

MAPS ON QUANTUM STATES PRESERVING BREGMAN AND JENSEN DIVERGENCES

DÁNIEL VIROSZTEK

Dedicated to Anna Gelniczky

ABSTRACT. We describe the structure of the bijective transformations on the set of density operators which preserve the Bregman f -divergence for an arbitrary differentiable strictly convex function f . Furthermore, we determine the preservers of the Jensen f -divergence in the case when the generating function f belongs to a recently introduced function class called Matrix Entropy Class.

1. INTRODUCTION

1.1. Motivations and overview of the literature. The investigation of certain measures of dissimilarity between two objects (numbers, vectors, matrices, functions and so on) plays an essential role in several areas of mathematics and mathematical physics. Some of the widely investigated measures are distance functions, 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* [7] has been applied to compare two probability density functions, the *Itakura-Saito divergence* [6] is used as a measure of the perceptual difference between spectra. The *Bregman divergence* was introduced by Lev Bregman [3] for convex functions $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ as the ϕ -depending nonnegative measure of discrepancy of elements of \mathbb{R}^d . Originally his motivation was the problem of convex programming, but it became widely researched both from theoretical and practical viewpoints. The remarkable fact that all the aforementioned divergences — let alone *Stein's loss* and *Umegaki's relative entropy* — are special cases of the Bregman divergence shows its importance [1].

2010 *Mathematics Subject Classification.* Primary: 47B49, 46L30.

Key words and phrases. Density operators, Bregman divergences, Jensen divergences, Preserver transformations.

The author was supported by the “Lendület” Program (LP2012-46/2012) of the Hungarian Academy of Sciences and by the Hungarian Scientific Research Fund (OTKA) Reg. No. K104206. The author was also supported by the “For the Young Talents of the Nation” scholarship program (NTP-EFÖ-P-15-0481) of the Hungarian State.

Moreover, in the recent volume [15] 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 [15]). Furthermore, the celebrated *Jensen-Shannon divergence* and its non-commutative counterpart the *Jensen-von Neumann divergence* are particular Jensen divergences.

For some recent results concerning Bregman divergences of operators we refer to [8] and [16]. In [8], M. Lewin and J. Sabin characterized a certain monotonicity property of the Bregman divergence by the operator monotonicity of the derivative of the corresponding scalar function. In [16], J. Pitrik and the author gave a characterization of the joint convexity of the Bregman divergence in terms of the generating function.

The Bregman and Jensen divergences are generalized distance measures. This latter notion stands for any function $d : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ on any set \mathcal{X} with the mere property that for $x, y \in \mathcal{X}$ we have $d(x, y) = 0$ if and only if $x = y$. Transformations which preserve generalized distance measures are called generalized isometries.

In a recent paper L. Molnár, J. Pitrik and the author determined the structure of the generalized isometries of the cone of positive definite matrices with respect to Bregman divergences and Jensen divergences [13]. Let us mention that in the papers [10, 14] L. Molnár and his coauthors considered a certain family of generalized distance measures on the cone of positive definite matrices and the structure of the isometries was obtained by Mazur-Ulam type arguments. The aforementioned family of divergences is almost disjoint from the families of Bregman divergences and Jensen divergences. The intersection contains only the Stein's loss and the Chebbi-Moakher log-determinant α -divergences, respectively.

1.2. Goals. In this paper we describe the structure of the generalized isometries of the set of density matrices — which used to represent the state space of a finite quantum system — with respect to Bregman and Jensen divergences. It turns out that every bijective transformation which leaves the Bregman or Jensen divergence invariant is implemented by a unitary or antiunitary operator on the underlying Hilbert space. Such a result may be considered as a Wigner type result. For other closely related Wigner type results we refer to [9] and [12]. In fact, at several points of our argument we use ideas and techniques of the latter two papers.

1.3. Basic notions and notations. Throughout this paper the following notations will be used. \mathcal{H} stands for a finite dimensional complex Hilbert space. \mathbb{R}^+ (\mathbb{R}^{++}) consists of all nonnegative (positive) numbers and

$\mathcal{B}(\mathcal{H})$ ($\mathcal{B}^{sa}(\mathcal{H})$, $\mathcal{B}^+(\mathcal{H})$, $\mathcal{B}^{++}(\mathcal{H})$) denotes the set of bounded (self-adjoint, positive semidefinite, positive definite) linear operators on the Hilbert space \mathcal{H} . $\mathcal{S}(\mathcal{H})$ stands for the state space of \mathcal{H} (the set of positive semidefinite operators with unit trace) and $\mathcal{P}_1(\mathcal{H})$ denotes the set of rank-one projections on \mathcal{H} .

If $f : I \rightarrow \mathbb{R}$ is a function defined on an interval $I \subset \mathbb{R}$ then the corresponding *standard operator function* is the following map:

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

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

where $\sigma(A)$ is the spectrum and P_a is the spectral projection corresponding to the eigenvalue a .

1.4. Bregman divergences on positive definite operators. Let f be a differentiable strictly convex function on $(0, \infty)$. (Note that the derivative of f is necessarily continuous.) The Bregman f -divergence of the positive definite operators $A, B \in \mathcal{B}^{++}(\mathcal{H})$ is defined by

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

see e. g. formula (5) in [16]. Easy computation shows that if the spectral decompositions are

$$A = \sum_{a \in \sigma(A)} a P_a \text{ and } B = \sum_{b \in \sigma(B)} b Q_b.$$

then we have

$$H_f(A, B) = \sum_{a \in \sigma(A), b \in \sigma(B)} (f(a) - f(b) - f'(b)(a - b)) \text{Tr} P_a Q_b.$$

1.4.1. The extension to positive semidefinite operators. If f can be extended to 0 by continuity, then the Bregman f -divergence can be extended to positive semidefinite operators by continuity the following way. Let

$$(1) \quad H_f(X, Y) := \lim_{\varepsilon \rightarrow 0} (X + \varepsilon I, Y + \varepsilon I)$$

for positive semidefinite operators X and Y . In the followings we show that the limit (1) always exists and takes values in $\mathbb{R}^+ \cup \{+\infty\}$. The argument is based on [16, Sec. 2.1].

