SOME INEQUALITIES FOR CENTRAL MOMENTS OF MATRICES

ZOLTÁN LÉKA

Abstract. In this paper we shall study noncommutative central moment inequalities with a main focus on whether the commutative bounds are tight in the noncommutative case, or not. We prove that the answer is affirmative for the fourth central moment and several particular results are given in the general case. As an application, we shall present some lower estimates of the spread of Hermitian and normal matrices as well.

1. Introduction

Let $X$ be a random variable on a probability space $(\Omega, P)$. Then its $p$th (fractional) central moments are defined by the formula

$$\mu_p(X) = \int_{\Omega} \left| X - \int_{\Omega} X \, dP \right|^p \, dP.$$ 

The most studied noncommutative analogue of these quantities is the noncommutative variance or quantum variance. Let $M_n(\mathbb{C})$ be the algebra of $n \times n$ complex matrices. Whenever $\Phi: M_n(\mathbb{C}) \to M_m(\mathbb{C})$ is a positive unital linear map, the variance of a matrix $A$ can be defined as $\Phi(A^*A) - \Phi(A)^{*}\Phi(A)$. For several interesting properties of this variance, we refer the reader to Bhatia’s book [4]. For instance, special choices of $\Phi$ and applications of variance estimates provided simple new proofs of spread estimates of normal and Hermitian matrices as well, see [5] and [6]. On the other hand, the first sharp estimate of the noncommutative variance appeared in K. Audenaert’s paper [1] in connection with the Böttcher–Wenzel commutator estimate. For several different proofs of his result, we refer to [8], [6] and [23]. Recently, extremal properties of the quantum variance were studied in [20].

It is simple to see that if $\omega$ is a state (i.e. positive linear functional of norm 1) of the algebra $M_n(\mathbb{C})$, then one has the upper bound

$$\omega(|A - \omega(A)|^2) = \omega(|A|^2) - |\omega(A)|^2 \leq \|A\|^2$$

(see [3, Theorem 3.1] for positive linear maps). A careful look of the previous inequality says that the noncommutative variance cannot be larger than the ordinary variance of random variables. In fact, if $X$ is a Bernoulli variable, that is, $P(X = 0) = p$ and $P(X = 1) = 1 - p$ ($0 \leq p \leq 1$), then $\mu_2(X) \leq 1/4$ holds. Furthermore, for any (complex-valued) random variable $Z: \Omega \to \mathbb{C}$ the inequality

$$\sqrt{\mu_2(Z)} \leq 2 \max\{\sqrt{\mu_2(X)}: X \text{ Bernoulli random variable}\} \|Z\|_\infty$$

readily follows, see [1, Theorem 7] in the discrete case and [13, Theorem 2] in the general case, for instance. Furthermore, one can have the following upper bound for
2. General moment inequalities

2.1. A moment estimate of partial isometries. Let $X$ be a Bernoulli random variable with parameter $p$. Then one has clearly the inequality $\mu_4(X) \leq \frac{1}{12}$, while $\mu_4(Z) \leq 4 \min_{\lambda \in \mathbb{C}} \max_i |z_i - \lambda|^4$ comes true for any finite–valued random variable $Z$. From a geometric point of view, the quantity $\min_{\lambda \in \mathbb{C}} \max_i |z_i - \lambda|$ is the radius of the smallest enclosing circle of the values of $Z$. For several inequalities in connection with it and vector norms, the reader might see [11, Section 4].

Our first result gives the corresponding noncommutative moment estimate for partial isometries. Recall that an $n \times n$ matrix $V$ is partial isometry if $V$ is an isometry on the orthogonal complement of its kernel. A useful characterization says that $V$ is a partial isometry if and only if $V^*V$ is an orthogonal projection (to the subspace $(\ker V)^\perp$), or, which is the same, $V = VV^*V$ (see [11], [21, page 95]). Hence $V^*$ is a partial isometry and $VV^*$ is an orthogonal projection as well.

We start with a technical lemma.

**Lemma 1.** Let $x_1, x_2, x_3, x_4 \in \mathbb{R}$. Then
\[
\max_{x} (2x_1^2x_3 - 2x_1x_4 + x_2^2) = 1
\]
subject to the constraints $0 \leq x_3 \leq x_2 \leq 1$ and $x_1 \leq x_4 + \sqrt{(1 - x_2^2)(x_2^2 - x_1^2)}$.

**Proof.** With a change of variables $y_2 = \sqrt{1 - x_2^2}$, $y_3 = \sqrt{x_2^2 - x_1^2}$ and $y_1 = x_1, y_4 = x_4$, we have $2x_2^2x_3 - 2x_1x_4 + x_2^2 - 1 = 2y_1^2\sqrt{1 - y_2^2 - y_3^2} - 2y_1y_4 - y_2^2$. Notice that the last function is convex in $y_1$, hence it attains its maximum when $y_1$ is the largest, i.e. $y_1 = y_4 + y_2y_3$. Therefore, it is enough to prove the general statement
\[
\max_{y_1} G(y_1, y_2, y_3, y_4) = \max_{y_1} 2y_1^2\sqrt{1 - y_2^2 - y_3^2} - 2y_1y_4 - y_2^2 = 0
\]
subject to $y_1 = y_4 + y_2y_3$.

