Trace inequalities for completely monotone functions and Bernstein functions

Koenraad M.R. Audenaert

Department of Mathematics, Royal Holloway, University of London, Egham
TW20 0EX, United Kingdom

Abstract

We prove a matrix trace inequality for completely monotone functions and for Bernstein functions. As special cases we obtain non-trivial trace inequalities for the power function $x \mapsto x^q$, which for certain values of $q$ complement McCarthy’s trace inequality and for others strengthen it.

Key words: Matrix Inequalities, Subadditivity, Superadditivity, Positive Semidefinite Matrix, Partitioned matrix

1991 MSC: 15A60

1 Introduction

Completely monotone functions play an important role in many branches of applied mathematics, and probability theory. They are defined as the Laplace transforms of measures on the half-line $[0, \infty)$. Closely related to these functions are the so-called Bernstein functions, which are the primitives of the positive completely monotone functions. Bernstein functions are important in probability theory as well, appearing for example in the study of random Markov processes.

In Section 3 of this paper we prove a matrix trace inequality for completely monotone functions, Bernstein functions, and primitive functions thereof. Our interest in these classes of functions does not stem from the applications just mentioned but from the fact that they contain the fractional power function $x \mapsto x^q$, for various ranges of $q$. 

Email address: koenraad.audenaert@rhul.ac.uk (Koenraad M.R. Audenaert).
By specialising our trace inequality to the fractional power function we obtain in Section 4 a number of non-trivial trace inequalities related to McCarthy’s trace inequality [6]. The latter states that the matrix function $A \mapsto \text{Tr} A^q$ is subadditive on the set of positive semidefinite matrices for $0 < q \leq 1$, and superadditive for $q \geq 1$:

\begin{align}
\text{Tr}(A + B)^q &\leq \text{Tr} A^q + \text{Tr} B^q, & 0 < q \leq 1 \\
\text{Tr}(A + B)^q &\geq \text{Tr} A^q + \text{Tr} B^q, & q \geq 1.
\end{align}

The inequalities we obtain are complementary to McCarthy’s in particular regions for $q$, and are strengthenings in others.

As a further application of these inequalities we obtain in Section 5 a simple proof of a norm inequality for partitioned positive semidefinite matrices that was first proven in [1] by other means.

2 Completely monotone functions and Bernstein functions

In this section, we collect a number of definitions and theorems about completely monotone functions and Bernstein functions that will be needed later on. For an in-depth treatment, including proofs and applications, we refer to the excellent monograph [8].

2.1 Completely monotone functions

Definition 1 A function $f : (0, \infty) \mapsto \mathbb{R}$ is completely monotone if it is infinitely differentiable, non-negative, and $(-1)^n f^{(n)}(x) \geq 0$ for $n = 1, 2, \ldots$ and $x > 0$.

An integral representation of completely monotone functions is provided by Bernstein’s theorem:

Theorem 1 (Bernstein) A function $f : (0, \infty) \mapsto \mathbb{R}$ is completely monotone if and only if it is the Laplace transform of a positive measure $\mu$ on $[0, \infty)$, i.e.

$$f(x) = \int_{[0, \infty)} \exp(-xt)\mu(dt) = a + \int_{(0, \infty)} \exp(-xt)\mu(dt),$$

where $a$ is given by $a = f(0^+)$. 

2
In addition, we will also define the bare completely monotone functions as those completely monotone functions for which \( a = f(0^+) = 0 \). We will denote the class of bare completely monotone functions by \( \text{CM}_0 \).

An important class of completely monotone functions are the negative power functions \( x \mapsto x^q, q < 0 \). That these functions are completely monotone follows from the integral representation

\[
x^q = \frac{1}{\Gamma(-q)} \int_0^\infty \exp(-xt)t^{-q-1}dt, \quad q < 0.
\]

This fits the representation of Bernstein’s theorem via \( \mu(dt) = t^{-q-1}dt/\Gamma(-q) \).

**Lemma 1** Any function \( f \in \text{CM}_0 \) is convex, monotonically decreasing and non-negative.

**Proof.** Obvious from the integral representation

\[
f(x) = \int_{(0,\infty)} \exp(-xt)\mu(dt),
\]

since \( \exp(-x) \) is convex, monotonically decreasing and non-negative. \( \square \)

2.2 Bernstein functions

**Definition 2** A function \( f : (0,\infty) \mapsto \mathbb{R} \) is a Bernstein function if it is infinitely differentiable, non-negative, and \((-1)^nf^{(n)}(x) \leq 0 \) for \( n = 1, 2, \ldots \) and \( x > 0 \).

Again, this class of functions admits an integral representation.