If X and Y admit the spectral decompositions $X = \sum_{x \in \sigma(X)} x P_x$ and $Y = \sum_{y \in \sigma(Y)} y Q_y$ then

$$(2) \quad \begin{aligned} & H_f(X + \varepsilon I, Y + \varepsilon I) \\ &= \sum_{x \in \sigma(X), y \in \sigma(Y)} (f(x + \varepsilon) - f(y + \varepsilon) - f'(y + \varepsilon)(x - y)) \text{Tr} P_x Q_y. \end{aligned}$$

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

Clearly, if $\lim_{\varepsilon \rightarrow 0} f'(\varepsilon) \in \mathbb{R}$, then the limit of (2) is a real number. If $\text{supp}(X) \subseteq \text{supp}(Y)$, then for every $x \in \sigma(X)$ we have $\text{Tr } P_x Q_0 = 0$ or $x = 0$, hence the limit of (2) is finite in this case, as well. It is easy to see that if $\text{supp}(X) \not\subseteq \text{supp}(Y)$ and $\lim_{\varepsilon \rightarrow 0} f'(\varepsilon) = -\infty$, then the limit of (2) is $+\infty$.

1.4.2. *Computation rules.* By the above argument, if $\lim_{x \rightarrow 0} f'(x) = -\infty$, for positive semidefinite operators X and Y the following computation rule holds.

$$(3) \quad H_f(X, Y) = \sum_{x \in \sigma(X), y \in \sigma(Y) \setminus \{0\}} (f(x) - f(y) - f'(y)(x - y)) \text{Tr } P_x Q_y$$

$$(4) \quad = \text{Tr}_{\text{supp}(Y)} (f(X) - f(Y) - f'(Y)(X - Y))$$

if $\text{supp}(X) \subseteq \text{supp}(Y)$ and

$$H_f(X, Y) = \infty$$

if $\text{supp}(X) \not\subseteq \text{supp}(Y)$. (For any $\mathcal{K} \subset \mathcal{H}$, $\text{Tr}_{\mathcal{K}}$ means that we take the trace only on the subspace \mathcal{K} .)

If $\lim_{x \rightarrow 0} f'(x) \in \mathbb{R}$ then the computation rule is simply the following.

$$(5) \quad H_f(X, Y) = \sum_{x \in \sigma(X), y \in \sigma(Y)} (f(x) - f(y) - f'(y)(x - y)) \text{Tr } P_x Q_y$$

$$(6) \quad = \text{Tr} (f(X) - f(Y) - f'(Y)(X - Y)).$$

Example. For the standard entropy function $f(x) = x \log x$ the induced Bregman f -divergence on density matrices is the *Umegaki relative entropy*

$$H_f(A, B) = \text{Tr } A (\log A - \log B)$$

which is one of the most important numerical quantities in quantum information theory. Therefore, Bregman f -divergences may be considered as generalized relative entropies [8].

For any $q > 1$, the function $f_q : x \mapsto f_q(x) := \frac{x^q - x}{q-1}$ is convex, and the induced Bregman divergence is

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

see [16]. In the particular case $q = 2$ the latter quantity is just the square of the Hilbert-Schmidt norm,

$$H_{f_2}(A, B) = \text{Tr} (A - B)^2.$$

1.5. Jensen divergences on positive semidefinite operators. Let f be a strictly convex function on $(0, \infty)$ such that the limit $\lim_{x \rightarrow 0^+} f(x) =: f(0)$ exists. The Jensen f -divergence of the positive semidefinite operators $A \in \mathcal{B}^+(\mathcal{H})$ and $B \in \mathcal{B}^+(\mathcal{H})$ is defined by

$$J_f(A, B) = \text{Tr} \left(\frac{1}{2} (f(A) + f(B)) - f \left(\frac{1}{2} (A + B) \right) \right),$$

see e. g. [13]. We investigate the preservers of the Jensen f -divergence in the case when the generating function f belongs to the *Matrix Entropy Class*.

In the recent paper [5] *Tropp and Chen* defined the Matrix Entropy Class the following way.

Definition. *The Matrix Entropy Class consists of the real valued 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 \mathcal{H} the map

$$\mathcal{B}(\mathcal{H})^{++} \rightarrow \mathcal{B}(\mathcal{B}(\mathcal{H})^{sa}); X \mapsto (\mathbf{D}f'[X])^{-1}$$

is concave with respect to the semidefinite order, where $\mathbf{D}f[Y]$ denotes the Fréchet derivative of the standard operator function $f : \mathcal{B}(\mathcal{H})^{++} \rightarrow \mathcal{B}(\mathcal{H})^{sa}$ at the point Y .

Example. The standard entropy function $f(x) = x \log x$ is an important element of the Matrix Entropy Class [5]. The induced Jensen f -divergence is the well-known *Jensen-von Neumann divergence*

$$J_f(A, B) = \frac{1}{2} (\text{Tr } A \log A + \text{Tr } B \log B) - \text{Tr} \left(\frac{A+B}{2} \log \left(\frac{A+B}{2} \right) \right).$$

If $1 < q \leq 2$, then the function $f_q : x \mapsto f_q(x) := \frac{x^q - x}{q-1}$ belongs to the Matrix Entropy Class [5], and the induced Jensen divergence is

$$J_{f_q}(A, B) = \frac{1}{q-1} \left(\frac{\text{Tr } A^q + \text{Tr } B^q}{2} - \text{Tr} \left(\frac{A+B}{2} \right)^q \right).$$

In particular, if $q = 2$, then we have

$$J_{f_2}(A, B) = \text{Tr} \left(\frac{A-B}{2} \right)^2.$$

2. THE MAIN RESULTS

It is clear that any unitary or antiunitary conjugation leaves the Bregman divergences and Jensen divergences invariant. The main result of this paper is that the converse statement is also true, i. e., the preservers

of Bregman and Jensen divergences are necessarily unitary or antiunitary conjugations.

Theorem 1. *Let $f \in C^1((0, \infty)) \cap C^0([0, \infty))$ be a strictly convex function. Let $\phi : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$ be a bijection which preserves the Bregman f -divergence, that is,*

$$H_f(\phi(A), \phi(B)) = H_f(A, B) \quad (A, B \in \mathcal{S}(\mathcal{H})).$$

Then there exists a unitary or antiunitary transformation $U : \mathcal{H} \rightarrow \mathcal{H}$ such that

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

Theorem 2. *Let f be a strictly convex element of the Matrix Entropy Class. Let $\phi : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$ be a bijection which preserves the Jensen f -divergence, that is,*

$$J_f(\phi(A), \phi(B)) = J_f(A, B) \quad (A, B \in \mathcal{S}(\mathcal{H})).$$

Then there exists a unitary or antiunitary transformation $U : \mathcal{H} \rightarrow \mathcal{H}$ such that

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

3. PROOFS

Remark. Affine perturbation of the generating function does not change the Bregman or the Jensen divergence, that is

$$H_{f+a}(\cdot, \cdot) = H_f(\cdot, \cdot)$$

and

$$J_{f+a}(\cdot, \cdot) = J_f(\cdot, \cdot)$$

for any convex function f and affine function $a(x) = \alpha x + \beta$. Therefore in the followings we may and do assume that $f(0) = f(1) = 0$.

