PHD THESIS

QUANTUM ENTROPIES, RELATIVE ENTROPIES, AND RELATED PRESERVER PROBLEMS

DÁNIEL VIROSZTEK

DEPARTMENT OF ANALYSIS
BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
HUNGARY

SUPERVISOR: PROF. DÉNES PETZ

2016

Contents

Index of Notation 2

Chapter 1. Introduction 3
1. C^*-algebras, von Neumann algebras 4
2. Tensor product 9
3. Continuous functional calculus 11
4. Commutative von Neumann algebras 13

Chapter 2. Quantum variances, generalized entropies and relative entopies 15
1. Decomposable quantum variances 16
2. Generalizations of the strong subadditivity inequality 21
3. Bregman divergences and their use 34

Chapter 3. Preserver problems 51
1. Maps preserving Bregman and Jensen divergences 53
2. Jordan triple endomorphisms of the positive cone 66
3. An application: the endomorphisms of the Einstein gyrogroup 78

Chapter 4. Summary 83
Ackowledgement 84

Bibliography 85

Index of Notation

$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$\mathbb{Q}$	the set of rational numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$x \in X$	x is an element of the set X
$A \times B$	the Cartesian product of the sets A and B
A^n	the nth Cartesian power of the set A
$\mathscr{H}$	a complex Hilbert space
$\langle x, y \rangle$	the inner product of the elements x and y of the Hilbert space $\mathscr{H}$ — we use the convention that the inner product is linear in the second variable and conjugate-linear in the first variable.
$\mathscr{B}(\mathscr{H})$	the set of bounded linear operators on $\mathscr{H}$
$\mathscr{B}^{sa}(\mathscr{H})$	the set of self-adjoint linear operators on $\mathscr{H}$
$\mathscr{B}^{+}(\mathscr{H})$	the set of positive semidefinite linear operators on $\mathscr{H}$
$\mathscr{B}^{++}(\mathscr{H})$	the set of positive definite linear operators on $\mathscr{H}$
ran(.)	the range of a linear operator
ker(.)	the kernel of a linear operator
$\mathbf{M}_n$	the set of $n \times n$ complex matrices
$\mathbf{M}_n^{sa}$	the set of $n \times n$ self-adjoint complex matrices
$\mathbf{M}_n^{+}$	the set of $n \times n$ positive semidefinite complex matrices
$\mathbf{M}_n^{++}$	the set of $n \times n$ positive definite complex matrices
$[A]_{i,j}$	the entry in the i-th row and j-th column of a matrix A
A^{tr}	the transpose of the matrix A
$I_{\mathscr{A}}$	the identity element of the unital algebra $\mathscr{A}$

CHAPTER 1

Introduction

The classical work [**44**] of *Andrey Nikolaevich Kolmogorov* laid the foundations of probability theory in 1933. In Kolmogorov's approach, the basic concept of probability theory is the *probability space*. A probability space is a triplet $(X, \mathscr{A}, \mathbf{P})$, where X is an arbitrary set, $\mathscr{A} \subseteq P(X)$ is a σ-algebra — $P(X)$ denotes the power set of X — and $\mathbf{P}$ is a finite measure on $\mathscr{A}$ which is normalized, that is, $\mathbf{P}(X) = 1$. This means that a probability space is nothing else but a measure space with total measure one, so one may consider probability theory as a branch of measure theory. On the other hand, probability theory is a richer structure than measure theory in the sense that several measure theoretical notions gain intuitive meanings from the viewpoint of a probability theorist. Without the requirement of generality, let us mention some of the intuitions which are associated with the notions of measure theory. The most basic concept is that the *measurable sets* — that is, the elements of the σ-algebra $\mathscr{A}$ — are considered to be *events*. A *measurable function* $f : (X, \mathscr{A}) \to (\mathbb{K}, \mathscr{B})$ is called a real/complex *random variable* if $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, respectively. Therefore, the *Lebesgue integral* $\int_X f \mathrm{d}\mathbf{P}$ of the measurable function f is called the *expected value* — if it exists. As $\mathbf{P}$ is a finite measure, it is quite easy to guarantee the existence of the integral of a measurable function. If f is essentially bounded, that is, $\mathbf{P}\left(\{x \in X : \left|f(x)\right| > K\}\right) = 0$ for some $K > 0$, then f is integrable, moreover, any power of f is integrable. This latter fact is remarkable as the integral $\int_X f^k \mathrm{d}\mathbf{P}$ is called the kth *moment* of the random variable f and plays an important role in probability theory. Let us denote by $L^\infty(X, \mathscr{A}, \mathbf{P})$ the set of essentially bounded measurable complex valued functions on the probability space $(X, \mathscr{A}, \mathbf{P})$. Let us introduce the notation

$$L^2(X, \mathscr{A}, \mathbf{P}) = \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is measurable and } \int_X |f|^2 \mathrm{d}\mathbf{P} < \infty \right\},$$

as well. Clearly, $L^2(X, \mathscr{A}, \mathbf{P})$ is a Hilbert space with the inner product $\langle f, g \rangle = \int_X \overline{f} g \mathrm{d}\mathbf{P}$. Every bounded measurable function $f : X \mapsto \mathbb{C}$ determines a bounded linear operator on the Hilbert space $L^2(X, \mathscr{A}, \mathbf{P})$ in the following way. Set $f \in L^{\infty}(X,\mathscr{A},\mathbf{P})$. Let us define the *multiplication operator* M_f by

$$M_f : L^2(X,\mathscr{A},\mathbf{P}) \to L^2(X,\mathscr{A},\mathbf{P}),\ g \mapsto M_f(g) := fg.$$

Straightforward computations show that M_f is linear, and the proof of the boundedness of M_f is quite easy, as well. So, $M_f \in \mathscr{B}\left(L^2(X,\mathscr{A},\mathbf{P})\right)$ for any $f \in L^{\infty}(X,\mathscr{A},\mathbf{P})$. Moreover, the operator norm of M_f coincides with the supremum norm of f, that is, $\left\|M_f\right\| = \left\|f\right\|_{\infty}$. This latter fact is also rather easy to prove. The map

$$M : L^{\infty}(X,\mathscr{A},\mathbf{P}) \to \mathscr{B}\left(L^2(X,\mathscr{A},\mathbf{P})\right),\ f \mapsto M_f \tag{1}$$

is a canonical isometric embedding of the commutative normed algebra $L^{\infty}(X,\mathscr{A},\mathbf{P})$ into the normed algebra $\mathscr{B}\left(L^2(X,\mathscr{A},\mathbf{P})\right)$, which is far from being commutative in general. This embedding is the starting point of the noncommutative generalization of probability theory. In the following section we give a brief introduction to the theory of *von Neumann algebras,* which are the appropriate mathematical objects to formalize the concepts of noncommutative probability theory. It is fair to remark that all the results of this thesis concern finite dimensional von Neumann algebras.

1. C^*-algebras, von Neumann algebras

DEFINITION 1 (Normed algebra). *A unital complex algebra $\mathscr{A}$ endowed with the norm $\|.\|$ is said to be a* normed algebra, *if the norm is submultiplicative, i. e., $\|ab\| \leq \|a\|\,\|b\|$ for any $a,b \in \mathscr{A}$ and the identity element is of norm one, that is, $\|1_{\mathscr{A}}\| = 1$.*

The reader who claims that not only unital algebras are said to be normed algebras is right. However, any algebra which appears in this work is unital, so for the sake of convenience, we incorporated the requirement of unitality in the definition.

DEFINITION 2 (Banach algebra). *A normed algebra which is a Banach space — that is, a complete normed space — is called a* Banach algebra.

DEFINITION 3 (Involution). *Let $\mathscr{A}$ be a complex algebra. A map $* : \mathscr{A} \to \mathscr{A}$, $a \mapsto a^*$ is called an* involution *if it satisfies the following properties.*

- *$*$ is antilinear: $(\lambda a + b)^* = \overline{\lambda} a^* + b^*$ for any $a,b \in \mathscr{A}$ and $\lambda \in \mathbb{C}$.*
- *$*^2 = \mathrm{id}$, that is, $(a^*)^* = a$ for any $a \in \mathscr{A}$.*
- *$*$ is an antihomomorphism with recpect to the product: $(ab)^* = b^* a^*$ for any $a,b \in \mathscr{A}$.*

DEFINITION 4 (C^*-algebra). *A Banach algebra endowed with an involution* $* : \mathscr{A} \to \mathscr{A}$ *which satisfies* $\|a^*a\| = \|a\|^2$ *for any* $a \in \mathscr{A}$ *is called a* C^*-algebra.

The above definition of C^*-algebras is rather abstract. However, we do not loose any generality if we consider the elements of a C^*-algebra as bounded operators on an appropriate Hilbert space. Indeed, any C^*-algebra is isomorphic to a closed (in the operator norm topology) unital *-subalgebra (that is, it is closed under the involution) of the operator algebra $\mathscr{B}(\mathscr{H})$ for a suitable Hilbert space $\mathscr{H}$. Furthermore, any commutative C^*-algebra is isomorphic to $C(X)$ for some compact Hausdorff space X. (The symbol $C(X)$ denotes the algebra of all continuous complex-valued functions defined on X endowed with the supremum norm.)

Despite the above remarkable facts, the C^*-algebra is still a bit too general notion to formalize the concepts of noncommutative probability theory. With an extra topological assumption we achieve the desired level of generality.

DEFINITION 5 (von Neumann algebra). *A* C^**-algebra which is closed not just in the operator norm but also in the weak operator topology is called a* von Neumann algebra.

Note that the above definition is correct as any C^*-algebra is isomorphic to an algebra of bounded linear operators on a Hilbert space $\mathscr{H}$, hence the condition about the closedness in the weak operator topology makes sense. The weak operator topology on $\mathscr{B}(\mathscr{H})$ is defined by the family of seminorms $\left\{p_{x,y} : x, y \in \mathscr{H}\right\}$ where $p_{x,y}(A) = \left|\left\langle Ax, y\right\rangle\right|$ $(A \in \mathscr{B}(\mathscr{H}))$.

A beautiful result of von Neumann shows that the purely topological condition of being closed in the weak operator topology can be characterized by a purely algebraic condition. In order to present von Neumann's theorem we define the notion of the *commutant.*

DEFINITION 6 (Commutant). *Let* S *be any subset of the algebra* $\mathscr{B}(\mathscr{H})$*. The commutant of* S *is denoted by* S' *and consists of bounded linear operators on* $\mathscr{H}$ *which commute with each element of* S*. That is,*

$$S' = \{B \in \mathscr{B}(\mathscr{H}) : AB = BA \text{ for any } A \in S\}.$$

The bicommutant S'' of the set S is defined as the commutant of its commutant. It is clear that $S'' \subseteq S$. The following theorem called *bicommutant theorem* asserts that the equality $S'' = S$ holds if and only if S is a von Neumann algebra.

THEOREM 7 (von Neumann's bicommutant theorem). *Let* $\mathscr{C}$ *be a unital *-subalgebra of the operator algebra* $\mathscr{B}(\mathscr{H})$ *for some Hilbert space* $\mathscr{H}$.

The closure of $\mathscr{C}$ *in the weak operator topology is equal to the bicommutant* $\mathscr{C}''$ *of* $\mathscr{C}$*. The algebra* $\mathscr{C}''$ *is the von Neumann algebra generated by* $\mathscr{C}$*.*

So, any von Neumnann algebra coincides with its bicommutant. There is another useful property of von Neumann algebras, namely, that they are isomorphic to the topological dual of some Banach spaces [**81**].

PROPOSITION 8. *A* C^**-algebra* $\mathscr{M}$ *is a von Neumann algebra if and only if there exists a Banach space* $\mathscr{B}$ *such that* $\mathscr{M}$ *is isomorphic (as a Banach space) to the Banach dual of* $\mathscr{B}$*, which is the space of all continuous linear functionals on* $\mathscr{B}$*.*

If there exists such a Banach space, then it is called the *predual* of the von Neumann algebra $\mathscr{M}$ and it is denoted by $\mathscr{M}_*$. There is a canonical embedding of $\mathscr{M}_*$ into the Banach dual of $\mathscr{M}$ which is denoted by $\mathscr{M}^*$. This embedding is the map

$$\hat{}\colon \mathscr{M}_* \to \mathscr{M}^*,\ \phi \mapsto \hat{\phi},$$

where $\hat{\phi} \in \mathscr{M}^*$ is defined by the equation

$$\hat{\phi}(A) = A(\phi) \qquad (A \in \mathscr{M}).$$

DEFINITION 9 (State). *Let* $\mathscr{C}$ *be a* C^**-algebra. A continuous linear functional* $\phi : \mathscr{C} \to \mathbb{C}$ *is said to be a* state *on the* C^**-algebra* $\mathscr{C}$ *if it is normalized, that is,* $\phi(I_{\mathscr{C}}) = 1$*, and positive, that is,* $\phi(A^* A) \geq 0$ *for any* $A \in \mathscr{C}$*.*

In the following, the set of all states on a C^*-algebra $\mathscr{C}$— which is clearly a convex set — will be denoted by $\mathscr{S}_{\mathscr{C}}$.

A state ϕ on a von Neumann algebra $\mathscr{M}$ is called *normal state* if for any countable collection $\{P_n\}_{n\in\mathbb{N}}$ of mutually orthogonal projections in $\mathscr{M}$ we have

$$\phi\left(\sum_{n\in\mathbb{N}} P_n\right) = \sum_{n\in\mathbb{N}} \phi(P_n).$$

(An element P of the von Neumann algebra $\mathscr{M}$ is called projection if $P^2 = P = P^*$, and the projections P and Q are orthogonal if $PQ = 0$.) Moreover, a rather interesting fact is that a state on a von Neumann algebra is normal if and only if it is in the embedded image of the predual.

EXAMPLE 10 (Normal states on the von Neumann algebra $\mathscr{B}(\mathscr{H})$). For a given separable Hilbert space $\mathscr{H}$, the algebra $\mathscr{B}(\mathscr{H})$ of all bounded linear operators is clearly a von Neumann algebra, as it is closed in $\mathscr{B}(\mathscr{H})$ in any topology. An operator $A \in \mathscr{B}(\mathscr{H})$ is said to be *trace class opearator,* if the sum

$$\sum_j \left\langle \left(A^* A\right)^{\frac{1}{2}} e_j, e_j \right\rangle$$

is finite for some (and hence for any) orthonormal basis $\{e_j\}_j \subset \mathscr{H}$. The trace of a trace class operator is defined by

$$\operatorname{Tr} A := \sum_j \langle Ae_j, e_j \rangle .$$

The definition is correct as the above sum is absolutely convergent by the definition of the trace class operators.

So, let ϕ be a normal state on $\mathscr{B}(\mathscr{H})$. Then there exists a unique self-adjoint trace class operator $\rho \in \mathscr{B}(\mathscr{H})$ with $\operatorname{Tr} \rho = 1$ such that

$$\phi(A) = \operatorname{Tr} \rho A \qquad (A \in \mathscr{B}(\mathscr{H}))\,.$$

This means that the predual of $\mathscr{B}(\mathscr{H})$ is the Banach space of all trace class operators on $\mathscr{H}$. This latter space is usually denoted by $\mathscr{T}(\mathscr{H})$.

It is an important corollary of the bicommutant theorem that any von Neumann algebra $\mathscr{M}$ is determined by its projection lattice $\mathscr{P}(\mathscr{M}) = \{P \in \mathscr{M} : P^2 = P = P^*\}$ in the sense that $\mathscr{M} = \mathscr{P}(\mathscr{M})''$ [**81**]. Therefore, the investigation of the projection lattice $\mathscr{P}(\mathscr{M})$ is of particular importance.

REMARK 11. The set of the projections $\mathscr{P}(\mathscr{M})$ of a von Neumann algebra $\mathscr{M}$ is called projection *lattice* as it is a lattice indeed. As any C^*-algebra is isomorphic to a subalgebra of $\mathscr{B}(\mathscr{H})$ for a suitable Hilbert space $\mathscr{H}$, we may consider the elements of $\mathscr{P}(\mathscr{M})$ as Hilbert space operators. The *Loewner ordering* ($P \le Q$ if and only if $\langle x, Px \rangle \le \langle x, Qx \rangle$ for any element x of the Hilbert space $\mathscr{H}$) is a partial ordering on $\mathscr{P}(\mathscr{M})$. Moreover, every two elements have a unique supremum (least upper bound) and a unique infimum (greatest lower bound). The infimum of the projections P and Q is the projection onto the intersection of the ranges of P and Q and it is denoted by $P \wedge Q$. The supremum is the projection onto the closed subspace generated by the ranges of P and Q — in notation: $P \vee Q$.

The classification of von Neumann algebras developed by *F. J. Murray* and von Neumann [**68**] is based on the following equivalence relation on the projection lattice of a von Neumann algebra.

DEFINITION 12 (Equivalence of projections). *Let $\mathscr{M}$ be a von Neumann algebra. The projections $P, Q \in \mathscr{P}(\mathscr{M})$ are said to be* equivalent *(in notation: $P \sim Q$), if there is a partial isometry $U \in \mathscr{M}$ such that $UU^* = P$ and $U^*U = Q$.*

Partial isometries can be defined even in C^*-algebras, as follows.

DEFINITION 13 (Partial isometry). *An element U of a C^*-algebra is called* partial isometry, *if it satisfies the equation $UU^*U = U$.*

The above defined relation on the projections is indeed an equivalence relations as one can convince oneself easily. Now we introduce a preorder with the use of this equivalence relation as follows. We say that $P \preceq Q$ if there exists some $P' \in \mathscr{P}(\mathscr{M})$ such that $P' \sim P$ and $P' \leq Q$, that is, the range of P' is contained in the range of Q. (Here and throughout this work, the symbol $\leq$ between operators refers to the Loewner partial ordering.) A projection P is *finite* if $Q \leq P$ and $Q \sim P$ imply $Q = P$. A projection which is not finite is called *infinite*. Let us mention an important result about the preordering $\preceq$.

THEOREM 14 (Comparsion Theorem). *Let $\mathscr{P}(\mathscr{M})$ be the projection lattice of a von Neumann algebra $\mathscr{M}$. Then, for any pair of elements $P, Q \in \mathscr{P}(\mathscr{M})$ there exists a projection $R \in \mathscr{P}(\mathscr{M}) \cap \mathscr{P}(\mathscr{M})'$ such that*

$$RPR \preceq RQR \text{ and } (I_{\mathscr{M}} - R)\, Q\, (I_{\mathscr{M}} - R) \preceq (I_{\mathscr{M}} - R)\, P\, (I_{\mathscr{M}} - R).$$

The comparsion theorem has an important consequence concerning algebras with trivial commutant. These algebras are called *factors*.

DEFINITION 15 (Factor). *A von Neumann algebra $\mathscr{M}$ is called* factor *if its commutant is trivial, that is, $\mathscr{M} \cap \mathscr{M}' = \mathbb{C} I_{\mathscr{M}}$.*

The consequence is the following. If $\mathscr{M}$ is a factor then $\mathscr{P}(\mathscr{M})$ is totally pre-ordered with respect to $\preceq$, that is, if $P, Q \in \mathscr{P}(\mathscr{M})$ then either $P \preceq Q$ or $Q \preceq P$ holds.

Furthermore, for any factor $\mathscr{M}$, there exists a function $d : \mathscr{P}(\mathscr{M}) \to [0, \infty]$ defined on the projection lattice with the following properties.

- $d(P) = 0$ if and only if $P = 0$.
- $d(P+Q) = d(P) + d(Q)$ for any $P, Q \in \mathscr{P}(\mathscr{M})$ which are orthogonal (i.e., $PQ = 0$.)
- $d(P) \leq d(Q)$ if and only if $P \preceq Q$.
- $d(P)$ is finite if and only if P is a finite projection.
- $d(P) = d(Q)$ if and only if $P \sim Q$
- $d(P) + d(Q) = d(P \wedge Q) + d(P \vee Q)$

The function d is called *dimension function*. For any given factor $\mathscr{M}$, the dimension function is unique up to harmless normalization, and the range of the dimension function describes the factor $\mathscr{M}$. Therefore, the *type of the factor* is determined by the range of the dimension function defined on its projection lattice. One of the main achievements of Murray and von Neumann was that they described the possible ranges of the dimension function on the projection lattice of a factor [**68**]. The possibilities are the followings.

the range of d	the type of the factor $\mathscr{M}$	in notation
$\{0,1,2,\dots,n\}$	discrete, finite	I_n
$\{0,1,2,\dots,n,\dots,\infty\}$	discrete, infinite	I_∞
$[0,1]$	continuous, finite	II_1
$[0,\infty]$	continuous, infinite	II_∞
$\{0,\infty\}$	purely infinite	III

Any factor is exactly one of the above types. Moreover, any von Neumann algebra over a separable Hilbert space can be written as a direct integral of factors. Therefore, the result of Murray and von Neumann is a classification of von Neumann algebras.

EXAMPLE 16. The algebra $\mathscr{B}(\mathscr{H})$ of all bounded linear operators on a separable Hilbert space $\mathscr{H}$ is of type I_n if $\mathscr{H}$ is n-dimensional and is of type I_∞ if $\mathscr{H}$ is infinite dimensional.

2. Tensor product

2.1. Tensor product of linear spaces. Let X and Y be arbitrary sets. Let $\mathscr{F}(X)$ and $\mathscr{F}(Y)$ denote the set of all complex valued functions on X and Y, respectively. Clearly, $\mathscr{F}(X)$ and $\mathscr{F}(Y)$ are linear spaces over the field $\mathbb{C}$. Let $M \subseteq \mathscr{F}(X)$ and $N \subseteq \mathscr{F}(Y)$ be arbitrary linear subspaces. One can define the *tensor product* operation on $M \times N$ the following way.

$$\otimes : M \times N \to \mathscr{F}(X \times Y);\ (m,n) \mapsto m \otimes n,$$

where $m \otimes n$ is defined by

$$(m \otimes n)(x,y) = m(x)n(y) \qquad \big(x \in X, y \in Y\big).$$

Note that as m and n are not necessarily linear functions — X and Y are not necessarily linear spaces!— the tensor product $m \otimes n$ is not a bilinear function on $X \times Y$ in general. However, the map $(m,n) \mapsto m \otimes n$ is clearly bilinear for arbitrary sets X, Y and arbitrary function spaces $M \subseteq \mathscr{F}(X), N \subseteq \mathscr{F}(Y)$.

The tensor product space $M \otimes N$ is defined by

$$M \otimes N = \left\{ \sum_{j=1}^{J} m_j \otimes n_j : J \in \mathbb{N}, m_j \in M, n_j \in N \right\}.$$

Let us denote by L^* the set of linear functionals on a linear space L. Clearly, any $l \in L$ may be considered as an element of $\mathscr{F}(L^*)$ by

$$l(\alpha) := \alpha(l) \qquad \big(l \in L,\ \alpha \in L^*\big).$$

That is, $L \subseteq \mathscr{F}(L^*)$ Furthermore, straightforward computation shows that any $l \in L$ is a linear function on L^*. Let U and V be complex vector spaces. Then $U \subseteq \mathscr{F}(U^*)$ and $V \subseteq \mathscr{F}(V^*)$.

According to the above definition, the tensor product map

$$\otimes : U \times V \to \mathscr{F}\left(U^* \times V^*\right);\ (u,v) \mapsto u \otimes v,$$

is defined by

$$(u \otimes v)\left(\alpha,\beta\right) = \alpha(u)\beta(v) \qquad \left(u \in U, v \in V, \alpha \in U^*, \beta \in V^*\right).$$

In this setting the map $u \otimes v : \left(\alpha,\beta\right) \mapsto (u \otimes v)\left(\alpha,\beta\right)$ is also bilinear, not just the map $\otimes : (u,v) \mapsto u \otimes v$.

Naturally, the tensor product space of the linear spaces U and V is

$$U \otimes V = \left\{ \sum_{j=1}^{J} u_j \otimes v_j : J \in \mathbb{N}, u_j \in U, v_j \in V \right\}.$$

2.2. Tensor product of linear operators. In the previous subsection we defined the tensor product of linear spaces, hence now we are in the position to define the tensor product of linear operators.

Let $\mathscr{A} \subseteq \mathrm{Lin}(U)$ and $\mathscr{B} \subseteq \mathrm{Lin}(V)$ denote linear subspaces which consist of linear operators on the vector spaces U and V, respectively. The tensor product of elements of $\mathscr{A}$ and $\mathscr{B}$ are defined as

$$\otimes : \mathscr{A} \times \mathscr{B} \to \mathrm{Lin}\left(U \otimes V\right);\ (A,B) \mapsto A \otimes B$$

where $A \otimes B$ is the linear extension of the map

$$(u \otimes v) \mapsto Au \otimes Bv \qquad (u \in U, v \in V).$$

2.3. Tensor product of C^*-algebras. As any C^*-algebra is a subalgebra of $\mathscr{B}(\mathscr{H})$ for some Hilbert space $\mathscr{H}$, the elements of a C^*-algebra may be viewed as linear operators on a linear space. Therefore, the tensor product $A \otimes B$ makes sense for any $A \in \mathscr{A}$ and $B \in \mathscr{B}$ if $\mathscr{A}$ and $\mathscr{B}$ are C^*-algebras. For the sake of simplicity, we consider only finite dimensional algebras. The tensor product of the finite dimensional C^*-algebras $\mathscr{A}$ and $\mathscr{B}$ is defined as

$$\mathscr{A} \otimes \mathscr{B} = \left\{ \sum_{j=1}^{J} A_j \otimes B_j : J \in \mathbb{N}, A_j \in \mathscr{A}, B_j \in \mathscr{B} \right\}.$$

Note that $\mathscr{A} \otimes \mathscr{B}$ is a C^*-algebra.

2.4. Reduced states and the partial trace. Let ϕ be a state (see Definition 9) on the C^*-algebra $\mathscr{A} \otimes \mathscr{B}$. Let us define the following functions:

$$\phi_1(A) := \phi\left(A \otimes I_{\mathscr{B}}\right) \qquad (A \in \mathscr{A})$$

and

$$\phi_2(B) := \phi\left(I_{\mathscr{A}} \otimes B\right) \qquad (B \in \mathscr{B}).$$

The functions ϕ_1 and ϕ_2 are clearly continuous linear functionals on $\mathscr{A}$ and $\mathscr{B}$, respectively. On the other hand,

$$\phi_1(I_{\mathscr{A}}) = \phi_2(I_{\mathscr{B}}) = \phi(I_{\mathscr{A}} \otimes I_{\mathscr{B}}) = \phi(I_{\mathscr{A}\otimes\mathscr{B}}) = 1$$

and

$$\begin{aligned}\phi_1\left(A^* A\right) &= \phi\left(A^* A \otimes I_{\mathscr{B}}\right) = \phi\left(\left(A^* \otimes I_{\mathscr{B}}\right)(A \otimes I_{\mathscr{B}})\right) \\ &= \phi\left((A \otimes I_{\mathscr{B}})^* (A \otimes I_{\mathscr{B}})\right) \geq 0,\end{aligned}$$

similarly,

$$\begin{aligned}\phi_2\left(B^* B\right) &= \phi\left(I_{\mathscr{A}} \otimes B^* B\right) = \phi\left(\left(I_{\mathscr{A}} \otimes B^*\right)(I_{\mathscr{A}} \otimes B)\right) \\ &= \phi\left((I_{\mathscr{A}} \otimes B)^* (I_{\mathscr{A}} \otimes B)\right) \geq 0,\end{aligned}$$

hence both ϕ_1 and ϕ_2 are states. These functionals are called *reduced states.*

Let $\mathscr{H}$ and $\mathscr{K}$ be separable Hilbert spaces and let us consider the von Neumann algebras $\mathscr{B}(\mathscr{H}), \mathscr{B}(\mathscr{K})$ and $\mathscr{B}(\mathscr{H}\otimes\mathscr{K}) = \mathscr{B}(\mathscr{H})\otimes\mathscr{B}(\mathscr{K})$. Any normal state ϕ on $\mathscr{B}(\mathscr{H})\otimes\mathscr{B}(\mathscr{K})$ can be represented by a density operator $\rho \in \mathscr{T}(\mathscr{H}\otimes\mathscr{K})$, see Example 10. Let the representing density operators of the reduced states ϕ_1 and ϕ_2 (which are also normal) be denoted by ρ_1 and ρ_2. The maps

$$\mathrm{Tr}_1 : \mathscr{T}(\mathscr{H}\otimes\mathscr{K}) \to \mathscr{T}(\mathscr{K});\ \rho \mapsto \mathrm{Tr}_1\,\rho := \rho_2$$

and

$$\mathrm{Tr}_2 : \mathscr{T}(\mathscr{H}\otimes\mathscr{K}) \to \mathscr{T}(\mathscr{K});\ \rho \mapsto \mathrm{Tr}_2\,\rho := \rho_1$$

are called *partial traces.*

3. Continuous functional calculus

If $\mathscr{A}$ is a unital complex algebra then the *spectrum* of an element $A \in \mathscr{A}$ is denoted by $\sigma(A)$ and defined by

$$\sigma(A) = \{\lambda \in \mathbb{C} :\ A - \lambda I_{\mathscr{A}} \text{ is not invertible}\}.$$

Let $\mathscr{A}$ be a C^*-algebra and let A be a normal element of $\mathscr{A}$, that is, $A^* A = AA^*$. Let $C(\sigma(A))$ denote the C^*-algebra of all continuous complex valued functions defined on the compact set $\sigma(A)$. Then there exists a unique isometric C^*-algebra homomorphism $\Phi_A : C(\sigma(A)) \to \mathscr{A}$ such that $\Phi_A(1) = I_{\mathscr{A}}$ and $\Phi_A(\mathrm{id}_{\sigma}(A)) = A$, see, e.g. [**41**, 4.4.5. Theorem]. The map Φ_A is called the *continuous functional calculus* corresponding to the normal element A and the notation $f(A) \equiv \Phi_A(f)$ is often used — it will be used also in this work. The most important consequence of this fact is that bounded normal operators on a Hilbert space do have a continuous functional calculus. The next example shows that if we assume that the Hilbert space $\mathscr{H}$ is finite dimensional, then the continuous functions of the normal elements of $\mathscr{B}(\mathscr{H})$ can be computed quite easily.

EXAMPLE 17. If $\mathscr{H}$ is finite dimensional Hilbert space then for any operator $A \in \mathscr{B}(\mathscr{H})$ the spectrum $\sigma(A)$ consists of finitely many point, and every element of the spectrum is an *eigenvalue* — that is, $\ker\left(A - aI_{\mathscr{B}(\mathscr{H})}\right) \subset \mathscr{H}$ in nontrivial for any $a \in \sigma(A)$.

If $f : G \to \mathbb{C}$ is a function defined on a set $G \subset \mathbb{C}$ then the corresponding *standard operator function* is the following map:

$$f : \{A \in \mathscr{B}(\mathscr{H}) : \; A^* A = AA^* \text{ and } \sigma(A) \subseteq G\} \to \mathscr{B}(\mathscr{H})$$

$$A = \sum_{a \in \sigma(A)} a P_a \mapsto f(A) := \sum_{a \in \sigma(A)} f(a) P_a,$$

where $\sigma(A)$ is the spectrum of A and P_a is the spectral projection corresponding to the eigenvalue a — which is nothing else but the orthogonal projection onto the subspace $\ker\left(A - aI_{\mathscr{B}(\mathscr{H})}\right) \subset \mathscr{H}$.

The mapping which maps the standard operator function f to the complex function f is exactly the continuous functional calculus on $\mathscr{B}(\mathscr{H})$.

3.1. Operator monotonicity and convexity.

DEFINITION 18 (Operator monotone function). *The real function $f : \mathbb{R} \supset I \to \mathbb{R}$ is said to be* operator monotone *if for any finite dimensional complex Hilbert space $\mathscr{H}$ the corresponding standard operator function*

$$f : \{A \in \mathscr{B}^{sa}(\mathscr{H}) : \sigma(A) \subseteq I\} \to \mathscr{B}(\mathscr{H})$$

$$A = \sum_{a \in \sigma(A)} a P_a \mapsto f(A) := \sum_{a \in \sigma(A)} f(a) P_a,$$

is monotone with respect to the Loewner partial ordering on self-adjoint operators, that is,

$$f(A) \leq f(B) \text{ whenever } A \leq B.$$

DEFINITION 19 (Operator convex function). *The real function $f : \mathbb{R} \supset I \to \mathbb{R}$ is said to be* operator convex *if for any finite dimensional complex Hilbert space $\mathscr{H}$ the corresponding standard operator function is convex with respect to the Loewner partial ordering on self-adjoint operators, that is,*

$$f(\lambda A + (1 - \lambda) B) \leq \lambda f(A) + (1 - \lambda) f(B)$$

for any $A, B \in \mathscr{B}^{sa}(\mathscr{H})$ with $\sigma(A), \sigma(B) \subset I$.