**Theorem 2** A function \( f : (0,\infty) \mapsto \mathbb{R} \) is a Bernstein function if and only if there exist \( a, b \geq 0 \) and a positive measure \( \mu(dt) \) on \((0,\infty)\) such that

\[
f(x) = ax + b + \int_{(0,\infty)}(1 - \exp(-xt))\mu(dt).
\]

In the probability theory literature, this representation is known as the Lévy-Khintchine representation. The constants \( a \) and \( b \) are given by the limits \( a = \lim_{x \to \infty} f(x)/x \) and \( b = f(0^+) \). In addition, we define the bare Bernstein functions as those Bernstein functions for which \( a = b = 0 \), and denote this class by \( \text{BF}_0 \).
The kernel function $1 - \exp(-xt)$ has leading order degree 1 for $x$ tending to 0, and degree 0 for $x$ tending to $\infty$. Therefore, in order for the integral in (5) to converge the measure $\mu$ must satisfy in integrability condition, such as the following one:

$$\int_{(0,\infty)} \min(1, t) \mu(dt) < \infty$$

It is easy to see that the derivative of every Bernstein function is completely monotone. Indeed, representation (5) can be obtained from (3) by simple integration. However, not every completely monotone function is the derivative of a Bernstein function, because of the extra positivity requirement for the latter. The function $x \mapsto x^q$, $q \leq -1$, for example, is completely monotone but its primitive $x^{q+1}/(q + 1)$ is negative for all $x > 0$ and therefore not a Bernstein function.

An important subclass of the Bernstein functions are the non-negative operator monotone functions. They contain the fractional power functions $x \mapsto x^q$, for $0 < q < 1$, as can be seen from the integral representation

$$x^q = \frac{q}{\Gamma(1-q)} \int_0^\infty (1 - \exp(-tx))t^{-q-1} dt, \quad 0 < q < 1. \quad (6)$$

**Lemma 2** Any function $f \in BF_0$ is concave, monotonically increasing, non-negative and $f(0) = 0$.

**Proof.** Obvious from the integral representation

$$f(x) = \int_{(0,\infty)} (1 - \exp(-tx))\mu(dt),$$

since $1 - \exp(-x)$ is concave, monotonically increasing and non-negative, and $1 - \exp(0) = 0$. \quad \Box

### 2.3 Integrals of Bernstein functions

We will also consider functions whose first (second) derivative is a bare Bernstein function. More precisely, given any bare Bernstein function $f \in BF_0$, we consider the functions

$$g(y) = \int_0^y dx \, f(x) \quad \text{and} \quad h(z) = \int_0^z dy \, g(y).$$
From representation (5) it follows that the function $g$ is represented by

$$g(x) = \int_{(0, \infty)} \left( \exp(-xt) - (1 - xt) \right) \frac{1}{t} \mu(dt),$$  \hspace{1cm} (7)$$

with $\mu(dt)$ the measure appearing in Theorem 2. The class of these functions will be denoted by $BF_1$. For $x$ tending to 0, the leading order of the kernel function $\exp(-xt) - (1 - xt)$ is of degree 2, while for $x$ tending to $\infty$ it is of degree 1. Convergence of the integral is therefore not affected by the extra factor $1/t$.

Likewise, the function $h$ is represented by

$$h(x) = \int_{(0, \infty)} \left( (1 - xt + x^2t^2/2) - \exp(-xt) \right) \frac{1}{t^2} \mu(dt).$$  \hspace{1cm} (8)$$

The class of these functions will be denoted by $BF_2$. Here, the kernel function $\exp(-xt) - (1 - xt + x^2t^2/2)$ has leading order degree 3 for $x$ tending to 0, so that convergence is again not affected by the factor $1/t^2$.

Continuing in this way, we can inductively define the classes $BF_k$, $k \in \mathbb{N}$, as the classes of $k$-fold integrals of bare Bernstein functions. That is, $f \in BF_k$ if and only if there is a function $g \in BF(k - 1)$ such that

$$f(x) = \int_0^x g(t) dt.$$  \hspace{1cm} (9)$$

It is easy to see that such functions have the integral representation

$$f(x) = \int_{(0, \infty)} (-1)^{k+1} \left( \exp(-xt) - \sum_{j=0}^{k} (-xt)^j / j! \right) \frac{1}{t^k} \mu(dt).$$  \hspace{1cm} (10)$$

For $x$ tending to 0, the leading order of the kernel function is of degree $k + 1$, while for $x$ tending to $\infty$ it is of degree $k$.

The function $x \mapsto x^q$ is in $BF_k$ for $k < q < k + 1$.