3.1. The proof of Theorem 1.

Case I. First we investigate the case when $\lim_{x \rightarrow 0} f'(x) = -\infty$.

In this first part of the proof we basically follow the argument of [9], but the more general statement requires new techniques at some crucial points of the proof.

As $f'(x)$ is unbounded from below, the divergence $H_f(A, B)$ is finite if and only if $\text{supp}(A) \subseteq \text{supp}(B)$. Therefore, any divergence-preserving transformation ϕ has the following properties.

$$(7) \quad \text{supp}(A) \subseteq \text{supp}(B) \Leftrightarrow \text{supp}(\phi(A)) \subseteq \text{supp}(\phi(B)),$$

$$(8) \quad \text{supp}(A) = \text{supp}(B) \Leftrightarrow \text{supp}(\phi(A)) = \text{supp}(\phi(B))$$

and

$$(9) \quad \text{supp}(A) \subsetneq \text{supp}(B) \Leftrightarrow \text{supp}(\phi(A)) \subsetneq \text{supp}(\phi(B)).$$

As a consequence, ϕ preserves the rank — the reader should consult [9] for a more detailed argument. In particular, the image of a rank-one projection is a rank-one projection, as well. So ϕ restricted to $\mathcal{P}_1(\mathcal{H})$ is a bijection from $\mathcal{P}_1(\mathcal{H})$ to $\mathcal{P}_1(\mathcal{H})$.

Let P and Q be orthogonal elements of $\mathcal{P}_1(\mathcal{H})$, set $0 < \lambda < \mu < 1$ such that $\lambda + \mu = 1$, $S := \lambda P + \mu Q$ and let $R \in \mathcal{P}_1(\mathcal{H})$.

If $\text{supp}(R) \subseteq \text{supp}(S)$ then by the computation rule (3) we have

$$\begin{aligned} & H_f(R, S) \\ &= (f(1) - f(\lambda) - f'(\lambda)(1 - \lambda)) \text{Tr} RP + (f(0) - f(\lambda) - f'(\lambda)(0 - \lambda)) \text{Tr}(I - R)P \\ &+ (f(1) - f(\mu) - f'(\mu)(1 - \mu)) \text{Tr} RQ + (f(0) - f(\mu) - f'(\mu)(0 - \mu)) \text{Tr}(I - R)Q \\ &= -f'(\lambda) \text{Tr} RP - f'(\mu) \text{Tr} RQ + \lambda f'(\lambda) - f(\lambda) + \mu f'(\mu) - f(\mu) \\ (10) \quad &= -f'(\lambda) \text{Tr} RP - f'(\mu) \text{Tr} RQ + C_{f,\lambda,\mu} \end{aligned}$$

if we introduce the notation $C_{f,\lambda,\mu} = \lambda f'(\lambda) - f(\lambda) + \mu f'(\mu) - f(\mu)$.

As R runs through the set of rank-one projections which have their support contained in $\text{supp}(S)$, $\text{Tr} RP$ and $\text{Tr} RQ$ take all values such that $0 \leq \text{Tr} RP, \text{Tr} RQ \leq 1$ and $\text{Tr} RP + \text{Tr} RQ = 1$. f is strictly convex, hence f' is strictly monotone increasing. By the strict monotonicity of f'

$$\max_{\{R \in \mathcal{P}_1(\mathcal{H}) : \text{supp}(R) \subseteq \text{supp}(S)\}} H_f(R, S) = -f'(\lambda) + C_{f,\lambda,\mu}$$

which maximum is taken only at $R = P$ and

$$\min_{\{R \in \mathcal{P}_1(\mathcal{H}) : \text{supp}(R) \subseteq \text{supp}(S)\}} H_f(R, S) = -f'(\mu) + C_{f,\lambda,\mu}$$

which minimum is taken only at $R = Q$. Clearly,

$$\begin{aligned} & \max_{\{R \in \mathcal{P}_1(\mathcal{H}) : \text{supp}(R) \subseteq \text{supp}(S)\}} H_f(R, S) \\ & - \min_{\{R \in \mathcal{P}_1(\mathcal{H}) : \text{supp}(R) \subseteq \text{supp}(S)\}} H_f(R, S) \\ &= f'(\mu) - f'(\lambda). \end{aligned}$$

f' is strictly monotone increasing, so $f'(\mu) - f'(\lambda) = f'(1 - \lambda) - f'(\lambda)$ is a strictly monotone decreasing function of λ . Therefore, $f'(\mu) - f'(\lambda)$ uniquely determines λ and hence the spectrum of the rank-two density S .

This means that the Bregman f -divergence preserving property of ϕ implies that $\phi(S)$ is a rank-two density with eigenvalues λ and μ (and possibly zero). Hence $\phi(S) = \lambda P' + \mu Q'$ with some rank-one projections P' and Q' which are orthogonal to each other.

Therefore, we can conclude that

$$\begin{aligned} R = P &\Leftrightarrow H_f(R, S) = -f'(\lambda) + C_{f,\lambda,\mu} \Leftrightarrow H_f(\phi(R), \phi(S)) = -f'(\lambda) + C_{f,\lambda,\mu} \Leftrightarrow \\ &\Leftrightarrow H_f(\phi(R), \lambda P' + \mu Q') = -f'(\lambda) + C_{f,\lambda,\mu} \Leftrightarrow \phi(R) = P'. \end{aligned}$$

We deduced that $\phi(P) = P'$. Similarly, $\phi(Q) = Q'$. So

$$(11) \quad \phi(S) = \phi(\lambda P + \mu Q) = \lambda\phi(P) + \mu\phi(Q),$$

and the mutual orthogonality of rank-one projections is preserved.

