimes Y - Y \otimes X) \\
&= \operatorname{tr}(X^*X \otimes Y^*Y - X^*Y \otimes Y^*X - Y^*X \otimes X^*Y + Y^*Y \otimes X^*X) \\
&= \|X\|_{\mathsf{F}}^2\|Y\|_{\mathsf{F}}^2 - (Y,X)(X,Y) - (X,Y)(Y,X) + \|X\|_{\mathsf{F}}^2\|Y\|_{\mathsf{F}}^2 \\
&= 2 - 2\,|(X,Y)|^2 = 2 - 2\,|\operatorname{tr}(Y^*X)|^2. \qquad (10)
\end{aligned}
$$

Both sides of the inequality $\|XY - YX\|_{\mathsf{F}} \le \|X \otimes Y - Y \otimes X\|_{\mathsf{F}}$ are invariant under the change of Y to

$$
Y' := Y - \frac{\overline{(X,Y)}}{\|X\|_{\mathsf{F}}^2} X.
$$

Since $(X, Y') = 0$, it therefore suffices to prove this inequality for matrices X, Y satisfying $(X,Y) = 0$. But in that case we obtain from (9) and (10) that

$$
\|XY - YX\|_{\mathsf{F}}^2 \le 2 = \|X \otimes Y - Y \otimes X\|_{\mathsf{F}}^2.
$$

Furthermore, for arbitrary X, Y of Frobenius norm 1 we have

$$
(2 - 2\,|(X,Y)|^2)^2 = 4 - 4\,|(X,Y)|^2 + r,
$$

with $r = 4\,|(X,Y)|^4 - 4\,|(X,Y)|^2 \le 0$ by Cauchy's inequality, which yields

$$
\begin{aligned}
(2 - 2\,|(X,Y)|^2)^2 &\le 4 - 4\,|(X,Y)|^2 \le 4 - 4\,(\operatorname{Re}(X,Y))^2 \\
&= (2 + 2\,\operatorname{Re}(X,Y))(2 - 2\,\operatorname{Re}(X,Y)). \qquad (11)
\end{aligned}
$$

Since

$$
\begin{aligned}
&(2 + 2\,\operatorname{Re}(X,Y))(2 - 2\,\operatorname{Re}(X,Y)) \\
&= (X+Y, X+Y)(X-Y, X-Y) = \|X+Y\|_{\mathsf{F}}^2\|X-Y\|_{\mathsf{F}}^2, \qquad (12)
\end{aligned}
$$

we see that $\|X+Y\|_{\mathsf{F}}^2\|X-Y\|_{\mathsf{F}}^2 = 4 - 4(\operatorname{Re}(X,Y))^2 \le 4$, while (11) and (12) imply that $2 - 2\,|(X,Y)|^2 \le \|X+Y\|_{\mathsf{F}}\|X-Y\|_{\mathsf{F}}$. $\square$

For a matrix $A \in M_n(\mathbf{C})$, the set $\mathcal{O}_A := \{gAg^{-1} : g \in GL(n, \mathbf{C})\}$ is called the similarity orbit of A. The vector product of two vectors $x = (x_1, x_2, x_3)^\top$ and $y = (y_1, y_2, y_3)^\top$ in $\mathbf{C}^3$ is defined as the vector

$$
x \times y := (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)^\top \in \mathbf{C}^3.
$$

Theorem 3.2 *The following statements are equivalent:*

(i) $\|XY - YX\|_{\mathsf{F}} \le \sqrt{2}$ *for all* $X, Y \in M_n(\mathbf{C})$ *with* $\|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1$*;*

(ii) $\|I \otimes X - X^\top \otimes I\|_\infty \le \sqrt{2}\,\|X\|_{\mathsf{F}}$ *for all* $X \in M_n(\mathbf{C})$*, where* $\|\cdot\|_\infty$ *denotes the spectral norm and* $X^\top$ *is the transpose of* X*;*

(iii) $\|XY - YX\|_{\mathsf{F}} \le \|X \otimes Y - Y \otimes X\|_{\mathsf{F}}$ *for all* $X, Y \in M_n(\mathbf{C})$*;*

(iv) $\|XY - YX\|_{\mathsf{F}} \le \sqrt{\|X+Y\|_{\mathsf{F}}\|X-Y\|_{\mathsf{F}}}$ *for all* $X, Y \in M_n(\mathbf{C})$ *with* $\|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1$*;*

(v) *if* $g : (-\varepsilon, \varepsilon) \to GL(n, \mathbf{C})$ *is any differentiable curve with* $g(0) = I$*, if* A *is any matrix in* $M_n(\mathbf{C})$ *with* $\|A\|_{\mathsf{F}} = 1$*, and if the curve* $h : (-\varepsilon, \varepsilon) \to \mathcal{O}_A$ *is defined by* $h(t) = g(t)Ag(t)^{-1}$*, then* $\|h'(0)\|_{\mathsf{F}} \le \sqrt{2}\,\|g'(0)\|_{\mathsf{F}}$*;*

(vi) *if* $v^{(jk)} \in \mathbf{C}^3$ *is any collection of* n^2 *vectors, then their vector products satisfy the inequality*

$$\sum_{i,j=1}^{n} \left\| \sum_{k=1}^{n} v^{(ik)} \times v^{(kj)} \right\|^2 \le \sum_{i,k,\ell,j=1}^{n} \left\| v^{(ik)} \times v^{(\ell j)} \right\|^2;$$

(vii) *if* f *and* g *are arbitrary complex-valued functions in* $L^2((-\pi,\pi)^2)$ *then*

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\int_{-\pi}^{\pi}\Big(f(x,t)g(-t,y) - g(x,t)f(-t,y)\Big)dt\right|^2 dx\,dy$$
$$\le 2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|f(x,y)|^2 dx\,dy \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|g(x,y)|^2 dx\,dy.$$

Proof. (i) $\Rightarrow$ (iii) and (i) $\Rightarrow$ (iv). When proving Theorem 3.1 we showed how (iii) and (iv) can be derived from (i).

(iii) $\Rightarrow$ (i). From (10) we get $\|X \otimes Y - Y \otimes X\|_{\mathsf{F}} \le \sqrt{2}$, which together with (iii) implies (i).

(iv) $\Rightarrow$ (i). Equality (12) shows that $\|X+Y\|_{\mathsf{F}}\|X-Y\|_{\mathsf{F}} \le 2$, which in conjunction with (iv) gives (i).

(i) $\Leftrightarrow$ (ii). Stacking matrices in $M_n(\mathbf{C})$ column by column, the linear map $Y \mapsto XY - YX$ of $M_n(\mathbf{C})$ into itself becomes multiplication by the matrix $I \otimes X - X^\top \otimes I$ in $\mathbf{C}^{n^2}$. Consequently, (i) is equivalent to the inequality

$$\|(I \otimes X - X^\top \otimes I)y\| \le \sqrt{2}\,\|X\|_{\mathsf{F}}\|y\|, \quad y \in \mathbf{C}^{n^2},$$

which is just (ii).

(i) $\Leftrightarrow$ (v). We may without loss of generality assume that $g(t) = e^{tX}$ with some $X \in M_n(\mathbf{C})$ (see [1, p. 189]). It follows that $g'(0) = X$ and $h'(0) = XA - AX$. The equivalence of (i) and (v) is therefore immediate.

(iii) $\Leftrightarrow$ (vi). Let $v^{(ik)} = (x_{ik}, y_{ik}, z_{ik})^\top \in \mathbf{C}^3$ and consider the $n \times n$ matrices $X = (x_{jk})$, $Y = (y_{jk})$, $Z = (z_{jk})$. The definition of the vector product then turns the left-hand side of the inequality in (vi) into

$$\|YZ - ZY\|_{\mathsf{F}}^2 + \|ZX - XZ\|_{\mathsf{F}}^2 + \|XY - YX\|_{\mathsf{F}}^2.$$

In the same way the right-hand side becomes

$$\|Y \otimes Z - Z \otimes Y\|_{\mathsf{F}}^2 + \|Z \otimes X - X \otimes Z\|_{\mathsf{F}}^2 + \|X \otimes Y - Y \otimes X\|_{\mathsf{F}}^2.$$

Hence, we obtain (vi) by applying (iii) three times. Conversely, (vi) with $Z = 0$ is exactly (iii).