**Lemma 3** Any function $f \in BF_k$, $k \geq 1$, is convex, monotonically increasing, non-negative, and $f(0) = 0$.

**Proof.** For $k = 1$, this is obvious from integral representation (7), since the function $x \mapsto \exp(-x) - 1 + x$ is convex, monotonically increasing, non-negative and $\exp(0) - 1 + 0 = 0$. 
For $k > 1$, this follows inductively from the defining integral (9), from which we get $f \geq 0$, $f(0) = 0$, $f' = g$ and $f'' = g'$, for $g \in \text{BF}(k-1)$. By the induction hypothesis, $g$ is non-negative and increasing, hence $f' \geq 0$ and $f'' \geq 0$. □

3 Main Results

As stated in the Introduction, we will exploit the integral representations of functions in $\text{CM}_0$ and $\text{BF}_k$ to extend inequalities for the exponential function to those classes of functions.

3.1 Scalar inequalities

In this section, we restrict ourselves to the scalar case, leaving the matrix case for the next section.

The following lemma concerns (scalar) sub- and superadditivity. Recall that a real-valued function $g$ is subadditive on $I$ if and only if $\forall x, y \in I : g(x + y) \leq g(x) + g(y)$; it is superadditive on $I$ if and only if $\forall x, y \in I : g(x + y) \geq g(x) + g(y)$.

**Lemma 4** Let $g$ be a function $g : [0, \infty) \to \mathbb{R}$. If $g \in \text{CM}_0 \cup \text{BF}_0$ then $g$ is subadditive on $[0, \infty)$. If $g \in \text{BF}_k$, $k \geq 1$, then $g$ is superadditive on $[0, \infty)$.

**Proof.** For $x, y \geq 0$, we have $e^x + e^y \geq 2$, hence $e^{-x-y} \leq 2e^{-x-y} \leq e^{-y} + e^{-x}$. Thus, the function $e^{-xt}$ is subadditive for all $t \geq 0$. Therefore, all functions in $\text{CM}_0$ are subadditive too.

The same is true for functions in $\text{BF}_0$, as can be seen from subadditivity of the function $1 - e^{-x}$. The latter follows from positivity of $(1 - e^{-x})(1 - e^{-y})$ for $x, y \geq 0$.

Superadditivity of functions in $\text{BF}_1$ follows from superadditivity of the function $x \mapsto e^{-x} - 1 + x$, which in turn follows from subadditivity of $1 - e^{-x}$ and additivity of $x \mapsto x$.

Superadditivity of functions in $\text{BF}_k$ for $k > 1$ follows inductively from superadditivity of functions in $\text{BF}(k-1)$. By definition, any function $h \in \text{BF}_k$ is given by the integral $h(x) = \int_0^x dtg(t)$ of a function $g \in \text{BF}(k-1)$. Therefore,

$$h(x + y) - h(x) - h(y) = \int_x^{x+y} dtg(t) - \int_0^y dtg(t)$$
\[
\begin{align*}
&= \int_0^y dt (g(t + x) - g(t)) \\
&\geq \int_0^y dt g(x) = yg(x) \geq 0.
\end{align*}
\]

In the last line we exploited superadditivity of \( g \) in the form \( g(t + x) - g(t) \geq g(x) \). \( \square \)

To obtain inequalities that complement the subadditivity (superadditivity) inequalities of the previous lemma, we need a not very well-known property of the exponential function. This property relies on the so-called geometrical concavity of the function \( 1 - \exp(-x) \), a concept that can be traced back to Montel [7].

**Definition 3** A function \( f(x) : \mathbb{R}^+ \to \mathbb{R}^+ \) is geometrically concave iff for all \( x, y \geq 0 \),

\[
f(\sqrt{xy}) \geq \sqrt{f(x)f(y)}.
\]

**Lemma 5** The function \( f(x) = 1 - \exp(-x) \) is geometrically concave.

*Proof.* Geometrical concavity of \( f \) is equivalent to concavity of \( g(x) = \log(1 - \exp(-\exp x)) \). The second order derivative of \( g \) is

\[
g''(x) = \frac{\exp x}{(\exp \exp x - 1)^2}(\exp \exp x - \exp(x + \exp x) - 1).
\]

The factor that determines the sign is clearly \( \exp \exp x - \exp(x + \exp x) - 1 \), which is non-positive. Indeed, substituting \( a = \exp x \), and noting that \( \exp(-a) \geq 1 - a \), yields

\[
\exp \exp x - \exp(x + \exp x) - 1 = \exp a - a \exp a - 1 \\
= (1 - a) \exp a - 1 \\
\leq \exp(-a) \exp a - 1 = 0.
\]

\( \square \)

This property of the function \( 1 - \exp(x) \) translates to a property of the exponential function. The connection to subadditivity (superadditivity) is immediate.