An exhaustive description of the results of matrix analysis related to operator monotonicity and convexity can be found in the monography of R. Bhatia [**9**].

4. Commutative von Neumann algebras

Now, we are in the position to answer the question *why we call von Neumann algebra theory sometimes noncommutative probability theory?*

It is clear by Definition 2 that the function space $L^\infty(X,\mathscr{A},\mathbf{P})$ is a commutative Banach algebra for any probability space $(X,\mathscr{A},\mathbf{P})$. Furthermore, it is easy to see that these Banach algebras are also C^*-algebras with the complex conjugation as involution. It is folklore that $L^\infty(X,\mathscr{A},\mathbf{P})$ is the Banach dual of the Banach space $L^1(X,\mathscr{A},\mathbf{P})$, which is defined as follows:

$$L^1(X,\mathscr{A},\mathbf{P}) = \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is measurable and } \int_X |f| \,\mathrm{d}\mathbf{P} < \infty \right\}.$$

So, $L^\infty(X,\mathscr{A},\mathbf{P})$ is a commutative C^*-algebra which is the dual of the Banach space $L^1(X,\mathscr{A},\mathbf{P})$. By Proposition 8, this means that $L^\infty(X,\mathscr{A},\mathbf{P})$ is a commutative von Neumann algebra. The interesting fact is that the converse statement is also true. That is, every abelian von Neumann algebra is isomorphic to $L^\infty(X,\mathscr{S},\mu)$ for some localizable measure space $(X,\mathscr{S},\mu)$, see [**81**]. (A localizable measure space is the direct sum of finite measure spaces.)

We can deduce that every probability space determines a commutative von Neumann algebra — the algebra of the bounded random variables — and every commutative von Neumann algebra determines a probability space, up to harmless normalization. That is the reason why the theory of von Neumann algbras may be considered as noncommutative probability theory.

CHAPTER 2

Quantum variances, generalized entropies and relative entopies

We finished the previous chapter with the description of the correspondence between probability spaces and abelian von Neumann algebras — see Section 4 in Chapter 1. Fortunately, several interesting and useful notions of probability theory can be extended to the general von Neumann algebra setting. In this thesis, we focus on two distinguished concepts of probability theory, namely the *(co)variance* and the *entropy*.

In the first section we investigate the following problem. Can we characterize those sets of *observables* for which the induced covariance mapping is a *roof*? (See Def. 21 for the definition of roof.) Note that this question does not make sense in the case of abelian von Neumann algebra for the following reason. It is known that every *pure state* is multiplicative on a commutative von Neumann algebra, see, e.g., [**41**, 4.4.1. Prop.]. Therefore, the covariance of any two observables is zero in any pure state. So, the covariance mapping is a roof if and only if it is identically zero which is clearly not the case.

In the second section the strong subadditivity inequality of the entropy is investigated. Fairly nontrivial but rather easy computations show that the *Shannon entropy* is strongly subadditive. In my opinion, a much more sophisticated argument shows that its noncommutative counterpart, the *von Neumann entropy* is also strongly subadditive. The latter statement is a celebrated result of Lieb and Ruskai [**49**]. We consider a one-parameter generalization of the von Neumann entropy which is called *Tsallis entropy*. We show — in particular — that the Tsallis entropy is not strongly subadditive for noncommutative von Neumann algebras in spite of the facts that it is strongly subadditive in the commutative case [**22**, Thm 3.4] and that it is subadditive in the noncommutative case, as well [**5**].

In the third section we introduce the *Bregman divergences* which may be considered as certain generalizations of the *Umegaki relative entropy*. We characterize those Bregman divergences which are jointly convex, and we use this result to derive a sharp inequality for Tsallis entropy which can be considered as a generalization of the strong subadditivity inequality of the von Neumann entropy.

1. Decomposable quantum variances

Let $\mathscr{A}$ be a von Neumann algebra of type I_n and let ϕ be a — necessarily normal — state on $\mathscr{A}$. The *covariance* of the self-adjoint elements $A, B \in \mathscr{A}$ is defined by

$$\mathrm{Cov}_\phi(A,B) = \phi(AB) - \phi(A)\phi(B).$$

In particular, the variance of the observable (self-adjoint elements are often called observables) A in the state ϕ is given by

$$\mathrm{Var}_\phi(A) = \mathrm{Cov}_\phi(A,A) = \phi\left(A^2\right) - \left(\phi(A)\right)^2.$$

It is rather easy to check that

$$\mathrm{Var}_\phi(A + \lambda I_{\mathscr{A}}) = \mathrm{Var}_\phi(A) \qquad (A \in \mathscr{A}, \lambda \in \mathbb{R})$$

holds for any state ϕ.

It is useful to introduce the *covariance matrix* of several observables. If $A_1, \dots A_r$ are self-adjoint elements of $\mathscr{A}$, then their covariance matrix is defined as

$$\left[\mathbf{Cov}_\phi(A_1,\dots,A_r)\right]_{i,j} := \mathrm{Cov}_\phi\left(A_i, A_j\right) \qquad (1 \le i, j \le r).$$

Observe that the above defined covariance matrix is necessarily self-adjoint as $\phi\left(A_i A_j\right) = \overline{\phi\left(A_j A_i\right)}$.

One of the most important properties of the covariance is that it is a concave map on the set of states, that is, the mapping

$$\mathbf{Cov}_{(.)}(A_1,\dots,A_r) : \mathscr{S}_{\mathscr{A}} \to \mathbf{M}_r^{sa}; \phi \mapsto \mathbf{Cov}_\phi(A_1,\dots,A_r) \tag{2}$$

is concave with respect to the *Loewner ordering* on the final space $\mathbf{M}_r$. (For any $A, B \in \mathbf{M}_r^{sa}$ we say that $A \le B$ if $B - A$ is a positive semidefinite matrix.)

Indeed, assume that $0 \le \lambda_1, \lambda_2$ such that $\lambda_1 + \lambda_2 = 1$ and $\phi_1, \phi_2 \in \mathscr{S}_{\mathscr{A}}$. Then

$$\begin{aligned}\mathrm{Cov}_{\lambda_1\phi_1+\lambda_2\phi_2}\left(A_i, A_j\right) - \left(\lambda_1 \mathrm{Cov}_{\phi_1}\left(A_i, A_j\right) + \lambda_2 \mathrm{Cov}_{\phi_2}\left(A_i, A_j\right)\right)\\ = \lambda_1\lambda_2\left(\phi_1(A_i) - \phi_2(A_i)\right)\left(\phi_1\left(A_j\right) - \phi_2\left(A_j\right)\right).\end{aligned}$$

Therefore, the matrix

$$\mathbf{Cov}_{\lambda_1\phi_1+\lambda_2\phi_2}(A_1,\dots,A_r) - \left(\lambda_1 \mathbf{Cov}_{\phi_1}(A_1,\dots,A_r) + \lambda_2 \mathbf{Cov}_{\phi_2}(A_1,\dots,A_r)\right)$$

is positive semidefinite, as it is a nonnegative multiple of a rank-one projection. In particular, we obtain the concavity of the variance functional

$$\phi \mapsto \mathrm{Var}_\phi(A)$$

for any self-adjoint A.

As the von Neumann algebra $\mathscr{A}$ is of type I_n — that is, it is isomorphic to the operator algebra $\mathscr{B}(\mathscr{H})$ for an n-dimensional complex Hilbert space $\mathscr{H}$, — every state is represented by a unique density operator (see Example 10). For the sake of simplicity, we will use the following notation. If the state ϕ is represented by the density operator D, then we define $\mathrm{Cov}_D(.,.) := \mathrm{Cov}_\phi(.,.)$, and so on, $\mathrm{Var}_D(.) := \mathrm{Var}_\phi(.)$ and $\mathbf{Cov}_D(.,.,\ldots,.) := \mathbf{Cov}_\phi(.,.,\ldots,.)$.

Using this notation, the above derived concavity of the covariance matrix map (2) can be written as

$$\mathbf{Cov}_D(A_1,\ldots,A_r) \geq \sum_{k=1}^m \lambda_k \mathbf{Cov}_{D_k}(A_1,\ldots,A_r) \quad \text{if} \quad D = \sum_{k=1}^m \lambda_k D_k,$$

where $\lambda_k \geq 0$ and $\sum_{k=1}^m \lambda_k = 1$.

For any inequality, it is an interesting task to investigate the case of equality. For such an investigation, a useful tool is the recently introduced notion of *roof* which is defined as follows.

DEFINITION 20 (Roof point). *Let Ω be a compact convex set contained in a finite dimensional real linear space. Let G be a mapping from Ω into a partially ordered set. A point $\omega \in \Omega$ is called* roof point, *if there are some extremal points $\pi_1,\ldots,\pi_m$ of Ω and nonnegative numbers $p_1,\ldots,p_m$ with $\sum_{k=1}^m p_k = 1$ such that*

$$\sum_{k=1}^m p_k \pi_k = \omega$$

and

$$\sum_{k=1}^m p_k G(\pi_k) = G(\omega).$$

DEFINITION 21 (Roof). *A mapping G defined on Ω is called* roof *if every $\omega \in \Omega$ is a roof point.*

The reader should consult [**87**] for further details.

As $\mathscr{A}$ is finite dimensional, the set of the density operators is a compact convex subset of the real vector space of the self-adjoint elements of $\mathscr{A}$. We are interested in the following question. *Is the concave mapping* (2) *a roof on $\mathscr{S}_\mathscr{A}$?* It is well-known that the extremal points of the set of densities are exactly the rank-one projections. So we can reformulate our question. Given an arbitrary density D, can we find rank one projections $P_1,\ldots,P_m$ and nonnegative weights $p_1,\ldots,p_m$ (with $\sum_{k=1}^m p_k = 1$) such that

$$D = \sum_{k=1}^m p_k P_k \tag{3}$$

and

$$\mathbf{Cov}_D(A_1,\dots,A_r)=\sum_{k=1}^m p_k\mathbf{Cov}_{P_k}(A_1,\dots,A_r)?$$

We say that (3) is an *extremal convex decomposition* of D.

For $r=1$ the answer is positive, and this is the first result in this topic, made by *Petz* and *Tóth* [**77**]. An extension of the former result was given by *Petz* and *Léka* in [**46**]. They proved that the answer is positive even in the case $r=2$. In our paper [**78**] we gave a necessary and sufficient condition for the covariance mapping (2) being a roof in terms of the corresponding observables. Our result applies for any finite collection of observables, and it recovers all the aforementioned results easily.

LEMMA 22. *If $D_1,\dots,D_m$ are density operators, $A_1,\dots,A_r$ are self-adjoint elements of $\mathscr{A}$ and $D=\sum_{k=1}^m \lambda_k D_k$ with some $0<\lambda_1,\dots,\lambda_m$, $\sum_{k=1}^m \lambda_k=1$, then*

$$\mathbf{Cov}_D(A_1,\dots,A_r)=\sum_{k=1}^m \lambda_k\mathbf{Cov}_{D_k}(A_1,\dots,A_r) \tag{4}$$

if and only if

$$\operatorname{Tr} D_k A_j=\operatorname{Tr} DA_j \quad \text{for all} \quad 1\le k\le m \quad \text{and} \quad 1\le j\le r. \tag{5}$$

PROOF. The covariance has the shift invariance property

$$\mathbf{Cov}_D(A_1,\dots,A_r)=\mathbf{Cov}_D(A_1-\lambda_1 I,\dots,A_r-\lambda_r I)$$

for every reals $\lambda_1,\dots,\lambda_r$. Set $\lambda_j:=\operatorname{Tr} DA_j$. With this choice $\operatorname{Tr} D(A_j-\lambda_j I)=0$ holds for every j. Therefore

$$\begin{aligned}[\mathbf{Cov}_D(A_1,\dots,A_r)]_{ij}&=[\mathbf{Cov}_D(A_1-\lambda_1 I,\dots,A_r-\lambda_r I)]_{ij}\\&=\operatorname{Tr} D(A_i-\lambda_i I)(A_j-\lambda_j I).\end{aligned}$$

Because of the concavity of the covariance,

$$\mathbf{Cov}_D(A_1,\dots,A_r)-\sum_{k=1}^m \lambda_k\mathbf{Cov}_{D_k}(A_1,\dots,A_r)$$

is a positive semi-definite matrix, hence it is equal to zero if and only if the diagonal elements are zeros, that is,

$$\operatorname{Tr} D(A_j-\lambda_j I)^2-\left(\sum_{k=1}^m \lambda_k \operatorname{Tr} D_k(A_j-\lambda_j I)^2-\lambda_k\left(\operatorname{Tr} D_k(A_j-\lambda_j I)\right)^2\right)=0 \tag{6}$$

holds for every j. It is easy to check that (6) holds if and only if $\operatorname{Tr} D_k(A_j-\lambda_j I)=0$ for every k, j and this is equivalent to (5). □

In the next subsection we use Lemma 22 to characterize those sets of self-adjoint elements of $\mathscr{A}$ for which the decomposition of the covariance with projections is possible.

1.1. The main theorem. Recall that our von Neumann algebra $\mathscr{A}$ is (isomorphic to) the operator algebra $\mathscr{B}(\mathscr{H})$, where $\mathscr{H}$ is a Hilbert space of dimension n. For an arbitrary subspace $\mathscr{K} \subset \mathscr{H}$, we denote by $Q^{\mathscr{K}}$ the orthogonal projection onto $\mathscr{K}$. We define

$$A^{\mathscr{K}} := Q^{\mathscr{K}} A Q^{\mathscr{K}}$$

for every element $A \in \mathscr{A}$ and

$$\mathscr{B}(\mathscr{K}) := Q^{\mathscr{K}} \mathscr{B}(\mathscr{H}) Q^{\mathscr{K}}, \qquad \mathscr{B}^{sa}(\mathscr{K}) := Q^{\mathscr{K}} \mathscr{B}^{sa}(\mathscr{H}) Q^{\mathscr{K}},$$

$$\mathscr{B}^{+}(\mathscr{K}) := Q^{\mathscr{K}} \mathscr{B}^{+}(\mathscr{H}) Q^{\mathscr{K}}, \qquad \mathscr{S}(\mathscr{K}) := \{X \in \mathscr{B}^{+}(\mathscr{K}) : \operatorname{Tr} X = 1\}.$$

DEFINITION 23. *Let $\{A_1, \dots, A_r\}$ be a set of self-adjoint elements of $\mathscr{A} = \mathscr{B}(\mathscr{H})$. The set $\{A_1, \dots, A_r\}$ is said to be* variance-decomposable *if for every $D \in \mathscr{S}_{\mathscr{A}}$ there exists an extremal convex decomposition*

$$D = \sum_{k=1}^{m} \lambda_k P_k$$

of D such that

$$\mathbf{Cov}_D(A_1, \dots, A_r) = \sum_{k=1}^{m} \lambda_k \mathbf{Cov}_{P_k}(A_1, \dots, A_r)$$

In other words, $\{A_1, \dots, A_r\}$ is variance-decomposable if and only if the mapping $D \mapsto \mathbf{Cov}_D(A_1, \dots, A_r)$ is a roof. Our main result reads as follows.

THEOREM 24. *The set $\{A_1, \dots, A_r\} \subset \mathscr{A}$ is variance-decomposable if and only if*

$$\dim\left(\operatorname{span}\left\{I^{\mathscr{K}}, A_1^{\mathscr{K}}, \dots, A_r^{\mathscr{K}}\right\}\right) < (\dim \mathscr{K})^2 \tag{7}$$

for every subspace $\mathscr{K} \subset \mathscr{H}$ with $\dim \mathscr{K} > 1$.

PROOF. By Lemma 22, $\{A_1, \dots, A_r\} \subset \mathscr{A}$ is variance-decomposable if and only if for every density operator $D \in \mathscr{S}_{\mathscr{A}}$ there exist $P_1, \dots, P_m$ rank-one projections such that $D \in \mathbf{Conv}(\{P_1, \dots, P_m\})$ – where $\mathbf{Conv}(H)$ denotes the convex hull of the set H – and

$$\operatorname{Tr} P_k A_j = \operatorname{Tr} D A_j \text{ for all } 1 \le \mathrm{k} \le \mathrm{m} \text{ and } 1 \le \mathrm{j} \le \mathrm{r}. \tag{8}$$

First, we show that the condition (7) is sufficient. It is enough to show that for every $D \in \mathscr{S}_{\mathscr{A}}$, $\operatorname{rank}(D) > 1$ there exist $E_1, \dots, E_m \in \mathscr{S}_{\mathscr{A}}$ density operators such that

$$D \in \mathbf{Conv}(\{E_1, \dots, E_m\}) \tag{9}$$

and

$$\operatorname{Tr} E_k A_j = \operatorname{Tr} D A_j \text{ for all k and j} \tag{10}$$

and

$$\text{rank}(E_k) < \text{rank}(D). \tag{11}$$

Let D be an arbitrary element of $\mathscr{S}_{\mathscr{A}}$ with $\text{rank}(D) > 1$ and $\mathscr{K} := \text{range}(D)$. The set $\mathscr{B}^{sa}(\mathscr{K})$ is a $(\dim(\mathscr{K}))^2$ dimensional Hilbert space over the field $\mathbb{R}$ with the positive definite inner product $\langle X, Y\rangle = \text{Tr}\, XY$. Let us use the notation $\mathbf{A} = (A_1, \ldots, A_r)$. In the following we denote the identity of $\mathscr{A}$ simply by I instead of $I_{\mathscr{A}}$. Define

$$\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} := \{X \in \mathscr{B}^{sa}(\mathscr{K}) : \langle X, I\rangle = 1, \left\langle X, A_j\right\rangle = \left\langle D, A_j\right\rangle \text{ for all j}\}.$$

Clearly,

$$\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} = \{X \in \mathscr{B}^{sa}(\mathscr{K}) : \left\langle X, I^{\mathscr{K}}\right\rangle = 1, \left\langle X, A^{\mathscr{K}}_j\right\rangle = \left\langle D, A^{\mathscr{K}}_j\right\rangle \text{ for all j}\}.$$

Because of the assumption $\dim\left(\text{span}\left\{I^{\mathscr{K}}, A^{\mathscr{K}}_1, \ldots, A^{\mathscr{K}}_r\right\}\right) < (\dim \mathscr{K})^2$, $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}}$ is an affine subspace of $\mathscr{B}^{sa}(\mathscr{K})$ with positive dimension.

It is well-known that $\mathscr{S}(\mathscr{K})$ is a bounded convex set (for example, $\|P\|_2 \leq 1$ if $P \in \mathscr{S}(\mathscr{K})$, where $\|\cdot\|_2$ denotes the Hilbert-Schmidt norm). Therefore, $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} \cap \mathscr{S}(\mathscr{K})$ is a bounded convex set and

$$\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} \cap \mathscr{S}(\mathscr{K}) \subset \mathbf{Conv}\left(\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} \cap \partial\mathscr{S}(\mathscr{K})\right),$$

where $\partial\mathscr{S}(\mathscr{K})$ denotes the relative boundary of $\mathscr{S}(\mathscr{K})$.

By definition, $D \in \mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} \cap \mathscr{S}(\mathscr{K})$, and hence

$$D = \sum_{k=1}^{m} \lambda_k E_k \text{ with some } \{\mathrm{E_k}\}_{\mathrm{k=1}}^{\mathrm{m}} \subset \mathscr{L}^{\mathscr{K}}_{\mathrm{D},\mathbf{A}} \cap \partial\mathscr{S}(\mathscr{K}) \text{ and } 0 \leq \lambda_{\mathrm{k}}, \sum \lambda_{\mathrm{k}} = 1.$$

This is exactly the statemant we wanted to prove, because $E_k \in \partial\mathscr{S}(\mathscr{K})$ implies that $\text{rank}(E_k) < \dim(\mathscr{K}) = \text{rank}(D)$, that is, (11) holds, and $E_k \in \mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}}$ implies that (10) holds.

Note that D has maximal rank in $\mathscr{S}(\mathscr{K})$, hence it is a (relative) interior point of $\mathscr{S}(\mathscr{K})$. On the other hand, $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}}$ lies in the affine hull of $\mathscr{S}(\mathscr{K})$ and has positive dimension. Therefore, the intersection $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} \cap \mathscr{S}(\mathscr{K})$ is not a single point.

To show that the condition (7) is necessary as well, assume that

$$\dim\left(\text{span}\left\{I^{\mathscr{K}}, A^{\mathscr{K}}_1, \ldots, A^{\mathscr{K}}_r\right\}\right) = (\dim \mathscr{K})^2$$

for some subspace $\mathscr{K} \subset \mathscr{H}$ with $\dim \mathscr{K} > 1$. Set $D \in \mathscr{S}(\mathscr{K})$, $\text{rank}(D) > 1$. Because of the assumption $\dim\left(\text{span}\left\{I^{\mathscr{K}}, A^{\mathscr{K}}_1, \ldots, A^{\mathscr{K}}_r\right\}\right) = (\dim \mathscr{K})^2$, $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}}$ is a 0 dimensional affine subspace of $\mathscr{B}^{sa}(\mathscr{K})$, that is, $\mathscr{L}^{\mathscr{K}}_{D,\mathbf{A}} = \{D\}$. Therefore, we have by Lemma 22 that the decomposition of D is impossible. □

The next example shows that for an arbitrary large n there exists a set of self-adjoint operators in $\mathscr{B}(\mathscr{H})$ — where $\dim(\mathscr{H}) = n$ — with only three elements which is not variance-decomposable.

EXAMPLE. For every $n \geq 2$ we can find self-adjoint operators $A_1, A_2, A_3 \in \mathscr{B}(\mathscr{H}) = \mathscr{A}$ and a density operator $D \in \mathscr{S}_{\mathscr{A}}$ with the following property. If $P_1, \ldots, P_m$ are rank-one projections such that $D = \sum_{k=1}^m \lambda_k P_k$ with some $0 < \lambda_1, \ldots, \lambda_m$, $\sum_{k=1}^m \lambda_k = 1$, then

$$\mathbf{Cov}_D(A_1, A_2, A_3) \neq \sum_{k=1}^m \lambda_k \mathbf{Cov}_{P_k}(A_1, A_2, A_3). \tag{12}$$

Let $\{e_1, \ldots, e_n\}$ be an arbitrary orthonormal basis in $\mathscr{H}$ and let us identify the elements of $\mathscr{B}(\mathscr{H})$ with their matrices in this basis. Let us use the Pauli matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

to define A_1, A_2, A_3 the following way

$$A_1 := \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 := \begin{bmatrix} \sigma_2 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_3 := \begin{bmatrix} \sigma_3 & 0 \\ 0 & 0 \end{bmatrix},$$

and $D := \mathrm{Diag}\left(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0\right)$. By Lemma 22,

$$\mathbf{Cov}_D(A_1, A_2, A_3) = \sum_{k=1}^m \lambda_k \mathbf{Cov}_{P_k}(A_1, A_2, A_3)$$

if and only if $\mathrm{Tr}\, P_k A_j = 0$ for every k and j, but in this case we have $P_k^{\mathscr{K}} = 0$ for $\mathscr{K} = \mathrm{range}(D)$, hence D can not be a convex combination of the P_k's. Therefore, (12) holds.

The proof of the statement of the previous example is shorter if we use the Theorem. The only thing we have to observe is that

$$\dim\left(\mathrm{span}\left\{I^{\mathscr{K}}, A_1^{\mathscr{K}}, A_2^{\mathscr{K}}, A_3^{\mathscr{K}}\right\}\right) = (\dim \mathscr{K})^2$$

for $\mathscr{K} = \mathrm{range}(D)$.

2. Generalizations of the strong subadditivity inequality

Let $\mathscr{A}$ be a von Neumann algebra of type I_n and let us denote by $\mathscr{H}$ the underlying n-dimensional Hilbert space — that is, $\mathscr{A} = \mathscr{B}(\mathscr{H})$. Let ρ be a density operator which represents a state on $\mathscr{A}$. Note that in this case $\rho \in \mathscr{A}$ and the expression $f(\rho)$ makes sense by the continuous functional calculus for any complex function f which is continuous on the spectrum of ρ — see Section 3.

The von Neumann entropy of the density operator ρ is defined by

$$S(\rho) = -\mathrm{Tr}\rho \ln \rho \tag{13}$$

see, e.g., [**9, 35, 71**]. Let the Hilbert space $\mathcal{H}$ be the tensor product of three finite dimensional Hilbert spaces, that is, $\mathcal{H} := \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. (Note that tensor products are defined in Section 2.) Let $\rho_{123} \in \mathcal{B}(\mathcal{H})$ be a density operator. The *reduced densities* are defined by partial traces — see Subsection 2.4. Let us use the following notation.

$$\rho_{12} := \mathrm{Tr}_3\, \rho_{123}, \qquad \rho_2 := \mathrm{Tr}_1\, \rho_{12}, \qquad \rho_{23} := \mathrm{Tr}_1\, \rho_{123}. \tag{14}$$

As in our case the states and the density operators are in one-to-one correspondence, densities will be called sometimes states, and we will refer to reduced densities sometimes by the expression *reduced state*.

One of the most important results in quantum information theory is the strong subadditivity of the von Neumann entropy, which is the following inequality.

$$S(\rho_{123}) + S(\rho_2) \le S(\rho_{12}) + S(\rho_{23}).$$

This result was made by E. Lieb and M. B. Ruskai in 1973 [**49, 71**]. Our aim is to generalize this inequality in various ways. The key object of our investigations is a certain generalization of the von Neumann entropy which is called

2.1. Tsallis entropy. The Tsallis entropy is a one-parameter extension of the von Neumann entropy. For any real q, one can define the deformed logarithm (or q-logarithm) function $\ln_q : (0,\infty) \to \mathbb{R}$ by

$$\ln_q x := \int_1^x t^{q-2}\,\mathrm{d}t = \begin{cases} \frac{x^{q-1}-1}{q-1} & \text{if } q \neq 1, \\ \ln x & \text{if } q = 1. \end{cases} \tag{15}$$

The corresponding entropy

$$S_q(\rho) = -\mathrm{Tr}\rho \ln_q \rho$$

is called Tsallis entropy [**2, 17**]. It is reasonable to restrict ourselves to the $0 < q$ case, because $\lim_{x\to 0^+} -x \ln_q x = 0$ if and only if $0 < q$. If we introduce the notation $f_q(x) = x \ln_q x$ we can write $S_q(\rho) = -\mathrm{Tr} f_q(\rho)$.

2.2. The Tsallis entropy is subadditive, but not strongly subadditive. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be finite dimensional Hilbert spaces. If ρ_{12} is a state on a Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ — that is, $\rho_{12} \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ such that $0 \le \rho_{12}$ and $\mathrm{Tr}\,\rho_{12} = 1$, — then it has reduced states $\rho_1 := \mathrm{Tr}_2\,\rho_{12}$ and $\rho_2 := \mathrm{Tr}_1\,\rho_{12}$ on the spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The subadditivity inequality of the Tsallis entropy is

$$S_q(\rho_{12}) \le S_q(\rho_1) + S_q(\rho_2), \tag{16}$$

and it has been proved for $q > 1$ by Audenaert in 2007 [**5**].

First, we have to note an easy consequence of Audenaert's theorem. Assume that $\mathscr{H} = \mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3$ for some finite dimensional Hilbert spaces $\mathscr{H}_i$ $(i = 1,2,3)$. Let $\rho_{123} \in \mathscr{B}(\mathscr{H})$ be a density operator and let the reduced densities be labeled according to the notation introduced in (14).

The subadditivity of the von Neumann entropy implies that

$$S_q(\rho_{123}) \le S_q(\rho_{12}) + S_q(\rho_3)$$

and

$$S_q(\rho_{123}) \le S_q(\rho_1) + S_q(\rho_{23}).$$

It follows that

$$S_q(\rho_{123}) + S_q(\rho_2) - S_q(\rho_{12}) - S_q(\rho_{23})$$

$$\le \min\{S_q(\rho_1) + S_q(\rho_2) - S_q(\rho_{12}), S_q(\rho_2) + S_q(\rho_3) - S_q(\rho_{23})\}.$$

The Tsallis entropy is nonnegative and takes its maximum at the completely mixed state — that is, in the state $\frac{1}{\dim \mathscr{H}} I_{\mathscr{B}(\mathscr{H})}$ — and the maximal value is $-\ln_q \frac{1}{d}$, where d is the dimension of the underlying Hilbert space. Therefore,

$$S_q(\rho_{123}) + S_q(\rho_2) - S_q(\rho_{12}) - S_q(\rho_{23}) \le -\ln_q \frac{1}{d_2} - \ln_q \frac{1}{\min\{d_1, d_3\}},$$

where d_i is the dimension of $\mathscr{H}_i$.

However, the strong subadditivity inequality

$$S_q(\rho_{123}) + S_q(\rho_2) \le S_q(\rho_{12}) + S_q(\rho_{23}) \tag{17}$$

does not hold in general.

PROPOSITION 25. *The only strongly subadditive Tsallis entropy is the von Neumann entropy, that is, the strong subadditivity of the Tsallis entropy holds if and only if* $q = 1$.

PROOF. It is known that if ρ_i is a state on a Hilbert space $\mathscr{H}_i$ for $i = 1,2$ then $\rho_1 \otimes \rho_2$ is a state on $\mathscr{H}_1 \otimes \mathscr{H}_2$ and the relation

$$S_q(\rho_1 \otimes \rho_2) = S_q(\rho_1) + S_q(\rho_2) + (1-q)S_q(\rho_1)S_q(\rho_2) \tag{18}$$

holds. Therefore, the Tsallis enrtopy can not be subadditive for $q < 1$. In fact, it is neither subadditive, nor superadditive [**22**]. So we have shown that the Tsallis entropy is not strongly subadditive for $q < 1$. On the other hand, the next examples show that (17) does not hold for $1 < q$ neither.

Set $q > 1$ and consider the operator

$$\mathscr{B}\left(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2\right) \ni \rho_{123} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{19}$$

ρ_{123} is positive and $\mathrm{Tr}\rho_{123} = 1$. We have

$$\rho_{12} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \rho_{23} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}, \quad \rho_2 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Straightforward computations show that

$$S_q(\rho_{123}) + S_q(\rho_2) = \frac{1}{q-1}\left(1 - 2\left(\frac{1}{2}\right)^q + 1 - 2\left(\frac{1}{2}\right)^q\right) = \frac{1}{q-1}\left(2 - 4\left(\frac{1}{2}\right)^q\right)$$

and

$$S_q(\rho_{12}) + S_q(\rho_{23}) = S_q(\rho_{23}) = \frac{1}{q-1}\left(1 - 4\left(\frac{1}{4}\right)^q\right).$$

By the inequality of geometric and arithmetic means, we have $2 \cdot 2^{1-q} < 1 + 4^{1-q}$, which immediately shows that $S_q(\rho_{123}) + S_q(\rho_2) > S_q(\rho_{12}) + S_q(\rho_{23})$. □

This is not so surprising, if we consider a bit more general example.

EXAMPLE. If ρ_{12} is an *entangled* pure state (i.e., $\mathrm{rank}\left(\rho_{12}\right) = 1$ but $\mathrm{rank}\left(\rho_1\right), \mathrm{rank}\left(\rho_2\right) > 1$), let ρ_3 be a state on $\mathscr{H}_3$ such that $1 < \mathrm{rank}(\rho_3)$ and let $\rho_{123} := \rho_{12} \otimes \rho_3$. Then

$$S_q(\rho_{123}) + S_q(\rho_2) > S_q(\rho_{12}) + S_q(\rho_{23})$$

holds for every $1 < q$. Indeed, if we use (18) we get

$$\begin{aligned} &S_q(\rho_{123}) + S_q(\rho_2) \\ &= S_q(\rho_{12}) + S_q(\rho_3) + (1-q) S_q(\rho_{12}) S_q(\rho_3) + S_q(\rho_2) = S_q(\rho_2) + S_q(\rho_3) \end{aligned}$$

and

$$S_q(\rho_{12}) + S_q(\rho_{23}) = S_q(\rho_{23}) = S_q(\rho_2) + S_q(\rho_3) + (1-q) S_q(\rho_2) S_q(\rho_3),$$

because $S_q(\rho_{12}) = 0$. The density ρ_{12} is entangled, hence $S_q(\rho_2) > 0$, and this verifies the statement.