(i) $\Leftrightarrow$ (vii). Let

$$f(x,y) = \sum_{m,n\in\mathbf{Z}} f_{mn} e^{imx} e^{iny}, \quad g(x,y) = \sum_{m,n\in\mathbf{Z}} g_{mn} e^{imx} e^{iny}$$

be the Fourier series of f and g. By Parseval's equality, the right-hand side of the inequality in (vii) is $(2\pi)^4$ times

$$2\left(\sum_{m,n\in\mathbf{Z}} |f_{mn}|^2\right)\left(\sum_{m,n\in\mathbf{Z}} |g_{mn}|^2\right).$$

On the other hand,

$$\begin{aligned}
\int_{-\pi}^{\pi} f(x,t)g(-t,y)\frac{dt}{2\pi} &= \sum_{m,n,j,k} f_{mn}g_{jk}e^{imx}e^{iky}\int_{-\pi}^{\pi} e^{i(n-j)t}\frac{dt}{2\pi}\\
&= \sum_{m,k}\Big(\sum_{j} f_{mj}g_{jk}\Big)e^{imx}e^{iky},\\
\int_{-\pi}^{\pi} g(x,t)f(-t,y)\frac{dt}{2\pi} &= \sum_{m,n,j,k} g_{mn}f_{jk}e^{imx}e^{iky}\int_{-\pi}^{\pi} e^{i(n-j)t}\frac{dt}{2\pi}\\
&= \sum_{m,k}\Big(\sum_{j} g_{mj}f_{jk}\Big)e^{imx}e^{iky}.
\end{aligned}$$

Thus, again by Parseval's equality, the left-hand side of the inequality in (vii) equals $(2\pi)^4$ times

$$\sum_{m,k\in\mathbf{Z}}\left|\sum_{j\in\mathbf{Z}}(f_{mj}g_{jk} - g_{mj}f_{jk})\right|^2.$$

The inequality in (vii) is therefore just inequality (1) for the infinite matrices $X = (f_{jk})_{j,k\in\mathbf{Z}}$ and $Y = (g_{jk})_{j,k\in\mathbf{Z}}$. We have $\sum_{j,k}|f_{jk}|^2 < \infty$ and $\sum_{j,k}|g_{jk}|^2 < \infty$. Moreover, given any infinite matrices $X = (f_{jk})_{j,k\in\mathbf{Z}}$ and $Y = (g_{jk})_{j,k\in\mathbf{Z}}$ such that $\sum_{j,k}|f_{jk}|^2 < \infty$ and $\sum_{j,k}|g_{jk}|^2 < \infty$, there are functions f and g in $L^2((-\pi,\pi)^2)$ such that $\{f_{jk}\}$ and $\{g_{jk}\}$ are the Fourier coefficients of f and g. Thus, assertion (vii) is equivalent to (1) for infinite matrices. But if (1) holds for all $n \times n$ matrices, passage to the limit $n \to \infty$ gives (1) for infinite matrices. Conversely, if (1) is true for all infinite matrices, it is all the more valid for arbitrary $n \times n$ matrices. $\square$

Remark 3.3 In connection with Theorem 3.2(iii) we first remark that the inequality $\|XY+YX\|_{\mathsf{F}} \le \|X\otimes Y + Y\otimes X\|_{\mathsf{F}}$ is in general not true. Indeed, taking $X = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $Y = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ we get

$$\|XY+YX\|_{\mathsf{F}} = \sqrt{12} \quad \text{and} \quad \|X\otimes Y + Y\otimes X\|_{\mathsf{F}} = \sqrt{10}.$$

Secondly, for arbitrary $X, Y \in M_n(\mathbf{C})$ we obviously have

$$\|XY\|_{\mathsf{F}} \le \|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}} = \|X\otimes Y\|_{\mathsf{F}}. \tag{13}$$

This inequality expresses some kind of monotonicity between the usual matrix product and the tensor product. The inequality in Theorem 3.2(iii) can be interpreted in a similar fashion: with $[X,Y] := XY - YX$ being the usual Lie bracket and defining $\{X,Y\} := X\otimes Y - Y\otimes X$ as a tensor product based Lie bracket analog, we have the monotonicity

$$\|[X,Y]\|_{\mathsf{F}} \le \|\{X,Y\}\|_{\mathsf{F}}$$

for Lie brackets. $\square$

Remark 3.4 Suppose X and Y are real matrices and $\|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1$. Then $X+Y$ and $X-Y$ are orthogonal and hence $\|X+Y\|_{\mathsf{F}}\|X-Y\|_{\mathsf{F}}$ is twice the area of the rhomb spanned by X and Y. Thus,

$$\|XY-YX\|_{\mathsf{F}}^2 \le 2\,\text{area}\,\text{rhomb}(X,Y). \tag{14}$$

From Theorem 3.1 and (14) we deduce that without any constraint on the norms of X and Y we have

$$\begin{aligned}\|XY-YX\|_{\mathsf{F}}^2 &\le 2\,\|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}}(\|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}} - |\operatorname{tr}(Y^*X)|^2) \\ &\le \Big\| \, \|Y\|_{\mathsf{F}}X + \|X\|_{\mathsf{F}}Y \Big\| \cdot \Big\| \, \|Y\|_{\mathsf{F}}X - \|X\|_{\mathsf{F}}Y \Big\| \\ &= 2\,\text{area}\,\text{rhomb}(\|Y\|_{\mathsf{F}}X, \|X\|_{\mathsf{F}}Y),\end{aligned}$$

the last equality for real matrices only. $\square$

Given $Z = A + iB$ with $A, B \in M_n(\mathbf{R})$, the real and imaginary parts are defined by $\operatorname{Re} Z = A$ and $\operatorname{Im} Z = B$. From (13) we infer that

$$\|Z\overline{Z}\|_{\mathsf{F}} \le \|Z\otimes\overline{Z}\|_{\mathsf{F}},$$

that is,

$$\|\operatorname{Re}(Z\overline{Z})\|_{\mathsf{F}}^2 + \|\operatorname{Im}(Z\overline{Z})\|_{\mathsf{F}}^2 \le \|\operatorname{Re}(Z\otimes\overline{Z})\|_{\mathsf{F}}^2 + \|\operatorname{Im}(Z\otimes\overline{Z})\|_{\mathsf{F}}^2.$$

In connection with this inequality, the following is quite curious.

Corollary 3.5 *For all $Z \in M_n(\mathbf{C})$,*

$$\| \operatorname{Im}(Z\overline{Z})\|_{\mathsf{F}} \le \| \operatorname{Im}(Z \otimes \overline{Z})\|_{\mathsf{F}}, \tag{15}$$

but there are $Z \in M_n(\mathbf{C})$ such that

$$\| \operatorname{Re}(Z\overline{Z})\|_{\mathsf{F}} > \| \operatorname{Re}(Z \otimes \overline{Z})\|_{\mathsf{F}}. \tag{16}$$

Proof. Since $\| \operatorname{Im}(Z\overline{Z})\|_{\mathsf{F}} = \|BA - AB\|_{\mathsf{F}}$ and $\| \operatorname{Im}(Z \otimes \overline{Z})\|_{\mathsf{F}} = \|B \otimes A - A \otimes B\|_{\mathsf{F}}$, inequality (15) is straightforward from Theorem 3.1 (or Theorem 2.2 in conjunction with Theorem 3.2(iii)). Letting

$$Z = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + i \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

we get $\| \operatorname{Re}(Z\overline{Z})\|_{\mathsf{F}} = \sqrt{21}$ and $\| \operatorname{Re}(Z \otimes \overline{Z})\|_{\mathsf{F}} = \sqrt{19}$, which gives (16). □

Remark 3.6 Theorems 2.2, 3.1, the equivalence of the first four statements in Theorem 3.2, and Corollary 3.5 remain true for Hilbert-Schmidt operators on arbitrary infinite-dimensional separable Hilbert spaces, because in every orthonormal basis every such operator is given by an infinite matrix $Z = (z_{jk})$ with $\sum_{j,k} |z_{jk}|^2 < \infty$ and the principal finite sections $Z_n := (z_{jk})_{|j| \le n, |k| \le n}$ converge to Z in the Hilbert-Schmidt norm. An observation of this kind was already employed in the proof of the equivalence (i) ⇔ (vii) of Theorem 3.2. □

4 Matrix pairs with maximal commutator

This section is devoted to the cases of equality in the inequality of Theorem 2.2. We call a pair (X, Y) of matrices in $M_n(\mathbf{C})$ maximal if $X \neq 0$, $Y \neq 0$, and $\|XY - YX\|_{\mathsf{F}} = \sqrt{2}\, \|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}$. In [4] we observed that if X and Y are chosen at random, then the ratio of $\|XY - YX\|_{\mathsf{F}}$ and $\|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}$ concentrates tightly around a number that goes to zero as $n \to \infty$. The following result may serve as another explanation for the phenomenon that maximal pairs are very rare and thus difficult to find on the off-chance.