**Lemma 6** The inequality

\[
g(a + b) - g(a) - g(b) \leq g(2\sqrt{ab}) - 2g(\sqrt{ab}) \tag{11}
\]
holds for any $a, b \geq 0$ when $g(x) = \exp(-x)$. If $g(x)$ is a quadratic polynomial then it holds with equality.

Proof. That equality holds for quadratic polynomials is immediate.

Geometrical concavity of $f(x) = 1 - \exp(-x)$ amounts to the inequality

$$(1 - \exp(-a))(1 - \exp(-b)) \leq (1 - \exp(-\sqrt{ab}))^2.$$ 

Hence, for all $a, b \geq 0$,

$$\exp(-(a+b)) - \exp(-a) - \exp(-b) \leq \exp(-2\sqrt{ab}) - 2\exp(-\sqrt{ab}).$$

$\square$

This inequality can be extended to completely monotone functions and Bernstein functions, using their integral representations.

**Theorem 3** Let $g$ be a function $g : [0, \infty) \to \mathbb{R}$. For $a, b \geq 0$,

$$g(a + b) - g(a) - g(b) \leq g(2\sqrt{ab}) - 2g(\sqrt{ab}),$$

(12)

holds if $g \in \text{CM}_0 \cup \text{BF}_1$. The inequality holds in the reversed sense if $g \in \text{BF}_0 \cup \text{BF}_2$. Equality holds when $g$ is a quadratic polynomial.

Proof. This follows immediately from Lemma 6 and from the integral representations of functions in $\text{CM}_0$, $\text{BF}_0$, $\text{BF}_1$ and $\text{BF}_2$. The sign with which $\exp(-tx)$ occurs in these representations determines whether the inequality holds in the stated sense or in the reversed sense. $\square$

It will be shown below that this inequality does not hold for functions in $\text{BF}_k$, $k > 2$.

The results from Lemma 6 and Theorem 3 can be summarised by the following inequalities:

$$g(x+y) - g(x) - g(y) \leq g(2\sqrt{xy}) - 2g(\sqrt{xy}) \leq 0, \quad g \in \text{CM}_0$$

$$0 \geq g(x+y) - g(x) - g(y) \geq g(2\sqrt{xy}) - 2g(\sqrt{xy}) \geq 0, \quad g \in \text{BF}_0$$

$$0 \leq g(x+y) - g(x) - g(y) \leq g(2\sqrt{xy}) - 2g(\sqrt{xy}) \leq 0, \quad g \in \text{BF}_1$$

$$g(x+y) - g(x) - g(y) \geq g(2\sqrt{xy}) - 2g(\sqrt{xy}) \geq 0, \quad g \in \text{BF}_2.$$ 

Thus, Theorem 3 is a stronger statement than subadditivity (superadditivity) for $g \in \text{CM}_0$ ($g \in \text{BF}_2$), while for $g \in \text{BF}_0$ ($g \in \text{BF}_1$) it provides a complementary inequality to subadditivity (superadditivity).
3.2 A matrix trace inequality

Theorem 4 is easy to extend to the matrix case via a simple application of the Golden-Thompson theorem, yielding our main trace inequality:

**Theorem 4** Let $A$ and $B$ be $d$-dimensional positive semidefinite matrices, with spectral decompositions $A = \sum_k a_k A_k$ and $B = \sum_k b_k B_k$, where $a_k, b_k \geq 0$ and $\{A_k\}$ and $\{B_k\}$ are two complete sets of mutually orthogonal projectors. Let $g$ be a function, $g : [0, \infty) \rightarrow \mathbb{R}$. The inequality

$$\text{Tr}(g(A + B) - g(A) - g(B)) \leq \sum_{k,l} (g(2\sqrt{a_kb_l}) - 2g(\sqrt{a_kb_l})) \text{Tr} A_k B_l$$

holds if $g \in BF_1$ and (for $A, B > 0$) if $g \in CM_0$.

The inequality holds in the reversed sense if $g \in BF_0 \cup BF_2$.

Equality holds if $g$ is a quadratic polynomial.

**Proof.** It is easy to check that the inequality reduces to an equality for $g(x) = 1$, $g(x) = x$ and $g(x) = x^2$. For $g(x) = 1$, the LHS is $-\text{Tr} I$, and the RHS is $-\sum_{k,l} \text{Tr} A_k B_l = -\text{Tr}(\sum_k A_k)(\sum_l B_l) = -\text{Tr} I$, due to completeness of the sets $\{A_k\}$ and $\{B_k\}$.

For $g(x) = x$, the LHS and RHS are both 0, and for $g(x) = x^2$ the LHS is $2 \text{Tr} AB$ and the RHS is $\sum_{k,l} 2a_kb_l \text{Tr} A_k B_l = 2 \text{Tr}(\sum_k a_k A_k)(\sum_l b_l B_l) = 2 \text{Tr} AB$.