First, we compute the extrema of $G$ when $(y_2, y_3, y_4)$ is in the open cylinder $\mathbb{D} \times \mathbb{R}$, where $\mathbb{D}$ denotes the open unit disk of the plane. To do this, let us find the constrained critical points of the Lagrangian $L(y, \lambda) = G(y) - \lambda c(y)$, where
\[
\lambda c(y)
\]
the constraint function is \( c(y) = y_1 - y_4 - y_2 y_3 \). A little calculation gives for the gradient equation \( \nabla \mathcal{L}(y, \lambda) = 0 \) that

\[
-\lambda - 2y_4 + 4y_1 \sqrt{1 - y_2^2 - y_3^2} = 0 \\
\lambda y_3 - \frac{2y_1 y_3}{\sqrt{1 - y_2^2 - y_3^2}} - 2y_2 = 0 \\
\lambda y_2 - \frac{2y_1^2 y_3}{\sqrt{1 - y_2^2 - y_3^2}} = 0 \\
\lambda - 2y_1 = 0 \\
y_1 + y_2 y_3 + y_4 = 0.
\]

To solve this system, note that if \( y_1 = \lambda/2 = 0 \) then \( y_2 = y_4 = 0 \) and \(-1 < y_3 < 1\). From \( \lambda = y_4 = 0 \), it is simple to check that \( G \leq 0 \). Indeed, \( y_1 = y_2 y_3 \) and

\[
2y_2^2 \sqrt{1 - y_2^2 - y_3^2} \leq 2y_2 \sqrt{1 - y_2^2} \leq 2 \sqrt{y_2^2(1 - y_2^2)} \leq 1.
\]

On the other hand, if \( \lambda \neq 0 \), from the third and fourth equation

\[
2y_2 \sqrt{1 - y_2^2 - y_3^2} = \lambda y_3.
\]

Substitute this to the second one and we obtain that

\[
y_2(1 - y_2^2 - y_3^2) = y_4^2 y_2 + y_2 \sqrt{1 - y_2^2 - y_3^2}.
\]

Since \( 1 - y_2^2 - y_3^2 < y_4^2 + \sqrt{1 - y_2^2 - y_3^2} \), if \((y_2, y_3) \in \mathbb{D}\) and \( y_1 \neq 0 \), it follows that \( y_2 = 0 \). Clearly, \( y_2 = y_3 = 0 \) and \( y_1 = y_4 = \lambda/2 \) hold. The corresponding Hessian of \( \mathcal{L} \) at a stationary point \((y_*, \lambda_*) = (t, 0, 0, t, 2t)\) is

\[
\nabla_{yy} \mathcal{L}(y_*, \lambda_*) = \begin{bmatrix}
4 & 0 & 0 & -2 \\
0 & -2(t^2 + 1) & 2t & 0 \\
0 & 2t & -2t^2 & 0 \\
-2 & 0 & 0 & 0
\end{bmatrix}.
\]

Now let us consider an \( y_4 \)-sections of the cylinder \( \mathbb{D} \times \mathbb{R} \); that is, add the constraint \( y_4 = t (\neq 0) \) to the optimization problem. Then \( c_2(y) = y_4 - t \) and \( \nabla c_2(y_*) = [0, 0, 0, 1]^* \). Note that \( \nabla c(y_*) = [1, 0, 0, -1]^* \), hence the tangent plane of the constraints at \( y_* \) is

\[
\mathcal{T}(\lambda_*):= \{ w: w^* \nabla c(y_*) = 0 \text{ and } w^* \nabla c_2(y_*) = 0 \} = \{ [0, w_2, w_3, 0]^*: w_i \in \mathbb{R} \}.
\]

Furthermore, we obtain

\[
w^* \nabla_{yy} \mathcal{L}(y_*, \lambda_*) w = -2t^2 w_2^2 - 2(w_2 - tw_3)^2 < 0, \quad 0 \neq w \in \mathcal{T}(\lambda_*).
\]

Thus \( y_* \) is a strict local maximum of \( G \) subject to \( c \) and \( c_2 \), see [19] Theorem 12.6], and \( G(y_*) = 0 \).

For \( y_4 \neq 0 \), all \( y_4 \)-sections of the cylinder \( \mathbb{D} \times \mathbb{R} \) contain exactly one local maximum point, hence \( 0 \) is the global maximum of \( G \) on its domain (s.t. the constraint).

\[\square\]

**Theorem 1.** Let \( V \) be a partial isometry in \( M_n(\mathbb{C}) \). Let \( Q \in M_n(\mathbb{C}) \) be a rank-one orthogonal projection. Then

\[
\text{Tr } [QV - VQ]|V|^{1/2} \leq \frac{4}{3}.
\]

Moreover, if the equality holds then \( |V|Q = Q|V| \).
Proof. Without loss of generality, we can assume that $\alpha := \text{Tr} \, QV = \text{Tr} \, QV^* \geq 0$. Let $V = UP$ be a polar decomposition of $V$, where $P = V^*V$ is an orthogonal projection and $U$ is unitary. Choose a unit vector $x$ such that $Q = x^*x$. First, one has that

$$\text{Tr} \, [QV - \alpha I]^4 = \text{Tr} \, [Q(V^*V - \alpha V^*V^2 - \alpha V^*V^* + \alpha^2 V^2V - \alpha^3 V)
- \alpha V^*V + \alpha^2 V^2 + \alpha^2 V^*V - \alpha^3 V)
- \alpha V^*V + \alpha^2 V^*V + \alpha^2 V^*V^* - \alpha^3 V^*
+ \alpha^2 V^*V - \alpha^3 V - \alpha^3 V^* - \alpha^4 I)]$$

and applying the identities $V^*V^* = V^*$ and $VV^*V = V$,

$$= \|Px\|^2 - 2\alpha \text{Re}(V^*V^2, x) + \alpha^2 (3\|V^*Vx\|^2 + \|VV^*x\|^2 - 2)
+ 2\alpha^2 \text{Re}(V^2x, x) - 3\alpha^4$$

and since $\|V^*Vx\|^2 + \|VV^*x\|^2 \leq 2$,

$$\leq \|Px\|^2 - 2\alpha \text{Re}(Vx, Px) + 2\alpha^2 \|Px\|^2 + 2\alpha^2 |\text{Re}(V^2x, x)| - 3\alpha^4.$$