Let P and Q be arbitrary mutually orthogonal elements of $\mathcal{P}_1(\mathcal{H})$ and set $R \in \mathcal{P}_1(\mathcal{H})$ such that $\text{supp}(R) \subseteq \text{supp}(P) + \text{supp}(Q)$. Let $0 < \lambda < \mu < 1$ with $\lambda + \mu = 1$. Then by (10), (11) and by the preserver property of ϕ

$$\begin{aligned} &H_f(R, \lambda P + \mu Q) \\ &= (-f'(\lambda) + C_{f,\lambda,\mu}) \text{Tr} RP + (-f'(\mu) + C_{f,\lambda,\mu}) \text{Tr} RQ \\ &= H_f(\phi(R), \lambda\phi(P) + \mu\phi(Q)) \\ &= (-f'(\lambda) + C_{f,\lambda,\mu}) \text{Tr} \phi(R)\phi(P) + (-f'(\mu) + C_{f,\lambda,\mu}) \text{Tr} \phi(R)\phi(Q) \end{aligned}$$

We used that $\phi(R) \in \mathcal{P}_1(\mathcal{H})$ such that

$$\text{supp}(\phi(R)) \subseteq \text{supp}(\lambda\phi(P) + \mu\phi(Q)) = \text{supp}(\phi(P)) + \text{supp}(\phi(Q)).$$

Any element of a nontrivial compact real interval is a unique convex combination of the endpoints, hence we get that

$$(12) \quad \text{Tr} RP = \text{Tr} \phi(R)\phi(P) \text{ and } \text{Tr} RQ = \text{Tr} \phi(R)\phi(Q).$$

Wigner's theorem states that any bijection $\xi : \mathcal{P}_1(\mathcal{H}) \rightarrow \mathcal{P}_1(\mathcal{H})$ which preserves the *transition probability* — i. e., for which $\text{Tr} \xi(P)\xi(Q) = \text{Tr} PQ$ holds for any $P, Q \in \mathcal{P}_1(\mathcal{H})$ — is implemented by a unitary or antiunitary operator — see e. g. [11]. So, by Wigner's theorem, we get that

$$(13) \quad \phi(R) = URU^* \quad (R \in \mathcal{P}_1(\mathcal{H}))$$

for some unitary or antiunitary operator U acting on \mathcal{H} .

Now let

$$\psi(D) = U^* \phi(D)U \quad (D \in \mathcal{S}(\mathcal{H})).$$

Then ψ is the identity on $\mathcal{P}_1(\mathcal{H})$ and it preserves the Bregman f -divergence. Note that by (7) for any $D \in \mathcal{S}(\mathcal{H})$ and $P \in \mathcal{P}_1(\mathcal{H})$

$$\text{supp}(P) \subseteq \text{supp}(D) \Leftrightarrow \text{supp}(\psi(P)) \subseteq \text{supp}(\psi(D))$$

and

$$P = \psi(P) \Rightarrow \text{supp}(P) = \text{supp}(\psi(P)),$$

hence

$$\text{supp}(D) = \text{supp}(\psi(D)).$$

By (4), for any $D \in \mathcal{S}(\mathcal{H})$ and $P, Q \in \mathcal{P}_1(\mathcal{H})$ with $\text{supp}(P) \subseteq \text{supp}(D)$, $\text{supp}(Q) \subseteq \text{supp}(D)$ we have

$$H_f(P, D) = \text{Tr}_{\text{supp}(D)} (f(P) - f(D) - f'(D)(P - D))$$

$$H_f(Q, D) = \text{Tr}_{\text{supp}(D)} (f(Q) - f(D) - f'(D)(Q - D))$$

Using that $f(0) = f(1) = 0$ and hence $f(P) = f(Q) = 0$, we get

$$(14) \quad H_f(P, D) - H_f(Q, D) = \text{Tr}_{\text{supp}(D)} f'(D)(Q - P).$$

Similarly,

$$(15) \quad H_f(P, \psi(D)) - H_f(Q, \psi(D)) = \text{Tr}_{\text{supp}(D)} f'(\psi(D))(Q - P).$$

ψ preserves the Bregman f -divergence, hence subtracting (15) from (14) one gets that that

$$(16) \quad \text{Tr}_{\text{supp}(D)} (f'(D) - f'(\psi(D)))(Q - P) = 0$$

for any $P, Q \in \mathcal{P}_1(\mathcal{H})$ with $\text{supp}(P) \subseteq \text{supp}(D)$, $\text{supp}(Q) \subseteq \text{supp}(D)$.

From now, unless stated otherwise, we restrict ourselves to $\text{supp}(D)$. It follows from (16) that

$$(17) \quad f'(D) - f'(\psi(D)) = cI$$

for some $c \in \mathbb{R}$ (in particular, $f'(D)$ and $f'(\psi(D))$ commute). Indeed, if P and Q are projections corresponding to different eigenvectors of $f'(D) - f'(\psi(D))$ (which is self-adjoint), then (16) shows that the eigenvalues are the same. So all the eigenvalues of $f'(D) - f'(\psi(D))$ are equal.

Suppose that $c \neq 0$, for example, $c > 0$. Then by (17),

$$f'(D) > f'(\psi(D)).$$

$f'(D)$ and $f'(\psi(D))$ commute and f' is monotone, hence we get that

$$D > \psi(D).$$

This is a contradiction, so $c = 0$, that is, $f'(D) = f'(\psi(D))$. By the strict monotonicity of f' this implies $D = \psi(D)$.

So we deduced that $D = \psi(D)$ on $\text{supp}(D) = \text{supp}(\psi(D))$ which means that without any restriction, we have

$$D = \psi(D).$$

We deduced that ψ is the identity of $\mathcal{S}(\mathcal{H})$, the proof is done. \square

Case II. Now we investigate the case when $\lim_{x \rightarrow 0} f'(x) \in \mathbb{R}$. In this case the Bregman divergence of any two states is finite, hence it is reasonable to define

$$(18) \quad M(X) := \max \{H_f(X, D) \mid D \in \mathcal{S}(\mathcal{H})\}.$$

Let us note that f is continuously differentiable, hence the map $D \mapsto H_f(X, D)$ is continuous on the compact set $\mathcal{S}(\mathcal{H})$. Therefore, the above definition is correct as $\max\{H_f(X, D) \mid D \in \mathcal{S}(\mathcal{H})\}$ exists. In the followings we show that for $P \in \mathcal{S}(\mathcal{H})$

$$M(P) = \max\{M(X) \mid X \in \mathcal{S}(\mathcal{H})\}$$

if and only if P is a pure state, i. e., a rank-one projection.