To prove the main statement of the theorem, we look again at the exponential function. The Golden-Thompson theorem states $\text{Tr} \exp(A + B) \leq \text{Tr} \exp A \exp B$, for any two Hermitian matrices $A$ and $B$. In particular, we have, for any $t$,

$$\text{Tr} \exp(-(A + B)t) \leq \text{Tr} \exp(-At) \exp(-Bt)$$
$$= \sum_{k,l} \exp(-a_k t) \exp(-b_l t) \text{Tr} A_k B_l$$
$$= \sum_{k,l} \exp(-(a_k + b_l)t) \text{Tr} A_k B_l.$$

Also, for any function $g$,

$$\text{Tr} g(A) = \sum_k g(a_k) \text{Tr} A_k = \sum_{k,l} g(a_k) \text{Tr} A_k B_l,$$
$$\text{Tr} g(B) = \sum_l g(b_l) \text{Tr} B_l = \sum_{k,l} g(b_l) \text{Tr} A_k B_l.$$
Therefore, for \( g(x) = \exp(-xt) \),

\[
\text{Tr}(g(A + B) - g(A) - g(B)) \leq \sum_{k,l} (g(a_k + b_l) - g(a_k) - g(b_l)) \text{Tr} A_k B_l.
\]

Using the same reasoning as in the proof of Theorem 3, we then find that this inequality holds for all \( g \in BF_1 \) and (for \( A, B > 0 \)) \( g \in CM_0 \), and in the reversed sense for all \( g \in BF_0 \cup BF_2 \).

Combining this with the scalar inequality of Theorem 3 applied to \( g(a_k + b_l) - g(a_k) - g(b_l) \) yields the stated inequalities. □

For completeness, we also state the extension of Lemma 4 to the matrix case. Let \( P \) denote the set of positive semidefinite matrices.

**Lemma 7** Let \( g \) be a function \( g : [0, \infty) \to \mathbb{R} \), and extended to \( P \) in the usual way. If \( g \in CM_0 \cup BF_0 \) then the function \( A \mapsto \text{Tr}g(A) \) is subadditive on \( P \), i.e. for all \( A, B \geq 0 \),

\[
\text{Tr}g(A + B) \leq \text{Tr}(g(A) + g(B)).
\]

If \( g \in BF_k, k \geq 1 \), then \( A \mapsto \text{Tr}g(A) \) is superadditive on \( P \), i.e. for all \( A, B \geq 0 \),

\[
\text{Tr}g(A + B) \geq \text{Tr}(g(A) + g(B)).
\]

**Proof.** To show the statement for \( g \in CM_0 \), we only need to show it for \( g(x) = \exp(-x) \), i.e. that \( \text{Tr} e^{-A-B} \leq \text{Tr} e^{-A} + e^{-B} \). By the Golden-Thompson inequality, we have \( \text{Tr} e^{-A-B} \leq \text{Tr} e^{-A} e^{-B} \). Since \( B \geq 0 \), we also have \( e^{-B} \leq I \). Thus, \( \text{Tr} e^{-A} e^{-B} \leq \text{Tr} e^{-A} \leq \text{Tr} e^{-A} + e^{-B} \), so that indeed \( \text{Tr} e^{-A-B} \leq \text{Tr} e^{-A} + e^{-B} \).

Next, to cover the case \( g \in BF_0 \), we just note that \( \text{Tr}g(A + B) \leq \text{Tr}(g(A) + g(B)) \) is a special case of Bourin and Uchiyama’s norm subadditivity inequality [4]. Indeed, by Lemma 2 functions in \( BF_0 \) satisfy the conditions of their Theorem.

Likewise, to cover the case \( g \in BF_k, k \geq 1 \), we note that \( \text{Tr}g(A + B) \geq \text{Tr}(g(A) + g(B)) \) is a special case of Kosem’s norm superadditivity inequality [5]. By Lemma 3 functions in \( BF_k, k \geq 1 \), satisfy the conditions of his Theorem. □
4 Inequalities for the power function

The inequality of Theorem 4 achieves its most elegant form when \( g(x) \) is the fractional power function \( x \mapsto x^q \).

**Corollary 1** For \( A, B \geq 0 \) and \( 0 < q \leq 1 \) or \( 2 \leq q \leq 3 \),

\[
\text{Tr}(A + B)^q - \text{Tr}(A^q + B^q) \geq (2^q - 2) \text{Tr} A^{q/2} B^{q/2}.
\] (13)

For \( q < 0 \) (in which case we require \( A, B > 0 \)) and \( 1 \leq q \leq 2 \), the inequality holds in the reversed sense.