It is an interesting fact that for classical probability distributions, the Tsallis entropy is strongly subadditive for $1 \le q$ [**22**], and this result has an elegant and short proof. The only thing necessary for the proof is that for any positive x, y and q, the identity

$$\ln_q x - \ln_q y = -\ln_q \left(\frac{y}{x}\right) x^{q-1} \tag{20}$$

holds. Now we restate the proof of Furuichi.

PROOF. Let $\{p_{jkl}\}_{j=1,k=1,l=1}^{m,\ n,\ r}$ be a discrete probability distribution. Let us introduce the notation $p_{jk} = \sum_{l=1}^r p_{jkl}, p_{kl} = \sum_{j=1}^m p_{jkl}$ and $p_k = \sum_{j=1}^m p_{jk}$. If $\mathscr{B}(\mathbb{C}^m \otimes \mathbb{C}^n \otimes \mathbb{C}^r) \ni \rho_{123} = \mathrm{Diag}(\{p_{jkl}\})$, then by (20) we have

$$S_q(\rho_{123}) - S_q(\rho_{12}) = -\sum_{j,k,l} p_{jkl}(\ln_q(p_{jkl}) - \ln_q(p_{jk})) = \sum_{j,k,l} p_{jkl}^q \ln_q\left(\frac{p_{jk}}{p_{jkl}}\right)$$

$$= \sum_{j,k,l} p_{jk}^q \left(\frac{p_{jkl}}{p_{jk}}\right)^q \ln_q\left(\frac{p_{jk}}{p_{jkl}}\right).$$

Observe that $x^q \ln_q\left(\frac{1}{x}\right) = -f_q(x)$, hence this expression can be written as

$$-\sum_{j,k,l} p_{jk}^q f_q\left(\frac{p_{jkl}}{p_{jk}}\right) = -\sum_{j,k,l}\left(\frac{p_{jk}}{p_k}\right)^q p_k^q f_q\left(\frac{p_{jkl}}{p_{jk}}\right)$$

$$\le -\sum_{k,l} p_k^q \left(\sum_j \frac{p_{jk}}{p_k} f_q\left(\frac{p_{jkl}}{p_{jk}}\right)\right).$$

f_q is convex, hence

$$-\sum_{k,l} p_k^q \left(\sum_j \frac{p_{jk}}{p_k} f_q\left(\frac{p_{jkl}}{p_{jk}}\right)\right) \le -\sum_{k,l} p_k^q f_q\left(\sum_j \frac{p_{jk}}{p_k}\frac{p_{jkl}}{p_{jk}}\right) = -\sum_{k,l} p_k^q f_q\left(\frac{p_{kl}}{p_k}\right) =$$

$$= \sum_{k,l} p_k^q \left(\frac{p_{kl}}{p_k}\right)^q \ln_q\left(\frac{p_k}{p_{kl}}\right) = \sum_{k,l} p_{kl}^q \ln_q\left(\frac{p_k}{p_{kl}}\right) = -\sum_{k,l} p_{kl}(\ln_q(p_{kl}) - \ln_q(p_k))$$

$$= S_q(\rho_{23}) - S_q(\rho_2).$$

□

2.3. Relative entropy and monotonicity. Let $\mathcal{H}$ be a finite dimensional complex Hilbert space and $n := \dim \mathcal{H}$. If $\phi \in \mathcal{H}$ then the symbols $|\phi\rangle$ and $\langle\phi|$ denote certain mappings which are defined as follows.

$$|\phi\rangle : \mathbb{C} \to \mathcal{H};\ \alpha \mapsto |\phi\rangle\,\alpha := \alpha\phi.$$

$$\langle\phi| : \mathcal{H} \to \mathbb{C};\ \psi \mapsto \langle\phi|\,\psi := \langle\phi,\psi\rangle.$$

Let f be a $(0,\infty) \to \mathbb{R}$ function, let $\rho,\sigma \in \mathcal{B}^{++}(\mathcal{H})$ be positive definite operators and $A \in \mathcal{B}(\mathcal{H})$. We define the relative quasi-entropy by

$$S_f^A\left(\rho\,||\,\sigma\right) := \left\langle A\rho^{\frac{1}{2}}, f\left(\Delta(\sigma/\rho)\right)\left(A\rho^{\frac{1}{2}}\right)\right\rangle, \tag{21}$$

where $\langle A,B\rangle = \operatorname{Tr} A^* B$ is the Hilbert-Schmidt inner product and $\Delta(\sigma/\rho)$ is the relative modular operator introduced by Araki [**4**]:

$$\Delta(\sigma/\rho) : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})\,, \quad X \mapsto \sigma X \rho^{-1}.$$

If $A = I$, we simply write $S_f\left(\rho||\sigma\right)$. Quasi entropies were introduced by *Dénes Petz,* see [**76**] and [**75**].

The following statement — which can be verified by direct computations — appeared in the paper [**83**] and it makes the relative quasi-entropy easy to compute in some cases.

LEMMA 26. *Let the spectral decomposition of the positive definite operators ρ and σ be given by*

$$\rho = \sum_{j=1}^{n} \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right| \ \textit{and}\ \sigma = \sum_{k=1}^{n} \mu_k \left|\psi_k\right\rangle\left\langle\psi_k\right|.$$

Then we have

$$S_f^A\left(\rho\,||\,\sigma\right) = \sum_{j,k} \lambda_j f\left(\frac{\mu_k}{\lambda_j}\right)\left|\left\langle\psi_k\right| A \left|\varphi_j\right\rangle\right|^2. \tag{22}$$

PROOF. As $\{\varphi_j\}_{j=1}^n$ and $\{\psi_k\}_{k=1}^n$ are orthonormal bases in $\mathcal{H}$, the set $\left\{\left|\psi_k\right\rangle\left\langle\varphi_j\right|\right\}_{j,k=1}^n$ forms an orthonormal basis of $\mathcal{B}(\mathcal{H})$ (with respect to the Hilbert-Schmidt inner product.) It is easy to check that with the notation $v_{jk} := \left|\psi_k\right\rangle\left\langle\varphi_j\right|$ we can write the relative modular operator as

$$\Delta(\sigma/\rho) = \sum_{j,k} \frac{\mu_k}{\lambda_j} \left|v_{jk}\right\rangle\left\langle v_{jk}\right|, \tag{23}$$

where $\left|v_{jk}\right\rangle\left\langle v_{jk}\right| : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is defined by

$$\left|v_{jk}\right\rangle\left\langle v_{jk}\right|(X) := v_{jk} \operatorname{Tr} v_{jk}^* X.$$

Therefore,

$$f\left(\Delta(\sigma/\rho)\right) = \sum_{j,k} f\left(\frac{\mu_k}{\lambda_j}\right)\left|v_{jk}\right\rangle\left\langle v_{jk}\right|.$$

Direct computation shows that

$$\left\langle A\rho^{\frac{1}{2}}, \left|v_{jk}\right\rangle\left\langle v_{jk}\right|\left(A\rho^{\frac{1}{2}}\right)\right\rangle =$$

$$= \operatorname{Tr}\left(\sum_a \lambda_a^{\frac{1}{2}} \left|\varphi_a\right\rangle\left\langle\varphi_a\right| A^* \left|\psi_k\right\rangle\left\langle\varphi_j\right| \operatorname{Tr}\left(\left|\varphi_j\right\rangle\left\langle\psi_k\right| A \sum_b \lambda_b^{\frac{1}{2}} \left|\varphi_b\right\rangle\left\langle\varphi_b\right|\right)\right)$$

$$= \sum_{a,b} \lambda_a^{\frac{1}{2}} \lambda_b^{\frac{1}{2}} \operatorname{Tr}\left(\left|\varphi_a\right\rangle\left\langle\varphi_a\right| A^* \left|\psi_k\right\rangle\left\langle\varphi_j\right| \operatorname{Tr}\left(\left\langle\varphi_b|\varphi_j\right\rangle\left\langle\psi_k\right| A \left|\varphi_b\right\rangle\right)\right)$$

$$= \sum_{a,b} \lambda_a^{\frac{1}{2}} \lambda_b^{\frac{1}{2}} \delta_{bj} \left\langle\psi_k\right| A \left|\varphi_b\right\rangle \operatorname{Tr}\left(\left|\varphi_a\right\rangle\left\langle\varphi_a\right| A^* \left|\psi_k\right\rangle\left\langle\varphi_j\right|\right)$$

$$= \sum_{a,b} \lambda_a^{\frac{1}{2}} \lambda_b^{\frac{1}{2}} \delta_{bj}\delta_{aj} \left\langle\psi_k\right| A \left|\varphi_b\right\rangle\left\langle\varphi_a\right| A^* \left|\psi_k\right\rangle = \lambda_j \left|\left\langle\psi_k\right| A \left|\varphi_j\right\rangle\right|^2,$$

therefore

$$S_f^A(\rho\,||\,\sigma)$$

$$= \left\langle A\rho^{\frac{1}{2}}, \sum_{j,k} f\left(\frac{\mu_k}{\lambda_j}\right) \left|v_{jk}\right\rangle\left\langle v_{jk}\right| \left(A\rho^{\frac{1}{2}}\right)\right\rangle = \sum_{j,k} \lambda_j f\left(\frac{\mu_k}{\lambda_j}\right) \left|\left\langle\psi_k\right| A \left|\varphi_j\right\rangle\right|^2.$$

□

A short and elegant proof of the monotonicity of the relative entropy is given by Nielsen and Petz in [**70**]. The statement is that if $A, B \in \mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathscr{K})$ are positive definite operators and we set $A_1 = \operatorname{Tr}_2 A,\ B_1 = \operatorname{Tr}_2 B$, then for an operator convex function f the following inequality holds:

$$S_f(A||B) \geq S_f(A_1||B_1). \tag{24}$$

There is a useful extension of this monotonicity in [**83**]. Now we restate Sharma's proof, which is essentially the same as the proof of Nielsen and Petz [**70**].

LEMMA 27. *Suppose that $A, B \in \mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathscr{K})$ are positive definite operators and let us use the notations $A_1 = \operatorname{Tr}_2 A,\ B_1 = \operatorname{Tr}_2 B$. If f is an operator convex function then for any $T \in \mathscr{B}(\mathscr{H})$ the following inequality holds:*

$$S_f^{T\otimes I_{\mathscr{B}(\mathscr{K})}}(A||B) \geq S_f^T(A_1||B_1), \tag{25}$$

where $I_{\mathscr{B}(\mathscr{K})}$ is the identity in $\mathscr{B}(\mathscr{K})$.

PROOF. Let us consider the linear map

$$\mathscr{U} : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathscr{K});\ X \mapsto \mathscr{U}(X) := \left(XA_1^{-\frac{1}{2}} \otimes I_{\mathscr{B}(\mathscr{K})}\right) A^{\frac{1}{2}}.$$

We can check that $\mathscr{U}$ is an isometry. For $X, Y \in \mathscr{B}(\mathscr{H})$,

$$\begin{aligned}\langle \mathscr{U}(X), \mathscr{U}(Y)\rangle &= \operatorname{Tr}\left(A^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}X^*\otimes I_{\mathscr{B}(\mathscr{K})}\right)\left(YA_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)A^{\frac{1}{2}}\right)\\ &= \operatorname{Tr}\left(A\left(A_1^{-\frac{1}{2}}X^*YA_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right) = \operatorname{Tr}A_1\left(A_1^{-\frac{1}{2}}X^*YA_1^{-\frac{1}{2}}\right)\\ &= \operatorname{Tr}X^*Y = \langle X, Y\rangle .\end{aligned}$$

The short computation

$$\begin{aligned}&\langle Y, \mathscr{U}(X)\rangle\\ &= \operatorname{Tr}Y^*\left(XA_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)A^{\frac{1}{2}} = \operatorname{Tr}\left(YA^{\frac{1}{2}}\right)^*\left(X\otimes I_{\mathscr{B}(\mathscr{K})}\right)\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\\ &= \operatorname{Tr}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\left(YA^{\frac{1}{2}}\right)^*\left(X\otimes I_{\mathscr{B}(\mathscr{K})}\right)\\ &= \operatorname{Tr}\left(YA^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right)^*\left(X\otimes I_{\mathscr{B}(\mathscr{K})}\right)\\ &\operatorname{Tr}\left(\operatorname{Tr}_2\left(YA^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right)\right)^*X = \left\langle \operatorname{Tr}_2\left(YA^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right), X\right\rangle\end{aligned}$$

shows that the adjoint of $\mathscr{U}$ (which will be denoted by $\mathscr{U}^*$) is the map

$$Y \mapsto \operatorname{Tr}_2\left(YA^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right).$$

One can see that $\mathscr{U}$ admits the beautiful relation

$$\mathscr{U}^*\Delta(B/A)\mathscr{U} = \Delta(B_1/A_1). \tag{26}$$

If $X \in \mathscr{B}(\mathscr{H})$, then

$$\begin{aligned}\mathscr{U}^*\Delta(B/A)\mathscr{U}(X) &= \operatorname{Tr}_2\left(B\left(XA_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)A^{\frac{1}{2}}A^{-1}A^{\frac{1}{2}}\left(A_1^{-\frac{1}{2}}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right)\\ &= \operatorname{Tr}_2\left(B\left(XA_1^{-1}\otimes I_{\mathscr{B}(\mathscr{K})}\right)\right) = B_1XA_1^{-1} = \Delta(B_1/A_1)(X).\end{aligned}$$

By definition of the relative entropy and by (26), the right-hand-side of (25) can be written as

$$\begin{aligned}S_f^T(A_1\|B_1) &= \left\langle TA_1^{\frac{1}{2}}, f(\Delta(B_1/A_1))\left(TA_1^{\frac{1}{2}}\right)\right\rangle\\ &= \left\langle TA_1^{\frac{1}{2}}, f\left(\mathscr{U}^*\Delta(B/A)\mathscr{U}\right)\left(TA_1^{\frac{1}{2}}\right)\right\rangle .\end{aligned}$$

The operator convexity of f implies that

$$f\left(\mathscr{U}^*\Delta(B/A)\mathscr{U}\right) \le \mathscr{U}^*f(\Delta(B/A))\mathscr{U}$$

(see Chapter 5 of [**9**]), and $\mathcal{U}\left(TA_1^{\frac{1}{2}}\right) = \left(T\otimes I_{\mathcal{B}(\mathcal{K})}\right)A^{\frac{1}{2}}$ is immediate. Therefore,

$$\left\langle TA_1^{\frac{1}{2}}, f\left(\mathcal{U}^*\Delta(B/A)\mathcal{U}\right)\left(TA_1^{\frac{1}{2}}\right)\right\rangle \leq \left\langle \mathcal{U}\left(TA_1^{\frac{1}{2}}\right), f(\Delta(B/A))\left(\mathcal{U}\left(TA_1^{\frac{1}{2}}\right)\right)\right\rangle$$
$$= \left\langle \left(T\otimes I_{\mathcal{B}(\mathcal{K})}\right)A^{\frac{1}{2}}, f(\Delta(B/A))\left(\left(T\otimes I_{\mathcal{B}(\mathcal{K})}\right)A^{\frac{1}{2}}\right)\right\rangle = S_f^{T\otimes I_{\mathcal{B}(\mathcal{K})}}(A||B),$$

and the proof is complete. □

2.4. From the relative entropy to the strong subadditivity. As we have seen in Subsubection 2.2, the strong subadditivity of the Tsallis entropy holds if and only if $q = 1$. Therefore, our goal is to find an inequality

$$S_q(\rho_{123}) + S_q(\rho_2) \leq S_q(\rho_{12}) + S_q(\rho_{23}) + g_q(\rho_{123}), \tag{27}$$

where $g_1(\rho_{123}) = 0$. Such a result may be considered as a generalization of the strong subadditivity inequality.

We collect here some elementary facts that will be useful in the following.

(1) For any positive real numbers x, y and q the identity

$$\ln_q x - \ln_q y = -\ln_q\left(\frac{y}{x}\right) - (q-1)\ln_q\left(\frac{y}{x}\right)\ln_q x \tag{28}$$

holds.

(2) If f and g are $\mathbb{R}\to\mathbb{R}$ functions, $\rho,\sigma \in \mathcal{B}^{sa}(\mathcal{H})$ and the spectral decompositions are $\rho = \sum_j \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ and $\sigma = \sum_k \mu_k \left|\psi_k\right\rangle\left\langle\psi_k\right|$, and domain of f and g contains the spectrum of ρ and σ, respectively, then

$$\operatorname{Tr} f(\rho)g(\sigma) = \sum_{j,k} f(\lambda_j)g(\mu_k)\left|\left\langle\varphi_j|\psi_k\right\rangle\right|^2. \tag{29}$$

(3) If $A \in \mathcal{B}(\mathcal{H})\otimes\mathcal{B}(\mathcal{K})$ and $B \in \mathcal{B}(\mathcal{H})$, then

$$\operatorname{Tr}_2\left(A\cdot B\otimes I_{\mathcal{B}(\mathcal{K})}\right) = (\operatorname{Tr}_2 A)\cdot B. \tag{30}$$

The strong subadditivity of the von Neumnann entropy can be derived from the monotonicity of the Umegaki relative entropy, which is a particular quasi-entropy [**13, 70**]. Therefore, it seems to be useful to reformulate the strong subadditivity of the Tsallis entropy as an inequality of certain quasi-entropies.

THEOREM 28. *Let ρ_{123} be an element of $\mathcal{B}^{++}(\mathcal{H}_1\otimes\mathcal{H}_2\otimes\mathcal{H}_3)$. The strong subadditivity inequality of the Tsallis entropy (17) is equivalent to*

$$S^U_{-\ln_q}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right) \geq S^V_{-\ln_q}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right), \tag{31}$$

where

$$U = \rho_{123}^{\frac{1}{2}(q-1)}, \qquad V = \rho_{23}^{\frac{1}{2}(q-1)}. \tag{32}$$

PROOF. By (30), we have

$$\mathrm{Tr}_3\left(\rho_{123}\ln_q(\rho_{12}\otimes I_3)\right) = \rho_{12}\ln_q(\rho_{12})$$

and

$$\mathrm{Tr}_3\left(\rho_{23}\ln_q(\rho_2\otimes I_3)\right) = \rho_2\ln_q(\rho_2),$$

hence the inequality

$$-S_q(\rho_{123}) + S_q(\rho_{12}) \geq -S_q(\rho_{12}) + S_q(\rho_2),$$

which is obviously equivalent to (17), can be written in the form

$$\mathrm{Tr}\,\rho_{123}\left(\ln_q(\rho_{123}) - \ln_q(\rho_{12}\otimes I_3)\right) \geq \mathrm{Tr}\,\rho_{23}\left(\ln_q(\rho_{23}) - \ln_q(\rho_2\otimes I_3)\right). \tag{33}$$

By (29), if $\rho_{123} = \sum_j \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ and $\rho_{12}\otimes I_3 = \sum_k \mu_k \left|\psi_k\right\rangle\left\langle\psi_k\right|$, then the left hand side of (33) is

$$\sum_{j,k}\lambda_j\left(\ln_q\lambda_j - \ln_q\mu_k\right)\left|\langle\varphi_j|\psi_k\rangle\right|^2.$$

By (20), this expression can be written as

$$\sum_{j,k}\lambda_j\left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\lambda_j^{q-1}\right)\left|\langle\varphi_j|\psi_k\rangle\right|^2. \tag{34}$$

In addition, if $U = \rho_{123}^{\frac{1}{2}(q-1)}$ then

$$\left|\left\langle\psi_k\right|U\left|\varphi_j\right\rangle\right|^2 = \lambda_j^{q-1}\left|\langle\psi_k|\varphi_j\rangle\right|^2,$$

hence by the result of Lemma 26, we can write (34) as the following relative entropy

$$\sum_{j,k}\lambda_j\left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\right)\left|\left\langle\psi_k\right|U\left|\varphi_j\right\rangle\right|^2 = S^U_{-\ln_q}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right). \tag{35}$$

The observation that the right hand side of (33) can be written as a relative entropy similarly to (35) completes the proof. □

Note that in the special case $q = 1$, Theorem 28 states the equivalence of the monotonicity of the Umegaki relative entropy and the strong subadditivity of the von Neumann entropy.

The following theorem provides an inequality which is of the form (27).

THEOREM 29. *Using the notation of the previous theorem, for* $0 < q \leq 2$ *the inequality*

$$S_q(\rho_{12}) + S_q(\rho_{23}) - S_q(\rho_{123}) - S_q(\rho_2)$$
$$\geq (q-1)\left(S_{\ln_q}^{(-\ln_q \rho_{123})^{\frac{1}{2}}}\left(\rho_{123}||\rho_{12}\otimes I_3\right) - S_{\ln_q}^{(-\ln_q \rho_{23})^{\frac{1}{2}}}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right)\right)$$

holds.

As we have claimed that the above inequality is of the form (27), it is not surprising that this statemant recovers the strong subadditivity of the von Neumann entropy if $q = 1$.

PROOF. We noted that if

$$\rho_{123} = \sum_j \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$$

and

$$\rho_{12}\otimes I_3 = \sum_k \mu_k \left|\psi_k\right\rangle\left\langle\psi_k\right|,$$

then the left hand side of (33) is

$$\sum_{j,k} \lambda_j \left(\ln_q \lambda_j - \ln_q \mu_k\right)\left|\left\langle\varphi_j|\psi_k\right\rangle\right|^2.$$

According to (28), it is equal to

$$\sum_{j,k} \lambda_j \left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right) - (q-1)\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\ln_q \lambda_j\right)\left|\left\langle\varphi_j|\psi_k\right\rangle\right|^2, \tag{36}$$

and by Lemma 26, (36) has the form

$$S_{-\ln_q}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right) + (q-1)S_{\ln_q}^{(-\ln_q \rho_{123})^{\frac{1}{2}}}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right). \tag{37}$$

If we rewrite the right hand side of (33) similarly to (37), we get that the strong subadditivity is equivalent to

$$S_{-\ln_q}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right) + (q-1)S_{\ln_q}^{(-\ln_q \rho_{123})^{\frac{1}{2}}}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right)$$

$$\geq S_{-\ln_q}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right) + (q-1)S_{\ln_q}^{(-\ln_q \rho_{23})^{\frac{1}{2}}}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right). \tag{38}$$

It is easy to derive from the Loewner-Heinz theorem [**9, 13, 35**] that $\ln_q x$ is operator monotone, if $0 < q \leq 2$. Any operator monotone function is operator concave [**35**], hence $-\ln_q x$ is operator convex.

By the monotonicity property (25), for $0 < q \leq 2$ we have

$$S_{-\ln_q}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right) \geq S_{-\ln_q}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right),$$

and by (38), this is equivalent to

$$-S_q(\rho_{123})+S_q(\rho_{12})-(q-1)\left(S_{\ln_q}^{(-\ln_q\rho_{123})^{\frac{1}{2}}}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right)\right)$$

$$\geq -S_q(\rho_{23})+S_q(\rho_2)-(q-1)\left(S_{\ln_q}^{(-\ln_q\rho_{23})^{\frac{1}{2}}}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right)\right). \tag{39}$$

This is the statement of the theorem. □

The notation (32) will be used again. Because of the monotonicity property (25), for $0<q\leq 2$ and $f(x)=-\ln_q x$, $A=\rho_{123}$, $B=\rho_{12}\otimes I_3$, $T=V$ we have

$$S_{-\ln_q}^{I_1\otimes V}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right)\geq S_{-\ln_q}^{V}\left(\rho_{23}\,||\,\rho_2\otimes I_3\right). \tag{40}$$

This formula is quite similar to the strong subadditivity inequality (31). By (40),

$$S_{-\ln_q}^{U}\left(\rho_{123}\,||\,\rho_{12}\otimes I_3\right)\geq S_{-\ln_q}^{I_1\otimes V}\left(\rho_{123}||\rho_{12}\otimes I_3\right) \tag{41}$$

implies the strong subadditivity (31). We try to find a sufficient condition for (41).

THEOREM 30. *If ρ_{123} and $I_1\otimes\rho_{23}$ commute, and (using the usual notation $\rho_{123}=\sum_j\lambda_j\left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ and $\rho_{12}\otimes I_3=\sum_k\mu_k\left|\psi_k\right\rangle\left\langle\psi_k\right|$) we have $\lambda_j\leq\mu_k$ whenever $\langle\psi_k|\varphi_j\rangle\neq 0$, then for any $1\leq q\leq 2$ the strong subadditivity inequality*

$$S_q(\rho_{123})+S_q(\rho_2)\leq S_q(\rho_{12})+S_q(\rho_{23})$$

holds.

Note that if ρ_{123} is a classical probability distribution (that is, $\rho_{123}=\mathrm{Diag}(\{p_{jkl}\})$), then the conditions of Theorem 30 are clearly satisfied.

PROOF. By Lemma 26, the inequality (41) is equivalent to

$$\sum_{j,k}\left|\left\langle\psi_k\right|\rho_{123}^{\frac{1}{2}(q-1)}\left|\varphi_j\right\rangle\right|^2\left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\right)\lambda_j$$

$$\geq\sum_{j,k}\left|\left\langle\psi_k\right|I_1\otimes\rho_{23}^{\frac{1}{2}(q-1)}\left|\varphi_j\right\rangle\right|^2\left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\right)\lambda_j.$$

If $\lambda_j\leq\mu_k$, then $\left(-\ln_q\left(\frac{\mu_k}{\lambda_j}\right)\right)\lambda_j\leq 0$. On the other hand,

$$\left|\left\langle\psi_k\right|\rho_{123}^{\frac{1}{2}(q-1)}\left|\varphi_j\right\rangle\right|^2=\lambda_j^{q-1}\left|\left\langle\psi_k|\varphi_j\right\rangle\right|^2.$$

If ρ_{123} and $I_1 \otimes \rho_{23}$ commute, then $I_1 \otimes \rho_{23}$ is diagonal in the basis $\{\varphi_j\}_{j\in J}$ and $\rho_{123} \leq I_1 \otimes \rho_{23}$ holds, that is, $I_1 \otimes \rho_{23} = \sum_j \nu_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ with some $\nu_j \geq \lambda_j$.

If $1 \leq q$, the map $t \mapsto t^{(q-1)}$ is monotone on $\mathbb{R}_+$, hence we have

$$\left|\left\langle\psi_k\right| I_1 \otimes \rho_{23}^{\frac{1}{2}(q-1)} \left|\varphi_j\right\rangle\right|^2 = \nu_j^{q-1} \left|\left\langle\psi_k|\varphi_j\right\rangle\right|^2 \geq \lambda_j^{q-1} \left|\left\langle\psi_k|\varphi_j\right\rangle\right|^2.$$

We concluded that if the conditions of Theorem 30 are satisfied, then (41) holds, and hence the proof is complete. □

The following example shows that one can apply Theorem 30 in essentially non-classical cases, as well.

EXAMPLE. Set $p, q \in \left[\frac{1}{2}, 1\right]$ such that $pq \leq 1-q$ and $t \in \mathbb{R}$. Let us define V and Λ by

$$V = \begin{bmatrix} \cos t & 0 & 0 & -\sin t \\ 0 & \cos t & -\sin t & 0 \\ 0 & \sin t & \cos t & 0 \\ \sin t & 0 & 0 & \cos t \end{bmatrix}$$

and

$$\Lambda = \mathrm{Diag}(pq, (1-p)q, p(1-q), (1-p)(1-q)).$$

V describes a family of orthonormal bases, this family can be considered as a one-parameter extension of the *Bell basis*. Let $\rho_1 \in \mathscr{B}(\mathscr{H})$ be an arbitrary density, and $\rho_{23} \in \mathscr{B}\left(\mathbb{C}^2 \otimes \mathbb{C}^2\right)$ be defined by

$$\rho_{23} = V\Lambda V^{-1} = \begin{bmatrix} A_{11} & 0 & 0 & A_{12} \\ 0 & B_{11} & B_{12} & 0 \\ 0 & B_{21} & B_{22} & 0 \\ A_{21} & 0 & 0 & A_{22} \end{bmatrix},$$

where

$$A_{11} = pq\cos^2 t + (1-p)(1-q)\sin^2 t,$$
$$A_{12} = (pq - (1-p)(1-q))\sin t \cos t,$$
$$A_{21} = (pq - (1-p)(1-q))\sin t \cos t,$$
$$A_{22} = (1-p)(1-q)\cos^2 t + pq\sin^2 t$$

and

$$B_{11} = (1-p)q\cos^2 t + p(1-q)\sin^2 t,$$
$$B_{12} = ((1-p)q - p(1-q))\sin t \cos t,$$
$$B_{21} = ((1-p)q - p(1-q))\sin t \cos t,$$
$$B_{22} = p(1-q)\cos^2 t + (1-p)q\sin^2 t.$$

One can easily compute that

$$\rho_2 = \mathrm{Tr}_3\, \rho_{23} = \begin{bmatrix} q\cos^2 t + (1-q)\sin^2 t & 0 \\ 0 & q\sin^2 t + (1-q)\cos^2 t \end{bmatrix}.$$

Let us take the density $\rho_{123} = \rho_1 \otimes \rho_{23}$. The spectrum of ρ_{123} is

$$\bigcup_{j=1}^{m} \{\nu_j pq, \nu_j (1-p)q, \nu_j p(1-q), \nu_j (1-p)(1-q)\},$$

where ν_j's are the eigenvalues of ρ_1. The spectrum of ρ_{12} (or $\rho_{12} \otimes I_3$) is

$$\bigcup_{j=1}^{m} \{\nu_j (q\cos^2 t + (1-q)\sin^2 t), \nu_j (q\sin^2 t + (1-q)\cos^2 t)\}.$$

The assumption $pq \le 1-q$ guarantees that the eigenvalues of ρ_{123} are smaller than the eigenvalues of $\rho_{12} \otimes I_3$, whenever the corresponding eigenvectors are not orthogonal. ρ_{123} and $I_1 \otimes \rho_{23}$ obviously commute, hence the conditions of Theorem 30 are satisfied by ρ_{123} despite the fact that ρ_{23} can not be diagonized in any product basis.

3. Bregman divergences and their use

The strong subadditivity of the von Neumann entropy is not the only celebrated result in the area of finite dimensional functional analysis which was made — partially or in its entirety — by *Eliott Lieb*. For example, the famous concavity theorem which states that the map

$$A \mapsto \mathrm{Tr} \exp\left(H + \log A\right)$$

is concave on $\mathscr{B}^{++}(\mathscr{H})$ for any $H \in \mathscr{B}^{sa}(\mathscr{H})$ if $\mathscr{H}$ is a finite dimensional Hilbert space is also due to Lieb.

In a recent paper, *J. A. Tropp* used the joint convexity of the Umegaki relative entropy to give a succint proof of this celebrated concavity theorem of Lieb [**86**]. This result inspires us to prove the joint convexity property for *Bregman divergences* which may be considered as generalized relative entropies as we will see in the following.

Once we are done with the proof of the convexity for a certain family of Bregman divergences, we are in the position to derive some consequences. It is quite surprising that we managed to generalize the strong subadditivity inequality (and not the concavity theorem) with the use of the convexity of certain Bregman divergences.