Theorem 4.1 *If (X, Y) is a maximal pair of matrices in $M_n(\mathbf{C})$, then*

(a) $\operatorname{rank} X \le 2$, $\operatorname{rank} Y \le 2$, (b) $X \perp \operatorname{Com} Y$, $Y \perp \operatorname{Com} X$,

where $\operatorname{Com} W$, *the commutant of* W, *is the algebra* $\{Z \in M_n(\mathbf{C}) : ZW = WZ\}$.

Proof. We use the notation of the proofs of Lemma 2.1 and Theorem 2.2. Suppose (X, Y) is a maximal pair. Then equality must hold in (8), which implies that $\|Bp\| = \|B\|_{\mathsf{F}}$ and $\|B^*q\| = \|B^*\|_{\mathsf{F}}$. It follows that $\|B\|_\infty = \|B\|_{\mathsf{F}}$ and hence that B has at most one nonzero singular value. Thus, $B = 0$ or $\operatorname{rank} B = 1$. In the first case, the matrix A in (6) has rank at most 2, yielding $\operatorname{rank} Y \le 2$, as desired. So assume $\operatorname{rank} B = 1$. (This already gives $\operatorname{rank} Y = \operatorname{rank} A \le 3$.) Writing $B = \tau r s^*$ with $\|r\| = \|s\| = 1$, we get $\|Bp\|^2 = |\tau|^2 |s^*p|^2 \|r\|^2 = \|B\|_\infty^2$. Consequently, $|s^*p| = 1$ and hence $s = \lambda p$ with $|\lambda| = 1$. Analogously, $r = \mu q$ with $|\mu| = 1$. We therefore obtain that $B = \tau\overline{\lambda}\mu q p^* =: \varkappa q p^*$.

We must further have equality in Lemma 2.1 with $z = \overline{y}$, $a = Bp$, $b = \overline{B^*q}$, $u = -\sigma q$, $v = -\sigma\overline{p}$. For this it is necessary that (5) is an equality, which is only possible if $\xi = \eta \in W$. In the case at hand,

$$a = \varkappa q p^* p = \varkappa q, \quad b = \overline{\overline{\varkappa} p q^* q} = \varkappa \overline{p}.$$

This shows that $w_\perp = 0$. Thus, ξ is a linear combination of w_+ and w_-,

$$\xi = \begin{pmatrix} x \\ \overline{y} \end{pmatrix} = \varepsilon_0 \begin{pmatrix} -q \\ -\overline{p} \end{pmatrix} + \delta_0 \begin{pmatrix} -q \\ \overline{p} \end{pmatrix},$$

which gives $x = \varepsilon q$ and $y = \delta p$ with complex numbers ε and δ. The matrix A in (6) therefore becomes

$$A = e^{i\varphi} \begin{pmatrix} \sigma & \overline{\varepsilon} p^* \\ \delta q & \varkappa q p^* \end{pmatrix}.$$

As this is a matrix of rank at most 2, we arrive at the conclusion that $\operatorname{rank} Y \le 2$. Interchanging Y with X we obtain that $\operatorname{rank} X \le 2$.

Now let $Z \in \operatorname{Com} Y \setminus \{0\}$. Since $\sqrt{2}\, \|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}} = \|XY - YX\|_{\mathsf{F}}$ by assumption and

$$\|XY - YX\|_{\mathsf{F}} = \|(X + \lambda Z)Y - Y(X + \lambda Z)\|_{\mathsf{F}} \le \sqrt{2}\, \|X + \lambda Z\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}$$

for every $\lambda \in \mathbf{C}$ by Theorem 2.2, we conclude that

$$\|X\|_{\mathsf{F}}^2 \le \|X + \lambda Z\|_{\mathsf{F}}^2 = \|X\|_{\mathsf{F}}^2 + 2 \operatorname{Re}[\overline{\lambda}(X, Z)] + |\lambda|^2 \|Z\|_{\mathsf{F}}^2.$$

For $\lambda = -(X, Z)/\|Z\|_{\mathsf{F}}^2$ the right-hand side becomes

$$\|X\|_{\mathsf{F}}^2 - 2\, \frac{|(X, Z)|^2}{\|Z\|_{\mathsf{F}}^2} + \frac{|(X, Z)|^2}{\|Z\|_{\mathsf{F}}^4} \|Z\|_{\mathsf{F}}^2 = \|X\|_{\mathsf{F}}^2 - \frac{|(X, Z)|^2}{\|Z\|_{\mathsf{F}}^2},$$

which implies that $(X, Z) = 0$. Thus, $X \perp \operatorname{Com} Y$. Analogously one gets that $Y \perp \operatorname{Com} X$. $\square$.

Corollary 4.2 *If (X, Y) is a maximal pair of matrices in $M_n(\mathbf{C})$, then necessarily*

$$\operatorname{rank} X \leq 2, \quad \operatorname{rank} Y \leq 2, \quad \operatorname{tr} X = \operatorname{tr} Y = 0, \quad (X, Y^m) = (X^m, Y) = 0$$

for all natural numbers m.

Proof. This is immediate from Theorem 4.1 along with the observations that polynomials of Z are in $\operatorname{Com} Z$ and that $\operatorname{tr} Z = (Z, I)$. $\square$

Remark 4.3 The matrices of a maximal pair need not to have the same rank: the pair

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is maximal, but $\operatorname{rank} X = 1$ and $\operatorname{rank} Y = 2$. Furthermore, conditions (a) and (b) of Theorem 4.1 are not sufficient for (X, Y) to be a maximal pair. Indeed, let

$$X = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

If $Z = (z_{jk})_{j,k=1}^3 \in \operatorname{Com} X$ then $z_{21} = z_{31} = 0$, and if $Z = (z_{jk})_{j,k=1}^3 \in \operatorname{Com} Y$ then $z_{12} = z_{13} = 0$. Thus, $X \perp \operatorname{Com} Y$ and $Y \perp \operatorname{Com} X$. It follows that conditions (a) and (b) of Theorem 4.1 are satisfied. However, $\|XY - YX\|_{\mathsf{F}}^2 = 4$ and $2\,\|X\|_{\mathsf{F}}^2 \|Y\|_{\mathsf{F}}^2 = 8$, that is, (X, Y) is not a maximal pair. $\square$

The following results characterize maximal pairs subject to additional constraints.