Next, from the Cauchy–Schwarz inequality

$$|\text{Re}(Vx, V^*x)| = |\text{Re}(PU Px, PU^*x)| \leq \|PU Px\| \|P^*x\| \leq \|PU Px\|.$$

Therefore, we obtain

$$\text{Tr} \, [QV - \alpha I]^4 \leq \|Px\|^2 + 2\alpha^2 \|Px\|^2 - 3\alpha^4 - 2\alpha \text{Re}(UPx, Px) + 2\alpha^2 \|PU Px\|$$

$$\leq \max_{0 \leq \alpha \leq 1} 1 + 2\alpha^2 - 3\alpha^4$$

$$+ \max(\|Px\|^2 - 1 - 2\alpha \text{Re}(UPx, Px) + 2\alpha^2 \|PU Px\|).$$

It is simple to check that

$$\max_{0 \leq \alpha \leq 1} 1 + 2\alpha^2 - 3\alpha^4 = \frac{4}{3}.$$

For the remaining part of the previous inequality, from the Cauchy–Schwarz inequality we have the following constraint for $0 \leq \alpha$

$$\alpha = \text{Re} \langle UPx, x \rangle$$

$$= \text{Re} \langle PU Px, Px \rangle + \text{Re} \langle P^+ U Px, P^+ x \rangle$$

$$\leq \text{Re} \langle UPx, Px \rangle + \|P^+ U Px\| \|P^+ x\|$$

$$= \text{Re} \langle UPx, Px \rangle + (\|UPx\|^2 - \|PU Px\|^2)^{1/2}(1 - \|Px\|^2)^{1/2}$$

$$= \text{Re} \langle UPx, Px \rangle + (\|Px\|^2 - \|PU Px\|^2)^{1/2}(1 - \|Px\|^2)^{1/2}.$$

Hence we can apply Lemma 1 to obtain that

$$2\alpha^2 \|PU Px\| - 2\alpha \text{Re}(UPx, Px) + \|Px\|^2 \leq 1.$$

Thus the inequality

$$\text{Tr} \, [QV - \alpha I]^4 \leq \frac{4}{3}$$

follows.

Note that when the equality occurs $\|Px\|^2 = 1$ must hold. This means that $Px = x$, which is the same as $x \in \text{ran} \, P$, or, $QV^*V = VV^*Q$. \qed
Example 1. We give a matrix example, which is not normal, to see that the previous inequality is tight. Set the partial isometry
\[
V = \frac{1}{\sqrt{3}} \begin{bmatrix}
1 & 1 & 1 & 0 \\
1 & -1/2 & -1/2 & 1 \\
1 & -1/2 & -1/2 & -1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
and define
\[
Q = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]
In fact, \(V\) is a partial isometry because \(V^*V\) is an orthogonal projection. Moreover, the spectrum of \(V\) is \(\sigma(V) = \{1, 1/\sqrt{3}, 0, -1\}\), hence \(\min_{\lambda \in \mathbb{C}} \|V - \lambda I\| = 1\). Furthermore, it is simple to see that
\[
\text{Tr} [QV - \text{Tr} [QV]^4] = \frac{4}{3}.
\]

2.2. Convex sets of density matrices. Let \(A_1, A_2, \ldots, A_k \in M_n(\mathbb{C})\) be Hermitian matrices and let \(\alpha_1, \alpha_2, \ldots, \alpha_k\) be real numbers. Let us consider the convex, compact set
\[
\mathcal{D}(A_1, A_2, \ldots, A_k) := \{X \geq 0 : \text{Tr} X = 1 \text{ and } \text{Tr} [XA_i] = \alpha_i, \ i = 1, 2, \ldots, k\}.
\]
Note that \(\mathcal{D}(A_1, A_2, \ldots, A_k) = \mathcal{D}(A_1 - \alpha_1 I, A_2 - \alpha_2 I, \ldots, A_k - \alpha_k I)\). The geometry of \(\mathcal{D}(A_1, A_2, \ldots, A_k)\) is strongly related to that of the elliptope; i.e. the set of real \(n \times n\) symmetric positive semidefinite matrices with an all-one diagonal (briefly, correlation matrices) [13, 10, Chapter 31.5]. Additionally, we used the set \(\mathcal{D}(A_1, A_2, \ldots, A_k)\) to provide a description of the extreme non-commutative covariance matrices associated to Hermitian tuples (see [13]).

We recall that whenever \(D\) is an extreme point of \(\mathcal{D}(A_1, A_2, \ldots, A_k)\), one has the rank estimate [13, Corollary 1]
\[
\text{rank } D \leq \sqrt{k} + 1.
\]

We remark that the proof of the previous inequality is closely related to a method invented by C.K. Li and B.S. Tam [13] in order to describe the extreme correlation matrices.