3.1. Introduction to Bregman divergences. In applications that involve measuring the dissimilarity between two objects (numbers, vectors, matrices, functions and so on) the definition of a divergence becomes essential. One such measure is a distance function, but there are many important measures which do not satisfy the properties of distance. For instance, the square loss function has been used widely for regression analysis, Kullback-Leibler divergence [**45**] has been applied to compare two probability density functions, the Itakura-Saito divergence [**38**] is used as a measure of the perceptual difference between spectra, or the Mahalonobis distance [**52**] is to measure the dissimilarity between two random vectors of the same distribution. The Bregman divergence was introduced by Lev Bregman [**11**] for convex functions $\phi : \mathbb{R}^d \to \mathbb{R}$ with gradient $\nabla\phi$, as the ϕ-dependent nonnegative measure of discrepancy

$$D_\phi(p,q) = \phi(p) - \phi(q) - \langle \nabla\phi(q), p-q\rangle \tag{42}$$

of d-dimensional vectors $p, q \in \mathbb{R}^d$. Originally his motivation was the problem of convex programming, but it became widely researched both from theoretical and practical viewpoints. For example the remarkable fact that all the aforementioned divergences are special cases of the Bregman divergence shows its importance [**6**]. In some literature it is applied under the name Bregman distance, in spite of that it is not in general a metric. Indeed, D_ϕ is definite, but does not satisfy the triangle inequality nor symmetry. In addition to the wide range of applications in information theory, statistics and computer science, Dénes Petz suggested the extension of the concept of Bregman divergence to operators [**72**]. If C denotes a convex set in a Banach space and $\mathscr{B}(\mathscr{H})$ denotes the bounded linear operators on the Hilbert space $\mathscr{H}$, for an operator valued smooth function $\Psi : C \to \mathscr{B}(\mathscr{H})$ the Bregman operator divergence is defined by

$$D_\Psi(x,y) = \Psi(x) - \Psi(y) - \lim_{t\to+0} \frac{\Psi(y + t(x-y)) - \Psi(y)}{t} \tag{43}$$

for all $x, y \in C$. Since the Bregman operator divergence can be written as

$$D_\Psi(x,y) = \lim_{t\to+0} \frac{t\Psi(x) + (1-t)\Psi(y) - \Psi(tx + (1-t)y)}{t},$$

for operator convex functions Ψ the inequality $D_\Psi(x,y) \geq 0$ remains true for the Loewner partial ordering between self-adjoint operators.

The most important preliminary of our investigations is the work of *H. Bauschke* and *J. Borwein*. They gave a necessary and sufficient condition for the joint convexity of the Bregman divergences on $\mathbb{R}^d$ [**7**]. However, the question about the joint convexity of the trace of Bregman operator divergence has been left open.

3.2. Definition and basic properties. Let the Hilbert space $\mathscr{H}$ be finite dimensional, as usual. Let $f : (0,\infty) \to \mathbb{R}$ be a convex function. Then the induced map

$$\varphi_f : \mathscr{B}^{++}(\mathscr{H}) \to \mathbb{R}, \quad X \mapsto \varphi_f(X) := \operatorname{Tr} f(X)$$

is convex, as well [**13**]. A differentiable convex function is bounded from below by its first-order Taylor polynomial, no matter what the base point is. Therefore, the expression

$$\varphi_f(X) - \varphi_f(Y) - \mathbf{D}\varphi_f[Y](X-Y),$$

where $\mathbf{D}\varphi_f[Y]$ denotes the Fréchet derivative of φ_f at the point Y, is non-negative for any $X, Y \in \mathscr{B}^{++}(\mathscr{H})$. By the linearity of the trace, for any $Y \in \mathscr{B}^{++}(\mathscr{H})$ we have $\mathbf{D}\varphi_f[Y] = \mathrm{Tr} \circ \mathbf{D}f[Y]$, where $\mathbf{D}f[Y]$ denotes the Fréchet derivative of the standard operator function $f : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{sa}(\mathscr{H})$ at Y. Let us define the central object of this section precisely.

DEFINITION. *Let $f \in C^1((0,\infty))$ be a convex function and $X, Y \in \mathscr{B}^{++}(\mathscr{H})$. The Bregman f-divergence of X and Y is defined by*

$$H_f(X,Y) = \mathrm{Tr}\left(f(X) - f(Y) - \mathbf{D}f[Y](X-Y)\right). \tag{44}$$

Note that this definition of the Bregman f-divergence coincides with the trace of the Bregman operator divergence (43), if Ψ is the standard operator function f and $C = \mathscr{B}^{++}(\mathscr{H})$, $\mathscr{H} = \mathbb{C}^n$.

Consider the spectral decomposition $A = \sum_j \lambda_j |\varphi_j\rangle\langle\varphi_j|$ of the positive definite operator A and denote the corresponding *matrix units* by $E_{ij} := |\varphi_i\rangle\langle\varphi_j|$. The Fréchet derivative of the standard operator function $f : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{sa}(\mathscr{H})$, $X \mapsto f(X)$ at the point $A \in \mathscr{B}^{++}(\mathscr{H})$ is

$$\mathbf{D}f[A] = \sum_{i,j} \int_0^1 f'\left(\lambda_j + t(\lambda_i - \lambda_j)\right) \mathrm{d}t \, |E_{ij}\rangle\langle E_{ij}|, \tag{45}$$

where Hermite's formula is used for the divided difference matrix [**35**, Thm. 3.33]. Remark that the Fréchet derivative $\mathbf{D}f[A]$ is an $\mathscr{B}^{sa}(\mathscr{H}) \to \mathscr{B}^{sa}(\mathscr{H})$ map, so the formula (45) holds in the sense that the left hand side of (45) is equal to the right hand side of (45) restricted to $\mathscr{B}^{sa}(\mathscr{H})$.

If f is differentiable at A, then the identities

$$\mathbf{D}f[A](B) = \left.\frac{\mathrm{d}}{\mathrm{d}t} f(A+tB)\right|_{t=0}$$

and

$$\mathrm{Tr}\left(\left.\frac{\mathrm{d}}{\mathrm{d}t} f(A+tB)\right|_{t=0}\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Tr}\, f(A+tB)\right|_{t=0} = \mathrm{Tr}\, f'(A)B$$

hold and — in particular — show that

$$H_f(X,Y) = \mathrm{Tr}\left(f(X) - f(Y) - f'(Y)(X-Y)\right). \tag{46}$$

LEMMA 31. *If $f \in C^2((0,\infty))$, the Bregman divergence admits the integral representation*

$$H_f(X,Y) = \int_{s=0}^1 (1-s)\,\mathrm{Tr}(X-Y)\mathbf{D}f'[Y+s(X-Y)](X-Y)\mathrm{d}s. \tag{47}$$

PROOF. Remark that

$$\begin{aligned}\operatorname{Tr} f(X) - \operatorname{Tr} f(Y) &= \int_{t=0}^{1} \frac{\mathrm{d}}{\mathrm{d}t} \operatorname{Tr} f(Y + t(X-Y))\mathrm{d}t \\ &= \int_{t=0}^{1} \operatorname{Tr} f'(Y + t(X-Y))(X-Y)\mathrm{d}t\end{aligned}$$

and

$$\begin{aligned}f'(Y + t(X-Y)) - f'(Y) &= \int_{s=0}^{t} \frac{\mathrm{d}}{\mathrm{d}s} f'(Y + s(X-Y))\mathrm{d}s \\ &= \int_{s=0}^{t} \mathbf{D}f'[Y + s(X-Y)](X-Y)\mathrm{d}s,\end{aligned}$$

hence

$$\begin{aligned}H_f(X,Y) &= \int_{t=0}^{1} \operatorname{Tr}\left(\left(\int_{s=0}^{t} \mathbf{D}f'[Y + s(X-Y)](X-Y)\mathrm{d}s\right)(X-Y)\right)\mathrm{d}t \\ &= \int_{t=0}^{1}\int_{s=0}^{t} \operatorname{Tr}(X-Y)\mathbf{D}f'[Y+s(X-Y)](X-Y)\mathrm{d}s\mathrm{d}t \\ &= \int_{s=0}^{1} (1-s)\operatorname{Tr}(X-Y)\mathbf{D}f'[Y+s(X-Y)](X-Y)\mathrm{d}s.\end{aligned}$$

□

3.3. A characterization of the joint convexity. In this section we investigate the Bregman f-divergence from the viewpoint of joint convexity, which is essential in the further applications. Since f is convex, it is clear that the Bregman divergence is convex in the first variable. For the original Bregman divergence (42) Bauschke and Borwein show [**7**] that D_ϕ is jointly convex - i. e.

$$D_\phi(tp_1 + (1-t)p_2, tq_1 + (1-t)q_2) \le tD_\phi(p_1,q_1) + (1-t)D_\phi(p_2,q_2),$$

where $p_1, p_2, q_1, q_2 \in \mathbb{R}^d$, $t \in [0,1]$ - if and only if the inverse of the Hessian of ϕ is concave in Loewner sense. Particularly, if ϕ is an $\mathbb{R} \supset I \to \mathbb{R}$ convex function, then D_ϕ is jointly convex if and only if $1/\phi''$ is concave. From this viewpoint the next characterization is rather interesting.

THEOREM 32. *Let $f \in C^2((0,\infty))$ be a convex function with $f'' > 0$ on $(0,\infty)$. Then the following conditions are equivalent.*

(A) The map

$$\mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}\left(\mathscr{B}^{sa}(\mathscr{H})\right); \quad X \mapsto \left(\mathbf{D}f'[X]\right)^{-1} \tag{48}$$

is operator concave.

(B) The Bregman f-divergence

$$H_f : \mathscr{B}^{++}(\mathscr{H}) \times \mathscr{B}^{++}(\mathscr{H}) \to [0,\infty); \quad (X,Y) \mapsto H_f(X,Y)$$

is jointly convex.

REMARK 33. For a convex function $f \in C^2((0,\infty))$ the property $f'' > 0$ is equivalent to the existence of $\left(\mathbf{D}f'[X]\right)^{-1}$ for every $X \in \mathscr{B}^{++}(\mathscr{H})$. On the one hand, $f'' > 0$ ensures that $\mathbf{D}f'[X] \in \mathscr{B}(\mathscr{B}^{sa}(\mathscr{H}))$ is a positive definite and hence invertible map — see formula (45). On the other hand, if $f''(\lambda) = 0$ for some $\lambda > 0$, then $\mathbf{D}f'[\lambda I] = 0 \in \mathscr{B}(\mathscr{B}^{sa}(\mathscr{H}))$.

In the recent paper [**16**] *Tropp* and *Chen* defined the *Matrix Entropy Class* the following way.

DEFINITION. *The Matrix Entropy Class consists of the real functions defined on* $[0,\infty)$ *that are either affine or satisfy the following conditions.*

- f *is convex and* $f \in C([0,\infty)) \cap C^2((0,\infty))$.
- *For every finite dimensional Hilbert space* $\mathscr{H}$ *the map* $\mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}(\mathscr{B}^{sa}(\mathscr{H}))$; $X \mapsto \left(\mathbf{D}f'[X]\right)^{-1}$ *is concave with respect to the Loewner (or semidefinite) order.*

By this definition, the statement of Theorem 32 is essentially the following: the set of those functions for which the corresponding Bregman divergence is jointly convex coincides with the Matrix Entropy Class defined by Tropp and Chen.

PROOF OF THEOREM 32: Let us prove the direction (A) $\Rightarrow$ (B) first. Let X_i and Y_i be positive definite operators on $\mathscr{H}$ ($i \in \{1,\dots,N\}$) and let α_i be reals such that $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$. Let us use the notations $X = \sum_i \alpha_i X_i$, $Y = \sum_i \alpha_i Y_i$. By the operator concavity of the map $X \mapsto \left(\mathbf{D}f'[X]\right)^{-1}$, for any $0 \leq s \leq 1$ we have

$$\operatorname{Tr}(X-Y)\mathbf{D}f'[Y+s(X-Y)](X-Y)$$

$$= \operatorname{Tr}\left(\sum_i \alpha_i (X_i - Y_i)\right)\left(\left(\mathbf{D}f'\left[\sum_i \alpha_i (Y_i + s(X_i - Y_i))\right]\right)^{-1}\right)^{-1}\left(\sum_i \alpha_i (X_i - Y_i)\right)$$

$$\leq \operatorname{Tr}\left(\sum_i \alpha_i (X_i - Y_i)\right)\left(\sum_i \alpha_i \left(\mathbf{D}f'\left[Y_i + s(X_i - Y_i)\right]\right)^{-1}\right)^{-1}\left(\sum_i \alpha_i (X_i - Y_i)\right).$$

We used that taking the inverse of an operator reverses the semidefinite order. If $\mathscr{H}$ is a Hilbert space, then the map

$$\mathscr{H} \times \mathscr{B}^{++}(\mathscr{H}) \to \mathbb{R}; \quad (x,T) \mapsto \left\langle x, T^{-1}x\right\rangle$$

is convex (see [**28**, Prop. 4.3], which may be obtained as a consequence of [**50**, Thm. 1]). If we apply this property to the Hilbert space $\mathscr{B}^{sa}(\mathscr{H})$ with the Hilbert-Schmidt inner product we get that

$$\operatorname{Tr}\left(\sum_i \alpha_i (X_i - Y_i)\right)\left(\sum_i \alpha_i \left(\mathbf{D}f'\left[Y_i + s(X_i - Y_i)\right]\right)^{-1}\right)^{-1}\left(\sum_i \alpha_i (X_i - Y_i)\right)$$

$$\le \sum_i \alpha_i \operatorname{Tr}(X_i - Y_i)\left(\left(\mathbf{D}f'\left[(Y_i + s(X_i - Y_i))\right]\right)^{-1}\right)^{-1}(X_i - Y_i)$$
$$= \sum_i \alpha_i \operatorname{Tr}(X_i - Y_i)\mathbf{D}f'\left[(Y_i + s(X_i - Y_i))\right](X_i - Y_i).$$

The result of Lemma 31 (eq. (47)) clearly shows that the obtained inequality

$$\operatorname{Tr}(X - Y)\mathbf{D}f'[Y + s(X - Y)](X - Y)$$
$$\le \sum_i \alpha_i \operatorname{Tr}(X_i - Y_i)\mathbf{D}f'\left[(Y_i + s(X_i - Y_i))\right](X_i - Y_i)$$

implies the joint convexity of the Bregman divergence.

The proof of (B) $\Rightarrow$ (A) is the following. The conditon (B) means that if $A_i \in \mathscr{B}^{++}(\mathscr{H})$, $B_i \in \mathscr{B}^{sa}(\mathscr{H})$ and $\alpha_i \ge 0$, $\sum_i \alpha_i = 1$, then

$$H_f\left(\sum_i \alpha_i(A_i + \varepsilon B_i), \sum_i \alpha_i A_i\right) \le \sum_i \alpha_i H_f\left(A_i + \varepsilon B_i, A_i\right), \tag{49}$$

where $\varepsilon < \varepsilon_0$ for some $\varepsilon_0 > 0$. By the integral representation (47), the right hand side of (49) can be written as

$$\sum_i \alpha_i H_f\left(A_i + \varepsilon B_i, A_i\right) = \sum_i \alpha_i \int_{s=0}^{1} (1-s)\operatorname{Tr}\varepsilon B_i \mathbf{D}f'[A_i + s\varepsilon B_i](\varepsilon B_i)\mathrm{d}s$$
$$= \varepsilon^2 \int_{s=0}^{1} (1-s)\sum_i \alpha_i \operatorname{Tr} B_i \mathbf{D}f'[A_i + s\varepsilon B_i](B_i)\mathrm{d}s.$$

Similarly, the left hand side is

$$H_f\left(\sum_i \alpha_i\left(A_i + \varepsilon B_i\right), \sum_i \alpha_i A_i\right)$$
$$= \varepsilon^2 \int_{s=0}^{1} (1-s)\operatorname{Tr}\left(\sum_i \alpha_i B_i\right)\mathbf{D}f'\left[\sum_i \alpha_i(A_i + s\varepsilon B_i)\right]\left(\sum_i \alpha_i B_i\right)\mathrm{d}s.$$

The assumption $f \in C^2((0,\infty))$ ensures that the map $\mathbf{D}f' : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}\left(\mathscr{B}^{sa}(\mathscr{H})\right)$ is continuous. Therefore, $\lim_{\varepsilon\to 0}\mathbf{D}f'[A_i + s\varepsilon B_i] = \mathbf{D}f'[A_i]$ etc. After division by ε^2 and taking the limit $\varepsilon \to 0$ we obtain from (49) that

$$\operatorname{Tr}\left(\sum_i \alpha_i B_i\right)\mathbf{D}f'\left[\sum_i \alpha_i A_i\right]\left(\sum_i \alpha_i B_i\right) \le \sum_i \alpha_i \operatorname{Tr} B_i \mathbf{D}f'[A_i](B_i),$$

that is, the map

$$\mathscr{B}^{++}(\mathscr{H}) \times \mathscr{B}^{sa}(\mathscr{H}) \ni (A, B) \mapsto \operatorname{Tr} B\mathbf{D}f'[A](B) \tag{50}$$

is jointly convex. This is sufficient to show the opearator concavity of the map $X \mapsto \left(\mathbf{D}f'[X]\right)^{-1}$ by the followings. Let $A_i \in \mathscr{B}^{++}(\mathscr{H})$ ($i \in \{1,\ldots,N\}$)

and $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$. Let us use the short notation $T_i = \mathbf{D}f'[A_i]$. For any $C \in \mathscr{B}^{sa}(\mathscr{H})$ we can define

$$B_i := \left(\mathbf{D}f'[A_i]\right)^{-1} \circ \left(\sum_j \alpha_j \left(\mathbf{D}f'[A_j]\right)^{-1}\right)^{-1}(C) \equiv T_i^{-1} \circ \left(\sum_j \alpha_j T_j^{-1}\right)^{-1}(C).$$

Observe that by this definition $\sum_i \alpha_i B_i = C$. On the one hand,

$$\sum_i \alpha_i \operatorname{Tr} B_i \mathbf{D}f'[A_i](B_i) = \sum_i \alpha_i \operatorname{Tr} B_i T_i(B_i)$$

$$= \sum_i \alpha_i \operatorname{Tr}\left(T_i^{-1} \circ \left(\sum_j \alpha_j T_j^{-1}\right)^{-1}(C) \cdot T_i \circ T_i^{-1} \circ \left(\sum_j \alpha_j T_j^{-1}\right)^{-1}(C)\right)$$

$$= \operatorname{Tr}\left(\left(\sum_i \alpha_i T_i^{-1}\right) \circ \left(\sum_j \alpha_j T_j^{-1}\right)^{-1}(C) \cdot \left(\sum_j \alpha_j T_j^{-1}\right)^{-1}(C)\right)$$

$$= \operatorname{Tr} C \cdot \left(\sum_i \alpha_i T_i^{-1}\right)^{-1}(C) = \operatorname{Tr} C \left(\sum_i \alpha_i \left(\mathbf{D}f'[A_i]\right)^{-1}\right)^{-1}(C).$$

On the other hand,

$$\operatorname{Tr}\left(\sum_i \alpha_i B_i\right) \mathbf{D}f'\left[\sum_i \alpha_i A_i\right]\left(\sum_i \alpha_i B_i\right) = \operatorname{Tr} C \mathbf{D}f'\left[\sum_i \alpha_i A_i\right](C)$$

By the joint convexity of (50),

$$\operatorname{Tr} C \mathbf{D}f'\left[\sum_i \alpha_i A_i\right](C) \leq \operatorname{Tr} C \left(\sum_i \alpha_i \left(\mathbf{D}f'[A_i]\right)^{-1}\right)^{-1}(C)$$

holds, and C was an arbitrary element of $\mathscr{B}^{sa}(\mathscr{H})$, hence the operator inequality

$$\mathbf{D}f'\left[\sum_i \alpha_i A_i\right] \leq \left(\sum_i \alpha_i \left(\mathbf{D}f'[A_i]\right)^{-1}\right)^{-1}$$

holds, which is equivalent to

$$\left(\mathbf{D}f'\left[\sum_i \alpha_i A_i\right]\right)^{-1} \geq \sum_i \alpha_i \left(\mathbf{D}f'[A_i]\right)^{-1}.$$

This is the desired concavity property. □

3.4. An extension of the Bregman divergence to singular operators. In quantum information theory, the singular density operators play a central role, therefore, we would like to extend the Bregman f-divergences from $\mathscr{B}^{++}(\mathscr{H})\times\mathscr{B}^{++}(\mathscr{H})$ to $\mathscr{B}^{+}(\mathscr{H})\times\mathscr{B}^{+}(\mathscr{H})$. It is a natural idea to define the Bregman divergence of the positive semidefinite operators X and Y as follows:

$$H_f(X,Y):=\lim_{\varepsilon\to 0} H_f\left(X+\varepsilon I, Y+\varepsilon I\right). \tag{51}$$

With the formula (46) at hand, easy computation shows that if X and Y admit the spectral decompositions $X=\sum_{j=1}^n \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ and $Y=\sum_{k=1}^n \mu_k \left|\psi_k\right\rangle\left\langle\psi_k\right|$, then

$$H_f(X+\varepsilon I, Y+\varepsilon I)$$

$$=\sum_{j,k=1}^n \left|\left\langle\varphi_j|\psi_k\right\rangle\right|^2\left(f(\lambda_j+\varepsilon)-f(\mu_k+\varepsilon)-f'(\mu_k+\varepsilon)(\lambda_j-\mu_k)\right). \tag{52}$$

Assume that $f\in C^0([0,\infty))\cap C^1((0,\infty))$, that is, $\lim_{x\to 0} f(x)\in\mathbb{R}$. The convexity of f gives that f' is monotone increasing, hence $\lim_{\varepsilon\to 0} f'(\varepsilon)\in\mathbb{R}$ or $\lim_{\varepsilon\to 0} f'(\varepsilon)=-\infty$.

Clearly, if $\lim_{\varepsilon\to 0} f'(\varepsilon)\in\mathbb{R}$, then the limit of (52) is a real number. If $\ker(Y)\subseteq\ker(X)$, then $\lambda_j=0$ whenever $\mu_k=0$ and $\left\langle\varphi_j|\psi_k\right\rangle\neq 0$, hence the limit is finite in this case, as well. If $\ker(Y)\nsubseteq\ker(X)$ and $\lim_{\varepsilon\to 0} f'(\varepsilon)=-\infty$, then the limit is $+\infty$.

So we conclude that if f is continuous at 0, then (51) is well-defined and takes values in $[0,\infty]$, that is, the Bregman f-divergences can be extended to $\mathscr{B}^{+}(\mathscr{H})\times\mathscr{B}^{+}(\mathscr{H})$ by continuity.

If $H_f(\cdot,\cdot)$ defined by a convex function $f\in C^2((0,\infty))$ is jointly convex on $\mathscr{B}^{++}(\mathscr{H})\times\mathscr{B}^{++}(\mathscr{H})$, then (assuming in addition that $f\in C^0([(0,\infty)))$ $H_f(\cdot,\cdot)$ is jointly convex on $\mathscr{B}^{+}(\mathscr{H})\times\mathscr{B}^{+}(\mathscr{H})$. Indeed, for $X_k, Y_k\in\mathscr{B}^{+}(\mathscr{H}), c_k\geq 0, \sum_k c_k=1$ we have

$$H_f\left(\sum_k c_k X_k, \sum_k c_k Y_k\right)=\lim_{\varepsilon\to 0} H_f\left(\sum_k c_k X_k+\varepsilon I, \sum_k c_k Y_k+\varepsilon I\right)$$

$$=\lim_{\varepsilon\to 0} H_f\left(\sum_k c_k(X_k+\varepsilon I), \sum_k c_k(Y_k+\varepsilon I)\right)\leq\lim_{\varepsilon\to 0}\sum_k c_k H_f\left(X_k+\varepsilon I, Y_k+\varepsilon I\right)$$

$$=\sum_k c_k\lim_{\varepsilon\to 0} H_f\left(X_k+\varepsilon I, Y_k+\varepsilon I\right)=\sum_k c_k H_f\left(X_k, Y_k\right).$$

Therefore, we can reformulate the main condition with a bit different conditions.

THEOREM 34. *Let $f \in C^0([0,\infty)) \cap C^2((0,\infty))$ be a convex function with $f'' > 0$ on $(0,\infty)$. Then the following conditions are equivalent.*

(i) The map

$$\mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}\left(\mathscr{B}^{sa}(\mathscr{H})\right); \quad X \mapsto \left(\mathbf{D}f'[X]\right)^{-1}$$

is operator concave.

(ii) The Bregman f-divergence

$$H_f : \mathscr{B}^+(\mathscr{H}) \times \mathscr{B}^+(\mathscr{H}) \to [0,\infty]]; \quad (X,Y) \mapsto H_f(X,Y)$$

is jointly convex.

3.5. A different condition and alternative proofs. In a recent preprint Hansen and Zhang investigated the connections between the condition (A) in Theorem 32 and the property that f'' is operator convex and numerically non-increasing. In an earlier version of their paper, these conditions were claimed to be equivalent [**30**, Thm 1.2]. Later the proof turned out to be incomplete. In the current version it is proved that if f'' is operator convex and numerically non-increasing, then the condition (A) in Theorem 32 is satisfied [**31**, Thm 1.3].

Now we give a direct proof of the fact that the operator convexity (and the non-increasing property) of f'' is sufficient to deduce the joint convexity of the Bregman f-divergence.

LEMMA 35. *Set $f \in C^1((0,\infty))$ and $A \in \mathscr{B}^{++}(\mathscr{H})$. Then the Fréchet derivative of the standard operator function f is*

$$\mathbf{D}f[A] = \int_0^1 f'(tL_A + (1-t)R_A)\,\mathrm{d}t,$$

where L_A (R_A) denotes the left (right) multiplication by A:

$$L_A : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H}),\ X \mapsto L_A(X) := AX,$$

$$R_A : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H}),\ X \mapsto R_A(X) := XA.$$

PROOF. Let us use the notations $A = \sum_j \lambda_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ and $E_{ij} = \left|\varphi_i\right\rangle\left\langle\varphi_j\right|$ again. It is easy to check that

$$L_A(E_{ij}) = \lambda_i E_{ij},\ R_A(E_{ij}) = \lambda_j E_{ij},$$

hence with $P_{ij} := \left|E_{ij}\right\rangle\left\langle E_{ij}\right|$ we have

$$L_A = \sum_{i,j} \lambda_i P_{ij},\ R_A = \sum_{i,j} \lambda_j P_{ij}.$$

Therefore

$$f'(tL_A + (1-t)R_A) = \sum_{i,j} f'(t\lambda_i + (1-t)\lambda_j) P_{ij},$$

and

$$\int_0^1 f'(tL_A + (1-t)R_A)\,\mathrm{d}t = \sum_{i,j} \int_0^1 f'\left(\lambda_j + t(\lambda_i - \lambda_j)\right)\mathrm{d}t \left|E_{ij}\right\rangle\left\langle E_{ij}\right|,$$

and this is exactly the formula that appeared in (45). □

Lemma 31 and Lemma 35 have an immediate consequence.

COROLLARY 36. *For $f \in C^2((0,\infty))$, the Bregman divergence can be written as*

$$H_f(X,Y)$$

(53)

$$= \int_{s=0}^1 \int_{t=0}^1 (1-s)\,\mathrm{Tr}(X-Y) f''\left(tL_{Y+s(X-Y)} + (1-t)R_{Y+s(X-Y)}\right)(X-Y)\,\mathrm{d}t\,\mathrm{d}s.$$

Therefore we can provide a sufficient condition for the joint convexity of the Bregman divergence.

THEOREM 37. *Let $f \in C^2((0,\infty))$ be a convex function. If f'' is operator convex and numerically non-increasing, then the Bregman f-divergence*

$$H_f : \mathscr{B}^{++}(\mathscr{H}) \times \mathscr{B}^{++}(\mathscr{H}) \to [0,\infty); \quad (X,Y) \mapsto H_f(X,Y)$$

is jointly convex.

FIRST PROOF OF THEOREM 37. On a Hilbert space $\mathscr{H}$ the map

$$\mathscr{B}(\mathscr{H})^{++} \times \mathscr{H} \to \mathbb{R} : \ (A,\xi) \mapsto \left\langle \xi, \varphi(A)(\xi)\right\rangle$$

is jointly convex if $\varphi : (0,\infty) \to \mathbb{R}$ is operator convex and numerically non-increasing. This fact relies on the joint convexity of the map $(A,\xi) \mapsto \left\langle \xi, A^{-1}\xi\right\rangle$ — which is stated in [**28**, Prop. 4.3] and may be derived from [**50**, Thm. 1] — and on an integral representation of the functions φ with the above property. This representation will be discussed in the second proof of this theorem. The maps

$$(X,Y) \mapsto tL_{Y+s(X-Y)} + (1-t)R_{Y+s(X-Y)}$$

and $(X,Y) \mapsto X - Y$ are affine, and with the Hilbert-Schmidt inner product (53) can be written as

$$H_f(X,Y)$$

(54)

$$= \int_{s=0}^1 \int_{t=0}^1 (1-s)\left\langle X-Y, f''\left(tL_{Y+s(X-Y)} + (1-t)R_{Y+s(X-Y)}\right)(X-Y)\right\rangle \mathrm{d}t\,\mathrm{d}s,$$

hence H_f is jointly convex if f'' is operator convex and non-increasing. □

We provide another proof of this theorem.

PROPOSITION 38. *Let $\mathscr{F}(A,B)$ denote the set of all $A \to B$ functions. The map*

$$H : C^2((0,\infty)) \to \mathscr{F}\left(\mathscr{B}^{++}(\mathscr{H}) \times \mathscr{B}^{++}(\mathscr{H}), \mathbb{R}\right),\ f \mapsto H_f(\cdot,\cdot) \tag{55}$$

is linear, and the kernel is the subspace of affine functions, that is,

$$\ker(H) = \{x \mapsto ax + b \,|\, a, b \in \mathbb{R}\}.$$

PROOF. The linearity is obvious, and from the integral formula (54) it is easy to see that the kernel of H is equal to the kernel of the operator

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} : C^2((0,\infty)) \to C((0,\infty)).$$

□

Therefore, if f can be written as $f = \sum_{j=1}^{k} f_j$, where the f_j's define jointly convex Bregman divergence, then $H_f(\cdot,\cdot)$ is jointly convex. The affine part of a function can be omitted.

SECOND PROOF OF THEOREM 37. If $f \in C^2((0,\infty))$ is convex function, then f'' is numerically non-increasing and operator convex if and only if

$$f''(x) = \gamma + \int_0^\infty \frac{1}{\lambda + x} \mathrm{d}\mu(\lambda), \tag{56}$$

where $\gamma \geq 0$ and μ is a nonnegative measure on $[0,\infty)$ such that

$$\int_0^\infty \frac{1}{1+\lambda} \mathrm{d}\mu(\lambda) < \infty.$$

This fact is stated in this form in [**3**, Thm. 3.1]. Now we intend to use the 'only if' direction, hence we outline the key steps of the proof of Ando and Hiai. A rather complex argument shows that a numerically non-increasing and operator convex function is operator monotone decreasing. At this point, Ando and Hiai provide a slightly simplified version of the original argument of Hansen [**29**] to verify the integral representation. This is the following. If f'' is operator monotone decreasing then $f''\left(\frac{1}{x}\right)$ is operator monotone, hence by [**9**, pp. 144-145] it has the form

$$f''\left(\frac{1}{x}\right) = \alpha + \beta x + \int_0^\infty \frac{(\lambda+1)x}{\lambda + x} \mathrm{d}\nu(\lambda) \tag{57}$$

where $\alpha, \beta \geq 0$ and ν is a nonnegative finite measure on $(0,\infty)$. Let $\mathrm{d}\tilde{\mu}(\lambda) := \mathrm{d}\nu(\frac{1}{\lambda})$ on $(0,\infty)$ and $\tilde{\mu}(\{0\}) := \beta$. Finally, $\mathrm{d}\mu(\lambda) := (\lambda+1)\mathrm{d}\tilde{\mu}(\lambda)$. By these operations, the representation (57) is transformed to (56).