Proposition 4.4 *Let X and Y be nonzero matrices in $M_2(\mathbf{C})$. Then (X, Y) is a maximal pair if and only if $\operatorname{tr} X = \operatorname{tr} Y = 0$ and $(X, Y) = 0$.*

Proof. Corollary 4.2 gives the "only if" part. To prove the reverse, take

$$X = \begin{pmatrix} c & a \\ b & -c \end{pmatrix}, \quad Y = \begin{pmatrix} z & x \\ y & -z \end{pmatrix}.$$

Since both sides of the equality $\|XY - YX\|_{\mathsf{F}}^2 = 2\,\|X\|_{\mathsf{F}}^2 \|Y\|_{\mathsf{F}}^2$ depend continuously on c and z, we may assume that $c \neq 0$ and $z \neq 0$ and hence that even $c = z = 1$. Under this assumption,

$$\|XY - YX\|_{\mathsf{F}}^2 = 2\,|bx - ay|^2 + 4\,|a - x|^2 + 4\,|b - y|^2, \tag{17}$$

$$2\,\|X\|_{\mathsf{F}}^2 \|Y\|_{\mathsf{F}}^2 = 2\,(2 + |a|^2 + |b|^2)(2 + |x|^2 + |y|^2). \tag{18}$$

The difference of (18) and (17) is $|2 + a\overline{x} + b\overline{y}|^2 = |(X,Y)|^2 = 0$, which completes the proof. $\square$

Proposition 4.5 *Let $X, Y \in M_n(\mathbf{C})$ and suppose $\|X\|_{\mathsf{F}} = \|Y\|_{\mathsf{F}} = 1$ and* $\operatorname{rank} X = \operatorname{rank} Y = 1$*. Then (X,Y) is a maximal pair if and only if* $\operatorname{tr} X = 0$ *and $Y = \varkappa X^*$ with some complex number of modulus 1.*

Proof. We have $X = ab^*$ and $Y = xy^*$ with $\|a\| = \|b\| = \|x\| = \|y\| = 1$. Hence

$$\begin{aligned}\|XY - YX\|_{\mathsf{F}}^2 &= \|(x,b)ay^* - (a,y)xb^*\|_{\mathsf{F}}^2\\ &= \operatorname{tr}\left[((b,x)ya^* - (y,a)bx^*)((x,b)ay^* - (a,y)xb^*)\right]\\ &= |(b,x)|^2 + |(a,y)|^2 - 2\operatorname{Re}\left[(b,x)(a,y)(x,a)(y,b)\right] \qquad (19)\end{aligned}$$

and $2\,\|X\|_{\mathsf{F}}^2\|Y\|_{\mathsf{F}}^2 = 2$. Suppose first that (X,Y) is a maximal pair. By Corollary 4.2, $\operatorname{tr} X = 0$ and $0 = (X,Y) = (y,b)(a,x)$. Thus, the real part in (19) vanishes and we get $|(b,x)|^2 + |(a,y)|^2 = 2$, which in turn implies that $x = \lambda b$ and $y = \mu a$ with $|\lambda| = |\mu| = 1$. It follows that $Y = \lambda\overline{\mu}ba^* = \lambda\overline{\mu}X^*$, as desired. Conversely, let $\operatorname{tr} X = 0$ and $Y = \varkappa X^*$ with $|\varkappa| = 1$. Then $Y = (\varkappa b)a^*$, and inserting $x = \varkappa b$, $y = a$ in (19) we obtain that

$$\|XY - YX\|_{\mathsf{F}}^2 = 2 - 2\operatorname{Re}\left[(b,a)(a,b)\right].$$

As $0 = \operatorname{tr} X = (a,b)$, we see that $\|XY - YX\|_{\mathsf{F}}^2 = 2$. $\square$

Proposition 4.6 *Suppose $X \in M_n(\mathbf{C})$ is normal. Then (X,Y) is a maximal pair if any only if there exist a unitary matrix $U \in M_n(\mathbf{C})$ and complex numbers λ, a, b such that $\lambda \ne 0$, $|a|^2 + |b|^2 > 0$, and*

$$X = U\begin{pmatrix} X_0 & 0 \\ 0 & 0 \end{pmatrix}U^* \quad \textit{with} \quad X_0 = \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix}, \qquad (20)$$

$$Y = U\begin{pmatrix} Y_0 & 0 \\ 0 & 0 \end{pmatrix}U^* \quad \textit{with} \quad Y_0 = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}. \qquad (21)$$

Proof. Suppose (X,Y) is a maximal pair. Since X is normal, we have $X = U\Lambda U^*$ with $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. By Corollary 4.2, at most two of the λ_j are nonzero and the sum of these two is zero. Thus, we may a priori assume X is of the form (20). The case $\lambda = 0$ gives the zero matrix. Hence $\lambda \ne 0$. Put $Z = U^*YU$. Then

$$\|XY - YX\|_{\mathsf{F}}^2 = \|\Lambda Z - Z\Lambda\|_{\mathsf{F}}^2 = \sum_{j\neq k} |\lambda_j - \lambda_k|^2 |z_{jk}|^2$$
$$= 4\,|\lambda|^2|z_{12}|^2 + 4\,|\lambda|^2|z_{21}|^2 + \sum_{k\geq 2} |\lambda|^2|z_{1k}|^2 + \sum_{j\geq 2} |\lambda|^2|z_{j1}|^2 \tag{22}$$

and

$$2\,\|X\|_{\mathsf{F}}^2\|Y\|_{\mathsf{F}}^2 = 4\,|\lambda|^2 \sum_{j,k=1}^{n} |z_{jk}|^2. \tag{23}$$

But if (22) and (23) are equal, then $z_{jk} = 0$ for $(j,k) \neq (1,2)$ and $(j,k) \neq (2,1)$. This implies that Y is of the form (21). As $Y \neq 0$, one of the numbers a and b is nonzero.

Conversely, let X and Y be as in (20) and (21). From (22) and (23) we infer that

$$\|XY - YX\|_{\mathsf{F}}^2 = 4\,|\lambda|^2(|a|^2 + |b|^2) = 2\,\|X\|_{\mathsf{F}}^2\|Y\|_{\mathsf{F}}^2,$$

which shows that (X, Y) is a maximal pair. □

Remark 4.7 From Proposition 4.6 we immediately obtain that a pair (X, Y) of normal (resp. Hermitian) matrices in $M_n(\mathbf{C})$ is maximal if and only if there exist a unitary matrix U and complex numbers λ, a, b such that (20) and (21) hold with $\lambda \neq 0$, $|a| = |b| \neq 0$ (resp. $\lambda = \overline{\lambda} \neq 0$, $a = \overline{b} \neq 0$). Theorem 4.1 implies that for $n \geq 3$ there are no maximal pairs in which at least one matrix is invertible. In particular, there are no maximal pairs with at least one unitary matrix. By Proposition 4.6, two matrices $X, Y \in U(2)$ form a maximal pair if and only if there is a $U \in U(2)$ such that

$$X = U \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix} U^*, \quad Y = U \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix} U^* \tag{24}$$

with $|\lambda| = |a| = |b| = 1$. These two matrices are in $SU(2)$ if and only if $\lambda \in \{i, -i\}$, $|a| = 1$, $b = -1/a$. It is easy to show by direct inspection that two matrices $X, Y \in O(2)$ are a maximal pair if and only if they are of the form (24) with $U \in O(2)$, $\lambda \in \{1, -1\}$, $a \in \{1, -1\}$, $b \in \{1, -1\}$. There do not exist maximal pairs in $SO(2)$. There are also no maximal pairs containing at least one positive semi-definite matrix. This follows from inequality (2) of paper [2] by Bhatia and Kittaneh, which implies that if $X \in M_n(\mathbf{C})$ is positive semi-definite, $X \geq 0$, then

$$\|XY - YX\|_{\mathsf{F}} \leq \|X\|_\infty \|Y\|_{\mathsf{F}} \leq \|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}}$$

for every $Y \in M_n(\mathbf{C})$ (see also Remark 5.1 of [4]). Moreover inequality (3) of [2] implies that if $X \geq 0$ and $Y \geq 0$, then

$$\|XY - YX\|_{\mathsf{F}} \leq \frac{1}{2}\,\|X\|_\infty\,\|Y \oplus Y\|_{\mathsf{F}} = \frac{1}{\sqrt{2}}\,\|X\|_\infty\,\|Y\|_{\mathsf{F}} \leq \frac{1}{\sqrt{2}}\,\|X\|_{\mathsf{F}}\,\|Y\|_{\mathsf{F}}.$$

Taking

$$X = \begin{pmatrix} X_0 & 0 \\ 0 & 0 \end{pmatrix}, \quad X_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_0 & 0 \\ 0 & 0 \end{pmatrix}, \quad Y_0 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

we get $\|XY - YX\|_{\mathsf{F}} = (1/\sqrt{2})\, \|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}$. Thus,

$$\sup \left\{ \frac{\|XY - YX\|_{\mathsf{F}}}{\|X\|_{\mathsf{F}}\, \|Y\|_{\mathsf{F}}} : X, Y \in M_n(\mathbf{C}) \setminus \{0\}, X \geq 0, Y \geq 0 \right\} = \frac{1}{\sqrt{2}}.$$