**Proof.** For \( g(x) = x^q \), the RHS of the inequality of Theorem 4 simplifies to

\[
\sum_{k,l} \left( 2 \sqrt{a_k b_l} \right)^q - 2 \left( \sqrt{a_k b_l} \right)^q \text{Tr} A_k B_l
\]

\[
= (2^q - 2) \sum_{k,l} a_k^{q/2} b_l^{q/2} \text{Tr} A_k B_l
\]

\[
= (2^q - 2) \text{Tr} A^{q/2} B^{q/2}.
\]

The corollary then follows by recalling that \( g(x) = x^q \) is in \( \text{CM}_0 \) for \( q < 0 \), and in \( \text{BF}_k \) for \( k < q < k + 1 \). \( \square \)

For \( q > 3 \) the inequality does no longer hold in general. Indeed, for scalars (e.g. \( A = 1, B = 2 \)) the inequality holds in the stated sense while, for example, with the choice \( A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, B = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \), the inequality holds in the reversed sense for all \( q > 3 \); the LHS is \((1 + \sqrt{2}/2)^q + (1 - \sqrt{2}/2)^q - 2 \) and the RHS \((2^q - 2)/2 \). More generally, this shows that the inequality of Theorem 4 does not hold for functions in \( \text{BF}_k \) for \( k > 2 \).

As in the scalar case, one sees that the inequality of Corollary 1 is stronger than McCarthy’s for \( 2 \leq q \leq 3 \). The corollary also implies that subadditivity holds for \( q < 0 \) too. On the other hand, for other parameter ranges the corollary complements McCarthy’s inequalities by providing a lower bound on \( \text{Tr}(A + B)^q - \text{Tr} A^q - \text{Tr} B^q \) for \( 0 < q \leq 1 \) and an upper bound for \( 1 \leq q \leq 2 \).

By replacing \( A \) and \( B \) by \( A^{1/q} \) and \( B^{1/q} \), and \( q \) by \( 1/p \), Corollary 1 can be reformulated as an inequality for the \( p \)-power means \( [2] \).
Corollary 2 For $A, B \geq 0$ and $p \geq 1$,

$$\text{Tr} \left( \frac{A^p + B^p}{2} \right)^{1/p} \geq 2^{1-1/p} \text{Tr} \frac{A + B}{2} + (1 - 2^{1-1/p}) \text{Tr} A^{1/2} B^{1/2}.$$  

Applying the Araki-Lieb-Thirring inequality, we obtain a closely related trace inequality, where the expression $\text{Tr} A^{q/2} B^{q/2}$ is replaced by $\text{Tr}(A^{1/2} B A^{1/2})^{q/2}$.

Corollary 3 For $A, B \geq 0$ and $0 < q \leq 1$ or $2 \leq q \leq 3$,

$$\text{Tr}(A + B)^q - \text{Tr}(A^q + B^q) \geq (2^q - 2) \text{Tr}(A^{1/2} B A^{1/2})^{q/2}.$$  \hspace{1cm} (14)

For $q \leq -2$ (in which case we require $A, B > 0$) and $1 \leq q \leq 2$, the inequality holds in the reversed sense.

Proof. As always, we require $A, B > 0$ for negative $q$.

By the Araki-Lieb-Thirring inequality for $A, B \geq 0$, we have, for $0 < q \leq 2$ and for $-2 \leq q < 0$,

$$\text{Tr} A^{q/2} B^{q/2} \leq \text{Tr}(A^{1/2} B A^{1/2})^{q/2},$$

while for $q \geq 2$ and for $q \leq -2$ the reversed inequality holds; In addition, $2^q - 2$ is positive for $q > 1$ and negative for $q < 1$. Thus, the corollary follows from Corollary 1 for all values of $q$ for which the latter holds (i.e. $q \leq 3$), except for $-2 < q < 0$. \Box

In contrast to inequality (13), which does not hold for $q > 3$, we have numerical evidence in support of the following conjecture:

Conjecture 1 Inequality (14) also holds for $q > 3$ and, in the reversed sense, for $-2 < q < 0$.

Additional evidence is given by:

Proposition 1 Inequality (14) holds for $q = 4$.

Proof. For $q = 4$, the left-hand side of (14) is

$$\text{Tr}(A + B)^4 - \text{Tr} A^4 - \text{Tr} B^4 = 4 \text{Tr}(A^3 B + A^2 B^2 + A B^3) + 2 \text{Tr}(AB)^2,$$

while the right-hand side is

$$(2^4 - 2) \text{Tr}(A^{1/2} B A^{1/2})^2 = 12 \text{Tr}(AB)^2.$$
Now, \( \text{Tr}(A^3B + AB^3) \geq 2 \text{Tr} A^2B^2 \). This follows from the scalar inequality \( x^3y + xy^3 = xy(x^2 + y^2) \geq xy(2xy) = 2x^2y^2 \) applied to the coefficients in the spectral decomposition of \( \text{Tr}(A^3B + AB^3) = \sum_{j,k}(a^3_j b_k + a_j b^3_k) \text{Tr} A_j B_k \) (in terms of the spectral decompositions \( A = \sum_j a_j A_j \) and \( B = \sum_k b_k B_k \)). Note that \( \text{Tr} A_j B_k \geq 0 \).