Integrating (56) two times with respect to x we get

$$f(x) = \alpha + \beta x + \frac{\gamma}{2}x^2 + \int_0^\infty \left((\lambda + x)\left(\log(\lambda + x) - \log(\lambda+1)\right) - (x-1)\right) \mathrm{d}\mu(\lambda), \tag{58}$$

where $\alpha, \beta \in \mathbb{R}$. One can see that $f(x)$ is the sum of the affine part

$$a(x) = \alpha + \beta x - \int_0^\infty \big(\log(\lambda+1)(\lambda+x) + (x-1)\big)\,\mathrm{d}\mu(\lambda),$$

the quadratic part $q(x) = \frac{\gamma}{2}x^2$ and the "entropic" part

$$e(x) = \int_0^\infty (\lambda+x)\log(\lambda+x)\mathrm{d}\mu(\lambda).$$

The quadratic part $H_q(X,Y) = \frac{\gamma}{2}\operatorname{Tr}(X-Y)^2$ is clearly jointly convex. By the result of [**51**], the same statement holds for the Bregman divergence induced by the standard entropy function $\varphi_0(x) = x\log x$,

$$H_{\varphi_0}(X,Y) = \operatorname{Tr}\big(X\big(\log X - \log Y\big) - (X-Y)\big).$$

On the other hand, one can check that the Bregman divergence induced by the shifted entropy function $\varphi_\lambda(x) = (x+\lambda)\log(x+\lambda)$ can be expressed as

$$H_{\varphi_\lambda}(X,Y) = H_{\varphi_0}(X+\lambda I, Y+\lambda I). \tag{59}$$

On the whole, if f'' is numerically decreasing and operator convex, then the Bregman divergence H_f can be written as $H_f = H_q + H_a + H_e$, where $H_a = 0$, H_q is obviously jointly convex and

$$H_e(X,Y) = \int_0^\infty H_{\varphi_\lambda}(X,Y)\mathrm{d}\mu(\lambda) = \int_0^\infty H_{\varphi_0}(X+\lambda I, Y+\lambda I)\mathrm{d}\mu(\lambda). \tag{60}$$

The map $(X,Y) \mapsto (X+\lambda I, Y+\lambda I)$ is affine, hence (60) is jointly convex, and this completes the proof. □

3.6. An application - the Tsallis entropy. The q-logarithm — or deformed logarithm — function $\ln_q : (0,\infty) \to \mathbb{R}$ was defined in (15). Recall that the definition was the following.

$$\ln_q x = \int_1^x t^{q-2}\mathrm{d}t = \begin{cases} \frac{x^{q-1}-1}{q-1} & \text{if } q \neq 1, \\ \ln x & \text{if } q = 1. \end{cases}$$

We also introduced the notation $f_q(x) = x\ln_q(x)$ and claimed that the *Tsallis entropy* of a density operator ρ is given by

$$S_q(\rho) = -\operatorname{Tr} f_q(\rho),$$

see [**2, 17**] for further details. We noted that if $q > 0$, then $\lim_{x\to 0} f_q(x) = 0$, hence f_q can be extended by continuity, thus the Tsallis entropy is well-defined for singular densities, as well. By the result of Tropp and Chen, f_q belongs to the Matrix Entropy Class for $1 \leq q \leq 2$ [**16**, Thm. 2.3]. Therefore, by Theorem 32, $H_{f_q}(\cdot,\cdot)$ is jointly convex. (Alternatively, we may use the well-known operator convexity of the function $x \mapsto x^{-r}$ $(0 \leq r \leq 1)$ and refer to Theorem 37.)

One can compute that for $q \neq 1$ we have

$$H_{f_q}(A,B) = \mathrm{Tr}\, B^q + \frac{1}{q-1}\left(\mathrm{Tr}\, A^q - q\, \mathrm{Tr}\, AB^{q-1}\right).$$

The Bregman divergence is clearly unitary invariant, that is,

$$H_f(UAU^*, UBU^*) = H_f(A,B) \tag{61}$$

for all unitary operators U. (Moreover, as it will turn out in the following, the Bregman divergences are invariant under *anti*unitary conjugations, as well.) Let $X \in \mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathscr{K})$ for some finite timensional Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$, and let us introduce the notation $m := \dim \mathscr{H}$ and $n := \dim \mathscr{K}$. Then there are some unitaries $U_1, \dots, U_{n^2}$ such that

$$X_1 \otimes \frac{1}{n} I_{\mathscr{B}(\mathscr{K})} = \sum_{k=1}^{n^2} \frac{1}{n^2} U_k X U_k^*,$$

where $X_1 = \mathrm{Tr}_2\, X$ and $I_{\mathscr{B}(\mathscr{K})}$ is the identity in $\mathscr{B}(\mathscr{K})$ (see e. g. [**8, 23**]). Therefore, from the joint convexity it follows that the Bregman divergence is monotone in the following sense:

$$H_f\left(X_1 \otimes \frac{1}{n} I_{\mathscr{B}(\mathscr{K})}, Y_1 \otimes \frac{1}{n} I_{\mathscr{B}(\mathscr{K})}\right) \leq H_f(X,Y) \tag{62}$$

if f satisfies the condition (A) in Theorem 32. Let us apply (62) to f_q with $1 < q \leq 2$ and

$$X = \rho_{123} \in \mathscr{B}^+(\mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3),\ Y = \frac{1}{d_1} I_1 \otimes \rho_{23}, \tag{63}$$

where $\mathscr{H}_i$ is a finite dimensional Hilbert space ($i \in \{1,2,3\}$), $d_i = \dim \mathscr{H}_i$ and $\rho_{23} = \mathrm{Tr}_1\, \rho_{123}$. Furthermore, we use the simplified notation $I_i \equiv I_{\mathscr{B}(\mathscr{H}_i)}$ for $i \in \{1,2,3\}$. The idea of this choice of X and Y comes from the tutorial [**70**]. With this choice we get

$$H_{f_q}\left(\rho_{12} \otimes \frac{1}{d_3} I_3, \frac{1}{d_1} I_1 \otimes \rho_2 \otimes \frac{1}{d_3} I_3\right) \leq H_{f_q}\left(\rho_{123}, \frac{1}{d_1} I_1 \otimes \rho_{23}\right). \tag{64}$$

Straightforward computations show that the left hand side of (64) equals to

$$\frac{1}{q-1}\left(d_3^{1-q}\, \mathrm{Tr}\, \rho_{12}^q - (d_1 d_3)^{1-q}\, \mathrm{Tr}\, \rho_2^q\right)$$

and the right hand side is

$$\frac{1}{q-1}\left(\mathrm{Tr}\, \rho_{123}^q - d_1^{1-q}\, \mathrm{Tr}\, \rho_{23}^q\right).$$

The result of this computation can be summarized as follows.

THEOREM 39. *If $\mathscr{H}_i$ is a finite dimensional Hilbert space for any $i \in \{1,2,3\}$, $d_i = \dim \mathscr{H}_i$, $1 \le q \le 2$, then for any $\rho_{123} \in \mathscr{B}^+(\mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3)$ the inequality*

$$d_3^{1-q} \operatorname{Tr} \rho_{12}^q + d_1^{1-q} \operatorname{Tr} \rho_{23}^q \le \operatorname{Tr} \rho_{123}^q + (d_1 d_3)^{1-q} \operatorname{Tr} \rho_2^q. \tag{65}$$

holds, where notations like ρ_{12} denote the appropriate reduced operators.

The fact that the strong subadditivity of the Tsallis entropy

$$\operatorname{Tr} \rho_{12}^q + \operatorname{Tr} \rho_{23}^q \le \operatorname{Tr} \rho_{123}^q + \operatorname{Tr} \rho_2^q$$

does not hold in general [**79**] (but holds for classical probability distributions [**22**]) makes Theorem 39 remarkable. Furthermore, one can not improve on (65), the inequality is sharp.

The example which shows the sharpness of the inequality (65) was used previously to demonstrate that the Tsallis entropy is not strongly subbadditive, see (19). So, the density operator

$$\rho_{123} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \in \mathscr{B}\left(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2\right) \tag{66}$$

has the reduced densities

$$\rho_{12} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \rho_{23} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix}, \quad \rho_2 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix},$$

hence

$$d_3^{1-q} \operatorname{Tr} \rho_{12}^q + d_1^{1-q} \operatorname{Tr} \rho_{23}^q = 2^{1-q} + 2^{1-q} 4^{1-q} = \operatorname{Tr} \rho_{123}^q + (d_1 d_3)^{1-q} \operatorname{Tr} \rho_2^q.$$

Note that (65) is equivalent to

$$(d_1 d_2 d_3)^{1-q} S_q(\rho_{123}) + d_2^{1-q} S_q(\rho_2) + \frac{(d_1 d_2 d_3)^{1-q} - 1}{q-1} + \frac{d_2^{1-q} - 1}{q-1}$$
$$\le (d_2 d_3)^{1-q} S_q(\rho_{23}) + (d_1 d_2)^{1-q} S_q(\rho_{12}) + \frac{(d_2 d_3)^{1-q} - 1}{q-1} + \frac{(d_1 d_2)^{1-q} - 1}{q-1},$$

which gives the strong subadditivity

$$S_q(\rho_{123}) + S_q(\rho_2) \leq S(\rho_{23}) + S(\rho_{12})$$

of the von Neumann entropy, if we take the limit $q \to 1$.

3.7. The relation of joint convexity and monotonicity under stochastic maps. For *homogeneous* relative entropy-type maps, the joint convexity and the monotonicity under stochastic maps is equivalent [**47**, remarks after Def. 2.3]. However, the Bregman divergence does not need to be homegeneous. For example,

$$H_{f_q}(\lambda A, \lambda B) = \lambda^q H_{f_q}(A, B) \tag{67}$$

for $0 < \lambda$ (and any positive q).

We show that a jointly convex Bregman divergence is not monotone in general. In order to see this surprising fact, we create an example that shows that a family of (jointly convex) Bregman divergences increases under the partial trace, which is a very important stochastic (that is, completely positive trace preserving - CPTP) map [**13, 47**].

Easy computations show that for any $A, B \in \mathcal{B}^+(\mathcal{H})$ we have

$$H_f\left(A \otimes \frac{1}{n} I_{\mathcal{B}(\mathcal{K})}, B \otimes \frac{1}{n} I_{\mathcal{B}(\mathcal{K})}\right) = n H_f\left(\frac{A}{n}, \frac{B}{n}\right), \tag{68}$$

where $I_{\mathcal{B}(\mathcal{K})}$ is the identity in $\mathcal{B}(\mathcal{K})$ and $\dim \mathcal{K} = n$. Recall that the density operator (66) saturates the inequality (64), that is,

$$H_{f_q}\left(\rho_{12} \otimes \frac{1}{2} I, \frac{1}{2} I \otimes \rho_2 \otimes \frac{1}{2} I\right) = H_{f_q}\left(\rho_{123}, \frac{1}{2} I \otimes \rho_{23}\right). \tag{69}$$

On the other hand, by (67) and (68),

$$H_{f_q}\left(\rho_{12} \otimes \frac{1}{2} I, \frac{1}{2} I \otimes \rho_2 \otimes \frac{1}{2} I\right) = 2^{1-q} H_{f_q}\left(\rho_{12}, \frac{1}{2} I \otimes \rho_2\right), \tag{70}$$

which means that

$$H_{f_q}\left(\rho_{12}, \frac{1}{2} I \otimes \rho_2\right) > H_{f_q}\left(\rho_{123}, \frac{1}{2} I \otimes \rho_{23}\right),$$

so the monotonicity under partial trace fails.

This means that the joint convexity does not imply monotonicity, but the converse is true. We summarize the results in the next theorem.

THEOREM 40. *Every monotone Bregman divergence is jointly convex. However, there are jointy convex Bregman divergences which are not monotone under stochastic maps. On the other hand, joint convexity implies the monotonicity under the stochastic maps of the form*

$$A \mapsto \sum_k c_k U_k A U_k^*, \tag{71}$$

where the U_k's are unitaries and $c_k \geq 0$, $\sum_k c_k = 1$.

PROOF. Set $X_1, X_2, Y_1, Y_2 \in \mathscr{B}^+(\mathscr{H})$ and let the block operators X, Y and $U \in \mathscr{B}(\mathscr{H} \oplus \mathscr{H})$ be defined by

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \; Y = \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}, \; U = \begin{bmatrix} 0 & I_{\mathscr{B}(\mathscr{H})} \\ I_{\mathscr{B}(\mathscr{H})} & 0 \end{bmatrix}.$$

The map

$$\mathscr{E} : \mathscr{B}(\mathscr{H} \oplus \mathscr{H}) \to \mathscr{B}(\mathscr{H} \oplus \mathscr{H}), \; X \mapsto \mathscr{E}(X) := \frac{1}{2}X + \frac{1}{2}UXU^*$$

is clearly stochastic, and

$$\mathscr{E}(X) = \frac{1}{2}\begin{bmatrix} X_1 + X_2 & 0 \\ 0 & X_1 + X_2 \end{bmatrix}, \mathscr{E}(Y) = \frac{1}{2}\begin{bmatrix} Y_1 + Y_2 & 0 \\ 0 & Y_1 + Y_2 \end{bmatrix}.$$

The Bregman divergence of block-diagonal operators is the sum of the Bregman divergence of the blocks, hence the monotonicity condition

$$H_f(\mathscr{E}(X), \mathscr{E}(Y)) \leq H_f(X, Y)$$

means that

$$2H_f\left(\frac{1}{2}(X_1 + X_2), \frac{1}{2}(Y_1 + Y_2)\right) \leq H_f(X_1, Y_1) + H_f(X_2, Y_2),$$

which is the midpoint convexity of $H_f(\cdot, \cdot)$. The Bregman f-divergence is continuous (by the assumption $f \in C^1((0, \infty))$), hence midpoint convexity implies convexity.

We have shown in this section that for $f_q(x) = \frac{x^q - q}{q-1}$ the corresponding Bregman divergence $H_{f_q}(\cdot, \cdot)$ is jointly convex but it is not monotone under stochastic maps ($1 < q \leq 2$).

In order to check the last statement of the theorem, suppose that $H_f(\cdot, \cdot)$ is jointly convex. If a map $\mathscr{E} : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ has the form (71), then by the unitary invariance of the Bregman divergence,

$$\begin{aligned} &H_f(\mathscr{E}(X), \mathscr{E}(Y)) \\ &= H_f\left(\sum_k c_k U_k X U_k^*, \sum_k c_k U_k Y U_k^*\right) \leq \sum_k c_k H_f\left(U_k X U_k^*, U_k Y U_k^*\right) \\ &= \sum_k c_k H_f(X, Y) = H_f(X, Y). \end{aligned}$$

□

CHAPTER 3

Preserver problems

We mentioned in Subsection 3.6 in the previous chapter of this thesis that the Bregman divergences of positive definite operators are invariant under unitary conjugations — see equation (61). We also noted that unitary conjugations are not the only transformations of the positive definite cone which preserve the Bregman divergences. It is a very natural goal to determine all the transformations on the set of positive definite operators which leave the Bregman divergences invariant. This question leads us to the topic of *preserver problems*.

A preserver problem consists of the following ingredients. Let H be a set. Let $\phi : H \to H$ be a mapping. Let m be a positive integer, let K be a set and let $X : H^m \to K$ be a map. We say that the transformation ϕ *preserves* X, if either

$$X\left(\phi(A_1),\ldots,\phi(A_m)\right) = X(A_1,\ldots,A_m) \qquad (A_1,\ldots,A_m \in H), \tag{72}$$

or

$$X\left(\phi(A_1),\ldots,\phi(A_m)\right) = \phi(X(A_1,\ldots,A_m)) \qquad (A_1,\ldots,A_m \in H) \tag{73}$$

holds, depending on the nature of the map X. (The equation (73) may play the role of the preserver equation only if $K = H$.) For any given sets H, K and mapping X, the solution of the preserver problem is the description of the structure of all the transformations ϕ which preserve X.

The following table enumerates some preserver problems.

H	m	K	X	Equation	Name of the problem
$\mathbb{R}^n$	2	$[0,\infty)$	$(a,b)\mapsto \|a-b\|$	(72)	isometries of $\mathbb{R}^n$
$\mathbf{M}_n$	1	$\mathbb{C}$	$A\mapsto \operatorname{Det} A$	(72)	determinant preserving maps
$\mathbf{M}_n^{sa}$	2	$\{0,1\}$	$(A,B)\mapsto \mathbb{1}_{A\leq B}$	(72)	order preserving maps
$\mathbf{M}_n^{++}$	2	$\mathbf{M}_n^{++}$	$(A,B)\mapsto ABA$	(73)	triple product preserving maps
$\mathbf{M}_n^{+}$	2	$[-\infty,\infty]$	$(A,B)\mapsto S_f(A,B)$	(72)	preservers of the quantum f-divergence
$\mathbf{M}_n^{++}$	m	$\mathbf{M}_n^{++}$	$(A_1,\dots,A_m)\mapsto M_G(A_1,\dots,A_m)$	(73)	preservers of the multivariable geometric mean

The above table makes it transparent that the topic of preserver problems covers a large area of mathematics. An exhaustive description such problems — including *Frobenius' theorem* on determinant preserving maps, the *Mazur-Ulam theorem* on isometries of real normed spaces and *Wigner's theorem* on the symmetry transformations of pure states with respect to the *transition probability* — can be found of in the monography [**60**] written by *Lajos Molnár*.

In this chapter we deal with several preserver problems. In the first section, we investigate the problem which has been already mentioned in the introduction of this chapter, namely the preservers of the Bregman (and Jensen) divergences on positive definite operators.

The second section is devoted to an algebraic preserver problem on the cone of positive definite operators. Namely, we determine the continuous *Jordan triple endomorphisms* of the positive definite operators acting on a two-dimensional Hilbert space. It is rather interesting that the problem was solved for any dimensions $n \geq 3$ several years ago, only the two-dimensional case remained unsolved.

In the third section we derive the description of the continuous endomorphisms of the *Einstein gyrogroup* from the previous result on Jordan triple products. Note that Einstein's velocity addition law was given more than a hundred years ago, but the preservers of that addition have not been determined yet.

1. Maps preserving Bregman and Jensen divergences

1.1. Introduction. In a series of papers [**33, 55, 65, 54**] Lajos Molnár and his coauthors described the structures of surjective maps of the positive definite cones in matrix algebras, or in operators algebras which can be considered generalized isometries meaning that they are transformations which preserve "distances" with respect to given so-called generalized distance measures. This latter notion stands for any function $d : \mathscr{X} \times \mathscr{X} \to [0,\infty)$ on any set $\mathscr{X}$ with the mere property that for $x, y \in \mathscr{X}$ we have $d(x, y) = 0$ if and only if $x = y$. We recall that in several areas of mathematics not only metrics are used to measure nearness of points but also more general functions of this latter kind.

In [**65, 54**] the considered generalized distance measures are of the form $d = d_{N,g}$, where $N(.)$ is a unitarily invariant norm on the underlying matrix algebra or operator algebra, $g : (0,\infty) \to \mathbb{C}$ is a continuous function with the properties

(a1) $g(y) = 0$ if and only if $y = 1$;
(a2) there exists a constant $K > 1$ such that $\left|g\left(y^2\right)\right| \geq K\left|g(y)\right|$, $y > 0$,

and the generalized distance measure $d_{N,g}$ is defined by

$$d_{N,g}(X, Y) = N\left(g\left(Y^{-1/2} X Y^{-1/2}\right)\right) \tag{74}$$

for all positive invertible elements X, Y of the underlying algebra. In the mentioned papers one can see several important examples of that sort of generalized distance measures, many of them having backgrounds in the differential geometry of positive definite matrices or operators. The basic tools in describing the structure of the corresponding generalized isometries have been so-called generalized Mazur-Ulam type theorems and descriptions of certain algebraic isomorphisms (Jordan triple isomorphisms) of the positive definite cones in question.

We determine the structures of generalized isometries with respect to other important types of generalized distance measures. Namely, here we consider Bregman divergences and Jensen divergences. These types of divergences have wide ranging applications in several areas of mathematics. For example, in the recent volume [**69**] on matrix information geometry 3 chapters are devoted to the study of Bregman divergences. One feature of Jensen divergences which justifies their importance is that Bregman divergences can be considered as asymptotic Jensen divergences (see Section 6.2 in [**69**]). We further mention that the famous

Stein's loss and Umegaki's relative entropy are among the most important Bregman divergences. Our basic tool to determine the corresponding preserver transformations is also algebraic in nature but rather different from what we have mentioned in the previous paragraph. Namely, here we use order isomorphisms.

We next define the two basic concepts what we consider now, Bregman divergence and Jensen divergence. Both concepts are connected to convex real functions. Let f be a convex function on the interval $(0,\infty)$. It is known that f is necessarily continuous and the set of points where f is differentiable has at most countable complement. It is a remarkable fact that if f is everywhere differentiable then its derivative f' is automatically continuous (see e.g. [**82**, Corollary 25.5.1]).

Let $\mathscr{H}$ be a finite dimensional Hilbert space, as usual. For a differentiable convex function f on $(0,\infty)$, the Bregman f-divergence on $\mathscr{B}^{++}(\mathscr{H})$ is

$$H_f(X,Y) = \operatorname{Tr}\left(f(X) - f(Y) - f'(Y)(X-Y)\right), \quad X,Y \in \mathscr{B}^{++}(\mathscr{H}),$$

see e.g. formula (5) in [**80**]. If $\lim_{x\to 0+} f(x)$ and $\lim_{x\to 0+} f'(x)$ exist, then f, f' have continuous extensions onto $[0,\infty)$ and the Bregman f-divergence is well-defined and finite for any pair of positive semidefinite operators, too.

For a convex function f on $(0,\infty)$ and for given $\lambda \in (0,1)$, the Jensen $\lambda - f$-divergence on $\mathscr{B}^{++}(\mathscr{H})$ is defined by

$$J_{f,\lambda}(X,Y) = \operatorname{Tr}\left(\lambda f(X) + (1-\lambda) f(Y) - f(\lambda X + (1-\lambda)Y)\right).$$

If $\lim_{x\to 0+} f(x)$ exists, then the Jensen $\lambda - f$-divergence is also well-defined and finite for any pair of positive semidefinite operators.

It is well-known that H_f and $J_{f,\lambda}$ are always nonnegative and they are generalized distance measures if and only if f is strictly convex.

Our main aim is to describe the "symmetries" of the positive definite cone $\mathscr{B}^{++}(\mathscr{H})$ that preserve above type of divergences. This means that we are looking for the structure of all bijective maps $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ for which

$$H_f\left(\phi(X), \phi(Y)\right) = H_f(X,Y), \quad X,Y \in \mathscr{B}^{++}(\mathscr{H})$$

or

$$J_{\lambda,f}\left(\phi(X), \phi(Y)\right) = J_{\lambda,f}(X,Y), \quad X,Y \in \mathscr{B}^{++}(\mathscr{H})$$

holds.

1.2. The results. It is useful to examine first the question that how different the present problem is from the ones that Lajos Molnár and his coauthors have considered in the papers [**33, 55, 65, 54**]. To see clearly the differences we determine below which Bregman divergences, respectively which Jensen divergences are of the form (74). Let us begin with the case of Bregman divergences.

One of the most important Bregman divergence is the one usually denoted by l which corresponds to the strictly convex function $f(x) = -\log x$, $x > 0$ and is called Stein's loss. Apparently, we have

$$l(X,Y) = -\operatorname{Tr}\left(\log X - \log Y - Y^{-1}(X-Y)\right) = \operatorname{Tr} XY^{-1} - \log \operatorname{Det} XY^{-1} - n$$

for all $X, Y \in \mathscr{B}^{++}(\mathscr{H})$, where we have used the identity $\operatorname{Tr}\circ\log = \log\circ\operatorname{Det}$ on $\mathscr{B}^{++}(\mathscr{H})$. On the other hand, the continuous function $g(y) = y - \log y - 1$, $y > 0$ is nonnegative, satisfies the conditions (a1), (a2) (with constant $K = 2$), and with the trace norm $\|.\|_1$ we easily get

$$d_{\|.\|_1,g}(X,Y) = \left\| g\left(Y^{-1/2}XY^{-1/2}\right)\right\|_1 = l(X,Y), \quad X, Y \in \mathscr{B}^{++}(\mathscr{H}).$$

In the next proposition we prove that Stein's loss is essentially the only Bregman divergence on $\mathscr{B}^{++}(\mathscr{H})$ which is a generalized distance measure of the form (74). Observe that any distance measure of the form (74) is invariant under multiplication of the variables X, Y by the same positive scalar t.

PROPOSITION 41. *Let f be a differentiable convex function on $(0,\infty)$. Assume that the Bregman f-divergence H_f on $\mathscr{B}^{++}(\mathscr{H})$ is homogeneous of degree 0, i.e., it satisfies*

$$H_f(tX, tY) = H_f(X,Y) \quad X, Y \in \mathscr{B}^{++}(\mathscr{H}), t > 0. \tag{75}$$

Then we have $f(x) = a\log x + bx + c$, $x > 0$ with some real scalars a, b, c and $a \le 0$.

PROOF. We plug scalar multiples of the identity $X = xI$, $Y = yI$, $x, y > 0$ into the equality (75) and obtain

$$f(tx) - f(ty) - f'(ty)t(x-y) = f(x) - f(y) - f'(y)(x-y), \quad t > 0. \tag{76}$$

Choosing $t = 1/y$ and reordering this equality we have

$$f'(y) = (1/(x-y))(f(x) - f(y) - f(x/y) + f(1) + f'(1)((x/y)-1)))$$

implying that f is twice continuously differentiable. Differentiating (76) with respect to x twice we obtain $f''(x) = f''(tx)t^2$, $x, t > 0$. In particular, it follows that $f''(t)$ is a constant multiple of t^{-2}, $t > 0$. We infer that $f(t) = a\log t + bt + c$, $t > 0$ holds with some real constants a, b, c. By the convexity of f we have $a \le 0$. □

Concerning the appearance of the function $x \mapsto bx + c$ in the above proposition we note that adding an affine function to a given convex function f does not change the corresponding Bregman divergence (the same holds for Jensen divergence, too), hence that part simply does not count.

We next examine the case of Jensen divergences. Again, let $f(x) = -\log x,\ x > 0$ and pick $\lambda \in (0,1)$. It is easy to see that the corresponding Jensen divergence is

$$J_{-\log,\lambda}(X,Y) = \log\mathrm{Det}\,(\lambda X + (1-\lambda)Y) - \log\mathrm{Det}\,X^{\lambda}Y^{1-\lambda}, \quad X,Y \in \mathscr{B}^{++}(\mathscr{H})$$

which can also be written as

$$J_{-\log,\lambda}(X,Y) =$$
$$\log\mathrm{Det}\left(\lambda Y^{-1/2}XY^{-1/2} + (1-\lambda)I\right) - \log\mathrm{Det}\left(Y^{-1/2}XY^{-1/2}\right)^{\lambda}$$
$$= \mathrm{Tr}\left(\log\left(\left(\lambda Y^{-1/2}XY^{-1/2} + (1-\lambda)I\right)\left(Y^{-1/2}XY^{-1/2}\right)^{-\lambda}\right)\right).$$

These divergences are scalar multiples of the Chebbi-Moakher log-determinant α-divergences, see [**15, 65**]. As mentioned in [**65**] (see pages 146-147) the continuous function $g_{\lambda}(y) = \log\left((\lambda y + (1-\lambda))/y^{\lambda}\right),\ y > 0$ is nonnegative, satisfies (a1), (a2) (with constant $K = 2$) and, moreover, we have

$$J_{-\log,\lambda}(X,Y) = \left\| g\left(Y^{-1/2}XY^{-1/2}\right)\right\|_1, \quad X,Y \in \mathscr{B}^{++}(\mathscr{H}).$$

This means that the Jensen divergences $J_{-\log,\lambda}$ are also of the form (74).

Let us insert the remark here that in the particular case $\lambda = 1/2$ the above Jensen divergence was considered and called S-divergence in the paper [**85**] of Sra. He proved the interesting fact there that the square root of this divergence is a true metric and proposed to use it as a convenient substitute for the geodesic distance $d_{\|.\|_2,\log}$ ($\|.\|_2$ standing for the Hilbert-Schmidt or Frobenius norm) originating from the natural Riemann geometric structure on $\mathscr{B}^{++}(\mathscr{H})$.

Continuing the discussion, above we have seen that the divergences $J_{-\log,\lambda}$ are of the form (74). In what follows we show that they are essentially the unique Jensen divergences with this property.

PROPOSITION 42. *Let f be a convex function on $(0,\infty)$ and pick a number $\lambda \in (0,1)$. Assume that the corresponding Jensen $\lambda - f$ divergence is homogeneous of order 0, i.e., it satisfies*

$$J_{f,\lambda}(tX,tY) = J_{f,\lambda}(X,Y) \quad X,Y \in \mathscr{B}^{++}(\mathscr{H}), t > 0. \tag{77}$$

Then we have $f(x) = a\log x + bx + c,\ x > 0$ with some real scalars a,b,c and $a \le 0$.

PROOF. Just as above, plugging scalar multiples of the identity $X = xI, Y = yI,\ x,y > 0$ into the equality (77) we have

$$\begin{aligned} &\lambda f(tx) + (1-\lambda)f(ty) - f(\lambda tx + (1-\lambda)ty) \\ &= \lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y), \quad t > 0. \end{aligned} \tag{78}$$

We assert that f is differentiable. In fact, choosing $x=1$ we have

$$f(t)=(1/\lambda)(\lambda f(1)+(1-\lambda)f(y)-f(\lambda+(1-\lambda)y)\\ -(1-\lambda)f(ty)+f(\lambda t+(1-\lambda)ty)).$$

The result [**39**, 11.3. Theorem] of Járai on the regularity of solutions of functional equations applies and tells us that f is necessarily continuously differentiable. Differentiating (78) with respect to t we have

$$0=\lambda f'(tx)x+(1-\lambda)f'(ty)y-f'(\lambda tx+(1-\lambda)ty)(\lambda x+(1-\lambda)y)$$

implying the equality

$$\lambda f'(tx)x+(1-\lambda)f'(ty)y=f'(\lambda tx+(1-\lambda)ty)(\lambda x+(1-\lambda)y)$$

for all positive t,x,y. Choosing $t=1$ we infer

$$\lambda f'(x)x+(1-\lambda)f'(y)y=f'(\lambda x+(1-\lambda)y)(\lambda x+(1-\lambda)y).$$

This means that the function $h(x)=f'(x)x$, $x>0$ is affine and hence it is of the form $h(x)=bx+a$, $x>0$. It follows that $f(x)=a\log x+bx+c$, $x>0$ with some real scalars a,b,c. Again, the convexity of f implies $a\leq 0$. $\square$

The above results can be considered as characterizations of the Stein's loss and the Chebbi-Moakher log-determinant α-divergences. Namely, these are the only Bregman (resp. Jensen) divergences which are homogeneous of degree 0.

We now turn to the descriptions of the corresponding generalized isometries, i.e., transformations preserving the considered divergences. We begin with a general result relating to Bregman divergences.

We first mention the easy fact that both kinds of divergences are clearly invariant under unitary and antiunitary congruence transformations. These are maps of the form $A\mapsto UAU^*$, where U is a unitary or an antiunitary operator on $\mathscr{H}$. In fact, this follows from the following. For any continuous function f on $(0,\infty)$ and for any unitary or antiunitary operator U on $\mathscr{H}$ we have that $f(UAU^*)=Uf(A)U^*$ holds for every positive definite A. This is the consequence of the fact that on the spectrum of A the function f coincides with a polynomial p and hence we have $f(UAU^*)=p(UAU^*)=Up(A)U^*=Uf(A)U^*$. Our theorems below show that in many cases only the unitary and antiunitary congruence transformations preserve the Bregman divergence.

THEOREM 43. *Let f be a differentiable convex function on* $(0,\infty)$ *such that f' is bounded from below and unbounded from above. Let $\phi:\mathscr{B}^{++}(\mathscr{H})\to\mathscr{B}^{++}(\mathscr{H})$ be a bijective map which satisfies*

$$H_f\left(\phi(A),\phi(B)\right)=H_f(A,B),\quad A,B\in\mathscr{B}^{++}(\mathscr{H}).$$

Then there exists a unitary or antiunitary operator $U : \mathcal{H} \to \mathcal{H}$ *such that* ϕ *is of the form*

$$\phi(A) = UAU^*, \quad A \in \mathcal{B}^{++}(\mathcal{H}).$$

PROOF. First we recall that since f is convex and everywhere differentiable, f' is continuous. By the convexity of f, its derivative f' is monotonically increasing. The assumption that f' is bounded from below implies that $\lim_{x\to 0+} f'(x)$ exists and is finite. Hence f' can be continuously extended onto $[0,\infty)$. The same holds for f, too. Indeed, $f(x) = f(1) + \int_1^x f'(t)\mathrm{d}t$, $x > 0$ and $\int_1^x f'(t)\mathrm{d}t$ is convergent as x tends to 0 by the boundedness of f' on $(0,1]$.

In the rest of the proof we shall need the following characterization of the usual order $\le$ on $\mathcal{B}^{++}(\mathcal{H})$.

Claim A. Let $B, C \in \mathcal{B}^{++}(\mathcal{H})$. The set

$$\left\{H_f(B,A) - H_f(C,A) | A \in \mathcal{B}^{++}(\mathcal{H})\right\} \tag{79}$$

is bounded from below if and only if $B \le C$.