Bloch and Iserles [3] studied the problem of determining

$$\sup \left\{ \frac{\|XY - YX\|_{\mathsf{F}}}{\|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}} : X, Y \in \mathfrak{g} \setminus \{0\} \right\} \tag{25}$$

where $\mathfrak{g}$ is a Lie algebra and proved that if $\mathfrak{g}$ is the Lie algebra $\mathfrak{so}(n)$ of skew-symmetric matrices in $M_n(\mathbf{R})$, then (25) is 0 for $n = 2$, $1/\sqrt{2}$ for $n = 3$, and 1 for $n \geq 4$. $\square$

Remark 4.8 Abbreviate $XY - YX$ to $[X, Y]$. Repeated application of Theorem 2.2 gives

$$\| \, [Z_1, [Z_2, \ldots [Z_{m-1}, Z_m]]] \, \|_{\mathsf{F}} \leq 2^{(m-1)/2} \|Z_1\|_{\mathsf{F}} \|Z_2\|_{\mathsf{F}} \ldots \|Z_m\|_{\mathsf{F}} \tag{26}$$

for arbitrary Z_j in $M_n(\mathbf{C})$. From Remark 4.7 we see that if (X, Y) is a maximal pair consisting of two normal matrices, then $(X, [X, Y])$ is also a maximal pair of two normal matrices. This implies that the constant $2^{(m-1)/2}$ in (26) is best possible. $\square$

5 Unitarily invariant norms

Let $\| \cdot \|$ be a unitarily invariant norm on $M_n(\mathbf{C})$ and put

$$\Phi(x_1, \ldots, x_n) = \| \operatorname{diag}(x_1, \ldots, x_n) \|$$

if $x_1, \ldots, x_n$ are real numbers. Throughout what follows we assume without loss of generality that $\Phi(1, 0, \ldots, 0) = 1$. The function Φ is a norm on $\mathbf{R}^n$ and it is invariant under the transformations $(x_1, \ldots, x_n) \mapsto (\pm x_1, \ldots, \pm x_n)$ and under permutations of $(x_1, \ldots, x_n)$. Conversely, given any function Φ with these properties, we obtain a unitarily invariant norm on $M_n(\mathbf{C})$ by defining $\|X\| = \|USV\| := \Phi(s_1, \ldots, s_n)$, where $X = USV$ with $S = \operatorname{diag}(s_1, \ldots, s_n)$ is the singular value decomposition. We refer to [1] for more on unitarily invariant norms. In what follows we order the singular values of a matrix X

in decreasing order, $s_1 \geq \ldots \geq s_n$, and we denote the vector $(s_1, \ldots, s_n)$ by $\Sigma(X)$.

Proposition 5.1 *Let $\|\cdot\|$ be a unitarily invariant norm on $M_n(\mathbf{C})$, and set $\mu = \Phi(1, 1, 0, \ldots, 0)$. Then*

$$\sup\left\{\frac{\|XY - YX\|}{\|X\|\,\|Y\|} : X, Y \in M_n(\mathbf{C}) \setminus \{0\}\right\} \geq \max\left(\mu, \frac{2}{\mu}\right) \geq \sqrt{2}.$$

Proof. It suffices to consider the case $n = 2$. For

$$X = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad XY - YX = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$$

we have $\Sigma(X) = \Sigma(Y) = (1, 1)$ and $\Sigma(XY - YX) = (2, 2)$, which gives

$$\frac{\|XY - YX\|}{\|X\|\,\|Y\|} = \frac{\Phi(2,2)}{\Phi(1,1)^2} = \frac{2\,\Phi(1,1)}{\Phi(1,1)^2} = \frac{2}{\Phi(1,1)} = \frac{2}{\mu},$$

while the singular values of

$$X = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \quad XY - YX = \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}$$

are $\Sigma(X) = \Sigma(Y) = (2, 0)$ and $\Sigma(XY - YX) = (4, 4)$, from which we obtain that

$$\frac{\|XY - YX\|}{\|X\|\,\|Y\|} = \frac{\Phi(4,4)}{\Phi(2,0)^2} = \frac{4\,\Phi(1,1)}{2^2\,\Phi(1,0)^2} = \Phi(1,1) = \mu.$$

Obviously, both μ and $2/\mu$ cannot be strictly less than $\sqrt{2}$. □

Theorem 2.2 and Proposition 5.1 imply that

$$\min_{\Phi} \sup\left\{\frac{\|XY - YX\|}{\|X\|\,\|Y\|} : X, Y \in M_n(\mathbf{C}) \setminus \{0\}\right\} = \sqrt{2},$$

the minimum over all unitarily invariant norms on $M_n(\mathbf{C})$, and that the minimum is attained for the Frobenius norm. In Example 5.7 we will show that the supremum in Proposition 5.1 may be strictly larger than $\max(\mu, 2/\mu)$.

Example 5.2 (Schatten norms) The pth Schatten norm $\|\cdot\|_p$ $(1 \leq p \leq \infty)$ is given by

$$\Phi_p(x_1, \ldots, x_n) := (|x_1|^p + \ldots + |x_n|^p)^{1/p}.$$

Thus, $\|\cdot\|_2 = \|\cdot\|_{\mathsf{F}}$ and $\|\cdot\|_\infty$ is the spectral norm. Since $\Phi_p(1, 1, 0, \ldots, 0) = 2^{1/p}$, we deduce from Proposition 5.1 that

$$\sup\left\{\frac{\|XY-YX\|_p}{\|X\|_p\,\|Y\|_p} : X,Y\in M_n(\mathbf{C})\setminus\{0\}\right\}$$
$$\geq \max(2^{1/p},2^{1/q}) = 2^{1/\min(p,q)}, \tag{27}$$

where $1/p+1/q=1$. We conjecture that in (27) actually equality holds:

$$\|XY-YX\|_p \leq 2^{1/\min(p,q)}\,\|X\|_p\,\|Y\|_p \tag{28}$$

for all $X,Y\in M_n(\mathbf{C})$. This is true for $p=2$ by Theorem 2.2 and trivial for $p=1$ and $p=\infty$. It is easy to prove (28) for $n=2$ and $1\leq p<2$. Indeed, letting $\Sigma(XY-YX)=:(s_1,s_2)$ we have

$$\begin{aligned}&\|XY-YX\|_p = (s_1^p+s_2^p)^{1/p} \leq 2^{1/p-1/2}(s_1^2+s_2^2)^{1/2}\\ &= 2^{1/p-1/2}\|XY-YX\|_2 \leq 2^{1/p-1/2}2^{1/2}\,\|X\|_2\|Y\|_2\\ &= 2^{1/p}\,\|X\|_2\|Y\|_2 \leq 2^{1/p}\,\|X\|_p\|Y\|_p;\end{aligned}$$

here we made use of Theorem 2.2 for $n=2$. We remark that the inequality $\|XY-YX\|_2\leq\sqrt{2}\,\|X\|_p\|Y\|_q$ is in general not true: taking

$$X=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad Y=\begin{pmatrix}1&1\\-1&-1\end{pmatrix},\quad XY-YX=\begin{pmatrix}-2&-2\\2&2\end{pmatrix}$$

we get $\sqrt{2}\,\|X\|_p\|Y\|_q = 2^{3/2+1/p} < 4 = \|XY-YX\|_2$ for $p>2$. □

Example 5.3 (Ky Fan norms) The kth Ky Fan norm $\|\cdot\|_{(k)}$ $(k=1,\ldots,n)$ is defined by

$$\Phi_{(k)}(x_1,\ldots,x_n) = |x_1|+\ldots+|x_k| \quad (|x_1|\geq\ldots\geq|x_n|).$$

Clearly, $\|\cdot\|_{(1)}=\|\cdot\|_\infty$ and $\|\cdot\|_{(n)}=\|\cdot\|_1$. Proposition 5.1 and the trivial estimate $\|XY-YX\|_{(k)}\leq 2\,\|X\|_{(k)}\|Y\|_{(k)}$ give

$$\sup\left\{\frac{\|XY-YX\|_{(k)}}{\|X\|_{(k)}\,\|Y\|_{(k)}} : X,Y\in M_n(\mathbf{C})\setminus\{0\}\right\} = 2. \quad \square$$

We don't know whether the Frobenius norm is the only unitarily invariant norm for which

$$\sup\left\{\frac{\|XY-YX\|}{\|X\|\,\|Y\|} : X,Y\in M_n(\mathbf{C})\setminus\{0\}\right\} = \sqrt{2}. \tag{29}$$

The rest of the paper is devoted to this question.