Also, by the Lieb-Thirring inequality, \( \text{Tr} A^2B^2 \geq \text{Tr}(AB)^2 \). Thus, indeed we find

\[
4 \text{Tr}(A^3B + A^2B^2 + AB^3) + 2 \text{Tr}(AB)^2 \geq 12 \text{Tr}(AB)^2
\]

\( \square \)

5 A new proof for a norm compression inequality

In this section, we consider an inequality that relates the Schatten \( q \)-norm, \( \|X\|_q := (\text{Tr}|X|^q)^{1/q} \), of a partitioned positive semidefinite matrix to the Schatten \( q \)-norms of its blocks. In particular, we compare it to the \( q \)-norm of the matrix that is obtained by replacing each block by its \( q \)-norm. An inequality of this type is sometimes called a norm compression inequality. The specific inequality presented here has first appeared in our [1], but had a long and intricate proof, and only for the case \( 1 \leq q \leq 2 \). Here we show how to prove it in a very simple way, using Corollary 3 from the previous section. Moreover, the proof given here extends the result to include the parameter range \( 0 < q < 1 \) and \( 2 < q \leq 3 \).

In the following, we consider the positive semidefinite block matrix

\[
A = \begin{pmatrix} B & C^* \\ C & D \end{pmatrix},
\]

where \( B \) and \( D \) are square blocks, and define the block norms \( \beta = \|B\|_q \), \( \delta = \|D\|_q \) and \( \gamma = \|C\|_q \).

An equivalent form of the inequalities in Corollary 3 is:

**Corollary 4** Let \( D > 0 \) and let \( C \) be any matrix. For \( q \leq -2 \) and for \( 1 \leq q \leq 2 \),

\[
\text{Tr} \left( \begin{pmatrix} C^* D^{-1} C & C^* \\ C & D \end{pmatrix} \right)^q - \text{Tr} \left( \begin{pmatrix} C^* D^{-1} C & 0 \\ 0 & D \end{pmatrix} \right)^q \leq (2^q - 2) \text{Tr}|C|^q.
\]

For \( 0 < q \leq 1 \) and for \( 2 \leq q \leq 3 \) the reversed inequality holds.
Proof. In the inequalities of Corollary 3 set \( A = D^{-1/2} C C^* D^{-1/2} \) and \( B = D \). Then
\[
A + B = (D^{-1/2} C \ D^{1/2}) \begin{pmatrix} C^* D^{-1/2} \\ D^{1/2} \end{pmatrix},
\]
which has the same non-zero eigenvalues as the block matrix
\[
Z = \begin{pmatrix} C^* D^{-1/2} \\ D^{1/2} \end{pmatrix} (D^{-1/2} C \ D^{1/2}) = \begin{pmatrix} C^* D^{-1} C \ C^* \\ C \ D \end{pmatrix}.
\]
Therefore \( \text{Tr}(A + B)^q \) is equal to \( \text{Tr} Z^q \).

Furthermore, \( A = D^{-1/2} C C^* D^{-1/2} \) and \( C^* D^{-1/2} D^{1/2} C = C^* D^{-1} C \) are unitarily equivalent, so that \( \text{Tr} A^q = \text{Tr}(C^* D^{-1} C)^q \).

Finally, because \( A = D^{-1/2} C C^* D^{-1/2} \) there exists a unitary matrix \( U \) such that \( C = D^{1/2} A^{1/2} U = B^{1/2} A^{1/2} U \). Thus, \( (A^{1/2} B A^{1/2})^{1/2} = U(C^* C)^{1/2} U^* = U|C| U^* \), whence
\[
\text{Tr}(A^{1/2} B A^{1/2})^{q/2} = \text{Tr} |C|^q.
\]
Substituting everything in the inequality of Corollary 3 yields the stated inequality.

Conversely, the inequality of Corollary 3 is obtained from the stated inequality by putting \( C = B^{1/2} A^{1/2} \) and \( D = B \). \( \square \)

We now present a new and much easier proof of the main result in [1]; moreover, we extend its validity to include the range \( 0 < q \leq 1 \) and \( 2 \leq q \leq 3 \).