To prove the claim we first compute

$$\begin{aligned} &H_f(B,A) - H_f(C,A) \\ &= \operatorname{Tr} f(B) - \operatorname{Tr} f(A) - \operatorname{Tr} f'(A)(B-A) \\ &- \operatorname{Tr} f(C) + \operatorname{Tr} f(A) + \operatorname{Tr} f'(A)(C-A) \\ &= \operatorname{Tr} f(B) - \operatorname{Tr} f(C) + \operatorname{Tr} f'(A)(C-B). \end{aligned} \tag{80}$$

Let k denote a lower bound of f'. Assume $B \le C$. Then by the inequality

$$f'(A) \ge kI, \quad A \in \mathcal{B}^{++}(\mathcal{H}),$$

we have that

$$\operatorname{Tr} f'(A)(C-B) \ge \operatorname{Tr} kI(C-B) = k\operatorname{Tr}(C-B)$$

holds for every $A \in \mathcal{B}^{++}(\mathcal{H})$ which shows that the set (79) is bounded from below. Conversely, if $B \not\le C$, then there exists a unit vector $x \in \mathcal{H}$ such that $\langle x, Bx\rangle > \langle x, Cx\rangle$. Let P_x denote the orthogonal projection onto the one-dimensional subspace generated by x. For any $t > 0$ we have

$$\operatorname{Tr} f'\left(tP_x + (I-P_x)\right)(C-B) = f'(t)\left\langle x, (C-B)x\right\rangle + f'(1)\operatorname{Tr}\left(I-P_x\right)(C-B). \tag{81}$$

Since $\langle x, (C-B)x\rangle < 0$ and $\left\{f'(t) | t > 0\right\}$ is unbounded from above, hence the first term on the right hand side of (81) is unbounded from below. By (80), it follows that that the set

$$\left\{H_f(B,A) - H_f(C,A) | A \in \mathcal{B}^{++}(\mathcal{H})\right\}$$

is unbounded from below. This proves our claim.

Since the bijective map $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ preserves the Bregman f-divergence, using the above characterization of the order we obtain that ϕ is an order automorphism, i.e., for any $B, C \in \mathscr{B}^{++}(\mathscr{H})$ we have $B \leq C$ if and only if $\phi(B) \leq \phi(C)$.

By the result [**58**, Theorem 1] of Lajos Molnár, ϕ, just as any order automorphism of $\mathscr{B}^{++}(\mathscr{H})$, is of the form

$$\phi(A) = TAT^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}),$$

where T is an invertible linear or conjugate-linear operator on $\mathscr{H}$. We may suppose that T is linear, since if we are done with this, the case of conjugate-linear T is not difficult to handle.

We show that T is unitary. Assume on the contrary that T is not unitary. Then $T^*T \neq I$. Consider the polar decomposition $T = UP$ of T where $P = \sqrt{T^*T}$ is positive definite and $U = TP^{-1}$ is unitary. Since the Bregman f-divergence is invariant under unitary congruences, hence we have

$$\begin{aligned} H_f(A,B) &= H_f\left(\phi(A), \phi(B)\right) = H_f\left(TAT^*, TBT^*\right) \\ &= H_f\left(UPAPU^*, UPBPU^*\right) = H_f\left(PAP, PBP\right) \end{aligned} \tag{82}$$

for all $A, B \in \mathscr{B}^{++}(\mathscr{H})$. Since T is not unitary, hence $P \neq I$, and thus P has an eigenvalue different from 1. Without serious loss of generality we may assume that P has an eigenvalue greater than 1 for the following reason. The map $A \mapsto PAP$ is a bijection of $\mathscr{B}^{++}(\mathscr{H})$ which preserves the Bregman f-divergence. Therefore, the inverse transformation $A \mapsto P^{-1}AP^{-1}$ preserves the Bregman f-divergence as well. If P does not have any eigenvalue greater than 1, P^{-1} must have one.

Suppose that $Pv = \lambda v$ for some $\lambda > 1$ and unit vector $v \in \mathscr{H}$. Let Q_v be the orthogonal projection onto the one-dimensional subspace generated by v. By (82), the transformation $A \mapsto PAP$ preserves the Bregman f-divergence and so does any of its powers $A \mapsto P^n A P^n$. Hence

$$\begin{aligned} H_f\left(\lambda^2 Q_v, Q_v\right) &= H_f\left(P^n \lambda^2 Q_v P^n, P^n Q_v P^n\right) \\ &= H_f\left(\lambda^{2(n+1)} Q_v, \lambda^{2n} Q_v\right) \end{aligned} \tag{83}$$

holds for any $n \in \mathbb{N}$.

We now consider the symmetrized Bregman f-divergence $H_f(A,B) + H_f(B,A)$ which can be written in the following convenient form

$$H_f(A,B) + H_f(B,A) = \operatorname{Tr} f(A) - \operatorname{Tr} f(B) - \operatorname{Tr} f'(B)(A-B)$$

$$+ \operatorname{Tr} f(B) - \operatorname{Tr} f(A) - \operatorname{Tr} f'(A)(B-A) = \operatorname{Tr}\left(f'(A) - f'(B)\right)(A-B).$$

By (83) we have

$$H_f\left(\lambda^2 Q_v, Q_v\right) + H_f\left(Q_v, \lambda^2 Q_v\right)$$

$$= H_f\left(\lambda^{2(n+1)}Q_v, \lambda^{2n}Q_v\right) + H_f\left(\lambda^{2n}Q_v, \lambda^{2(n+1)}Q_v\right)$$
$$= \mathrm{Tr}\left(f'(\lambda^{2(n+1)}Q_v) - f'(\lambda^{2n}Q_v)\right)\left(\lambda^{2(n+1)}Q_v - \lambda^{2n}Q_v\right)$$
$$= \left(f'\left(\lambda^{2(n+1)}\right) - f'\left(\lambda^{2n}\right)\right)\lambda^{2n}\left(\lambda^2 - 1\right)$$

for any $n \in \mathbb{N}$. This means that $\left(f'(\lambda^{2(n+1)}) - f'\left(\lambda^{2n}\right)\right)\lambda^{2n}$ is independent of n, that is,

$$f'\left(\lambda^{2(n+1)}\right) - f'\left(\lambda^{2n}\right) = \frac{c}{\left(\lambda^2\right)^n}$$

holds for some constant c. Therefore,

$$\lim_{x\to\infty} f'(x) = \lim_{n\to\infty} f'\left(\lambda^{2(n+1)}\right) = \lim_{n\to\infty}\left(f'(1) + \sum_{k=0}^{n}\left(f'\left(\lambda^{2(k+1)}\right) - f'\left(\lambda^{2k}\right)\right)\right)$$
$$= f'(1) + \lim_{n\to\infty}\sum_{k=0}^{n}\frac{c}{\left(\lambda^2\right)^k} = f'(1) + c\sum_{n=0}^{\infty}\left(\lambda^{-2}\right)^n < \infty.$$

which contradicts the assumption that f' is unbounded from above. The proof of the theorem is complete. □

The probably most important Bregman f-divergences on $\mathscr{B}^{++}(\mathscr{H})$ correspond to the following functions: $x \mapsto x^p$ $(p > 1)$, $x \mapsto x\log x - x$, $x \mapsto -\log x$. Unfortunately, the general theorem above covers only the case of the first type of functions. Indeed, the second function has derivative which is neither bounded from below nor unbounded from above and the derivative of the third one is not bounded from below. Fortunately, the Bregman divergence related to the third function, i.e., Stein's loss is of the form (74) and the corresponding preservers were characterized in [**65**]. By [**65**, Theorem 2] a surjective map $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ preserves the Stein's loss if and only if there is an invertible linear or conjugate linear operator $T : \mathscr{H} \to \mathscr{H}$ such that ϕ is of the form

$$\phi(A) = TAT^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

In what follows we characterize the preservers of Umegaki's relative entropy or, in other words, von Neumann divergence which is the Bregman divergence corresponding to the function $x \mapsto x\log x - x$. It is clear that this divergence equals

$$\mathrm{Tr}\left(A(\log A - \log B) - (A - B)\right), \quad A, B \in \mathscr{B}^{++}(\mathscr{H}).$$

The result reads as follows.

THEOREM 44. *Let $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ be a surjective map which satisfies*

$$\mathrm{Tr}\left(\phi(A)\left(\log\phi(A) - \log\phi(B)\right) - \left(\phi(A) - \phi(B)\right)\right)$$
$$= \mathrm{Tr}(A(\log A - \log B) - (A - B)), \quad A, B \in \mathscr{B}^{++}(\mathscr{H}). \tag{84}$$

Then there exists a unitary or antiunitary operator $U : \mathscr{H} \to \mathscr{H}$ such that ϕ is of the form

$$\phi(A) = UAU^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

PROOF. We begin with the following general observation. Assume that f is a strictly convex differentiable function on $(0,\infty)$. We assert that for any $B, C \in \mathscr{B}^{++}(\mathscr{H})$, the set

$$\left\{H_f(A,B) - H_f(A,C) | A \in \mathscr{B}^{++}(\mathscr{H})\right\}$$

is bounded from below if and only if $f'(B) \le f'(C)$. Indeed, we have

$$H_f(A,B) - H_f(A,C)$$

$$= \operatorname{Tr}\left(f'(C) - f'(B)\right) A + \operatorname{Tr}\left(f(C) - f(B) - \left(f'(C)C - f'(B)B\right)\right)$$

which is easily seen to be bounded from below if and only if $f'(C) - f'(B) \ge 0$.

Therefore, for any surjective (and hence, by the strict convexity of f, bijective) map $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ which preserves the Bregman f-divergence, we obtain that ϕ has the property that

$$f'(B) \le f'(C) \Longleftrightarrow f'\left(\phi(B)\right) \le f'\left(\phi(C)\right).$$

This means that the transformation $A \mapsto f'\left(\phi\left(f'^{-1}(A)\right)\right)$ is an order automorphism of the set of all self-adjoint operators on $\mathscr{H}$ with spectrum contained in the range of f'.

In our particular case we have $f' = \log$, therefore $A \mapsto \log\left(\phi\left(e^A\right)\right)$ is an order automorphism of the set of all self-adjoint operators on $\mathscr{H}$. By [**57**, Theorem 2] we have an invertible linear or conjugate-linear operator $T : \mathscr{H} \to \mathscr{H}$ and a self-adjoint linear operator $X : \mathscr{H} \to \mathscr{H}$ such that

$$\log\left(\phi\left(e^A\right)\right) = TAT^* + X$$

or

$$\phi(A) = e^{T \log A T^* + X}, \quad A \in \mathscr{B}^{++}(\mathscr{H}). \tag{85}$$

We assume that T is linear, the conjugate-linear case is similar, not difficult to handle. Consider the polar decomposition $T = UP$ of T where U is unitary and $P = \sqrt{T^*T}$ is positive definite. As already mentioned, the unitary similarity transformation $A \mapsto UAU^*$ preserves Bregman divergences. Since

$$\phi(A) = e^{UP(\log A)PU^* + X} = Ue^{P(\log A)P + U^*XU}U^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}),$$

it follows that in (85) we can and do assume that T is a positive definite operator. We prove that necessarily $T = I$ holds. To see this, first assume that T has an eigenvalue which is greater than 1. We know that

(86)
$$\operatorname{Tr}\left(e^{T(\log A)T+X}(T(\log A)T-T(\log B)T)-\left(e^{T(\log A)T+X}-e^{T(\log B)T+X}\right)\right)$$
$$=\operatorname{Tr}(A(\log A-\log B)-(A-B)),\quad A,B\in\mathscr{B}^{++}(\mathscr{H}).$$

Fixing A,B, write tB in the place of B where $t>0$ is arbitrary. The above equality gives us that

(87)
$$a\log t+\operatorname{Tr}e^{(\log t)T^2+T(\log B)T+X}+c=d\log t+et+f,\quad t>0$$

holds for some real constants a,b,c,d,e,f. Select a real number μ such that $\mu I\leq T(\log B)T+X$. Let λ be an eigenvalue of T which is greater than 1 and let x be a corresponding unit eigenvector. Denote by P_x the rank-one projection onto the subspace generated by x. Then $\lambda^2P_x\leq T^2$, so for $t\geq 1$ we have $(\log t)\lambda^2P_x+\mu I\leq(\log t)T^2+T(\log B)T+X$. By the monotonicity of trace functions (see [**13**, 2.10. Theorem]) this implies that

$$\operatorname{Tr}e^{(\log t)\lambda^2P_x+\mu I}\leq\operatorname{Tr}e^{(\log t)T^2+T(\log B)T+X},\quad t\geq 1.$$

Therefore, the function $t\mapsto\operatorname{Tr}e^{(\log t)T^2+T(\log B)T+X}$, $t\geq 1$ can be minorized by a function $\alpha t^{\lambda^2}+\beta$ with some positive α and real number β. Since $\lambda^2>1$, considering the equality (87) and letting t tend to infinity, we easily obtain a contradiction. Therefore, the eigenvalues of T are all less than or equal to 1. However, the inverse of ϕ also preserves the von Neumann divergence and we have

$$\phi^{-1}(A)=e^{T^{-1}(\log A)T^{-1}-T^{-1}XT^{-1}},\quad A\in\mathscr{B}^{++}(\mathscr{H})$$

It follows that the eigenvalues of T^{-1} are also not greater than 1. We conclude that $T=I$.

We finally prove that $X=0$. Let $A=I$ and B be any element of $\mathscr{B}^{++}(\mathscr{H})$ which commutes with X. We obtain from (86) that

$$\operatorname{Tr}e^X(-\log B+B-I)=\operatorname{Tr}(-\log B+B-I).$$

By the properties of the function $x\mapsto -\log x+x-1$, $x>0$ (strictly increasing for $x\geq 1$, and takes the value 0 at 1) it is easy to see that any positive semidefinite operator D which commutes with X can be written as $D=-\log B+B-I$ with some positive definite B which commutes with X. Consequently, we have

$$\operatorname{Tr}e^XD=\operatorname{Tr}D$$

for any positive semidefinite operator D on $\mathscr{H}$. This clearly implies that $e^X=I$, i.e., $X=0$. The proof of the theorem is complete. □

We now turn to the preservers of Jensen divergences. Our general result reads as follows.

THEOREM 45. *Let f be a differentiable strictly convex function on $(0,\infty)$, assume $\lim_{x\to 0+} f(x)$ exists and finite and f' is unbounded from above. Pick $\lambda \in (0,1)$. If $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ is a surjective map which satisfies*

$$J_{f,\lambda}\left(\phi(A),\phi(B)\right) = J_{f,\lambda}(A,B), \quad A,B \in \mathscr{B}^{++}(\mathscr{H}),$$

then there exists a unitary or antiunitary operator $U : \mathscr{H} \to \mathscr{H}$ such that ϕ is of the form

$$\phi(A) = UAU^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

PROOF. Observe first that, by the assumptions on the function, f is monotonically increasing for large enough values of its variable. Next, we verify the following.

Claim B. For any $B,C \in \mathscr{B}^{++}(\mathscr{H})$, the set

$$\left\{J_{f,\lambda}(A,B) - J_{f,\lambda}(A,C) | A \in \mathscr{B}^{++}(\mathscr{H})\right\}$$

is bounded from below if and only if $B \le C$.

To prove the claim first observe that

$$J_{f,\lambda}(A,B) - J_{f,\lambda}(A,C) = (1-\lambda)\left(\operatorname{Tr} f(B) - \operatorname{Tr} f(C)\right) +$$

$$+ \operatorname{Tr} f(\lambda A + (1-\lambda)C) - \operatorname{Tr} f(\lambda A + (1-\lambda)B), \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

Assume now that $B \le C$ and that there is a sequence (A_k) in $\mathscr{B}^{++}(\mathscr{H})$ such that

$$\operatorname{Tr} f(\lambda A_k + (1-\lambda)C) - \operatorname{Tr} f(\lambda A_k + (1-\lambda)B) \to -\infty.$$

Denote $\mu_i^{(k)}$, $i = 1,\dots,n$ the eigenvalues of $\lambda A_k + (1-\lambda)B$ and $\lambda_i^{(k)}$, $i = 1,\dots,n$ the eigenvalues of $\lambda A_k + (1-\lambda)C$ both ordered increasingly. By the Weyl inequality (see e.g. [**37**, 4.3.3. Corollary]) it follows that $\mu_i^{(k)} \le \lambda_i^{(k)}$ for all i and k. Moreover, we have $\sum_i \left(f\left(\lambda_i^{(k)}\right) - f\left(\mu_i^{(k)}\right)\right) \to -\infty$ which implies that the sequence $\left(f\left(\lambda_i^{(k)}\right) - f\left(\mu_i^{(k)}\right)\right)$ is not bounded from below for some $i = 1,\dots,n$ and hence it has a subsequence $\left(f\left(\lambda_i^{(k_l)}\right) - f\left(\mu_i^{(k_l)}\right)\right) \to -\infty$. Since f is bounded from below, we deduce that $f\left(\mu_i^{(k_l)}\right) \to \infty$. This implies that $\mu_i^{(k_l)} \to \infty$ and hence $f\left(\lambda_i^{(k_l)}\right) - f\left(\mu_i^{(k_l)}\right) \ge 0$ for all but finitely many indexes l which is a contradiction.

Conversely, if $B \not\le C$, then there exists a unit vector $x \in \mathscr{H}$ such that $\langle x, Bx\rangle > \langle x, Cx\rangle$. Set $\varepsilon = \langle x, (B-C)x\rangle$ and let P_x denote the orthogonal projection onto the one-dimensional subspace generated by x. Let m be a positive number such that $mI - (1-\lambda)C$ is positive definite. For any $t > 0$ set

$$A_t := \frac{1}{\lambda}\left(mI + tP_x - (1-\lambda)C\right).$$

Now

$$\begin{aligned} J_{f,\lambda}(A_t,B) - J_{f,\lambda}(A_t,C) &= (1-\lambda)\big(\operatorname{Tr} f(B) - \operatorname{Tr} f(C)\big) \\ &+ \operatorname{Tr} f(\lambda A_t + (1-\lambda)C) - \operatorname{Tr} f(\lambda A_t + (1-\lambda)B) \\ &= (1-\lambda)\big(\operatorname{Tr} f(B) - \operatorname{Tr} f(C)\big) \\ &+ \operatorname{Tr} f(mI + tP_x) - \operatorname{Tr} f(mI + tP_x + (1-\lambda)(B-C)). \end{aligned} \tag{88}$$

Let $\{x, y_2, y_3, \ldots, y_n\}$ be an orthonormal basis in $\mathscr{H}$. For $2 \le j \le n$, denote by a_{jj} the diagonal matrix elements of $mI + tP_x + (1-\lambda)(B-C)$ relative to this basis, that is,

$$a_{jj} := \langle (mI + tP_x + (1-\lambda)(B-C))\, y_j, y_j\rangle = \langle (mI + (1-\lambda)(B-C))\, y_j, y_j\rangle .$$

Note that a_{jj} is independent of t for $2 \le j$ and

$$\langle (mI + tP_x + (1-\lambda)(B-C))\, x, x\rangle = m + t + (1-\lambda)\varepsilon.$$

By Peierls inequality (see [**13**, 2.9. Theorem]), for the convex function f we have

$$f(m+t+(1-\lambda)\varepsilon) + \sum_{j=2}^{n} f(a_{jj}) \le \operatorname{Tr} f(mI + tP_x + (1-\lambda)(B-C)).$$

On the other hand, it is apparent that

$$\operatorname{Tr} f(mI + tP_x) = f(m+t) + (n-1)f(m).$$

Therefore, by (88),

$$J_{f,\lambda}(A_t,B) - J_{f,\lambda}(A_t,C) \le f(m+t) - f(m+t+(1-\lambda)\varepsilon) + K(B,C),$$

where

$$K(B,C) = (1-\lambda)\big(\operatorname{Tr} f(B) - \operatorname{Tr} f(C)\big) + (n-1)f(m) - \sum_{j=2}^{n} f(a_{jj})$$

is independent of the parameter t. Since f' is unbounded from above, hence

$$\lim_{t\to\infty} f(m+t) - f(m+t+(1-\lambda)\varepsilon) = -\infty,$$

and this completes the proof of our claim.

Using the characterization of the order given in Claim B, we see that the transformation ϕ is an order automorphism of $\mathscr{B}^{++}(\mathscr{H})$ and hence it is of the form

$$\phi(A) = TAT^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}),$$

where T is an invertible linear or conjugate-linear operator on $\mathscr{H}$. We consider only the case where T is linear and prove that then T is necessarily unitary.

As already mentioned, the unitary-antiunitary congruence transformations preserve the Jensen divergences. Hence, by polar decomposition, we can assume that T is a positive definite operator.

Any power of ϕ is also a divergence preserver, so for every positive integer n we have that

$$\begin{aligned} &\operatorname{Tr} \lambda f(T^n A T^n) + (1-\lambda) f(T^n B T^n) - f(T^n(\lambda A + (1-\lambda)B)T^n) \\ &= \operatorname{Tr} \lambda f(A) + (1-\lambda) f(B) - f(\lambda A + (1-\lambda)B), \quad A, B \in \mathscr{B}^{++}(\mathscr{H}). \end{aligned} \tag{89}$$

Since f is continuously extendible onto $[0,\infty)$, thus we can insert any positive semidefinite operators A, B in the equality above. Assume T has an eigenvalue, say s, which is greater than 1 and x is a corresponding unit eigenvector. As before, denote by P_x the orthogonal projection onto the subspace generated by x. Plug $A = P_x$ and $B = 0$ into (89). We have

$$\lambda f\left(s^{2n}\right) - f\left(\lambda s^{2n}\right) = c$$

with some constant c. One can easily check that the function $t \mapsto \lambda f(t) - f(\lambda t)$ is monotonically increasing (just differentiate and use that f' is increasing). Since $s^{2n} \to \infty$, it follows that

$$\lambda f(t) - f(\lambda t) = c, \quad t > 0.$$

We deduce $\lambda f'(t) = f'(\lambda t)\lambda$, $t > 0$ and this implies that f' is constant. We obtain that f is affine, a contradiction. It follows that T has no eigenvalue which is greater than 1. Since ϕ^{-1} also preserves the $\lambda - f$ Jensen divergence, we have that the eigenvalues of T^{-1} are also not greater than 1. It follows that $T = I$ and the proof of the theorem is complete. □

Let us again consider the three most important examples $x \mapsto x^p$ $(p > 1)$, $x \mapsto x \log x - x$, $x \mapsto -\log x$ of generating functions. The first two do satisfy the conditions in our theorem, hence the corresponding preservers are unitary-antiunitary congruence transformations. As for the third one, it does not satisfy the conditions (not bounded below), but the Jensen divergence in that case is of the form (74), see the discussion before Proposition 42. By [**65**, Theorem 2], a surjective map $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ preserves the corresponding Jensen divergence (i.e., Chebbi-Moakher log-determinant α-divergence) if and only if there is an invertible linear or conjugate linear operator $T : \mathscr{H} \to \mathscr{H}$ such that ϕ is of the form

$$\phi(A) = TAT^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

2. Jordan triple endomorphisms of the positive cone

Now we turn to another preserver problem on the cone of positive operators. Namely, we describe the structure of the Jordan triple endomorphisms of the positive definite cones in operator algebras. These endomorphisms are maps which are morphisms with respect to the operation of the Jordan triple product $(A,B) \mapsto ABA$ which is a well-known operation in ring theory. Our main reason for investigating these maps comes from the fact that they naturally appear in the study of surjective isometries and surjective maps preserving generalized distance measures between positive definite cones. For details see [**55, 54, 65**].

In the paper [**55**] Lajos Molnár proved the following statement. (In the following, as usual, $\mathcal{H}$ stands for a finite dimensional complex Hilbert space.)

THEOREM 46. *Assume* $\dim(\mathcal{H}) \geq 3$. *Let* $\phi : \mathcal{B}^{++}(\mathcal{H}) \to \mathcal{B}^{++}(\mathcal{H})$ *be a continuous map which is a Jordan triple endomorphism, i.e.,* ϕ *is a continuous map which satisfies*

$$\phi(ABA) = \phi(A)\phi(B)\phi(A), \quad A, B \in \mathcal{B}^{++}(\mathcal{H}).$$

Then there exist a unitary operator $U \in \mathcal{B}(\mathcal{H})$, *a real number* c, *a set* $\{P_1, \ldots, P_n\}$ *of mutually orthogonal rank-one projections in* $\mathcal{B}(\mathcal{H})$, *and a set* $\{c_1, \ldots, c_n\}$ *of real numbers such that* ϕ *is of one of the following forms:*

(a1) $\phi(A) = (\mathrm{Det} A)^c UAU^*, \ A \in \mathcal{B}^{++}(\mathcal{H})$;
(a2) $\phi(A) = (\mathrm{Det} A)^c UA^{-1}U^*, \ A \in \mathcal{B}^{++}(\mathcal{H})$;
(a3) $\phi(A) = (\mathrm{Det} A)^c UA^{tr}U^*, \ A \in \mathcal{B}^{++}(\mathcal{H})$;
(a4) $\phi(A) = (\mathrm{Det} A)^c UA^{tr^{-1}}U^*, \ A \in \mathcal{B}^{++}(\mathcal{H})$;
(a5) $\phi(A) = \sum_{j=1}^n (\mathrm{Det} A)^{c_j} P_j, \ A \in \mathcal{B}^{++}(\mathcal{H})$.

Observe that the converse statement in Theorem 46 is also true meaning that any transformation of any of the forms (a1)-(a5) is necessarily a continuous Jordan triple endomorphism of $\mathcal{B}^{++}(\mathcal{H})$.

One may immediately ask why we assume the condition $\dim(\mathcal{H}) \geq 3$, what happens in the case where $\dim(\mathcal{H}) = 2$. The fact is that in the proof of Theorem 46 Molnár used such tools which are applicable only if $\dim(\mathcal{H}) \geq 3$. The remaining case $\dim(\mathcal{H}) = 2$ was proposed as an open problem in the papers [**55**] (see Remark 11) and [**54**] (see Remark 23).

One may think that when $\dim(\mathcal{H}) = 2$, one can simply compute and obtain the solution straightaway. But this is far from being true as it will turn out below.

Our main result on Jordan triple endomorphisms reads as follows.

THEOREM 47. *Let $\mathscr{H}$ be a two-dimensional Hilbert space. Let $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ be a continuous Jordan-triple endomorphism. Then we have the following possibilities:*

(b1) *there is a unitary operator $U \in \mathscr{B}(\mathscr{H})$ and a real number c such that*

$$\phi(A) = (\mathrm{Det}A)^c UAU^*, \quad A \in \mathscr{B}^{++}(\mathscr{H});$$

(b2) *there is a unitary operator $V \in \mathscr{B}(\mathscr{H})$ and a real number d such that*

$$\phi(A) = (\mathrm{Det}A)^d VA^{-1}V^*, \quad A \in \mathscr{B}^{++}(\mathscr{H});$$

(b3) *there is a unitary operator $W \in \mathscr{B}(\mathscr{H})$ and real numbers c_1, c_2 such that*

$$\phi(A) = W\mathrm{Diag}[(\mathrm{Det}A)^{c_1}, (\mathrm{Det}A)^{c_2}]W^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

Before presenting the proof we introduce a few notation and make some useful observations.

We equip $\mathscr{B}^{sa}(\mathscr{H})$ with the inner product $\langle X, Y\rangle = (1/2)\,\mathrm{Tr}\,XY$. The induced norm is denoted by $\|\cdot\|$. Let $\{e_1, e_2\}$ be an arbitrary orthonormal basis in $\mathscr{H}$. In the following, whenever a 2×2 matrix appears as an element of $\mathscr{B}(\mathscr{H})$, it denotes the operator with the following (characterizing) property: its matrix expansion in the above fixed orthonormal basis coincides with the given matrix. The set

(90)
$$\left\{\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\right\}$$

is a convenient orthonormal basis in $\mathscr{B}^{sa}(\mathscr{H})$. Let $\mathscr{B}^{sa}_0(\mathscr{H})$ denote the traceless subspace of $\mathscr{B}^{sa}(\mathscr{H})$ (the subspace of all elements in $\mathscr{B}^{sa}(\mathscr{H})$ with zero trace).

In the proof of our theorem we shall use the following two observations. We first claim that for $X \in \mathscr{B}^{sa}_0(\mathscr{H})$, the equality $X^2 = I$ holds iff $\|X\| = 1$. Indeed, let us denote the eigenvalues of X by λ and $-\lambda$, $\lambda \geq 0$. We have

$$X^2 = I \Leftrightarrow \lambda^2 = 1 \Leftrightarrow \frac{1}{2}\left(\lambda^2 + (-\lambda)^2\right) = 1 \Leftrightarrow \|X\| = 1$$

verifying our first calim.

Next, we assert that for any $0 \neq X \in \mathscr{B}^{sa}_0(\mathscr{H})$ we have $e^X = (\cosh\|X\|)\,I + (\sinh\|X\|)\,(X/\|X\|)$. To see this, using $(X/\|X\|)^2 = I$, we compute

$$e^X = e^{\|X\|\frac{X}{\|X\|}} = \sum_{k=0}^{\infty} \frac{1}{k!}\,\|X\|^k \left(\frac{X}{\|X\|}\right)^k$$

$$= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \|X\|^{2k} I + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \|X\|^{2k+1} \frac{X}{\|X\|}.$$

This proves our assertion.

Now we turn to the proof of the main result.

PROOF OF THEOREM 47. Let $\mathscr{H}$ be a two-dimensional Hilbert space and let $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ be a continuous Jordan triple endomorphism. Then, by [**55**, Lemma 6] there exists a commutativity preserving linear transformation $f : \mathscr{B}^{sa}(\mathscr{H}) \to \mathscr{B}^{sa}(\mathscr{H})$ such that

$$\phi(A) = \exp(f(\log A)), \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

In fact, similar conclusion holds for all continuous Jordan triple endomorphisms between the positive definite cones of general C^*-algebras as it has been shown in [**54**, Lemma 16]. By a commutativity preserving linear map we simply mean a transformation which sends commuting elements to commuting elements.

We have two possibilities for $f(I)$: It is either a scalar multiple of the identity or it is not. We divide the argument accordingly.

Assume first that $f(I)$ is not a scalar multiple of the identity. Then up to unitary similarity we may and do assume that $f(I)$ is a diagonal operator with two different eigenvalues. By the commutativity preserving property of f, for every $A \in \mathscr{B}^{sa}(\mathscr{H})$ we have that $f(A)$ commutes with $f(I)$ and then it follows that $f(A)$ is diagonal, too. Therefore, we have linear functionals $\varphi, \psi : \mathscr{B}^{sa}(\mathscr{H}) \to \mathbb{R}$ such that

$$f(A) = \begin{bmatrix} \varphi(A) & 0 \\ 0 & \psi(A) \end{bmatrix}, \quad A \in \mathscr{B}^{sa}(\mathscr{H})$$

and hence

$$\phi(A) = \begin{bmatrix} e^{\varphi(\log A)} & 0 \\ 0 & e^{\psi(\log A)} \end{bmatrix}, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

Since ϕ is a Jordan triple endomorphism, we deduce easily that

$$\varphi(\log ABA) = 2\varphi(\log A) + \varphi(\log B), \quad A, B \in \mathscr{B}^{++}(\mathscr{H})$$

and similar equality holds for ψ as well. Since φ is a linear functional on $\mathscr{B}^{sa}(\mathscr{H})$, by Riesz representation theorem we have an element $T \in \mathscr{B}^{sa}(\mathscr{H})$ such that $\varphi(\cdot) = \langle \cdot, T \rangle$. It follows that we have

$$\mathrm{Tr}((\log ABA)T) = 2\,\mathrm{Tr}((\log A)T) + \mathrm{Tr}((\log B)T), \quad A, B \in \mathscr{B}^{++}(\mathscr{H}).$$

Following the argument given on p. 2844 in [**53**] from the displayed equality (2) on, one can verify that T is necessarily a scalar multiple of the identity and that means that $\varphi(A) = c\,\mathrm{Tr}\,A$, $A \in \mathscr{B}^{sa}(\mathscr{H})$ holds for some real number c. The same observation applies for ψ, too, and then we conclude that there are real numbers c_1, c_2 such that we have

$$\phi(A) = \begin{bmatrix} (\mathrm{Det} A)^{c_1} & 0 \\ 0 & (\mathrm{Det} A)^{c_2} \end{bmatrix}, \quad A \in \mathscr{B}^{++}(\mathscr{H}),$$

which gives us (b3).