Indeed, assume that ρ is not a pure state, that is, $\rho = \sum_{j=1}^k \lambda_j P_j$ for some $k \geq 2$, for some real numbers $0 < \lambda_1, \dots, \lambda_k < 1$ with $\sum_{j=1}^k \lambda_j = 1$ and for some rank-one projections P_1, \dots, P_k . Assume that $D^* \in \mathcal{S}(\mathcal{H})$ has the property that

$$M(\rho) = \max\{H_f(\rho, D) \mid D \in \mathcal{S}(\mathcal{H})\} = H_f(\rho, D^*).$$

By the strict convexity of f , the map

$$X \mapsto H_f(X, Y) = \text{Tr}(f(X) - f(Y) - f'(Y)(X - Y))$$

is strictly convex on $\mathcal{S}(\mathcal{H})$, see e. g. [4, 2.10. Theorem]. So

$$H_f(\rho, D^*) = H_f\left(\sum_{j=1}^k \lambda_j P_j, D^*\right) < \sum_{j=1}^k \lambda_j H_f(P_j, D^*).$$

Therefore, $H_f(\rho, D^*) < H_f(P_{j^*}, D^*)$ for some $j^* \in \{1, \dots, k\}$. This means that

$$M(\rho) = H_f(\rho, D^*) < H_f(P_{j^*}, D^*) \leq M(P_{j^*}).$$

On the other hand, it can be easily seen — for example, by the unitary invariance of the Bregman divergences — that $M(P) = M(Q)$ for all $P, Q \in \mathcal{P}_1(\mathcal{H})$. Therefore, $M(P)$ is maximal, if $P \in \mathcal{P}_1(\mathcal{H})$.

So we have the following characterization of the pure states.

$$(19) \quad P \in \mathcal{P}_1(\mathcal{H}) \Leftrightarrow M(P) = \max\{M(X) \mid X \in \mathcal{S}(\mathcal{H})\}.$$

If ϕ is a bijection that preserves the Bregman f -divergence, then for any $X \in \mathcal{S}(\mathcal{H})$

$$\begin{aligned} M(\phi(X)) &= \max\{H_f(\phi(X), \phi(D)) \mid D \in \mathcal{S}(\mathcal{H})\} \\ &= \max\{H_f(X, D) \mid D \in \mathcal{S}(\mathcal{H})\} = M(X). \end{aligned}$$

So ϕ restricted to $\mathcal{P}_1(\mathcal{H})$ is a $\mathcal{P}_1(\mathcal{H}) \rightarrow \mathcal{P}_1(\mathcal{H})$ bijection.

Now let $P, Q \in \mathcal{P}_1(\mathcal{H})$. By the computation rule (5)

$$\begin{aligned} H_f(P, Q) &= (f(0) - f(0) - f'(0)(0 - 0)) \text{Tr}(I - P)(I - Q) \\ &\quad + (f(1) - f(0) - f'(0)(1 - 0)) \text{Tr}(P)(I - Q) \\ &\quad + (f(0) - f(1) - f'(1)(0 - 1)) \text{Tr}(I - P)(Q) + (f(1) - f(1) - f'(1)(1 - 1)) \text{Tr}PQ \\ (20) \quad &= (1 - \text{Tr}PQ) (f'(1) - f'(0)). \end{aligned}$$

Similarly,

$$(21) \quad H_f(\phi(P), \phi(Q)) = (1 - \text{Tr} \phi(P)\phi(Q)) (f'(1) - f'(0)).$$

f' is strictly monotone, hence $(f'(1) - f'(0)) \neq 0$. So by (20) and (21), $H_f(\phi(P), \phi(Q)) = H_f(P, Q)$ gives us

$$(22) \quad \text{Tr} PQ = \text{Tr} \phi(P)\phi(Q) \quad (P, Q \in \mathcal{P}_1(\mathcal{H})).$$

From now on, our argument is very similar to the ending part of the discussion of *Case I*. By Wigner's theorem,

$$(23) \quad \phi(R) = URU^* \quad (R \in \mathcal{P}_1(\mathcal{H}))$$

for some unitary or antiunitary operator U .

Now let

$$\psi(D) = U^* \phi(D)U \quad (D \in \mathcal{S}(\mathcal{H})).$$

Then ψ is the identity on $\mathcal{P}_1(\mathcal{H})$ and it preserves the Bregman f -divergence. By (6), for any $D \in \mathcal{S}(\mathcal{H})$ and $P, Q \in \mathcal{P}_1(\mathcal{H})$

$$H_f(P, D) = \text{Tr}(f(P) - f(D) - f'(D)(P - D))$$

$$H_f(Q, D) = \text{Tr}(f(Q) - f(D) - f'(D)(Q - D))$$

It follows that

$$(24) \quad H_f(P, D) - H_f(Q, D) = \text{Tr} f'(D)(Q - P).$$

Similarly,

$$(25) \quad H_f(P, \psi(D)) - H_f(Q, \psi(D)) = \text{Tr} f'(\psi(D))(Q - P).$$

ψ preserves the Bregman f -divergence, hence (24) and (25) imply that

$$(26) \quad \text{Tr}(f'(D) - f'(\psi(D)))(Q - P) = 0$$

for any $P, Q \in \mathcal{P}_1(\mathcal{H})$.

It follows from (26) that

$$(27) \quad f'(D) - f'(\psi(D)) = cI$$

for some $c \in \mathbb{R}$, and it is easy to show that $c = 0$ by necessity.

That is, $f'(D) = f'(\psi(D))$. By the strict monotonicity of f' this implies

$$D = \psi(D).$$

We deduced that ψ is the identity of $\mathcal{S}(\mathcal{H})$, the proof is done. \square

3.2. The proof of Theorem 2. Our aim is to prove that any bijective transformation of $\mathcal{S}(\mathcal{H})$ which preserves the Jensen f -divergence is implemented by a unitary or an antiunitary operator. Recall that $J_f(.,.)$ denotes the Jensen f -divergence, which quantity was defined in Subsection 1.5.