Turning back to moment inequalities, Audenaert’s theorem [1] on the (quantum) standard deviation states that for any \(A \in M_n(\mathbb{C})\) there exists a rank-one orthogonal projection \(P\) such that
\[
\max_{\|D\| \geq 0, \text{Tr} D = 1} (\text{Tr} [DA - \text{Tr} [DA]^2])^{1/2} = (\text{Tr} [|P|A - \text{Tr} [PA]|^2])^{1/2} = \min_{\lambda \in \mathbb{C}} \|A - \lambda I\|.
\]

This result was proved directly in [7] and in [8] for \(C^*\)-algebras by means of a characterization of the Birkhoff–James orthogonality in matrix and operator algebras, respectively.

Throughout the paper we say that an \(n \times n\) matrix \(D\) is a density if \(D\) is positive semidefinite and \(\text{Tr} D = 1\). Exploiting the aforementioned rank estimate, now we can prove the following

Theorem 2. Let \(D \in M_n(\mathbb{C})\) be a density. For any \(1 \leq p < \infty\) and \(A \in M_n(\mathbb{C})\), there exists a rank-one orthogonal projection \(P \in M_n(\mathbb{C})\) such that
\[
\text{Tr} [DA - \text{Tr} [DA]^p] = \text{Tr} [|P|A - \text{Tr} [PA]|^p].
\]
Proof. Without loss of generality we can assume that \( \text{Tr} \ [DA] = \alpha \) is real, hence \( \text{Tr} \ D \frac{A + A^*}{2} = \alpha \) holds as well. Let us introduce the convex set
\[
\mathcal{D}([A - \alpha I_n]^p, \frac{A + A^*}{2}) := \left\{ X \text{ density} : \text{Tr} \ [X|A - \alpha I_n|^p] = \text{Tr} \ [D|A - \alpha I_n|^p] \right. \\
\left. \quad \text{and} \ \text{Tr} \ X \frac{A + A^*}{2} = \alpha \right\}.
\]
Obviously, \( \mathcal{D}([A - \alpha I_n]^p, \frac{A + A^*}{2}) \) is non-empty. Relying upon the inequality (1), the rank of extreme points of \( \mathcal{D} \) is at most \( \sqrt{3} \). Since any rank-1 density is an orthogonal projection, the proof is complete. \( \square \)

Now we can prove the main theorem of the section.

2.3. A 4-order moment estimate.

**Theorem 3.** Let \( A \in M_n(\mathbb{C}) \) and let \( D \in M_n(\mathbb{C}) \) be a density matrix. Then
\[
\text{Tr} \ [D|A - \text{Tr} \ [DA]|^4] \leq \frac{4}{3} \min_{\lambda \in \mathbb{C}} \| A - \lambda I_n \|^4.
\]

**Proof.** Without loss of generality, we can assume that \( \| A \| = 1 \) holds. Now form the partial isometry (see [11])
\[
V = \begin{bmatrix} A & (I - AA^*)^{1/2} \\ 0 & 0 \end{bmatrix} \in M_{2n}(\mathbb{C}).
\]
Let
\[
\tilde{D} := \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},
\]
which is a density matrix, of course. Then a straightforward calculation gives that
\[
|V - \text{Tr} \ [\tilde{D}V]|^4 = \begin{bmatrix} |A - \text{Tr} \ [DA]|^4 + X & * \\ * & * \end{bmatrix},
\]
where \( X = (A - \text{Tr} \ [DA])^*(I - AA^*)(A - \text{Tr} \ [DA]) \geq 0 \), hence
\[
|A - \text{Tr} \ [DA]|^4 \leq \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} |V - \text{Tr} \ [\tilde{D}V]|^4 \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.
\]
Therefore the inequality
\[
\text{Tr} \ [D|A - \text{Tr} \ [DA]|^4] \leq \text{Tr} \ [\tilde{D}|V - \text{Tr} \ [\tilde{D}V]|_{2n}|^4]
\]
follows. Additionally, relying on Theorem 2, one can assume that \( \tilde{D} = xx^* \) is a rank-one orthogonal projection with some unit vector \( x \). Then Theorem 1 immediately gives that
\[
\text{Tr} \ [D|A - \text{Tr} \ [DA]|^4] \leq \text{Tr} \ [\tilde{D}|V - \text{Tr} \ [\tilde{D}V]|_{2n}|^4]
\]
\[
\leq \frac{4}{3} \|V\|^4
\]
\[
= \frac{4}{3} \|A\|^4.
\]
Changing \( A \) to \( A - \lambda I \), we get the proof of the statement. \( \square \)

Surprisingly, the next example shows that if \( A \) is not normal the previous upper bound is not necessarily sharp. Please, compare it with Example 1 and Audenaert’s theorem on the noncommutative variance.

**Example 2.** Let \( A \) denote the Jordan block \( \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \). We calculate the value of
\[
\mu_4(A) := \max \{ \text{Tr} \ [D|A - \text{Tr} \ [DA]|^4] : 0 \leq D \in M_2(\mathbb{C}) \text{ and } \text{Tr} \ D = 1 \}.
\]
Indeed, from Theorem 2 we can find a projection \( P = zz^* \), \( z^* = [z_1, z_2] \in \mathbb{C}^2 \) and \( |z_1|^2 + |z_2|^2 = 1 \), such that

\[
\mu_4(A) = \text{Tr} |P| A - \text{Tr} [PA]|^4.
\]

Then a little computation gives that

\[
\text{Tr} |P| A - \text{Tr} [PA]|^4 = |z_1|^4|z_2|^4 + |z_1|^4|z_2|^2 + |z_1|^2|z_2|^2 + 2|z_1|^2|z_2|^2(2|z_1|^2|z_2|^2 + 1) + |z_2|^2(1 + |z_1|^2|z_2|^2)^2
\]

\[
= 4|z_2|^4 - 3|z_2|^6 = p(|z_2|).
\]

Note that \( \max_{1 \leq i \leq 1} p(|z_2|) = p(1) = 1 \). Furthermore, it is simple to check that \( \min_{\lambda \in \mathbb{C}} \| A - \lambda \| = \| A - I_2 \| = 1 \). Hence we get that

\[
\mu_4(A) = \min_{\lambda \in \mathbb{C}} \| A - \lambda \|^4.
\]

However, from \cite{13} Theorem 4] we have

\[
\mu_4(A) = \frac{4}{3} \min_{\lambda \in \mathbb{C}} \| A - \lambda \|^4
\]

for any normal \( A \).