In the remaining part of the proof we assume that $f(I)$ is a scalar multiple of the identity.

Let us define the linear functional $f_0 : \mathscr{B}^{sa}(\mathscr{H}) \to \mathbb{R}$ by $f_0(\cdot) = \langle \sigma_0, f(\cdot) \rangle$, that is, by $f_0(A) = (1/2)\operatorname{Tr} f(A)$, $A \in \mathscr{B}^{sa}(\mathscr{H})$.

The first crucial step in the proof follows.

Claim 1. *The linear functional f_0 vanishes on $\mathscr{B}_0^{sa}(\mathscr{H})$.*

The subspace $\mathscr{B}_0^{sa}(\mathscr{H})$ is generated by $\sigma_x, \sigma_y, \sigma_z$. We show that $f_0(\sigma_x) = f_0(\sigma_y) = 0$, the remaining equality $f_0(\sigma_z) = 0$ can be verified similarly. In what follows we consider arbitrary positive real parameters s, t. Direct calculations show that for all such s, t we have

$$\begin{aligned}
& e^{\frac{s}{2}\sigma_x} e^{t\sigma_y} e^{\frac{s}{2}\sigma_x} \\
&= \left(\cosh\left(\frac{s}{2}\right) I + \sinh\left(\frac{s}{2}\right)\sigma_x\right)\left(\cosh(t) I + \sinh(t)\sigma_y\right) \times \\
&\quad \times \left(\cosh\left(\frac{s}{2}\right) I + \sinh\left(\frac{s}{2}\right)\sigma_x\right) \\
&= \cosh(t)\cosh^2\left(\frac{s}{2}\right) I + 2\cosh(t)\cosh\left(\frac{s}{2}\right)\sinh\left(\frac{s}{2}\right)\sigma_x + \cosh^2\left(\frac{s}{2}\right)\sinh(t)\sigma_y \\
&\quad + \sinh\left(\frac{s}{2}\right)\sinh(t)\cosh\left(\frac{s}{2}\right)\sigma_x\sigma_y + \cosh\left(\frac{s}{2}\right)\sinh(t)\sinh\left(\frac{s}{2}\right)\sigma_y\sigma_x \\
&\quad + \cosh(t)\sinh^2\left(\frac{s}{2}\right)\sigma_x^2 + \sinh^2\left(\frac{s}{2}\right)\sinh(t)\sigma_x\sigma_y\sigma_x \\
&= \cosh(s)\cosh(t) I + \cosh(t)\sinh(s)\sigma_x + \sinh(t)\sigma_y.
\end{aligned}$$

Here we have used the equalities $\sigma_x\sigma_y + \sigma_y\sigma_x = 0$, $\sigma_x^2 = I$, $\sigma_x\sigma_y\sigma_x = -\sigma_y$ and some identities of the hyperbolic functions.

Since, by the multiplicativity of the determinant, we have

$$\mathrm{Det}\left(e^{\frac{s}{2}\sigma_x} e^{t\sigma_y} e^{\frac{s}{2}\sigma_x}\right) = 1,$$

hence

$$e^{\frac{s}{2}\sigma_x} e^{t\sigma_y} e^{\frac{s}{2}\sigma_x} = e^{rW}$$

holds for some $W \in \mathscr{B}_0^{sa}(\mathscr{H})$ with $\|W\| = 1$ and $r \geq 0$ (observe that W depends on s, t). Since $e^{rW} = \cosh(r) I + \sinh(r) W$ we obtain the equality

$$\cosh(r) I + \sinh(r) W = \cosh(s)\cosh(t) I + \cosh(t)\sinh(s)\sigma_x + \sinh(t)\sigma_y.$$

Taking trace we first deduce that

$$r = \cosh^{-1}\left(\cosh(s)\cosh(t)\right) \tag{91}$$

and next that

$$\sinh(r)W = \cosh(t)\sinh(s)\sigma_x + \sinh(t)\sigma_y.$$

Clearly, due to $s, t > 0$, the possibility $r = 0$ is ruled out and hence we infer that

$$W = \frac{1}{\sinh(r)}\left(\cosh(t)\sinh(s)\sigma_x + \sinh(t)\sigma_y\right)$$

$$= \frac{\cosh(t)\sinh(s)\sigma_x + \sinh(t)\sigma_y}{\sqrt{\cosh^2(s)\cosh^2(t) - 1}}. \tag{92}$$

Now, on the one hand, we compute

$$\mathrm{Det}\left(\phi\left(e^{\frac{s}{2}\sigma_x}e^{t\sigma_y}e^{\frac{s}{2}\sigma_x}\right)\right)$$

$$= \mathrm{Det}\left(e^{f\left(\log\left(e^{\frac{s}{2}\sigma_x}e^{t\sigma_y}e^{\frac{s}{2}\sigma_x}\right)\right)}\right) = e^{\mathrm{Tr}\, f\left(\log\left(e^{\frac{s}{2}\sigma_x}e^{t\sigma_y}e^{\frac{s}{2}\sigma_x}\right)\right)} = e^{2f_0(rW)}. \tag{93}$$

On the other hand, since ϕ is a Jordan triple endomorphism, the quantity (93) is equal to

$$\begin{aligned} &\mathrm{Det}\left(\phi\left(e^{\frac{s}{2}\sigma_x}\right)\phi\left(e^{t\sigma_y}\right)\phi\left(e^{\frac{s}{2}\sigma_x}\right)\right) = \mathrm{Det}\left(e^{\frac{s}{2}f(\sigma_x)}e^{tf(\sigma_y)}e^{\frac{s}{2}f(\sigma_x)}\right) \\ &= e^{\frac{s}{2}\mathrm{Tr}\, f(\sigma_x)}e^{t\,\mathrm{Tr}\, f(\sigma_y)}e^{\frac{s}{2}\mathrm{Tr}\, f(\sigma_x)} = e^{2\left(sf_0(\sigma_x) + tf_0(\sigma_y)\right)}. \end{aligned} \tag{94}$$

Let us introduce the auxiliary function

$$N(s,t) = \frac{\cosh^{-1}\left(\cosh(s)\cosh(t)\right)}{\sqrt{\cosh^2(s)\cosh^2(t) - 1}}, \quad 0 < s, t \in \mathbb{R}.$$

By (91), (92), (93), (94) we have

$$sf_0(\sigma_x) + tf_0(\sigma_y) = f_0(rW)$$

$$= N(s,t)\cosh(t)\sinh(s)f_0(\sigma_x) + N(s,t)\sinh(t)f_0(\sigma_y)$$

for all $0 < s, t \in \mathbb{R}$. It is not difficult to check that the two-variable functions $g(s,t) = N(s,t)\cosh(t)\sinh(s) - s$ and $h(s,t) = N(s,t)\sinh(t) - t$ are linearly independent. Indeed, one can see that the determinant of the matrix

$$\begin{bmatrix} g(1,1) & h(1,1) \\ g(2,2) & h(2,2) \end{bmatrix}$$

is nonzero (its value is close to -0.5) which implies the desired linear independence. It then follows that $f_0(\sigma_x) = f_0(\sigma_y) = 0$ and we obtain Claim 1.

As a consequence we infer that the subspace $f(\mathscr{B}_0^{sa}(\mathscr{H}))$ is orthogonal to $\sigma_0 = I$ meaning that it consists of traceless operators, $f(\mathscr{B}_0^{sa}(\mathscr{H})) \subset \mathscr{B}_0^{sa}(\mathscr{H})$. Since $f(I)$ is a scalar multiple of the identity, we also have

$f(\mathscr{B}_0^{sa}(\mathscr{H})^\perp) \subset \mathscr{B}_0^{sa}(\mathscr{H})^\perp$. We will use these facts in the second crucial step of the proof which follows.

Claim 2. *The restriction of f to $\mathscr{B}_0^{sa}(\mathscr{H})$ is a non-negative scalar multiple of an isometry.*

To see this, it is sufficient to show that $f(\sigma_x), f(\sigma_y), f(\sigma_z)$ are mutually orthogonal and of the same norm. Clearly, we are done if we verify this for any two elements of the collection $f(\sigma_x), f(\sigma_y), f(\sigma_z)$. We shall consider, for example, $f(\sigma_x)$ and $f(\sigma_y)$. Recalling that $f(W)$ is traceless, in the case where $f(W) \neq 0$, we compute

$$l(s,t) := \frac{1}{2}\operatorname{Tr}\phi\left(e^{\frac{s}{2}\sigma_x}e^{t\sigma_y}e^{\frac{s}{2}\sigma_x}\right) = \frac{1}{2}\operatorname{Tr}\left(e^{f\left(\log\left(e^{\frac{s}{2}\sigma_x}e^{t\sigma_y}e^{\frac{s}{2}\sigma_x}\right)\right)}\right) = \frac{1}{2}\operatorname{Tr}e^{rf(W)}$$

$$= \frac{1}{2}\operatorname{Tr}\left(\cosh\left(r\left\|f(W)\right\|\right)I + \sinh\left(r\left\|f(W)\right\|\right)\frac{f(W)}{\left\|f(W)\right\|}\right) = \cosh\left(r\left\|f(W)\right\|\right).$$

If $f(W) = 0$, then we again have $l(s,t) = \cosh\left(r\left\|f(W)\right\|\right)$ and, by (92), we can further compute

$$l(s,t) = \cosh\left(\left\|f(W)\right\|\cosh^{-1}\left(\cosh(s)\cosh(t)\right)\right)$$

$$= \cosh\Bigg(\frac{\cosh^{-1}(\cosh(s)\cosh(t))}{\sqrt{\cosh^2(s)\cosh^2(t)-1}} \times \sqrt{\sinh^2(s)\cosh^2(t)\left\|f(\sigma_x)\right\|^2}$$
$$\overline{+\left\langle f(\sigma_x), f(\sigma_y)\right\rangle 2\sinh(s)\sinh(t)\cosh(t) + \sinh^2(t)\left\|f(\sigma_y)\right\|^2}\Bigg)$$

$$= \cosh\Bigg(\frac{\cosh^{-1}(\cosh(s)\cosh(t))}{\sqrt{\cosh^2(s)\cosh^2(t)-1}} \times$$
$$\times\sqrt{\left(\cosh^2(s)\cosh^2(t) - \cosh^2(t)\right)\left\|f(\sigma_x)\right\|^2 +}$$
$$\overline{+\left\langle f(\sigma_x), f(\sigma_y)\right\rangle 2\sinh(s)\sinh(t)\cosh(t) + \left(\cosh^2(t)-1\right)\left\|f(\sigma_y)\right\|^2}\Bigg). \tag{95}$$

Since ϕ is a Jordan triple endomorphism, (95) is equal to

$$m(s,t) := \frac{1}{2}\operatorname{Tr}\left(\phi\left(e^{\frac{s}{2}\sigma_x}\right)\phi\left(e^{t\sigma_y}\right)\phi\left(e^{\frac{s}{2}\sigma_x}\right)\right)$$
$$= \frac{1}{2}\operatorname{Tr}\left(e^{\frac{s}{2}f(\sigma_x)}e^{tf(\sigma_y)}e^{\frac{s}{2}f(\sigma_x)}\right) = \frac{1}{2}\operatorname{Tr}\left(e^{sf(\sigma_x)}e^{tf(\sigma_y)}\right).$$

Assume $f(\sigma_x), f(\sigma_y) \neq 0$ and denote $X = f(\sigma_x)/\left\|f(\sigma_x)\right\|$ and $Y = f(\sigma_y)/\left\|f(\sigma_y)\right\|$. Then, since $f(\sigma_x), f(\sigma_y)$ are traceless, we can continue

$$m(s,t) = \frac{1}{2}\operatorname{Tr}\left(\cosh\left(s\left\|f(\sigma_x)\right\|\right)I + \sinh\left(s\left\|f(\sigma_x)\right\|\right)\frac{f(\sigma_x)}{\left\|f(\sigma_x)\right\|}\right)$$

$$\times \left(\cosh\left(t \left\| f(\sigma_y) \right\|\right) I + \sinh\left(t \left\| f(\sigma_y) \right\|\right) \frac{f(\sigma_y)}{\left\| f(\sigma_y) \right\|} \right)$$

$$= \cosh\left(s \left\| f(\sigma_x) \right\|\right) \cosh\left(t \left\| f(\sigma_y) \right\|\right) +$$

$$+ \langle X, Y \rangle \sinh\left(s \left\| f(\sigma_x) \right\|\right) \sinh\left(t \left\| f(\sigma_y) \right\|\right). \tag{96}$$

We show that $\left\| f(\sigma_x) \right\| = \left\| f(\sigma_y) \right\|$. To this, set $\alpha := \left\| f(\sigma_x) \right\|$, $\beta := \left\| f(\sigma_y) \right\|$, $\gamma := \langle X, Y \rangle$. It is easy to check that

$$\lim_{t\to\infty} \frac{1}{t} \cosh^{-1}\left(\cosh^2(t)\right) = 2$$

and

$$\lim_{t\to\infty} \sqrt{\frac{\left(\cosh^4(t) - \cosh^2(t)\right)\alpha^2 + 2\alpha\beta\gamma \sinh^2(t)\cosh(t) + \left(\cosh^2(t) - 1\right)\beta^2}{\cosh^4(t) - 1}}$$

$$= \alpha.$$

From these we get that for every $0 < \varepsilon (< 2)$ there exists some $0 < T_\varepsilon$ such that $l(t,t) \geq \cosh((2-\varepsilon)\alpha t)$ holds for $t > T_\varepsilon$. On the other hand, it is easy to see that

$$m(t,t) = \frac{1}{4} e^{(\alpha+\beta)t} \left(1 + \gamma + o(1)\right).$$

Therefore, the inequality

$$m(t,t) = l(t,t) \geq \cosh((2-\varepsilon)\alpha t)$$

is equivalent to

$$\frac{1}{4}\left(1 + \gamma + o(1)\right) \geq \frac{1}{2}\left(e^{((1-\varepsilon)\alpha - \beta)t} + e^{-((3-\varepsilon)\alpha+\beta)t}\right) \tag{97}$$

Taking the limit $t \to \infty$ in (97) we infer that $\beta \geq (1-\varepsilon)\alpha$. This is true for any $0 < \varepsilon < 2$, hence letting $\epsilon \to 0$ we obtain $\beta \geq \alpha$, that is, $\left\| f(\sigma_y) \right\| \geq \left\| f(\sigma_x) \right\|$. By changing the roles of σ_x and σ_y we get the desired equality $\left\| f(\sigma_y) \right\| = \left\| f(\sigma_x) \right\|$. Having this in mind, it is clear that the function $m(\cdot,\cdot)$ is symmetric in the sense that we have $m(s,t) = m(t,s)$ for all $0 < s, t \in \mathbb{R}$, see (96). It follows that $l(\cdot,\cdot)$ is also symmetric which can happen only when $\left\langle f(\sigma_x), f(\sigma_y) \right\rangle = 0$, see (95). Therefore, we have $\| f(\sigma_x) \| = \| f(\sigma_y) \|$, $\left\langle f(\sigma_x), f(\sigma_y) \right\rangle = 0$ and we are done in the case where $f(\sigma_x), f(\sigma_y) \neq 0$.

Assume now that $f(\sigma_x) = 0, f(\sigma_y) \neq 0$. By (95), (96) we have

$$\cosh\left(\frac{\cosh^{-1}(\cosh(s)\cosh(t))}{\sqrt{\cosh^2(s)\cosh^2(t) - 1}} \sqrt{\left(\cosh^2(t) - 1\right) \left\| f(\sigma_y) \right\|^2} \right)$$

$$= \cosh\left(t \left\| f(\sigma_y) \right\|\right).$$

It follows that

$$\frac{\cosh^{-1}\left(\cosh^2(t)\right)}{t}\sqrt{\frac{\left(\cosh^2(t)-1\right)\left\|f(\sigma_y)\right\|^2}{\cosh^4(t)-1}}=\left\|f(\sigma_y)\right\|.$$

Letting t tend to infinity, we obtain $f(\sigma_y)=0$, a contradiction.

Assume $f(\sigma_x)\neq 0, f(\sigma_y)=0$. Again, by (95), (96) we have

$$\cosh\left(\frac{\cosh^{-1}(\cosh(s)\cosh(t))}{\sqrt{\cosh^2(s)\cosh^2(t)-1}}\sqrt{\left(\cosh^2(s)\cosh^2(t)-\cosh^2(t)\right)\left\|f(\sigma_x)\right\|^2}\right)$$
$$=\cosh\left(s\left\|f(\sigma_x)\right\|\right).$$

It follows that

$$\frac{\cosh^{-1}\left(\cosh^2(t)\right)}{t}\sqrt{\frac{\left(\cosh^4(t)-\cosh^2(t)\right)\left\|f(\sigma_x)\right\|^2}{\cosh^4(t)-1}}=\left\|f(\sigma_x)\right\|.$$

Letting t tend to infinity, we deduce $2\left\|f(\sigma_x)\right\|=\left\|f(\sigma_x)\right\|$, i.e., $\left\|f(\sigma_x)\right\|=0$, a contradiction again. So it remains only the possibility $f(\sigma_x)=f(\sigma_y)=0$ and this proves Claim 2.

To complete the proof of our theorem, let us see what happens when the restriction of f onto $\mathscr{B}_0^{sa}(\mathscr{H})$ is zero. We have $f(I)=(2c)I$ with some real number c. Then $f(A)=c(\mathrm{Tr}\,A)I$, $A\in\mathscr{B}^{sa}(\mathscr{H})$ and we obtain $\phi(A)=(\mathrm{Det}A)^cI$, $A\in\mathscr{B}^{++}(\mathscr{H})$. This means that ϕ is of the form (b3).

Now assume that the restriction of f onto $\mathscr{B}_0^{sa}(\mathscr{H})$ is a positive scalar multiple of an isometry. It follows that in the orthonormal basis (90), the transformation f has the block-matrix form

$$f=p\begin{bmatrix} v & 0\\ 0 & M\end{bmatrix},$$

where p is a positive real number, v is a real number and M is a 3×3 orthogonal matrix.

If $M\in\mathbf{SO}(3)$, then

$$f=p\begin{bmatrix} 1+2c & 0\\ 0 & R\end{bmatrix}$$

for some $c\in\mathbb{R}$ and $R\in\mathbf{SO}(3)$. Similarly, if $-M\in\mathbf{SO}(3)$, then

$$f=p\begin{bmatrix} -1+2c & 0\\ 0 & -R\end{bmatrix}$$

for some $c\in\mathbb{R}$ and $R\in\mathbf{SO}(3)$.

For any $R\in\mathbf{SO}(3)$ there exists a $U\in\mathbf{SU}(2)$ such that the matrix of the transformation $A\mapsto UAU^*$ is

$$\begin{bmatrix} 1 & 0\\ 0 & R\end{bmatrix},$$

see [**84**, Proposition VII.5.7.]. Therefore, in the case where $M \in \mathbf{SO}(3)$ we have

$$\begin{aligned}\phi(A) &= \exp(f(\log(A))) = \exp(f(\log A - (\mathrm{Tr}(\log A)/2)I)) + f((\mathrm{Tr}(\log A)/2)I))\\ &= \exp(pU(\log A - (\mathrm{Tr}(\log A)/2)I)U^*)\exp(p(1+2c)\,\mathrm{Tr}(\log A)/2)\\ &= \exp(pU(\log A)U^*)\exp((pc)\,\mathrm{Tr}(\log A)) = (\mathrm{Det}A)^{pc}UA^pU^*.\end{aligned}$$

Since $\phi(ABA) = \phi(A)\phi(B)\phi(A)$, we infer $(ABA)^p = A^pB^pA^p$, $A,B \in \mathscr{B}^{++}(\mathscr{H})$ which holds only if $p = 1$. Consequently, we have $\phi(A) = (\mathrm{Det}A)^cUAU^*$, $A \in \mathscr{B}^{++}(\mathscr{H})$. This means that ϕ is of the form (b1). Similarly, in the case where $-M \in \mathbf{SO}(3)$ one can conclude $\phi(A) = (\mathrm{Det}A)^cUA^{-1}U^*$, $A \in \mathscr{B}^{++}(\mathscr{H})$, i.e, ϕ is of the form (b2). The proof of the theorem is complete. □

One can notice that in Theorem describing the structure of continuous Jordan triple endomorphisms of $\mathscr{B}^{++}(\mathscr{H})$, in the case where $\dim(\mathscr{H}) \geq 3$ the transpose operation and its composition with the inverse operation also appear and one may ask why it is not so in the case where $\dim(\mathscr{H}) = 2$. There is no contradiction here, it is easy to see that in fact those two possibilities do appear in Theorem 47 in a hidden way. Indeed, when $\dim(\mathscr{H}) = 2$, the transpose operation can be written in the form (a2) above. Namely, for the unitary operator

$$U = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

we have $A^{tr} = (\mathrm{Det}A)UA^{-1}U^*$ for all $A \in \mathscr{B}^{++}(\mathscr{H})$.

The following structural result concerning the continuous Jordan triple automorphisms of $\mathscr{B}^{++}(\mathscr{H})$ follows from the proof of Theorem 47.

THEOREM 48. *Assume that* $\dim(\mathscr{H}) = 2$. *If* $\phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ *is a continuous Jordan triple automorphism, then* ϕ *is of one of the following two forms:*

(c1) *there is a real number* $c \neq -1/2$ *and* $U \in \mathbf{SU}(2)$ *such that*

$$\phi(A) = (\mathrm{Det}A)^cUAU^*, \quad A \in \mathscr{B}^{++}(\mathscr{H});$$

(c2) *there is a real number* $d \neq 1/2$ *and* $V \in \mathbf{SU}(2)$ *such that*

$$\phi(A) = (\mathrm{Det}A)^dVA^{-1}V^*, \quad A \in \mathscr{B}^{++}(\mathscr{H}).$$

The result above has the following immediate consequence. In the case where $\dim(\mathscr{H}) \geq 3$, in [**65**, Theorem 1] a general result was obtained describing the possible structure of surjective maps on $\mathscr{B}^{++}(\mathscr{H})$ which preserve a generalized distance measure of a certain quite general kind.

It is easy to see that, following the proof of [**65**, Theorem 1] and applying Theorem 48, the result in [**65**] remains valid also in the case where $\dim(\mathscr{H}) = 2$.

Now, we present an application of Theorem 47 for the description of so-called sequential endomorphisms of effect algebras.

Effects play an important role in certain parts of quantum mechanics, for instance, in the quantum theory of measurement [**12**]. Mathematically, effects are represented by positive semi-definite Hilbert space operators which are bounded (in the natural order $\leq$ among self-adjoint operators) by the identity. The set of all Hilbert space effects are called the Hilbert space effect algebra (although it is clearly not an algebra in the classical algebraic sense). In [**27**] Gudder and Nagy introduced the operation $\circ$ called sequential product on effects which has an important physical a meaning and which is closely related the Jordan triple product. Namely, they defined

$$A \circ B = A^{1/2} B A^{1/2}$$

for arbitrary Hilbert space effects A, B. The corresponding endomorphisms, i.e., maps ϕ on Hilbert space effects which satisfy

$$\phi(A \circ B) = \phi(A) \circ \phi(B)$$

for all pairs A, B of effects are called sequential endomorphisms. In the literature one can find results related to sequential automorphisms or isomorphisms (bijective sequential endomorphisms). For example, Gudder and Greechie proved in [**25**, Theorem 1] that, supposing the dimension of the underlying Hilbert space is at least 3, the sequential automorphisms of the Hilbert space effect algebra are exactly the transformations ϕ which are of the form $\phi : A \mapsto UAU^*$, where U is either a unitary or an antiunitary operator on the underlying Hilbert space. Afterwards, in [**61**] Lajos Molnár substantially generalized the previous results and described the structure of sequential isomorphisms between von Neumann algebra effects (i.e., between sets of effects on Hilbert spaces belonging to given von Neumann algebras).

In the paper [**18**] Molnár and *Dolinar* studied sequential endomorphisms of effect algebras over finite dimensional Hilbert spaces of dimension at least 3. Anybody can easily be convinced that the problem of describing non-bijective morphisms is usually much harder than that of describing bijective ones. In [**18**, Theorem 1] they managed to give the precise description of all continuous sequential endomorphisms assuming the dimension is at least 3. However, the 2-dimensional case remained unresolved and in [**18**, Remark 6] they proposed it as an open problem. Now, using Theorem 47 we can present a solution of the problem.

For any positive integer n denote by $\mathbb{E}_n$ the set of all positive semidefinite operators A which act on an n-dimensional Hilbert space and satisfy $A \le I$.

THEOREM 49. *Assume that* $\dim(\mathcal{H}) = 2$ *and* $\phi : \mathbb{E}_2 \to \mathbb{E}_2$ *is a continuous sequential endomorphism. Then we have the following four possibilities:*

(d1) *there exists a unitary* $U \in \mathscr{B}(\mathcal{H})$ *and a non-negative real number* c *such that*

$$\phi(A) = (\mathrm{Det} A)^c U A U^*, \quad A \in \mathbb{E}_2;$$

(d2) *there exists a unitary* $V \in \mathscr{B}(\mathcal{H})$ *such that*

$$\phi(A) = V(\mathrm{adj}\, A)V^*, \quad A \in \mathbb{E}_2;$$

(d3) *there exists a unitary* $V \in \mathscr{B}(\mathcal{H})$ *and a real number* $d > 1$ *such that*

$$\phi(A) = \begin{cases} (\mathrm{Det} A)^d V A^{-1} V^*, & \text{if } A \in \mathbb{E}_2 \text{ is invertible;} \\ 0, & \text{otherwise;} \end{cases}$$

(d4) *there exists a unitary* $W \in \mathscr{B}(\mathcal{H})$ *and non-negative real numbers* c_1, c_2 *such that*

$$\phi(A) = W \mathrm{Diag}[(\mathrm{Det} A)^{c_1}, (\mathrm{Det} A)^{c_2}] W^*, \quad A \in \mathbb{E}_2.$$

Here, we mean $0^0 = 1$.

PROOF. First observe that every sequential endomorphism $\phi : \mathbb{E}_2 \to \mathbb{E}_2$ is automatically a Jordan triple map. Indeed, we clearly have $\phi(A^2) = \phi(A)^2$, $A \in \mathbb{E}_2$. It implies that $\phi(\sqrt{A}) = \sqrt{\phi(A)}$, $A \in \mathbb{E}_2$ and hence it follows that ϕ is a Jordan triple map, i.e., ϕ satisfies $\phi(ABA) = \phi(A)\phi(B)\phi(A)$, $A, B \in \mathbb{E}_2$. Moreover, we infer that ϕ sends projections to projections implying that $\phi(I)$ is a projection. If $\phi(I) = 0$, we easily have that ϕ is identically zero. If $\phi(I) = P$ is a rank-one projection, then by $\phi(A) = \phi(IAI) = P\phi(A)P$ it follows the map $A \mapsto \phi(A) + (I - P)$, $A \in \mathbb{E}_2$ is a sequential endomorphism of $\mathbb{E}_2$ which is unital, i.e., it sends I to I.

Therefore, in what follows we may and do assume that our original transformation ϕ is a continuous unital sequential endomorphism (and hence a Jordan triple map).

Consider the function $\lambda \mapsto \mathrm{Det}\phi(\lambda I)$, $\lambda \in [0,1]$. Clearly, this is a continuous multiplicative map of the interval $[0,1]$ into itself which sends 1 to 1. Lemma 3 in [**18**] tells us that such a function is either everywhere equal to 1 or it is a power function corresponding to a positive exponent. This means that $\phi(\lambda I)$ is invertible for all $0 < \lambda \le 1$. We claim that ϕ sends invertible elements of $\mathbb{E}_2$ to invertible elements. To see this, first observe that ϕ preserves the usual order $\le$. Indeed, by [**26**, Theorem 5.1] we know that for any $A, B \in \mathbb{E}_2$ we have $A \le B$ if and only if there is a $C \in \mathbb{E}_2$ such that $A = B \circ C$. This clearly shows that for any $A, B \in \mathbb{E}_2$ with $A \le B$ we have $\phi(A) \le \phi(B)$. Now, if $A \in \mathbb{E}_2$ is invertible, then there is a scalar $0 < \lambda \le 1$ such that $\lambda I \le A$ holds which implies that $\phi(\lambda I) \le \phi(A)$. Since $\phi(\lambda I)$ is invertible, it follows that $\phi(A)$ is also invertible.

The sequential endomorphism ϕ preserves commutativity. This follows from the fact that for any pair A, B of effects we have $A \circ B = B \circ A$ if and only A, B as operators commute (see, e.g., Corollary 2.2 in [**27**]). It follows that the effects $\phi(\lambda I)$, $\lambda \in [0,1]$ all commute and hence they are jointly diagonizable. This means that up to unitary similarity we can write

$$\phi(\lambda I) = \begin{bmatrix} \varphi(\lambda) & 0 \\ 0 & \psi(\lambda) \end{bmatrix}, \quad \lambda \in [0,1]$$

where $\varphi, \psi : [0,1] \to [0,1]$ are continuous multiplicative functions which send 1 to 1. Therefore, by [**18**, Lemma 3] again, we have real numbers $c, d \ge 0$ such that

$$\phi(\lambda I) = \begin{bmatrix} \lambda^c & 0 \\ 0 & \lambda^d \end{bmatrix}, \quad \lambda \in [0,1].$$

We now distinguish two cases. Assume first that there is $\phi(A)$ which is not diagonal. Since $\phi(A)$ necessarily commute with $\phi(\lambda I)$, $\lambda \in [0,1]$, one can easily deduce that we necessarily have $c = d$. It follows that $\phi(\lambda I) = \lambda^c I$ and hence we have $\phi(\lambda A) = \lambda^c \phi(A)$ for all $\lambda \in [0,1], A \in \mathbb{E}_2$.

We next define $\Phi : \mathscr{B}^{++}(\mathscr{H}) \to \mathscr{B}^{++}(\mathscr{H})$ by

$$\Phi(A) = \|A\|^c \phi(A/\|A\|), \quad A \in \mathbb{E}_2. \tag{98}$$

In contrast to the proof of our main result, $\|.\|$ stands here for the operator norm (spectral norm) of operators; we do hope it does not cause serious confusion. It follows that for any invertible effect $A \in \mathbb{E}_2$ we have

$$\Phi(A) = \|A\|^c \phi(A/\|A\|) = \phi(\|A\|(A/\|A\|)) = \phi(A).$$

We assert that Φ is a Jordan triple endomorphism of $\mathscr{B}^{++}(\mathscr{H})$. Indeed, for any $A, B \in \mathscr{B}^{++}(\mathscr{H})$ we compute

$$\Phi(A)\Phi(B)\Phi(A) = \|A\|^{2c}\|B\|\phi\Big(\frac{A}{\|A\|}\Big)\phi\Big(\frac{B}{\|B\|}\Big)\phi\Big(\frac{A}{\|A\|}\Big)$$

$$= \|A\|^{2c}\|B\|\phi\Big(\frac{ABA}{\|A\|\|B\|\|A\|}\Big) = \|A\|^{2c}\|B\|\phi\Big(\frac{\|ABA\|}{\|A\|\|B\|\|A\|}\frac{ABA}{\|ABA\|}\Big)$$

$$= \|A\|^{2c}\|B\|^c\Big(\frac{\|ABA\|}{\|A\|\|B\|\|A\|}\Big)^c\phi\Big(\frac{ABA}{\|ABA\|}\Big) = \|ABA\|^c\phi\Big(\frac{ABA}{\|ABA\|}\Big) = \Phi(ABA),$$

where we have used the facts that $\|ABA\|/(\|A\|\|B\|\|A\|) \le 1$ and that $(ABA)/\|ABA\|$ is an effect. Clearly, Φ is continuous and hence Theorem 47 applies and we obtain that Φ is of one of the forms (b1), (b2). In the case of (b1), we have that $\phi(A) = (\mathrm{Det}A)^c UAU^*$ holds for all invertible $A \in \mathbb{E}_2$ with a given unitary operator U and real number c. Since ϕ sends effects to effects, it follows easily that c is necessarily non-negative. By continuity we deduce

$$\phi(A) = (\mathrm{Det}A)^c UAU^*, \quad A \in \mathbb{E}_2$$

yielding the possibility (d1). Consider now the case where $\phi(A) = (\mathrm{Det}A)^d VA^{-1}V^*$ holds for all invertible $A \in \mathbb{E}_2$ with a given unitary operator V and real number d. Again, since ϕ sends effects to effects, one can easily verify that $d \geq 1$. If $d = 1$, then we have

$$\phi(A) = V(\mathrm{adj}\, A)V^*$$

for all invertible $A \in \mathbb{E}_2$ and by continuity it follows that the same formula remains valid for any $A \in \mathbb{E}_2$, too. This gives us (d2). Assume $d > 1$. Letting A be an invertible effect tending to some non-invertible one, it follows that $\phi(A) = (\mathrm{Det}A)^d VA^{-1}V^*$ tends to 0. Hence, we obtain that

$$\phi(A) = \begin{cases} (\mathrm{Det}A)^d VA^{-1}V^*, & \text{if } A \in \mathbb{E}_2 \text{ is invertible;} \\ 0, & \text{otherwise} \end{cases}$$

and this is the possibility (d3).