First of all, from Examples 5.2 and 5.3 we know that the Schatten norms $\|\cdot\|_p$ $(p\neq 2)$ and the Ky Fan norms $\|\cdot\|_{(k)}$ do not satisfy (29).

Let $\|\cdot\|$ be a unitarily invariant norm on $M_n(\mathbf{C})$. The set

$$B_\Phi := \{x \in \mathbf{R}^n : \Phi(x) \le 1\} \tag{30}$$

is closed and convex and invariant under the transformations $(x_1, \ldots, x_n) \mapsto (\pm x_1, \ldots, \pm x_n)$ and under permutations of $(x_1, \ldots, x_n)$. This set is the usual Euclidean unit ball of $\mathbf{R}^n$ if and only if $\|\cdot\|$ is the Frobenius norm. Here is the ultimate result for $n = 2$.

Theorem 5.4 *A unitarily invariant norm $\|\cdot\|$ on $M_2(\mathbf{C})$ satisfies the inequality $\|XY - YX\| \le \sqrt{2}\,\|X\|\,\|Y\|$ for all $X, Y \in M_2(\mathbf{C})$ if and only if it is the Frobenius norm.*

Proof. By virtue of Theorem 2.2, we are left with the "only if" part. Thus, we have to show that B_Φ is the closed unit disk, which is equivalent to proving that $\Phi(x, y) = 1$ for all (x, y) on the eighth of the unit circle between the points $(1, 0)$ and $(1/\sqrt{2}, 1/\sqrt{2})$.

Let $0 \le y \le x \le 1$, $x^2 + y^2 = 1$, and put

$$X = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -y \\ x & 0 \end{pmatrix}, \quad XY - YX = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The singular values of X and Y are x, y, while those of $XY - YX$ are $1, 1$. By assumption

$$\sqrt{2} \ge \frac{\|XY - YX\|}{\|X\|\,\|Y\|} = \frac{\Phi(1,1)}{\Phi(x,y)^2}. \tag{31}$$

Taking $x = 1$, $y = 0$ we get $\Phi(1, 1) \le \sqrt{2}$, and the choice $x = y = 1/\sqrt{2}$ yields $\Phi(1, 1) \ge \sqrt{2}$. Thus, $\Phi(1, 1) = \sqrt{2}$ and (31) implies that $\Phi(x, y) \ge 1$, which means that B_Φ is a subset of the closed unit disk.

To get the other half of the theorem, suppose $0 < s < c < 1$, $c^2 + s^2 = 1$, and let

$$X = \begin{pmatrix} s & -c \\ c & s \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix} = \begin{pmatrix} cs & -s^2 \\ c^2 & -cs \end{pmatrix},$$

$$Y = \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c & -s \\ c & s \end{pmatrix} = \begin{pmatrix} cs & -s^2 \\ s^2 & -cs \end{pmatrix}.$$

Since X and Y are unitarily equivalent to $\mathrm{diag}(1,0)$, we have $\|X\| = \|Y\| = \Phi(1,0) = 1$. Because

$$XY - YX = \begin{pmatrix} u & -v \\ v & -u \end{pmatrix}, \quad u = c^4 - s^4, \quad v = 2cs(c^2 - s^2),$$

the singular values of Z are

$$|u+v| = (c^2 - s^2)(c+s)^2, \quad |u-v| = (c^2 - s^2)(c-s)^2.$$

The inequality $\|XY - YX\| \le \sqrt{2}\,\|X\|\,\|Y\|$ therefore implies that

$$\Phi\left(\frac{|u+v|}{\sqrt{2}}, \frac{|u-v|}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} \frac{\|XY - YX\|}{\|X\|\,\|Y\|} \le 1. \tag{32}$$

Now let $0 \le y \le x \le 1$ and $x^2 + y^2 = 1$ and put

$$c = \frac{1 + \sqrt{y/x}}{\sqrt{2(1 + y/x)}}, \quad s = \frac{1 - \sqrt{y/x}}{\sqrt{2(1 + y/x)}}.$$

Then $0 < s < c < 1$, $c^2 + s^2 = 1$, and

$$\left(\frac{|u+v|}{\sqrt{2}}, \frac{|u-v|}{\sqrt{2}}\right) = \left(\frac{\sqrt{8xy}}{(x+y)^2}\, x, \frac{\sqrt{8xy}}{(x+y)^2}\, y\right).$$

Thus, (32) gives

$$\Phi(x,y) \le \frac{(x+y)^2}{\sqrt{8xy}} = \frac{1 + 2xy}{\sqrt{8xy}}. \tag{33}$$

For the next step, let a, b, c, s be any real numbers such that

$$\frac{1}{\sqrt{2}} < c < 1, \quad 0 < s < \frac{1}{\sqrt{2}}, \quad c^2 + s^2 = 1, \quad 0 < b < a < 1, \quad a^2 + b^2 = 1.$$

Consider

$$X = \begin{pmatrix} a & b \\ b & -a \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} b & -a \\ a & b \end{pmatrix}, \quad Y = \begin{pmatrix} b & a \\ a & -b \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

A straightforward computation delivers

$$XY - YX = (a^2 - b^2)(c+s) \begin{pmatrix} -u & v \\ -v & u \end{pmatrix}, \quad u = c - s, \quad v = 2ab(c+s).$$

We have $\Sigma(X) = \Sigma(Y) = (c, s)$ and the singular values of $XY - YX$ are

$$(a^2 - b^2)(c + s)|u + v|, \quad (a^2 - b^2)(c + s)|u - v|. \tag{34}$$

Choosing a and b so that

$$a^2 = \frac{c + s + 1}{2(c + s)}, \quad b^2 = \frac{c + s - 1}{2(c + s)}$$

we achieve that the numbers (34) become $\sqrt{2cs} + c - s$ and $|\sqrt{2cs} - (c - s)|$. The inequality $\|XY - YX\| \le \sqrt{2}\,\|X\|\,\|Y\|$ therefore yields

$$\Phi\left(\sqrt{cs} + \frac{c - s}{\sqrt{2}}, \left|\sqrt{cs} - \frac{c - s}{\sqrt{2}}\right|\right) \le \Phi(c, s)^2. \tag{35}$$

We remark that $\sqrt{cs} \ge (c-s)/\sqrt{2}$ if and only if $c \le \cos(\pi/12)$ and, accordingly, $s \ge \sin(\pi/12)$.

The function $f(\alpha) = \sqrt{\cos\alpha\,\sin\alpha} + (\cos\alpha - \sin\alpha)/\sqrt{2}$ maps the line segment $[\pi/12, \pi/4]$ bijectively onto the line segment $[1/\sqrt{2}, 1]$. It follows that if β is arbitrarily given between 0 and $\pi/4$, then there is a unique α between $\pi/12$ and $\pi/4$ such that

$$\sqrt{\cos\alpha\,\sin\alpha} + \frac{\cos\alpha - \sin\alpha}{\sqrt{2}} = \cos\beta,$$

which automatically implies that also

$$\sqrt{\cos\alpha\,\sin\alpha} - \frac{\cos\alpha - \sin\alpha}{\sqrt{2}} = \sin\beta.$$

Consequently, given any point (ξ, η) such that $0 < \eta < \xi < 1$ and $\xi^2 + \eta^2 = 1$, there is a unique point (c, s) such that

$$\begin{aligned} &\frac{1}{\sqrt{2}} < c < \cos\frac{\pi}{12}, \quad \sin\frac{\pi}{12} < s < \frac{1}{\sqrt{2}}, \\ &\sqrt{cs} + \frac{c - s}{\sqrt{2}} = \xi, \quad \sqrt{cs} - \frac{c - s}{\sqrt{2}} = \eta. \end{aligned} \tag{36}$$

Equalities (36) show that $2\sqrt{cs} = \xi + \eta$, whence $4cs = 1 + 2\xi\eta$ or equivalently,

$$cs = \frac{1}{4} + \frac{1}{2}\xi\eta. \tag{37}$$

From (35) we infer that $\Phi(\xi, |\eta|) \le \Phi(c, s)^2$.