Here we make a direct application of Theorem 3 to obtain a lower bound for the spread of normal and Hermitian matrices. If \( A \) is an \( n \times n \) matrix and \( \lambda_i(A) \) \((1 \leq i \leq n)\) denote its eigenvalues then the spread of \( A \) is

\[
\text{spd}(A) = \max_{i,j} |\lambda_i(A) - \lambda_j(A)|.
\]

Spread estimates were initiated by L. Mirsky in his seminal papers \cite{17} and \cite{18}. After that several author provided upper and lower bounds for it, see \cite{3, 6, 12} and \cite{10}, for instance, and the references therein.

For a normal \( A \) the spectral theorem gives that \( \min_{\lambda \in \mathbb{C}} \| A - \lambda I \| = r_A \), where \( r_A \) denotes the radius of the smallest disk that contains the eigenvalues of \( A \). Jung’s theorem on the plane says that if \( F \) is a finite set of points of diameter \( d \) then \( F \) must be contained in a closed disk of radius \( d/\sqrt{3} \), see \cite{22} Chapter 16]. Hence, for any normal \( A \),

\[
\min_{\lambda \in \mathbb{C}} \| A - \lambda I \| \leq \frac{1}{\sqrt{3}} \text{spd}(A)
\]

(see \cite{5}, p. 1567–1568).

**Corollary 1.** Let \( A \in M_n(\mathbb{C}) \) be a normal matrix. Then

\[
\text{Tr} |D|A - \text{Tr} [DA]|^4| \leq \frac{4}{27} \text{spd}(A)^4.
\]

Moreover, if \( A \) is Hermitian then

\[
\text{Tr} |D|A - \text{Tr} [DA]|^4| \leq \frac{\text{spd}(A)^4}{12}.
\]

For a different proof of the last statement, the reader might see \cite{14} p. 169 Remark, \cite{24} Theorem 3] and \cite{13} Theorem 2 for the normal case in Theorem 3.

We recall that the quantity \( \Delta(A) \equiv \min_{\lambda \in \mathbb{C}} \| A - \lambda I \| \) appeared in Stampfli’s well-known result \cite{25} for the derivation norm

\[
2\Delta(A) = \max_{\| X \| = 1} \| AX - XA \|
\]

while the diameter of the unitary orbit of \( A \) is given by the formula

\[
2\Delta(A) = \max \{ \| A - UAU^* \| : U \text{ is unitary} \},
\]
see [7]. Hence any lower estimate of $\Delta(A)$ in terms of central moments might have its own interest. In the case of the noncommutative variance, this method was first exploited by R. Bhatia and R. Sharma in a series of papers [5], [6]. Choosing different density matrices in the variance inequality, they got several interesting inequalities for $\Delta(A)$ and the spread of $A$, as well. Briefly, the idea of non-commutative variance estimates turned out to be fruitful and led to simple new proofs of known spread estimates, including Mirsky’s and Barnes–Hoffman’s lower bounds (see [5] for details).

2.4. Remark. Let $\omega$ be a positive linear functional of $M_n(\mathbb{C})$. Then the map $A \mapsto \omega(|A|^p)^{1/p}$, $p \neq 2$, is not a norm on $M_n(\mathbb{C})$, because the triangle inequality fails, in general. However, the monotonicity statement

$$\omega(|A|^p)^{1/p} \leq \omega(|A|^q)^{1/q}$$

clearly holds for all $1 \leq p \leq q < \infty$. In fact, $\omega(A) = \text{Tr} DA$ with some $D \succeq 0$ and $\text{Tr} D = 1$. Furthermore, we can assume that $|A| = (A^*A)^{1/2}$ is diagonal, hence the discrete Hölder-inequality gives that

$$\left(\sum_{i=1}^n d_{ii} a_{ii}^p\right)^{1/p} \leq \left(\sum_{i=1}^n d_{ii} a_{ii}^q\right)^{1/q}$$

which is exactly what we need. Moreover,

$$\omega(|A - \omega(A)|^2)^{1/2} = (\omega(|A|^2) - |\omega(A)|^2)^{1/2} \leq \omega(|A|^2)^{1/2} \leq \|A\|.$$ 

Therefore, we get for any $1 \leq p \leq 2$ and $A \in M_n(\mathbb{C})$ that

$$\omega(|A - \omega(A)|^p)^{1/p} \leq \min_{\lambda \in \mathbb{C}} \|A - \lambda I\|.$$ 

Note that a simple calculus gives that

$$2 \max\{(\mathbb{E}|X - \mathbb{E}(X)|^p)^{1/p} : X \text{ Bernoulli random variable }\} = 1$$

if $1 \leq p \leq 2$, hence the commutative bound turns out to be a tight bound in the non-commutative case as well. Actually, set $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Then

$$\text{Tr} [PA - \text{Tr} [PA]]^p = 1 = \min_{\lambda \in \mathbb{C}} \|A - \lambda I\|.$$ 

2.5. Remark. It would be interesting to know whether $\Phi: M_n(\mathbb{C}) \to M_k(\mathbb{C})$ is a positive unital linear map then the inequality

$$\Phi(|A - \Phi(A)|^4)^{1/4} \leq \frac{4}{3} \min_{\lambda \in \mathbb{C}} \|A - \lambda I\|$$

holds. The corresponding result for the noncommutative standard deviation was proved in [5, Theorem 3.1].