It remains to discuss the case where all $\phi(A)$ are diagonal, that is when we have

$$\phi(A) = \begin{bmatrix} \varphi(A) & 0 \\ 0 & \psi(A) \end{bmatrix}, \quad A \in \mathbb{E}_2$$

for continuous (unital) Jordan triple maps $\varphi, \psi : \mathbb{E}_2 \to [0,1]$. As in (98), we can extend φ, ψ from the set of all invertible elements of $\mathbb{E}_2$ to continuous Jordan triple functionals $\tilde{\varphi}, \tilde{\psi} : \mathscr{B}^{++}(\mathscr{H}) \to]0,\infty[$. Applying Theorem 47, it follows that $\tilde{\varphi}, \tilde{\psi}$ are non-negative powers of the determinant function. Hence we obtain that

$$\phi(A) = \begin{bmatrix} (\mathrm{Det}A)^c & 0 \\ 0 & (\mathrm{Det}A)^d \end{bmatrix}, \quad A \in \mathbb{E}_2$$

holds for some non-negative real numbers c, d. This gives (d4) and the proof of the theorem is complete. □

3. An application: the endomorphisms of the Einstein gyrogroup

3.1. Introduction. Velocity addition was defined by Einstein in his famous paper of 1905 which founded the special theory of relativity. In fact, the whole theory is essentially based on Einstein velocity addition law, see [**20**]. The algebraic structure corresponding to this operation is a particular example of so-called gyrogroups the general theory of which has been developed by Ungar [**88**].

The Einstein gyrogroup of dimension three is the pair $(\mathbf{B}, \oplus)$, where $\mathbf{B} = \{\mathbf{u} \in \mathbb{R}^3 : \|\mathbf{u}\| < 1\}$ and $\oplus$ is the binary operation on $\mathbf{B}$ given by

$$\oplus : \mathbf{B} \times \mathbf{B} \to \mathbf{B};\ (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} \oplus \mathbf{v} := \frac{1}{1 + \langle \mathbf{u}, \mathbf{v} \rangle} \left(\mathbf{u} + \frac{1}{\gamma_{\mathbf{u}}} \mathbf{v} + \frac{\gamma_{\mathbf{u}}}{1 + \gamma_{\mathbf{u}}} \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{u} \right), \tag{99}$$

where $\gamma_{\mathbf{u}} = \left(1 - \|\mathbf{u}\|^2\right)^{-\frac{1}{2}}$ is the so-called Lorentz factor. The operation $\oplus$ is called Einstein velocity addition or relativistic sum (cf. [**1, 43**]). Here and throughout this section, $\langle \cdot, \cdot \rangle$ stands for the usual Euclidean inner product and $\|\cdot\|$ denotes the induced norm.

The main result of this research is obtained as an application of the result on the Jordan triple endomorphisms of positive definite operators acting on two-dimensional Hilbert spaces. The other ingredient of our argument is the result [**43**, Theorem 3.4] of Kim.

The main statement reads as follows.

THEOREM 50. *Let $\beta : \mathbf{B} \to \mathbf{B}$ be a continuous map. We have β is an algebraic endomorphism with respect to the operation $\oplus$, i.e., β satisfies*

$$\beta(\mathbf{u} \oplus \mathbf{v}) = \beta(\mathbf{u}) \oplus \beta(\mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in \mathbf{B}$$

if and only if

(i) *either there is an orthogonal matrix $O \in \mathbf{M}_3(\mathbb{R})$ such that*

$$\beta(\mathbf{v}) = O\mathbf{v}, \quad \mathbf{v} \in \mathbf{B};$$

(ii) *or we have*

$$\beta(\mathbf{v}) = 0, \quad \mathbf{v} \in \mathbf{B}.$$

Here continuity refers to the usual topology on $\mathbf{B}$ inherited from the Euclidean space $\mathbb{R}^3$.

By the above result we have the interesting conclusion that the group of all (continuous) automorphisms of the Einstein gyrogroup $(\mathbf{B}, \oplus)$ coincides with the orthogonal group of $\mathbb{R}^3$.

3.2. Open Bloch ball, qubit density operators, and positive definite operators of determinant one. To prove our main result we need an important observation made by Kim what we present below.

We denote by $\mathbb{P}_2$ the set of all 2×2 positive definite complex matrices. Let $\mathbb{D}$ stand for the set of all 2×2 regular density matrices, i.e., the collection of all elements of $\mathbb{P}_2$ with trace 1,

$$\mathbb{D} = \{A \in \mathbb{P}_2 \mid \operatorname{Tr} A = 1\}.$$

From the quantum theoretical point of view, $\mathbb{D}$ is the set of all regular density matrices of the 2-level quantum system. One can define a binary operation $\odot$ on $\mathbb{D}$ as

$$\odot : \mathbb{D}\times\mathbb{D}\to\mathbb{D};\ (A,B)\mapsto A\odot B := \frac{1}{\operatorname{Tr} AB} A^{\frac{1}{2}} B A^{\frac{1}{2}}.$$

For certain reasons, we call $\odot$ the normalized sequential product.

The well-known Bloch parametrization of regular density matrices is the following map:

$$\rho : \mathbb{R}^3 \supset \mathbf{B} \to \mathbf{M}_2(\mathbb{C});\ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{v} \mapsto \rho(\mathbf{v}) := \frac{1}{2}\begin{bmatrix} 1+v_3 & v_1 - i v_2 \\ v_1 + i v_2 & 1 - v_3 \end{bmatrix}.$$

The transformation ρ is clearly a bijection between $\mathbf{B}$ and $\mathbb{D}$, and in fact, by [**43**, Theorem 3.4], much more is true.

THEOREM 51 (S. Kim). *The Bloch parametrization* $\rho : (\mathbf{B},\oplus) \to (\mathbb{D},\odot);\ \mathbf{v}\mapsto\rho(\mathbf{v})$ *is an isomorphism.*

Let us now consider a structure which is similar to the space of 2×2 regular density matrices equipped with the normalized sequential product. Namely, Let $\mathbb{P}_2^1$ be the set of all 2×2 positive definite matrices with determinant 1. The sequential product $\boxdot$ on $\mathbb{P}_2^1$ is defined as

$$\boxdot : \mathbb{P}_2^1 \times \mathbb{P}_2^1 \to \mathbb{P}_2^1;\ (A,B)\mapsto A\boxdot B := A^{\frac{1}{2}} B A^{\frac{1}{2}}.$$

We show that $(\mathbb{D},\odot)$ and $(\mathbb{P}_2^1,\boxdot)$ are isomorphic structures.

PROPOSITION 52. *The map* $\tau : (\mathbb{D},\odot) \to (\mathbb{P}_2^1,\boxdot);\ A\mapsto \tau(A) := \frac{1}{\sqrt{\operatorname{Det} A}} A$ *is an isomorphism.*

PROOF. To prove the injectivity assume that $\tau(A) = \tau(B)$ for some $A,B\in\mathbb{D}$. That is, $\frac{1}{\sqrt{\operatorname{Det} A}} A = \frac{1}{\sqrt{\operatorname{Det} B}} B$, which means that A is a positive scalar multiple of B. By $\operatorname{Tr} A = \operatorname{Tr} B = 1$ we can deduce that $\sqrt{\operatorname{Det} A} = \sqrt{\operatorname{Det} B}$ and therefore $A = B$.

For any $A\in\mathbb{P}_2^1$ we have $\frac{1}{\operatorname{Tr} A} A \in \mathbb{D}$. By $\operatorname{Det} A = 1$ it follows that

$$\tau\left(\frac{1}{\operatorname{Tr} A} A\right) = \frac{1}{\sqrt{\operatorname{Det}\left(\frac{1}{\operatorname{Tr} A} A\right)}} \frac{1}{\operatorname{Tr} A} A = \frac{1}{\frac{\sqrt{\operatorname{Det} A}}{\operatorname{Tr} A}} \frac{1}{\operatorname{Tr} A} A = A.$$

This shows the surjectivity of τ.

Finally, we need to show that τ respects the operations $\odot,\boxdot$. Using the properties of the determinant, for any $A,B\in\mathbb{D}$ we compute

$$\tau(A\odot B) = \frac{1}{\sqrt{\operatorname{Det}\left(\frac{1}{\operatorname{Tr} AB} A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)}} \frac{A^{\frac{1}{2}} B A^{\frac{1}{2}}}{\operatorname{Tr} AB} = \frac{\operatorname{Tr} AB}{\sqrt{\operatorname{Det}\left(A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)}} \frac{A^{\frac{1}{2}} B A^{\frac{1}{2}}}{\operatorname{Tr} AB}$$

$$= \left(\frac{A}{\sqrt{\operatorname{Det} A}}\right)^{\frac{1}{2}} \frac{B}{\sqrt{\operatorname{Det} B}} \left(\frac{A}{\sqrt{\operatorname{Det} A}}\right)^{\frac{1}{2}} = \frac{A}{\sqrt{\operatorname{Det} A}} \boxdot \frac{B}{\sqrt{\operatorname{Det} B}} = \tau(A) \boxdot \tau(B).$$

This completes the proof. □

Observe that the inverse of τ is given by $\tau^{-1}(A) = \frac{1}{\operatorname{Tr} A} A$, $A \in \mathbb{P}_2$.

3.3. Proof of the main result. Using the result of Theorem 47, the continuous sequential endomorphisms of $\mathbb{P}_2^1$ can be described as follows.

PROPOSITION 53. *Let $\phi : \mathbb{P}_2^1 \to \mathbb{P}_2^1$ be a continuous endomorphism with respect to the operation $\boxdot$ meaning that ϕ satisfies*

$$\phi(A \boxdot B) = \phi(A) \boxdot \phi(B), \quad A, B \in \mathbb{P}_2^1.$$

Then ϕ is of one of the following forms:

(1) *there is a unitary matrix $U \in \mathbf{M}_2(\mathbb{C})$ such that*

$$\phi(A) = UAU^*, \quad A \in \mathbb{P}_2^1;$$

(2) *there is a unitary matrix $V \in \mathbf{M}_2(\mathbb{C})$ such that*

$$\phi(A) = VA^{-1}V^*, \quad A \in \mathbb{P}_2^1;$$

(3) *we have*

$$\phi(A) = I, \quad A \in \mathbb{P}_2^1.$$

PROOF. If $\phi : \mathbb{P}_2^1 \to \mathbb{P}_2^1$ is a sequential endomorphism, then it is a Jordan triple endomorphism, as well. Indeed, $\phi(A^2) = \phi(A \boxdot A) = \phi(A) \boxdot \phi(A) = \phi(A)^2$ holds for all $A \in \mathbb{P}_2^1$. It follows that $\phi(ABA) = \phi(A^2 \boxdot B) = \phi(A^2) \boxdot \phi(B) = \left(\phi(A)^2\right)^{\frac{1}{2}} \phi(B) \left(\phi(A)^2\right)^{\frac{1}{2}} = \phi(A)\phi(B)\phi(A)$ for all $A, B \in \mathbb{P}_2^1$.

The map

$$\psi : \mathbb{P}_2 \to \mathbb{P}_2;\ A \mapsto \psi(A) := \sqrt{\operatorname{Det} A} \cdot \phi\left(\frac{A}{\sqrt{\operatorname{Det} A}}\right)$$

is clearly a continuous Jordan triple endomorphism of $\mathbb{P}_2$ which extends ϕ (the idea of the definition of ψ comes from [**65**, proof of Theorem 3]). Now, the statement is an immediate consequence of the previous theorem. □

Using the isomorphism τ defined in Proposition 52 which is clearly a homeomorphism, too, we can pull back the structural result on the continuous endomorphisms of $(\mathbb{P}_2^1, \boxdot)$ to $(\mathbb{D}, \odot)$. Namely, the continuous endomorphism of $(\mathbb{D}, \odot)$ are exactly the maps of the form

$$\tau^{-1} \circ \phi \circ \tau,$$

where ϕ is a continuous endomorphism of $(\mathbb{P}_2^1, \boxdot)$. The following corollary can be verified by straightforward computations.

COROLLARY 54. *Let $\alpha : \mathbb{D} \to \mathbb{D}$ be a continuous endomorphism with respect to the operation $\odot$. Then α is of one of the following forms:*

(1) *there is a unitary matrix $U \in \mathbf{M}_2(\mathbb{C})$ such that*
$$\alpha(A) = UAU^*, \quad A \in \mathbb{D};$$
(2) *there is a unitary matrix $V \in \mathbf{M}_2(\mathbb{C})$ such that*
$$\alpha(A) = \frac{VA^{-1}V^*}{\operatorname{Tr} A^{-1}}, \quad A \in \mathbb{D};$$
(3) *we have*
$$\alpha(A) = I/2, \quad A \in \mathbb{D}.$$

Putting all information we have together, the proof of the main result is now easy.

PROOF. We have learned from the result Theorem 51 due to Kim that the Bloch parametrization ρ is an isomorphism between $(\mathbf{B}, \oplus)$ and $(\mathbb{D}, \odot)$. Clearly, ρ is a homeomorphism, too. Therefore, the continuous endomorphisms of $(\mathbf{B}, \oplus)$ are exactly the maps of the form $\beta = \rho^{-1} \circ \alpha \circ \rho$, where α is a continuous endomorphism of $(\mathbb{D}, \odot)$.

By Corollary 54 there are three possibilities. Assume first that we have a unitary $U \in \mathbf{M}_2$ such that $\alpha(A) = UAU^*$, $A \in \mathbb{D}$. Denote by $\mathbf{H}_2^0(\mathbb{C})$ the linear space of all traceless self-adjoint 2×2 complex matrices and equip this space with the inner product $\langle A, B \rangle := \frac{1}{2} \operatorname{Tr} AB$, $A, B \in \mathbf{H}_2^0(\mathbb{C})$. Define
$$\gamma : \mathbb{R}^3 \to \mathbf{H}_2^0(\mathbb{C}); \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{v} \mapsto \gamma(\mathbf{v}) := \begin{bmatrix} v_3 & v_1 - i v_2 \\ v_1 + i v_2 & -v_3 \end{bmatrix}.$$
Clearly, γ is a linear isomorphism from $\mathbb{R}^3$ onto $\mathbf{H}_2^0(\mathbb{C})$ which preserves the inner product. Define $\tilde{\alpha} : \mathbf{H}_2^0(\mathbb{C}) \to \mathbf{H}_2^0(\mathbb{C})$ by $\tilde{\alpha}(A) = UAU^*$, $A \in \mathbf{H}_2^0(\mathbb{C})$. Then $O := \gamma^{-1} \circ \tilde{\alpha} \circ \gamma$ is an orthogonal linear transformation on $\mathbb{R}^3$ and using the relation $\gamma(\mathbf{v}) = 2\rho(\mathbf{v}) - I$, $\mathbf{v} \in \mathbf{B}$, we easily deduce that $\rho \circ \alpha \circ \rho = O$ holds.

If $\alpha(A) = \frac{VA^{-1}V^*}{\operatorname{Tr} A^{-1}}$, $A \in \mathbb{D}$, then the conclusion follows from the previous case. The only thing we have to observe is that $\frac{(\rho(\mathbf{v}))^{-1}}{\operatorname{Tr}(\rho(\mathbf{v}))^{-1}} = \rho(-\mathbf{v})$ holds which follows from [**43**, Remark 3.5].

Finally, if $\alpha(A) = I/2$, then we clearly have $\rho^{-1} \circ \alpha \circ \rho = 0$.

The converse statement that the formulas in (i) and (ii) define continuous endomorphisms of the Einstein gyrogroup is just obvious. □

CHAPTER 4

Summary

In this thesis we investigated noncommutative variances, certain generalized quantum entropies and relative entropies, and we presented the solutions of some related preserver problems.

In the mathematical model of quantum systems, the states are described by positive trace-class operators of trace one. These operators are called stochastic operators or — in the finite dimensional case — density matrices. These objects are non-commutative generalizations of the classical probability distributions. The quantities that play important roles in the classical probability theory — variance, entropies — can be defined in this non-commutative environment, as well. In the first part of the thesis, we gave an interesting characterization of the decomposability of certain variances.

One of the most important quantities in the classical probability theory is the *Shannon entropy*. The non-commutative generalization is the *von Neumann entropy*, which has a well-known one-parameter extension called *Tsallis entropy*. The strong subadditivity of the von Neumann entropy is a celebrated result in quantum information theory. In the thesis, we showed that the Tsallis entropy is not strongly subadditive and we derived further relevant inequalities concerning the Tsallis entropy and relative entropy.

The *Bregman divergences* are similar to the relative entropies in the sense that they measure the dissimilarity between density matrices, or more generally, positive matrices. In general, these quantities are not symmetric and they do not satisfy the triangle-inequality. We gave a necessary and sufficient condition for the joint convexity of the Bregman divergences concerning the generating functions of them. Using this characterization, we derived a sharp inequality for the Tsallis entropy, which is a generalization of the strong subadditivity of the von Neumann entropy.

Furthermore, we determined the transformations of the cone of positive matrices which preserve the Bregman divergences and the *Jensen divergences*. We considered other preserver problems concerning the

cone of positive matrices, as well. The description of the transformations of positive definite operators acting on a two-dimensional Hilbert space which preserve the *Jordan triple product* was an open problem for a while. In the thesis, we gave the solution of this problem. As an application, we determined the algebraic endomorphisms of the three-dimensional Einstein gyropgroup.

Ackowledgement

I would like to express my deep gratitude to my supervisor, *Professor Dénes Petz*, who introduced me to the topic of matrix analysis and provided me several interesting questions. Moreover, we managed to solve some of these questions together.

I am also very grateful to *Professor Lajos Molnár* for so many reasons that I do not have enough space to enumerate all of them. Let me mention the most important ones. He proposed me and my classmates a very interesting open problem while giving us an excellent intensive course at the BUTE as an invited guest professor form the University of Debrecen. Afther a while we solved with Lajos the aforementioned question and that was the starting point of our flourishing collaboration. Among others, quite the half of my thesis is based on common works with Lajos.

I am grateful to *József Pitrik* for providing interesting problems and for the great common work experience and to *Professor Miklós Horváth* for his support during my PhD studies.

I would like to thank the members of my family — in the order of appearance, — my mother, *Katalin Erdősi*, my father, *Attila Virosztek*, my brother, *Tamás Virosztek*, my wife, *Anna Gelniczky* and my son, *Márton Virosztek* for their continuous support.

It is a pleasure to thank Anna also for her patience and tolerance which properties turned out to be really useful in the periods when I was trying to do something in mathematics. I am also grateful to *Márton Virosztek* for his valuable comments and remarks on the thesis.

During the PhD period, my research was supported by the “Lendület” Program (LP2012-46/2012) of the Hungarian Academy of Sciences, by a grant of the Hungarian Scientific Research Fund (OTKA Reg. No. K104206) and by the “For the Young Talents of the Nation” scholarship program (NTP-EFÖ-P-15-0481) of the Hungarian State.

Bibliography

[1] T. Abe, *Gyrometric preserving maps on Einstein gyrogroups, Möbius gyrogroups and proper velocity gyrogroups*, Nonlinear Functional Analysis and Applications **19** (2014), 1-17.

[2] J. Aczél and Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, San Diego, 1975.

[3] T. Ando and F. Hiai, *Operator log-convex functions and operator means*, Mathematische Annalen **350(3)**(2011), 611-630.

[4] H. Araki, *Relative entropy of state of von Neumann algebras*, Publ. RIMS Kyoto Univ. **9**(1976), 809 - 833.

[5] K. M. R. Audenaert, *Subadditivity of q-entropies for $q > 1$*, J. Math. Phys. **48**(2007), 083507.

[6] A. Banerjee et al., *Clustering with Bregman divergences*, J. Mach. Learn. Res. **6**(2005), 1705-1749.

[7] H. Bauschke and J. Borwein, *Joint and separate convexity of the Bregman distance*, Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (Haifa 2000), D. Butnariu, Y. Censor, S. Reich (editors), Elsevier, pp. 23-36, 2001.

[8] Á. Besenyei and D. Petz, *Partial subadditivity of entropies*, Linear Algebra and its Applications, **439**(2013), 3297 - 3305.

[9] R. Bhatia, *Matrix Analysis,* Springer-Verlag, New York, 1996.

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

[11] L. M. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics **7(3)**(1967), 200-217.

[12] P. Busch, P.J. Lahti and P. Mittelstaedt, *The Quantum Theory of Measurement,* Springer-Verlag, 1991.

[13] E. Carlen, *Trace inequalities and quantum entropy: an introductory course,* Contemp. Math. **529** (2010), 73-140.

[14] G. Cassinelli, E. De Vito, P.J. Lahti and A. Levrero, *The Theory of Symmetry Actions in Quantum Mechanics,* Lecture Notes in Physics 654, Springer, 2004.

[15] Z. Chebbi and M. Moakher, *Means of hermitian positive-definite matrices based on the log-determinant α-divergence function,* Linear Algebra Appl. **436** (2012), 1872–1889.

[16] R. Y. Chen and J. A. Tropp, *Subadditivity of matrix φ-entropy and concentration of random matrices*, Electron. J. Probab. **19**(2014), 1-30.

[17] Z. Daróczi, *General information functions*, Information and Control **16**(1970), 36 - 51.

[18] G. Dolinar and L. Molnár, *Sequential endomorphisms of finite dimensional Hilbert space effect algebras,* J. Phys. A: Math. Theor. **45** (2012), 065207.

[19] E. Effros, *A matrix convexity approach to some celebrated quantum inequalities*, Proc. Natl. Acad. Sci. USA, **106**(2009), 1006 - 1008.

[20] A. Einstein, *Einstein's Miraculous Years: Five Papers That Changed the Face of Physics,* Princeton University, Princeton, NJ, 1998.

[21] W. Feller, *An Introduction to Probability Theory and its Applications I*, Oxford, England: Wiley, 1950.

[22] S. Furuichi, *Information theoretical properties of Tsallis entropies*, J. Math. Phys. **47**, 023302 (2006)

[23] S. Furuichi, K. Yanagi and K. Kuriyama, *Fundamental properties of Tsallis relative entropy*, J. Math.Phys. **45**(2004), 4868-4877.

[24] P. Gibilisco, F. Hiai and D. Petz, *Quantum covariance, quantum Fisher information and the uncertainty principle*, IEEE Trans. Inform. Theory **55**(2009), 439-443.

[25] S. Gudder and R. Greechie, *Sequential products on effect algebras,* Rep. Math. Phys. **49** (2002), 87–111.

[26] S. Gudder and R. Greechie, *Uniqueness and order in sequential effect algebras,* Int. J. Theor. Phys. **44** (2005), 755–770.

[27] S. Gudder and G. Nagy, *Sequential quantum measurements,* J. Math. Phys. **42** (2001), 5212–5222.

[28] F. Hansen, *Extensions of Lieb's Concavity Theorem*, J. Stat. Phys. **124**(2006), 87-101.

[29] F. Hansen, *Trace functions as Laplace transforms*, J. Math. Phys. **47**, 043504 (2006).

[30] F. Hansen and Z. Zhang, *Characterization of matrix entropies*, arXiv:1402:2118v2, 20 Mar., 2014.

[31] F. Hansen and Z. Zhang, *Characterization of matrix entropies*, arXiv:1402:2118v3, 16 Mar., 2015.

[32] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[33] O. Hatori and L. Molnár, *Isometries of the unitary groups and Thompson isometries of the spaces of invertible positive elements in C^*-algebras,* J. Math. Anal. Appl. **409** (2014), 158–167.

[34] F. Hiai and D. Petz, *From quasi-entropy to various quantum information quantities*, Publ. RIMS Kyoto University **48**(2012), 525 - 542.

[35] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications*, Hindustan Book Agency and Springer Verlag, 2014.

[36] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North-Holland, Amsterdam, 1982.

[37] R.A. Horn and C.R. Johnson, *Matrix Analysis,* Cambridge University Press, 1990.

[38] F. Itakura and S. Saito, *Analysis synthesis telephony based on the maximum likelihood method*, in 6th Int. Congr. Acoustics, Tokyo, Japan., pp. C-17-C-20 (1968)

[39] A. Járai, *Regularity Properties of Functional Equations in Several Variables,* Advances in Mathematics (Springer), Springer, New York, 2005.

[40] A. Jencová and M. B. Ruskai, *A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality*, Rev. Math. Phys. **22**(2010), 1099 - 1121.

[41] R. V. Kadison and J. R. Ringrose, *Fundamentals of the Theory of Operator Algebras,* Volumes I and II, Academic Press, Orlando, 1983 and 1986.

[42] I. H. Kim, *Operator extension of strong subadditivity of entropy,* J. Math. Phys. **53**(2012), 122204.

[43] S. Kim, *Distances of qubit density matrices on Bloch sphere*, J. Math. Phys. **52**, 102303 (2011).

[44] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung,* Springer Verlag, Berlin, 1933; English translation: *Foundations of the Theory of Probability,* Chelsea Publishing Co., New York, 1956.

[45] S. Kullback and R.A: Leibler, *On information and sufficiency*, Ann. Math. Statist. **22(1)**(1951), 79 - 86.

[46] Z. Léka and D. Petz, *Some decompositions of matrix variances*, Probability and Mathematical Statistics, **33**(2013), 191-199.

[47] A. Lesniewski, M. B. Ruskai, *Monotone Riemannian metrics and relative entropy on non-commutative probability spaces*, J. Math. Phys. **40**(1999), 5702-5724.

[48] M. Lewin and J. Sabin, *A family of monotone quantum relative entropies*, Lett. Math. Phys. **104**(2014) 691-705.

[49] E. Lieb and M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys. **14**(1973), 1938-1941.

[50] E. H. Lieb, M. B. Ruskai, *Some operator inequalities of the Schwarz type*, Adv. in Math. **12**(1974), 269-273.

[51] G. Linblad, *Expectations and entropy inequalities*, Commun. Math. Phys. **39**(1974), 111-119.

[52] P.C. Mahalonobis, *On the generalized distance in statistics*, Proceedings of National Institute of Science of India, **12**(1936), 49 - 55.

[53] L. Molnár, *A remark to the Kochen-Specker theorem and some characterizations of the determinant on sets of Hermitian matrices,* Proc. Amer. Math. Soc. **134** (2006), 2839-2848.

[54] L. Molnár, *General Mazur-Ulam type theorems and some applications,* in Operator Semigroups Meet Complex Analysis, Harmonic Analysis and Mathematical Physics, W. Arendt, R. Chill, Y. Tomilov (Eds.), Operator Theory: Advances and Applications, Vol. 250, pp. 311-342, Birkhäuser, 2015.

[55] L. Molnár, *Jordan triple endomorphisms and isometries of spaces of positive definite matrices,* Linear Multilinear Alg. **63** (2015), 12–33.

[56] L. Molnár, *Maps on states preserving the relative entropy* J. Math. Phys. **49,** 032114 (2008).

[57] L. Molnár, *Order-automorphisms of the set of bounded observables,* J. Math. Phys. **42** (2001), 5904-5909.

[58] L. Molnár, *Order automorphisms on positive definite operators and a few applications,* Linear Algebra Appl. **434** (2011), 2158-2169.

[59] L. Molnár, *Preservers on Hilbert space effects,* Linear Algebra Appl. **370** (2003), 287–300.

[60] L. Molnár, *Selected Preserver Problems on Algebraic Structures of Linear Operators and on Function Spaces,* Lecture Notes in Mathematics, Vol. 1895, p. 236, Springer, 2007.

[61] L. Molnár, *Sequential isomorphisms between the sets of von Neumann algebra effects,* Acta Sci. Math. (Szeged) **69** (2003), 755–772.

[62] L. Molnár and G. Nagy, *Transformations on density operators that leave the Holevo bound invariant,* Int. J. Theor. Phys. **53** (2014), 3273–3278.

[63] L. Molnár, G. Nagy and P. Szokol, *Maps on density operators preserving quantum f-divergences,* Quantum Inf. Process. **12** (2013), 2309-2323.

[64] L. Molnár, J. Pitrik and D. Virosztek, *Maps on positive definite matrices preserving Bregman and Jensen divergences,* Linear Algebra Appl. **495** (2016), 174–189.

[65] L. Molnár and P. Szokol, *Transformations on positive definite matrices preserving generalized distance measures,* Linear Algebra Appl. **466** (2015), 141–159.

[66] L. Molnár and D. Virosztek, *Continuous Jordan triple endomorphisms of* $\mathbb{P}_2$, J. Math. Anal. Appl. **438(2)** (2016), 828-839.

[67] L. Molnár and D. Virosztek, *On algebraic endomorphisms of the Einstein gyrogroup,* J. Math. Phys. **56**, 082302 (2015).

[68] F. J. Murray and J. von Neumann, *On rings of operators,* Annals of Mathematics **37** (1936), 116-229.

[69] F. Nielsen and R. Bhatia (Eds), *Matrix Information Geometry,* Springer, Heidelberg, 2013.

[70] M. Nielsen and D. Petz, *A simple proof of the strong subadditivity inequality*, Quantum Information & Computation, **6**(2005), 507 - 513.

[71] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993. Second edition 2004.

[72] D. Petz, *Bregman divergence as relative operator entropy*, Acta Math. Hungar, **116**(2007), 127-131.

[73] D. Petz, *Covariance and Fisher information in quantum mechanics*, J. Phys. A, Math. Gen. **35**(2003), 79-91.

[74] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer-Verlag, Heidelberg, 2008.

[75] D. Petz, *Quasi-entropies for finite quantum systems*, Rep. Math. Phys. **23**(1986), 57 - 65.

[76] D. Petz, *Quasi-entropies for states of a von Neumann algebra,* Publ. RIMS. Kyoto Univ. **21** (1985), 781–800.

[77] D. Petz and G. Tóth, *Matrix variances with projections*, Acta Sci. Math. (Szeged), **78**(2012), 683–688.

[78] D. Petz and D. Virosztek, *A characterization theorem for matrix variances*, Acta Sci. Math. (Szeged) **80** (2014), 681-687.

[79] D. Petz and D. Virosztek, *Some inequalities for quantum Tsallis entropy related to the strong subadditivity*, Math. Inequal. Appl. **18(2)**(2015), 555-568.

[80] J. Pitrik and D. Virosztek, *On the joint convexity of the Bregman divergence of matrices,* Lett. Math. Phys. **105** (2015), 675-692.

[81] M. Rédei and S.J. Summers, *Quantum probability theory,* Studies in the History and Philosophy of Modern Physics **38** (2007), 390-417.

[82] R.T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton, 1970.

[83] N. Sharma, *Equality conditions for quantum quasi-entropies under monotonicity and joint-convexity*, Nat. Conf. Commun. (NCC), 2014.

[84] B. Simon, *Representations of Finite and Compact Groups,* Graduate Studies in Math., 10, American Mathematical Society, Providence, R.I., 1996.

[85] S. Sra, *Positive definite matrices and the S-divergence,* preprint, arXiv math.FA-1110.1773v4. Shorter version of the manuscript is submitted to Proc. Amer. Math. Soc.

[86] J. A. Tropp, *From joint convexity of quantum relative entropy to a concavity theorem of Lieb*, Proc. Amer. Math. Soc. **140**(2012), 1757-1760.

[87] A. Uhlmann, *Roofs and convexity*, Entropy, **12(7)** (2010), 1799-1832.

[88] A.A. Ungar, *Analytic Hyperbolic Geometry and Albert Einstein's Special Theory of Relativity,* World Scientific, Singapore, 2008.