Lemma 3. *For any rank-one projections $P, Q \in \mathcal{P}_1(\mathcal{H})$ we have*

$$(28) \quad J_f(P, Q) = (-1) \left(f \left(\frac{1}{2} \left(1 + \sqrt{\text{Tr} PQ} \right) \right) + f \left(\frac{1}{2} \left(1 - \sqrt{\text{Tr} PQ} \right) \right) \right).$$

Proof. Let $\{e_1, \dots, e_n\} \subseteq \mathcal{H}$ be an orthonormal basis such that $\text{supp}(P) = \text{span}(\{e_1\})$ and $\text{supp}(Q) \subseteq \text{span}(\{e_1, e_2\})$. In this basis

$$P = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} p & \sqrt{p(1-p)}\bar{\alpha} & 0 & \dots & 0 \\ \sqrt{p(1-p)}\alpha & 1-p & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

where we introduced the notation $p := \text{Tr} PQ$ and $\alpha \in \mathbb{C}, |\alpha| = 1$. Let us denote by μ and $1 - \mu$ the eigenvalues of $\frac{1}{2}(P + Q)$ restricted to $\text{span}(\{e_1, e_2\})$. Easy computations show that

$$\det \left(\frac{1}{2}(P + Q)_{|\text{span}(\{e_1, e_2\})} \right) = \frac{1}{4}(1 - p).$$

Therefore, $\mu(1 - \mu) = \frac{1}{4}(1 - p)$. The solution of this quadratic equation is

$$\mu_{1,2} = \frac{1}{2} \left(1 \pm \sqrt{p} \right),$$

which gives the result of the lemma. (We used that by $f(0) = f(1) = 0$ we have $\text{Tr} f(P) = \text{Tr} f(Q) = 0$.) \square

Corollary 4. *By the convexity of f ,*

$$\max_{P, Q \in \mathcal{P}_1(\mathcal{H})} J_f(P, Q) = -2f \left(\frac{1}{2} \right) =: M_f$$

and by the strict convexity of f , for $P, Q \in \mathcal{P}_1(\mathcal{H})$ we have $J_f(P, Q) = M_f$ if and only if $\text{Tr} PQ = 0$, that is, $PQ = 0$.

Claim 5.

$$\max_{A, B \in \mathcal{S}(\mathcal{H})} J_f(A, B) = M_f$$

and for any $A, B \in \mathcal{S}(\mathcal{H})$, if $J_f(A, B) = M_f$, then $AB = 0$.

Proof. The function f is an element of the Matrix Entropy Class, hence by [16, Thm. 2] the induced Bregman f -divergence (denoted by H_f) is jointly convex. Observe that

$$J_f(A, B) = \frac{1}{2} \left(H_f \left(A, \frac{A+B}{2} \right) + H_f \left(B, \frac{A+B}{2} \right) \right).$$

Suppose that

$$A = \sum_i \lambda_i P_i, \text{ and } B = \sum_j \mu_j Q_j$$

where the P_i 's and the Q_j 's are rank-one projections and the λ_i 's and the μ_j 's are positive numbers such that $\sum_i \lambda_i = \sum_j \mu_j = 1$. The joint convexity of the Bregman divergence implies that

$$\begin{aligned} J_f(A, B) &= \frac{1}{2} \left(H_f \left(A, \frac{A+B}{2} \right) + H_f \left(B, \frac{A+B}{2} \right) \right) \\ &= \frac{1}{2} \left(H_f \left(\sum_i \lambda_i P_i, \sum_i \lambda_i \frac{P_i+B}{2} \right) + H_f \left(\sum_i \lambda_i B, \sum_i \lambda_i \frac{P_i+B}{2} \right) \right) \\ &\leq \sum_i \lambda_i \left(\frac{1}{2} H_f \left(P_i, \frac{P_i+B}{2} \right) + \frac{1}{2} H_f \left(B, \frac{P_i+B}{2} \right) \right) = \sum_i \lambda_i J_f(P_i, B). \end{aligned}$$

Similarly, $J_f(P_i, B) \leq \sum_j \mu_j J_f(P_i, Q_j)$ for any i . Therefore,

$$J_f(A, B) \leq \sum_i \sum_j \lambda_i \mu_j J_f(P_i, Q_j) \leq \max_{P, Q \in \mathcal{P}_1(\mathcal{H})} J_f(P, Q) = M_f.$$

We have $\lambda_i \mu_j > 0$ for any i and j , and $\sum_i \sum_j \lambda_i \mu_j = 1$, hence if $J_f(A, B) = M_f$, then $J_f(P_i, Q_j) = M_f$ for any i and j . By Corollary 4 this means that we always have $P_i Q_j = 0$, and therefore $AB = 0$. \square

Claim 6. If ϕ is a bijection on $\mathcal{S}(\mathcal{H})$ that preserves the Jensen f -divergence, then ϕ restricted to $\mathcal{P}_1(\mathcal{H})$ is a bijection from $\mathcal{P}_1(\mathcal{H})$ to $\mathcal{P}_1(\mathcal{H})$.

Proof. If $P \in \mathcal{P}_1(\mathcal{H})$, then there exists a set $H \subseteq \mathcal{S}(\mathcal{H})$ such that $P \in H$, $|H| = \dim(\mathcal{H})$ and $J_f(A, B) = M_f$ for any $A, B \in H$. (H contains orthogonal rank-one projections.) ϕ preserves the Jensen f -divergence, hence $J_f(X, Y) = M_f$ for any $X, Y \in \phi(H)$, that is, by Lemma 5, $\phi(H) \subseteq \mathcal{S}(\mathcal{H})$ has $\dim(\mathcal{H})$ pairwise orthogonal elements. This implies that all the elements of $\phi(H)$ are rank-one projections, in particular, $\phi(P) \in \mathcal{P}_1(\mathcal{H})$. \square

Claim 7. *There exists a unitary or antiunitary operator $U : \mathcal{H} \rightarrow \mathcal{H}$ such that*

$$\phi(R) = URU^* \quad (R \in \mathcal{P}_1(\mathcal{H})).$$

Proof. By the strict convexity of f , (28) shows that for $P, Q \in \mathcal{P}_1(\mathcal{H})$, $J_f(P, Q)$ is a strictly monotone decreasing function of $\text{Tr} PQ$. Therefore, if ϕ preserves the Jensen f -divergence, then it preserves the transition probability $\text{Tr} PQ$, as well. So ϕ restricted to $\mathcal{P}_1(\mathcal{H})$ is a bijection from $\mathcal{P}_1(\mathcal{H})$ to $\mathcal{P}_1(\mathcal{H})$ which preserves the transition probability ($\text{Tr} \phi(P)\phi(Q) = \text{Tr} PQ$ for any $P, Q \in \mathcal{P}_1(\mathcal{H})$). Therefore, by Wigner's theorem we obtain the statement of this Claim. \square

Claim 8. *Let us define the map $\psi : \mathcal{S}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$ by*

$$\psi(D) := U^* \phi(D) U.$$

Then $\psi(D) = D$ for any $D \in \mathcal{S}(\mathcal{H})$.