**Theorem 5** Let \( A \) be a positive semi-definite block matrix, partitioned as above. Then for \( 1 \leq q \leq 2 \), with \( \beta = ||B||_q \), \( \delta = ||D||_q \) and \( \gamma = ||C||_q \),
\[
\text{Tr} A^q \leq (2^q - 2) \gamma^q + \beta^q + \delta^q. \tag{15}
\]

For \( 0 < q \leq 1 \) and for \( 2 \leq q \leq 3 \), the reversed inequality holds.

Proof. As already noted in [1], it is enough to consider positive \( C \). In that case the inequality (15) can be rephrased as follows:
\[
\text{Tr} \begin{pmatrix} B & C \\ C & D \end{pmatrix}^q - \text{Tr} \begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix}^q \leq \text{Tr} \begin{pmatrix} C & C \\ C & C \end{pmatrix}^q - \text{Tr} \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}^q. \tag{16}
\]
Consider first the cases $0 < q \leq 1$ and $2 \leq q \leq 3$. Let us calculate the minimum value of the left-hand side of (16) over all allowed $B$. The constraint on $B$, originating from the requirement $A \geq 0$, is $B \geq CD^{-1}C$. We will show that the minimum over $B$ is obtained in $B = B_0 := CD^{-1}C$. Let us thereto put $B = B_0 + t\Delta$, with $\Delta \geq 0$, and define

$$f(t) := \text{Tr} \left( B_0 + t\Delta C \right)^q - \text{Tr} \left( B_0 + t\Delta D \right)^q.$$

The Fréchet derivative [3] of $f$ is given by

$$f'(t) = q \text{Tr} \left[ \left( \begin{pmatrix} B C \\ C D \end{pmatrix}^{q-1} - \begin{pmatrix} B 0 \\ 0 0 \end{pmatrix} \right) \begin{pmatrix} \Delta 0 \\ 0 0 \end{pmatrix} \right].$$

Introducing the projector $P = I \oplus 0$, we can write

$$f'(t) = q \text{Tr} \left[ \left( P \begin{pmatrix} B C \\ C D \end{pmatrix}^{q-1} P - \begin{pmatrix} P B C P \end{pmatrix}^{q-1} \right) \begin{pmatrix} \Delta 0 \\ 0 0 \end{pmatrix} \right].$$

For $0 < q \leq 1$ and for $2 \leq q \leq 3$, the function $x \mapsto g(x) = x^{q-1}$ is operator convex on $(0, +\infty)$. Therefore (3, Exercise V.2.2 applied to a compression to the upper left block)

$$P \begin{pmatrix} B C \\ C D \end{pmatrix}^{q-1} P \geq \begin{pmatrix} P B C P \end{pmatrix}^{q-1}.$$  

This shows that $f'(t) \geq 0$ and that $f(t)$ is indeed minimal in $t = 0$. Therefore, we can restrict to $B = CD^{-1}C$. The theorem now follows immediately from Corollary 4.

For the case $1 \leq q \leq 2$, we proceed in exactly the same way. The function $x \mapsto g(x) = x^{q-1}$ is now operator concave on $[0, +\infty)$, with $g(0) = 0$. Therefore (3, Theorem V.2.3) we now have

$$P \begin{pmatrix} B C \\ C D \end{pmatrix}^{q-1} P \leq \begin{pmatrix} P B C P \end{pmatrix}^{q-1}.$$  

Hence, $f(t)$ is maximal in $t = 0$. \(\Box\)

We believe that the reversed inequality holds for $q > 3$, but we haven’t been able to prove this yet. The proof given in [1], using a duality argument, is incorrect.
References

[1] K.M.R. Audenaert, “A norm compression inequality for block partitioned positive semidefinite matrices,” Lin. Alg. Appl. 413, 155–176 (2006).

[2] K.V. Bhagwat and A. Subramanian, “Inequalities between means of positive operators”, Math. Proc. Camb. Phil. Soc. 83, 393–401 (1978).

[3] R. Bhatia, Matrix Analysis, Springer, Heidelberg (1997).

[4] J.-C. Bourin, M. Uchiyama, “A matrix subadditivity inequality for $f(A + B)$ and $f(A) + f(B)$,” Lin. Alg. Appl. 423, 512–518 (2007).

[5] T. Kosem, “Inequalities between $\|f(A + B)\|$ and $\|f(A) + f(B)\|$,” Lin. Alg. Appl. 418, 153–160 (2006).

[6] C.A. McCarthy, “$c_p$”, Israel J. Math. 5, 249–271 (1967).

[7] P. Montel, “Sur les fonctions convexes et les fonctions sousharmoniques”, J. Math. Pures Appl. Neuvième Série, 7, 29–60 (1928).

[8] R.L. Schilling, R. Song and Z. Vondracek, Bernstein functions, theory and applications, Studies in Mathematics vol. 37, Walter de Gruyter, Berlin (2010).