Finally, let $0 < y < x < 1$ and $x^2 + y^2 = 1$ and put $(x_0, y_0) := (x, y)$. Having (x_k, y_k), we define (x_{k+1}, y_{k+1}) as in the preceding paragraph by

$$\sqrt{x_{k+1}y_{k+1}} + \frac{x_{k+1} - y_{k+1}}{\sqrt{2}} = x_k, \quad \sqrt{x_{k+1}y_{k+1}} - \frac{x_{k+1} - y_{k+1}}{\sqrt{2}} = y_k.$$

Note that

$$\frac{1}{\sqrt{2}} < x_k < \cos\frac{\pi}{12}, \quad \sin\frac{\pi}{12} < y_k < \frac{1}{\sqrt{2}}$$

for all $k \geq 1$ (though not necessarily for $k = 0$), which implies that $|y_k| = y_k$ for $k \geq 1$. The equality $|y_0| = y_0$ is satisfied by assumption. Thus, by virtue of (35),

$$\Phi(x, y) \leq \Phi(x_1, y_1)^2 \leq \Phi(x_2, y_2)^4 \leq \ldots \leq \Phi(x_k, y_k)^{2^k}.$$

Taking into account (33) we get

$$\Phi(x, y) \leq \left(\frac{1 + 2x_k y_k}{\sqrt{8x_k y_k}}\right)^{2^k}, \tag{38}$$

and from (37) we obtain that

$$\begin{aligned} x_k y_k &= \frac{1}{4} + \frac{1}{2}\, x_{k-1}y_{k-1} = \frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2^2}\, x_{k-2}y_{k-2} = \ldots \\ &= \frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4} + \ldots + \frac{1}{2^k}\cdot\frac{1}{4} + \frac{1}{2^k}\, xy = \frac{1}{2} + \frac{1}{2^k}\left(xy - \frac{1}{2}\right). \end{aligned}$$

Hence, letting $m = 2^k$ and $z = xy - 1/2$ we arrive at the estimate

$$\Phi(x, y) \leq \left(\frac{1 + 2\left(\frac{1}{2} + \frac{z}{m}\right)}{\sqrt{8\left(\frac{1}{2} + \frac{z}{m}\right)}}\right)^m = \left(\frac{1 + z/m}{\sqrt{1 + 2z/m}}\right)^m. \tag{39}$$

The right-hand side of (39) goes to 1 as $m \to \infty$, which proves that $\Phi(x, y) \leq 1$ and thus that B_Φ contains the entire closed unit disk. □

Remark 5.5 The idea of the previous proof may be interpreted geometrically. Inequality (33) says that the curve

$$\left\{\frac{\sqrt{8\cos\varphi\,\sin\varphi}}{1 + 2\,\cos\varphi\,\sin\varphi}\,(\cos\varphi, \sin\varphi) : \varphi \in [0, \pi/4]\right\} \tag{40}$$

is contained in B_Φ. This curve is the inner curve in Figure 1. Estimate (35) tells us that if a curve $\{\varrho(\varphi)(\cos\varphi, \sin\varphi) : \varphi \in [0, \pi/4]\}$ is a subset of B_Φ, then so also is the new curve

$$\left\{\varrho(\varphi)^2\left(\sqrt{\cos\varphi\,\sin\varphi} + \frac{\cos\varphi - \sin\varphi}{\sqrt{2}}, \sqrt{\cos\varphi\,\sin\varphi} - \frac{\cos\varphi - \sin\varphi}{\sqrt{2}}\right)\right\}, \tag{41}$$

where φ ranges over $[0, \pi/4]$ and where it would even be sufficient to take φ from the segment $[\pi/12, \pi/4]$ only. Finally, starting with the curve (40) and iteratively constructing new curves via (41) we arrive at the inequalities (38). The first few of these new curves are seen in Figure 1. The figure convincingly reveals that the iteratively obtained curves approximate the unit circle. That this is really the case was shown in the last step of the proof. $\square$

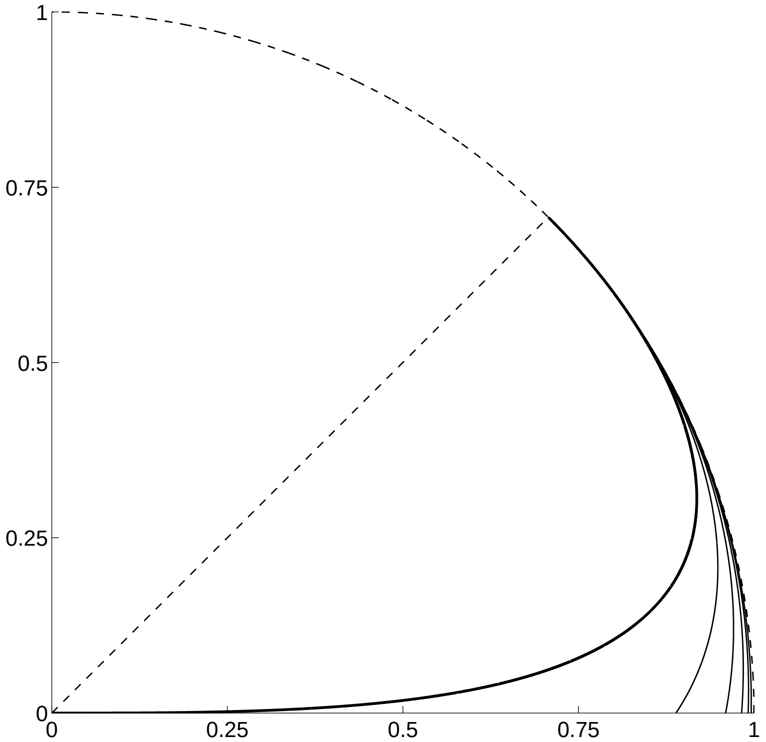

Fig. 1. A sequence of curves approximating an eighth of the unit circle.

Remark 5.6 In the proof of Theorem 5.4 we worked with real matrices only. This shows that if $\|\cdot\|$ is a unitarily invariant norm on $M_2(\mathbf{R})$ such that $\|XY - YX\| \le \sqrt{2}\,\|X\|\,\|Y\|$ for all $X, Y \in M_2(\mathbf{R})$, then $\|\cdot\|$ is necessarily the Fobenius norm. $\square$

Example 5.7 (Polyhedral norms) A unitarily invariant norm $\|\cdot\|$ on $M_n(\mathbf{C})$ is called a polyhedral norm if the set B_Φ defined by (30) is a (convex) polyhedron in $\mathbf{R}^n$. Suppose $\|\cdot\|$ to be a unitarily invariant polyhedral norm on $M_n(\mathbf{C})$ satisfying (29). From Theorem 5.4 we deduce that the intersection of the polyhedron B_Φ with the plane $\{(x_1, x_2, 0, \dots, 0) : x_1, x_2 \in \mathbf{R}\}$ is the closed unit disk, which is impossible. Consequently, there are no unitarily invariant polyhedral norms on $M_n(\mathbf{C})$ for which (29) is true.