3. Moment inequalities for the tracial state

We start this section with a moment estimate of matrices, determined by the tracial state.
3.1. Central moments for the tracial state. Let $A$ denote an $n \times n$ matrix and let us define its Schatten $p$-norm

$$
\|A\|_p = (\text{Tr} |A|^p)^{1/p},
$$

where $1 \leq p < \infty$ and $|A| = (A^*A)^{1/2}$ by definition. Then $\| \cdot \|_p$ is a norm on $M_n(\mathbb{C})$. We recall that the duality formula

$$
\|A\|_p = \max\{ |\text{Tr} [B^*A]| : |B| \leq 1 \}
$$

holds, where $b = 1/p + 1/q = 1$ (\cite[Theorem 7.1]{4}).

Let $X$ be a (real) discrete random variable on a finite set $\{1, \ldots, n\}$. We recall that the duality formula

$$
\mathbb{E}(|X - \mathbb{E}(X)|^p) = \mathbb{E}(\mathbb{E}(|X - \mathbb{E}(X)|^p))
$$

without loss of generality, we can assume that the center of this distribution is zero. Furthermore, we obtain with $b_p = \max_{0 \leq t \leq 1} t^p(1 - t) + t(1 - t)^p$.

**Lemma 2.** Let $A \in M_n(\mathbb{C})$ be a normal matrix and let $1 \leq p < \infty$. Then

$$
(\text{Tr} D |A - \text{Tr} DA|^p)^{1/p} \leq 2b_p^{1/p} \min_{\lambda \in \mathbb{C}} \|A - \lambda I_n\|
$$

holds, where $b_p$ denotes the largest $p$th central moment of the Bernoulli distribution.

**Proof.** By means of a diagonalization and the previous remarks, for any Hermitian matrix $H$ and density $D$ one has that

$$
(\text{Tr} [D |H - \text{Tr} DH|^p])^{1/p} \leq \frac{b_p}{p} \text{diam} \sigma(H) = 2b_p^{1/p} \min_{\lambda \in \mathbb{C}} \|H - \lambda I\|.
$$

For a normal $A$, $\min_{\lambda \in \mathbb{C}} \|A - \lambda I_n\|$ equals to the radius of the smallest enclosing circle of $\sigma(A)$. Without loss of generality, we can assume that the center of this circle is at the origin. Let us write $A$ as a diagonal matrix $A = \sum_{i=1}^n \lambda_i P_i$, where $\lambda_i$-s are the eigenvalues of $A$ and $P_i$-s are orthogonal projections. Set the diagonal matrices

$$
\tilde{A} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\
0 & \cdots & 0 \\
0 & \cdots & \lambda_n \end{bmatrix}
$$

and

$$
\tilde{H} = \begin{bmatrix} |\lambda_1| & \cdots & 0 \\
0 & \cdots & 0 \\
0 & \cdots & |\lambda_n| \end{bmatrix}
$$

in $M_{2n}(\mathbb{C})$. Clearly, $\min_{\lambda \in \mathbb{C}} \|A - \lambda I\| = \min_{\lambda \in \mathbb{C}} \|\tilde{H} - \lambda I\|$. Moreover, we obtain with $\tilde{D} = D \oplus 0 \in M_{2n}(\mathbb{C})$ that

$$
\text{Tr} [D |A - \text{Tr} DA|^p]^{1/p} = \text{Tr} [\tilde{D} |\tilde{A} - \text{Tr} \tilde{A}|^p]^{1/p}
$$

$$
= \text{Tr} \left[ \tilde{D} \left( \sum_{i=1}^n |\lambda_i|^p |P_i \oplus -P_i| - \text{Tr}[\tilde{D}(P_i \oplus -P_i)]|^p \right) \right]^{1/p}
$$

$$
= \text{Tr} \left[ \tilde{D} |\tilde{H} - \text{Tr} \tilde{H}|^p \right]^{1/p},
$$

and since $\tilde{H}$ is Hermitian

$$
\leq 2b_p^{1/p} \min_{\lambda \in \mathbb{C}} \|\tilde{H} - \lambda I\|
$$

$$
= 2b_p^{1/p} \min_{\lambda \in \mathbb{C}} \|A - \lambda I\|.
$$
which completes the proof. □

**Theorem 4.** Let $A \in M_n(\mathbb{C})$ and let $1 \leq p < \infty$. Then
\[
\left( \frac{1}{n} \text{Tr} \left| A - \frac{1}{n} \text{Tr} A \right|^p \right)^{1/p} \leq 2b_p^{1/p} \min_{\lambda \in \mathbb{C}} \| A - \lambda I_n \|
\]
holds, where $b_p$ denotes the largest $p$th central moment of the Bernoulli distribution.