Proof. Clearly, ψ preserves the Jensen f -divergence and $\psi(R) = R$ for any $R \in \mathcal{P}_1(\mathcal{H})$. Therefore, we have

$$J_f(R, D) = J_f(R, \psi(D)) \quad R \in \mathcal{P}_1(\mathcal{H}), D \in \mathcal{S}(\mathcal{H}),$$

which equation can be written as

$$(29) \quad \text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) - \text{Tr} f\left(\frac{1}{2}(R + D)\right) = \frac{1}{2} \text{Tr} f(\psi(D)) - \frac{1}{2} \text{Tr} f(D).$$

Let $D \in \mathcal{S}(\mathcal{H})$ be arbitrary but fixed. Suppose that the spectral decompositions of D and $\psi(D)$ are

$$(30) \quad D = \sum_{i=1}^m \lambda_i P_i \text{ and } \psi(D) = \sum_{j=1}^n \mu_j Q_j,$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_m$ and $\mu_1 > \mu_2 > \dots > \mu_n$.

f is strictly convex, hence the difference quotient function $h(a, b) := \frac{f(a) - f(b)}{a - b}$ is strictly monotone increasing in both a and b .

Now, we show that for any $R \in \mathcal{P}_1(\mathcal{H})$ the quantity $\text{Tr} f\left(\frac{1}{2}(R + D)\right)$ is maximal, that is,

$$\text{Tr} f\left(\frac{1}{2}(R + D)\right) = \max_{X \in \mathcal{P}_1(\mathcal{H})} \text{Tr} f\left(\frac{1}{2}(X + D)\right)$$

if and only if $R \leq P_1$, and in this case

$$\text{Tr} f\left(\frac{1}{2}(R + D)\right) = f\left(\frac{\lambda_1}{2} + \frac{1}{2}\right) - f\left(\frac{\lambda_1}{2}\right) + \sum_{i=1}^m \text{rank}(P_i) f\left(\frac{\lambda_i}{2}\right).$$

Indeed, let us denote the dimension of the Hilbert space \mathcal{H} by N and let $v_1 \geq v_2 \geq \dots \geq v_N$ denote the (not necessarily different) eigenvalues of D . As $R \geq 0$, by Weyl's inequality (see e. g. [2, Thm. III.2.1]) we get that the

eigenvalues of $R+D$ can be written in the form $\nu_1 + \varepsilon_1, \nu_2 + \varepsilon_2, \dots, \nu_N + \varepsilon_N$, where $\varepsilon_k \geq 0$ for any $k \in \{1, \dots, N\}$. The condition $\text{Tr} R = 1$ ensures that $\sum_{k=1}^N \varepsilon_k = 1$. Obviously,

$$\begin{aligned} \text{Tr} f\left(\frac{1}{2}(R+D)\right) - \text{Tr} f\left(\frac{1}{2}D\right) &= \sum_{k=1}^N f\left(\frac{\nu_k + \varepsilon_k}{2}\right) - f\left(\frac{\nu_k}{2}\right) \\ &= \sum_{k:\varepsilon_k > 0} \frac{\varepsilon_k}{2} \frac{f\left(\frac{\nu_k + \varepsilon_k}{2}\right) - f\left(\frac{\nu_k}{2}\right)}{\frac{\varepsilon_k}{2}}. \end{aligned}$$

If $R \not\leq P_1$, then $\varepsilon_l, \varepsilon_k > 0$ for some $1 \leq l \neq k \leq N$. In this case, by the strict monotonicity of the difference quotient function h we have

$$\begin{aligned} \sum_{k:\varepsilon_k > 0} \frac{\varepsilon_k}{2} \frac{f\left(\frac{\nu_k + \varepsilon_k}{2}\right) - f\left(\frac{\nu_k}{2}\right)}{\frac{\varepsilon_k}{2}} &< \sum_{k:\varepsilon_k > 0} \frac{\varepsilon_k}{2} \frac{f\left(\frac{(\nu_1 + \sum_{l=1}^{k-1} \varepsilon_l) + \varepsilon_k}{2}\right) - f\left(\frac{\nu_1 + \sum_{l=1}^{k-1} \varepsilon_l}{2}\right)}{\frac{\varepsilon_k}{2}} \\ &= \sum_{k=1}^N f\left(\frac{(\nu_1 + \sum_{l=1}^{k-1} \varepsilon_l) + \varepsilon_k}{2}\right) - f\left(\frac{\nu_1 + \sum_{l=1}^{k-1} \varepsilon_l}{2}\right) = f\left(\frac{\nu_1 + 1}{2}\right) - f\left(\frac{\nu_1}{2}\right). \end{aligned}$$

If $R \leq P_1$, then the eigenvalues of $R+D$ are $\nu_1 + 1, \nu_2, \dots, \nu_N$. Therefore, in this case we have

$$\text{Tr} f\left(\frac{1}{2}(R+D)\right) - \text{Tr} f\left(\frac{1}{2}D\right) = f\left(\frac{\nu_1 + 1}{2}\right) - f\left(\frac{\nu_1}{2}\right).$$

So we get that $\text{Tr} f\left(\frac{1}{2}(R+D)\right)$ is maximal if and only if $R \leq P_1$.

Similarly, $\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right)$ is maximal, that is,

$$\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) = \max_{X \in \mathcal{P}_1(\mathcal{H})} \text{Tr} f\left(\frac{1}{2}(X + \psi(D))\right)$$

if and only if $R \leq Q_1$, and in this case

$$\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) = f\left(\frac{\mu_1}{2} + \frac{1}{2}\right) - f\left(\frac{\mu_1}{2}\right) + \sum_{j=1}^n \text{rank}(Q_j) f\left(\frac{\mu_j}{2}\right).$$

Observe that the right hand side of (29) is independent of R , hence $\text{Tr} f\left(\frac{1}{2}(R+D)\right)$ is maximal if and only if $\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right)$ is maximal. That is, $R \leq P_1 \Leftrightarrow R \leq Q_1$, so $P_1 = Q_1$.