Let $\|\cdot\|_{p_m}$ be the polygonal norm on $M_2(\mathbf{C})$ for which the set (30) is the regular m-gon inscribed in the unit circle. Since $(\pm 1, 0)$ and $(0, \pm 1)$ must be vertices of this m-gon, the number m is necessarily divisible by 4. Put

$$C_m := \sup\left\{ \frac{\|XY - YX\|_{p_m}}{\|X\|_{p_m}\|Y\|_{p_m}} : X, Y \in M_2(\mathbf{C}) \setminus \{0\} \right\}.$$

It is easily seen that $\cos(\pi/m)\|Z\|_{p_m} \le \|Z\|_{\mathsf{F}} \le \|Z\|_{p_m}$ for all $Z \in M_2(\mathbf{C})$. From Theorem 2.2 (for $n = 2$) we therefore get

$$\cos\frac{\pi}{m}\,\|XY - YX\|_{p_m} \le \|XY - YX\|_{\mathsf{F}} \le \sqrt{2}\,\|X\|_{\mathsf{F}}\|Y\|_{\mathsf{F}} \le \sqrt{2}\,\|X\|_{p_m}\|Y\|_{p_m}.$$

Thus, $C_m \le \sqrt{2}/\cos(\pi/m)$ for all m. If $m = 8k + 4$ $(k = 0, 1, 2, \ldots)$, then $\Phi(1, 1) = \sqrt{2}/\cos(\pi/m)$ and hence Proposition 5.1 implies that

$$C_m = \sqrt{2}/\cos(\pi/m)$$

in this case. The case where $m = 8k$ $(k = 1, 2, 3, \ldots)$ is more complicated. From the proof of Theorem 5.4 we see that

$$\Phi\left(\frac{\sqrt{8xy}}{1 + 2xy}\,x, \frac{\sqrt{8xy}}{1 + 2xy}\,y\right) \le \frac{C_m}{\sqrt{2}} \tag{42}$$

whenever $0 \le y \le x \le 1$ and $x^2 + y^2 = 1$. If $\varphi = \pi/4 - \pi/m$, $x = \cos\varphi$, $y = \sin\varphi$, then the left-hand side of (42) is $1/\cos(\pi/m)$ times

$$\frac{\sqrt{8\,\cos\varphi\,\sin\varphi}}{1 + 2\,\cos\varphi\,\sin\varphi} = \frac{2\sqrt{\sin 2\varphi}}{1 + \sin 2\varphi} = \frac{2\sqrt{\cos\frac{2\pi}{m}}}{1 + \cos\frac{2\pi}{m}} = 1 - O\left(\frac{1}{m^4}\right).$$

Consequently, if m is divisible by 8 we have

$$\frac{\sqrt{2}}{\cos\frac{\pi}{m}}\left(1 - O\left(\frac{1}{m^4}\right)\right) \le C_m \le \frac{\sqrt{2}}{\cos\frac{\pi}{m}}. \tag{43}$$

We conjecture that in fact $C_m = \sqrt{2}/\cos(\pi/m)$. Note that the lower bound in (43) is strictly larger than $\max(\mu, 2/\mu) = \Phi(1, 1) = \sqrt{2}$ if m is large enough (actually even for all $m = 8k \ge 8$), which reveals that the bound provided by Proposition 5.1 is not sharp. $\square$

Remark 5.8 Let $\|\cdot\|$ again be a unitarily invariant norm on $M_n(\mathbf{C})$ subject to (29). By embedding $M_2(\mathbf{C})$ appropriately into $M_n(\mathbf{C})$, we obtain from Theorem 5.4 that the intersection of B_Φ with each of the $n(n-1)/2$ planes spanned by two of the coordinate axes is the closed unit disk. In particular, B_Φ is necessarily contained in the intersection of the $n(n-1)/2$ cylinders $x_j^2 + x_k^2 \le 1$.

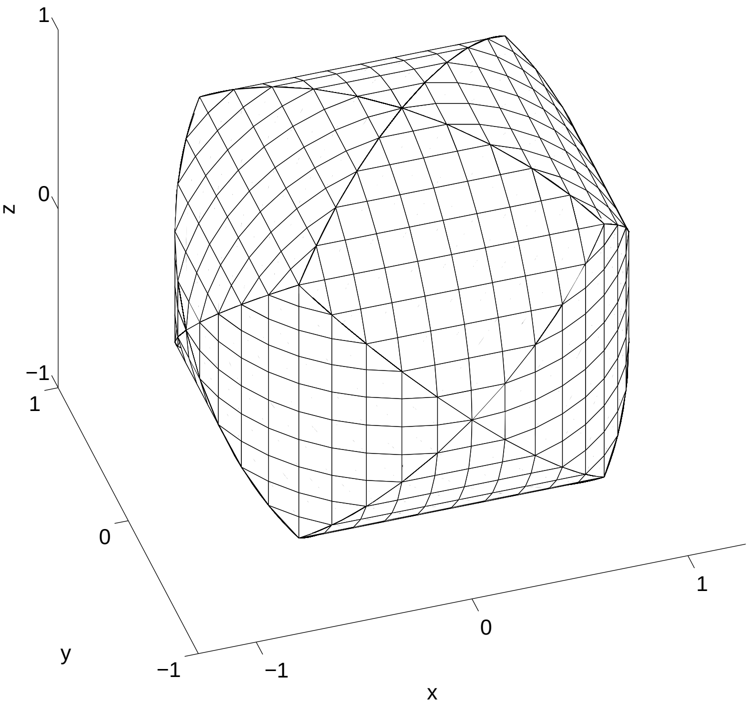

Fig. 2. The intersection of the cylinders $x^2 + y^2 \leq 1$, $x^2 + z^2 \leq 1$, $y^2 + z^2 \leq 1$.

Now let $n = 3$ and denote by B the intersection of the three cylinders given by $x^2 + y^2 \leq 1$, $x^2 + z^2 \leq 1$, $y^2 + z^2 \leq 1$; see Figure 2. Defining

$$\begin{aligned} &\Phi(x, y, z) = \min\{t > 0 : (x, y, z)/t \in B\} \\ &= \min\{t > 0 : (x^2 + y^2)/t^2 \leq 1, (x^2 + z^2)/t^2 \leq 1, (y^2 + z^2)/t^2 \leq 1\} \\ &= \max\left(\sqrt{x^2 + y^2}, \sqrt{x^2 + z^2}, \sqrt{y^2 + z^2}\right) \end{aligned}$$

we have $B = B_\Phi$. The unitarily invariant norm associated with Φ is given by $\|X\| = \sqrt{s_1^2 + s_2^2}$ where $s_1 \geq s_2 \geq s_3$ are the singular values of X. In the notation of [1, p. 95], this is the $\|\cdot\|_{(2)}^{(2)}$ norm, a mixture of the 2nd Ky Fan and the 2nd Schatten (= Frobenius) norms. Clearly, $\|\cdot\|_{(2)}^{(2)}$ is a good candidate for a norm satisfying (29). If we put $\Sigma(X) = (s_1, s_2, s_3)$, $\Sigma(Y) = (t_1, t_2, t_3)$, $\Sigma(XY - YX) = (z_1, z_2, z_3)$, the singular values always in decreasing order, then Theorem 2.2 is equivalent to the inequality

$$z_1^2 + z_2^2 + z_3^2 \leq 2(s_1^2 + s_2^2 + s_3^2)(t_1^2 + t_2^2 + t_3^2),$$

while the question whether $\|\cdot\|_{(2)}^{(2)}$ satisfies (29) amounts to the inequality

$$z_1^2 + z_2^2 \le 2(s_1^2 + s_2^2)(t_1^2 + t_2^2).$$

We don't know whether the last inequality is true or not. Notice that the inequality $z_1^2 \le 2\, s_1^2 t_1^2$ is not true, because it is equivalent to saying that $\|XY - YX\|_\infty \le \sqrt{2}\,\|X\|_\infty \|Y\|_\infty$ which, by Example 5.2, is only valid with $\sqrt{2}$ replaced by 2.

For $n \ge 4$ the number of candidates for unitarily invariant norms satisfying (29) increases. The candidates include the norms

$$\|X\|_{(k)}^{(2)} = \sqrt{s_1^2 + \ldots + s_k^2} \quad (s_1 \ge \ldots \ge s_n)$$

with $2 \le k \le n$. We remark that for all these norms the intersection of B_Φ with an arbitrary plane spanned by two of the coordinate axes is the unit disk. □

Acknowledgement. We thank all the many people who turned to us with remarks, comments, and suggestions pertaining to the topic since the appearance of [4]. We are in particular greatly indebted to Rajendra Bhatia, Lajos László, Ilya and Valentin Spitkovsky, Berthold Euler, Arieh Iserles, Seak-Weng Vong, and Xiao-Qing Jin.

References

[1] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York 1997.

[2] R. Bhatia and F. Kittaneh, Commutators, pinchings, and spectral variation, *Operators and Matrices* **2** (2008), pp. 143–151.

[3] A. M. Bloch and A. Iserles, Commutators of skew-symmetric matrices, *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **15** (2005), pp. 793–801.

[4] A. Böttcher and D. Wenzel, How big can the commutator of two matrices be and how big is it typically? *Linear Algebra Appl.* **403** (2005), pp. 216–228.

[5] 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.

[6] Seak-Weng Vong and Xiao-Qing Jin, Proof of Böttcher and Wenzel's conjecture, *Operators and Matrices*, to appear.