**Proof.** Without loss of generality one can assume that $A$ is a contraction, i.e. $\| A \| = 1$. From the singular value decomposition of $A$, one can find two unitaries $U_1$ and $U_2$ such that
\[
A = \frac{1}{2} U_1 + \frac{1}{2} U_2
\]
(see [4, p. 62-63], for instance). The convexity of the Schatten $p$-norms and the central moment estimates of normal matrices in Lemma 2 imply that
\[
\left( \frac{1}{n} \text{Tr} \left| A - \frac{1}{n} \text{Tr} A \right|^p \right)^{1/p} \leq \frac{1}{2} \left( \frac{1}{n} \text{Tr} \left| U_1 - \frac{1}{n} \text{Tr} U_1 \right|^p \right)^{1/p} + \frac{1}{2} \left( \frac{1}{n} \text{Tr} \left| U_2 - \frac{1}{n} \text{Tr} U_2 \right|^p \right)^{1/p}
\]
\[
\leq b_p^{1/p} \| U_1 \| + b_p^{1/p} \| U_2 \|
\]
\[
= 2b_p^{1/p} \| A \|.
\]

Changing $A$ to $A - \lambda I$ we get the proof of the statement. □

An application of Hölder’s inequality gives that the function $p \mapsto b_p^{1/p}$ is monotone increasing on $\mathbb{R}_+$, hence $\lim_{p \to \infty} b_p^{1/p} = b_\infty = 1$. Similarly, $\| \cdot \|_p \to \| \cdot \|$ follows for the Schatten $p$-norms, if $p \to \infty$. Therefore, we obtain with (2) at hand that

**Corollary 2.** Let $A \in M_n(\mathbb{C})$ be a normal matrix. Then
\[
\frac{\sqrt{3}}{2} \left\| A - \frac{1}{n} \text{Tr} A \right\| \leq \text{spd}(A).
\]
Moreover, if $A$ is Hermitian then
\[
\left\| A - \frac{1}{n} \text{Tr} A \right\| \leq \text{spd}(A).
\]

3.2. **Central moments of matrix elements.** In this section, we make some estimates of the moments of matrix elements.

A conditional expectation operator $\mathbb{E}_B$ is an orthogonal projection from the matrix algebra $M_n(\mathbb{C})$, endowed with the Hilbert–Schmidt inner product $\langle A, B \rangle = \text{Tr} B^* A$, onto the $*$-subalgebra $\mathcal{B}$ (see [9, Section 4.3]). Here we collect a few basic properties of the conditional expectation operator. First, we recall that for any $A \in M_n(\mathbb{C})$,
\[
\text{Tr} A = \text{Tr} \mathbb{E}_B(A).
\]
Moreover, for each $B \in \mathcal{B}$, it follows the module properties
\[
\mathbb{E}_B(BA) = B \mathbb{E}_B(A) \quad \text{and} \quad \mathbb{E}_B(AB) = \mathbb{E}_B(A)B.
\]
A useful property here is that the conditional expectation operators can be uniformly approximated by the convex sums of the unitary conjugates of $A$. That is, for all $\varepsilon > 0$, there exist unitary operators $U_1, \ldots, U_m$ in the commutant algebra of $\mathcal{B}$ such that
\[
\| \mathbb{E}_B(A) - \sum_{j=1}^m \lambda_j U_j^* A U_j \| \leq \varepsilon,
\]
\[ \sum_{j=1}^m \lambda_j = 1 \] and \[ 0 < \lambda_1, \ldots, \lambda_m < 1. \] For a proof of these statements, we refer the reader to [9, Theorem 4.13].

The following proposition might be known in the literature, however, we were unable to find any reference.

**Proposition 1.** Let \( \mathcal{E} \) be a unital \(*\)-subalgebra of \( M_n(\mathbb{C}) \) and let \( \mathbb{E}_\mathcal{E} \) be the conditional expectation operator onto \( \mathcal{E} \). Then
\[ \text{Tr} |\mathbb{E}_\mathcal{E}(A)|^p \leq \text{Tr} |A|^p, \]
for every \( 1 \leq p < \infty \).

**Proof.** The duality formula tells us
\[ (\text{Tr} |\mathbb{E}_\mathcal{E}(A)|^p)^{1/p} = \max_{B \in M_n(\mathbb{C})} \{ \text{Tr} |\mathbb{E}_\mathcal{E}(A)|^p : \| B \|_q \leq 1 \} \]
holds where \( 1/p + 1/q = 1 \). Furthermore, for any \( \varepsilon > 0 \), there exist unitary matrices \( W_1, \ldots, W_m \) such that
\[ \| \mathbb{E}_\mathcal{E}(A) - \sum_{j=1}^m \lambda_j W_j^* A W_j \| \leq \varepsilon \]
and \( \sum_{j=1}^m \lambda_j = 1 \) (\( \lambda_j \geq 0 \)). Hence
\[ (\text{Tr} |\mathbb{E}_\mathcal{E}(A)|^p)^{1/p} \leq \max_{B \in M_n(\mathbb{C})} \left\{ \text{Tr} \left[ \sum_{j=1}^m \lambda_j B W_j^* A W_j \right] : \| B \|_q \leq 1 \right\} + O(\varepsilon) \]
and applying again the duality formula,
\[ = \sum_{j=1}^m \lambda_j (\text{Tr} |W_j^* A W_j|^p)^{1/p} + O(\varepsilon) \]
which is what we intended to have. \( \Box \)

Now we can prove

**Theorem 5.** Let \( e_1, \ldots, e_n \) be an orthonormal basis of \( \mathbb{C}^n \). For any \( A \in M_n(\mathbb{C}) \) and \( 1 \leq p < \infty \),
\[ \left( \frac{1}{n} \sum_{i=1}^n |\langle A e_i, e_i \rangle - \frac{1}{n} \sum_{j=1}^n \langle A e_j, e_j \rangle|^p \right)^{1/p} \leq \left( \frac{1}{n} \text{Tr} \left[ A - \frac{1}{n} \text{Tr} A \right]^p \right)^{1/p} \]
\[ \leq 2 b_p^{1/p} \min_{c \in \mathcal{E}} \| A - c I_n \|. \]