Let us introduce the notation $S_1 := P_1 = Q_1$. For any $R \in \mathcal{P}_1(\mathcal{H})$ with $RS_1 = 0$ $\text{Tr} f\left(\frac{1}{2}(R+D)\right)$ is maximal, that is,

$$\text{Tr} f\left(\frac{1}{2}(R+D)\right) = \max_{X \in \mathcal{P}_1(\mathcal{H}): XS_1=0} \text{Tr} f\left(\frac{1}{2}(X+D)\right)$$

if and only if $R \leq P_2$, and in this case

$$\text{Tr} f\left(\frac{1}{2}(R+D)\right) = f\left(\frac{\lambda_2}{2} + \frac{1}{2}\right) - f\left(\frac{\lambda_2}{2}\right) + \sum_{i=1}^m \text{rank}(P_i) f\left(\frac{\lambda_i}{2}\right).$$

Similarly, $\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right)$ is maximal, that is,

$$\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) = \max_{X \in \mathcal{P}_1(\mathcal{H}): XS_1=0} \text{Tr} f\left(\frac{1}{2}(X + \psi(D))\right)$$

if and only if $R \leq Q_2$, and in this case

$$\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) = f\left(\frac{\mu_2}{2} + \frac{1}{2}\right) - f\left(\frac{\mu_2}{2}\right) + \sum_{j=1}^n \text{rank}(Q_j) f\left(\frac{\mu_j}{2}\right).$$

The right hand side of (29) is independent of R , hence $\text{Tr} f\left(\frac{1}{2}(R + D)\right)$ is maximal if and only if $\text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right)$ is maximal. That is, $R \leq P_2 \Leftrightarrow R \leq Q_2$, so $P_2 = Q_2$.

And so on, we can deduce that all the eigenprojections coincide, that is, $m = n$ and $P_i = Q_i$ for all $1 \leq i \leq m$.

Let us define the following function on the positive half line.

$$g(a) := f\left(\frac{a}{2} + \frac{1}{2}\right) - f\left(\frac{a}{2}\right)$$

By the strict convexity of f , g is strictly monotone increasing. It follows easily from the above variational formulas that

$$\begin{aligned} \max_{R \in \mathcal{P}_1(\mathcal{H}): R \leq P_l + P_{l+1}} \text{Tr} f\left(\frac{1}{2}(R + D)\right) - \min_{R \in \mathcal{P}_1(\mathcal{H}): R \leq P_l + P_{l+1}} \text{Tr} f\left(\frac{1}{2}(R + D)\right) \\ = g(\lambda_l) - g(\lambda_{l+1}) \end{aligned}$$

and

$$\begin{aligned} \max_{R \in \mathcal{P}_1(\mathcal{H}): R \leq P_l + P_{l+1}} \text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) - \\ - \min_{R \in \mathcal{P}_1(\mathcal{H}): R \leq P_l + P_{l+1}} \text{Tr} f\left(\frac{1}{2}(R + \psi(D))\right) \\ = g(\mu_l) - g(\mu_{l+1}) \end{aligned}$$

for all $1 \leq l \leq m-1$. The right hand side of (29) is independent of R , hence we have

$$g(\lambda_l) - g(\lambda_{l+1}) = g(\mu_l) - g(\mu_{l+1})$$

for all $1 \leq l \leq m-1$.

Indirectly assume that $\lambda_k > \mu_k$ for some $1 \leq k \leq m$. Then by $\sum_{i=1}^m \lambda_i = \sum_{i=1}^m \mu_i = 1$ we have $\lambda_s < \mu_s$ for some $1 \leq s \leq m$. Without loss of generality we may assume that $s > k$. Then there exists some $t \in \{k, k+1, \dots, s-1\}$ such that

$$[\lambda_{t+1}, \lambda_t] \supsetneq [\mu_{t+1}, \mu_t]$$

which implies by the strict monotonicity of g that

$$g(\lambda_t) - g(\lambda_{t+1}) > g(\mu_t) - g(\mu_{t+1}).$$

A contradiction.

So the proof of the claim is done, and hence the proof Theorem 2 is complete. \square

Acknowledgement. The author is grateful to Lajos Molnár for illuminating discussions.

REFERENCES

1. A. Banerjee *et al.*, *Clustering with Bregman divergences*, J. Mach. Learn. Res. **6** (2005), 1705-1749.
2. R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.
3. 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.
4. E. Carlen, *Trace inequalities and quantum entropy: an introductory course*, Contemp. Math. **529** (2010), 73-140.
5. R. Y. Chen and J. A. Tropp, *Subadditivity of matrix φ -entropy and concentration of random matrices*, Electron. J. Probab. **19** (2014), 1-30.
6. 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)
7. S. Kullback and R.A. Leibler, *On information and sufficiency*, Ann. Math. Statist. **22(1)** (1951), 79 - 86.
8. M. Lewin and J. Sabin, *A family of monotone quantum relative entropies*, Lett. Math. Phys. **104** (2014) 691-705.
9. L. Molnár, *Maps on states preserving the relative entropy* J. Math. Phys. **49**, 032114 (2008).
10. 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.
11. L. Molnár, *An algebraic approach to Wigner's unitary-antiunitary theorem*, J. Austral. Math. Soc. (Series A) **65** (1999), 354-369.
12. L. Molnár, G. Nagy and P. Szokol, *Maps on density operators preserving quantum f -divergences*, Quantum Inf. Process. **12** (2013), 2309-2323.
13. L. Molnár, J. Pitrik and D. Viosztek, *Maps on positive definite matrices preserving Bregman and Jensen divergences*, Linear Algebra Appl. **495** (2016), 174-189.
14. L. Molnár and P. Szokol, *Transformations on positive definite matrices preserving generalized distance measures*, Linear Algebra Appl. **466** (2015), 141-159.
15. F. Nielsen and R. Bhatia (Eds), *Matrix Information Geometry*, Springer, Heidelberg, 2013.
16. J. Pitrik and D. Viosztek, *On the joint convexity of the Bregman divergence of matrices*, Lett. Math. Phys. **105** (2015), 675-692.

DEPARTMENT OF ANALYSIS, INSTITUTE OF MATHEMATICS, BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS, H-1521 BUDAPEST, HUNGARY AND MTA-DE "LENDÜLET" FUNCTIONAL ANALYSIS RESEARCH GROUP, INSTITUTE OF MATHEMATICS, UNIVERSITY OF DEBRECEN, H-4002 DEBRECEN, P.O. BOX 400, HUNGARY

E-mail address: viosz@math.bme.hu

URL: <http://www.math.bme.hu/~viosz>