**Proof.** Let \( \mathcal{E} \) denote the commutative unital \(*\)-algebra generated by the orthogonal projections \( e_i^* e_i \) (\( 1 \leq i \leq n \)). From the previous proposition one obtains that
\[ \sum_{i=1}^n |\langle A e_i, e_i \rangle - \frac{1}{n} \sum_{j=1}^n \langle A e_j, e_j \rangle|^p = \text{Tr} \left| \mathbb{E}_\mathcal{E} \left( A - \frac{1}{n} \text{Tr} A \right) \right|^p \leq \text{Tr} \left| A - \frac{1}{n} \text{Tr} A \right|^p, \]
which is what we intended to have. \( \Box \)
Lastly, the next corollary gives some information about the spread of Hermitians and normal matrices in terms of the statistical dispersions of their diagonal elements.

**Corollary 3.** Let $1 \leq p < \infty$. Let $A \in M_n(\mathbb{C})$ be a normal matrix. Then

$$\left( \frac{1}{n} \sum_{i=1}^{n} |\langle Ae_i, e_i \rangle - \frac{1}{n} \sum_{j=1}^{n} \langle Ae_j, e_j \rangle|^p \right)^{1/p} \leq \frac{2}{\sqrt{3}} b_1^{1/p} \text{spd}(A).$$

Moreover, if $A$ is Hermitian then

$$\left( \frac{1}{n} \sum_{i=1}^{n} |\langle Ae_i, e_i \rangle - \frac{1}{n} \sum_{j=1}^{n} \langle Ae_j, e_j \rangle|^p \right)^{1/p} \leq b_1^{1/p} \text{spd}(A).$$

**References**

[1] K.M.R. Audenaert, Variance bounds, with an application to norm bounds for commutators, *Linear Algebra Appl.*, **432** (2010), 1126–1143.

[2] K.M.R. Audenaert, On the asymmetry of the relative entropy, *J. Math. Phys.*, **54**, 073506 (2013).

[3] E. R. Barnes, A. J. Hoffman, Bounds for the spectrum of normal matrices, *Linear Algebra Appl.*, **201** (1994) 79–90.

[4] R. Bhatia, *Positive Definite Matrices*, Princeton Univ. Press, 2007.

[5] R. Bhatia and R. Sharma, Some inequalities for positive linear maps, *Linear Algebra Appl.*, **436** (2012), 1562–1571.

[6] R. Bhatia and R. Sharma, Positive linear maps and spreads of matrices, *Amer. Math. Monthly*, **127** (2014), 619–624.

[7] R. Bhatia and P. Semrl, Orthogonality of matrices and some distance problem, *Linear Algebra Appl.*, **287** (1999), 77–85.

[8] T. Bhattacharyya and P. Grover, Characterization of Birkhoff–James orthogonality, *J. Math. Anal. Appl.*, **407** (2013), 350–358.

[9] E. Carlen, Trace inequalities and quantum entropy: An introductory course, *Contemp. Math.*, **529**, Amer. Math. Soc., 2010.

[10] M.M. Deza and M. Laurent, *Geometry of Cuts and Metrics*, Algorithm and Combinatorics 15, Springer, 1997.

[11] Halmos, P. R. and McLaughlin, J. E., Partial isometries, *Pacific J. Math.*, **13** (1963), 585–596.

[12] C.R. Johnson, R. Kumar and H. Wolkowicz, Lower bounds for the spread of a matrix, *Linear Algebra Appl.*, **71** (1985),161–173.

[13] Z. Léka, A note on extremal decompositions of covariances, to appear in *Rocky Mountain J. Math.*

[14] Z. Léka, A note on central moments in $C^*$-algebras, *J. Math. Inequal.*, **9** (2015), 165–175.

[15] C-K. Li and B-S. Tam, A note on extreme correlation matrices, *SIAM J. Matrix Anal. Appl.*, **15** (1994), 903–908.

[16] J.K. Merikoski and R. Kumar, Characterizations and lower bounds for the spread of a normal matrix, *Linear Algebra Appl.*, **364** (2003),13–31.

[17] L. Mirsky, The spread of a matrix, *Mathematics, 3* (1956), 127–130.

[18] L. Mirsky, Inequalities for normal and Hermitian matrices, *Duke Math. J.*, **24** (1957), 591–598.

[19] J. Nocedal and S. Wright, *Numerical Optimization*, Springer–Verlag, 1999.

[20] D. Petz and G. Tóth, Extremal properties of the variance and the quantum Fisher information, *Phys. Rev. A*, **87** (2013), 032324.

[21] G.K. Pedersen, *Analysis Now*, Springer–Verlag, 1989.

[22] H. Rademacher and O. Toeplitz, *The Enjoyment of Mathematics*, Dover, 1990.

[23] M.A. Rieffel, Standard deviation is a strongly Leibniz seminorm, *New York J. Math.*, **20** (2014), 35–56.

[24] R. Sharma, R. Kumar, R. Saini, G. Kapoor, Complementary upper bounds for fourth central moment with extensions and applications, arXiv preprint, arXiv:1503.03786, 2015.

[25] J.G. Stampfli, The norm of a derivation, *Pacific J. Math.*, **33** (1970), 737–747.