Acta Math. Hung.
59 (3–4) (1992), 449–455.

CHARACTERIZATION OF THE RELATIVE ENTROPY OF STATES OF MATRIX ALGEBRAS

D. PETZ (Budapest)

For two finite probability distributions $(p_1, p_2, \dots, p_n)$ and $(q_1, q_2, \dots, q_n)$ the quantity

$$\sum_{k=1}^{n} p_k(\log p_k - \log q_k) \tag{1}$$

was introduced in 1951 by Kulback and Leibler. They called it information for discrimination ([5,6]). Some years later Rényi suggested the name information gain ([12]). As a natural analogue of (1) Umegaki [14] defined the relative entropy of two density matrices in 1962 by the formula

$$\operatorname{Tr} \varrho(\log \varrho - \log \varphi). \tag{2}$$

Properties of the relative entropy functional were established in many papers and the highlight of this development was Lieb's convexity theorem ([7]). The notion received much attention in quantum mechanics ([8]). Concerning the details we refer to the survey papers [2] and [9].

The aim of the present paper is to characterize the relative entropy functional through its wellknown properties. As a frame we consider finite dimensional C^*-algebras ([13], p.50). Such algebras are decomposed into a direct sum of full matrix algebras and the commutative ones are isomorphic to the n-tuples of complex numbers for some positive integer n. By a state we mean a positive normalized functional. Each finite dimensional C^*-algebra possesses a natural "uniform distribution" which, more precisely, is a positive functional taking the value one on each minimal projection. It is unique and we denote it by Tr. In case of a matrix algebra Tr reduces to the usual matrix trace and on complex n-tuples

$$\operatorname{Tr}(c_1, \dots, c_n) = \sum_{n=1}^{n} c_i.$$

To each state φ one associates a density D_φ such that

$$\varphi(a) = \operatorname{Tr} D_\varphi a$$

and we call φ faithful if D_φ is invertible. For a faithful state φ and an arbitrary state ω on a finite dimensional C^*-algebra $\mathcal{A}$ their relative entropy is defined as

$$S(\varphi, \omega) = \operatorname{Tr} D_\omega(\log D_\omega - \log D_\varphi). \tag{3}$$

(Note that $D_\omega \log D_\omega$ is well defined even if D_ω is not invertible.)

Our crucial postulate for the relative entropy includes the notion of conditional expectation. Let us recall that in the setting of operator algebras conditional expectation (or projection of norm one) is defined as a positive unital idempotent linear mapping onto a subalgebra ([13], p. 131).

Now we list properties of the relative entropy functional which will be used in an axiomatic characterization:

(i) Conditional expectation property: Assume that $\mathcal{A}$ is a subalgebra of $\mathcal{B}$ and there exists a projection E of $\mathcal{B}$ onto $\mathcal{A}$ of norm one such that $\varphi \circ E = \varphi$. Then for every state ω of $\mathcal{B}$

$$S(\varphi,\omega) = S(\varphi|\mathcal{A},\omega|\mathcal{A}) + S(\omega \circ E,\omega)$$

holds.

(ii) Invariance property: For every automorphism α of $\mathcal{B}$ we have

$$S(\varphi,\omega) = S(\varphi \circ \alpha,\omega \circ \alpha).$$

(iii) Direct sum property: Assume that $\mathcal{B} = \mathcal{B}_1 \oplus \mathcal{B}_2$ and

$$\varphi_{12}(a \oplus b) = \lambda\varphi_1(a) + (1-\lambda)\varphi_2(b), \quad \omega_{12}(a \oplus b) = \lambda\omega_1(a) + (1-\lambda)\omega_2(b)$$

for every $a \in \mathcal{B}_1, b \in \mathcal{B}_2$ and some $0 < \lambda < 1$. Then

$$S(\varphi_{12},\omega_{12}) = \lambda S(\varphi_1,\omega_1) + (1-\lambda)S(\varphi_2,\omega_2).$$

(iv) Nilpotence property: $S(\varphi,\varphi) = 0$.

(v) Measurability property: The function

$$(\varphi,\omega) \mapsto S(\varphi,\omega)$$

is measurable on the state space of the finite dimensional C^*-algebra $\mathcal{B}$ (when φ is assumed to be faithful).

Properties (i)–(v) are wellknown for the relative entropy functional. Among them the conditional expectation property is the most crucial (it was obtained in [11] in full generality). There are plenty of information quantities sharing properties (ii)–(v), all the quasientropies discussed in [10] are so.

Our main result is the following.

THEOREM. *If a real valued functional $S'(\varphi,\omega)$ defined for faithful states φ and arbitrary states ω of finite dimensional C^*-algebras shares properties (i)–(v) then there exists a constant $C \in \mathbb{R}$ such that*

$$S'(\varphi,\omega) = C \operatorname{Tr} D_\omega(\log D_\omega - \log D_\varphi).$$

The proof consists of several steps. We show that for larger and larger class of states

$$S'(\varphi,\omega) = C\, S(\varphi,\omega)$$

must hold.

Consider the three dimensional commutative algebra $\mathbf{C}^3$. Its states correspond to probability distributions (p_1,p_2,p_3) (i.e. $0 \leqq p_i,\ p_1+p_2+p_3 = 1$).

LEMMA 1. *For probability distributions* (p_1,p_2,p_3) *and* (q_1,q_2,q_3) *the recursive relation*

$$\begin{aligned}(4)\qquad S'((q_1,q_2,q_3),(p_1,p_2,p_3)) &= S'((q_1+q_2,q_3),(p_1+p_2,p_3))+\\ &+(p_1+p_2)S'\left(\left(\frac{q_1}{q_1+q_2},\frac{q_2}{q_1+q_2}\right),\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right)\right)\end{aligned}$$

holds.

We benefit from the conditional expectation property in the situation $\mathbf{C}^2 \cong \{(c_1,c_1,c_2) : c_1,c_2 \in \mathbf{C}\} \subset \mathbf{C}^3$. There exists a conditional expectation $E : \mathbf{C}^3 \to \mathbf{C}^2$ preserving the state $\varphi = \varphi_{(q_1,q_2,q_3)}$ and it is given by

$$E : (c_1,c_2,c_3) \mapsto \left(\frac{q_1c_1+q_2c_2}{q_1+q_2},\ \frac{q_1c_1+q_2c_2}{q_1+q_2},\ c_3\right).$$

The state $\omega_{(p_1,p_2,p_3)} \circ E$ corresponds to the measure

$$\left(\frac{p_1+p_2}{q_1+q_2}q_1,\ \frac{p_1+p_2}{q_1+q_2}q_2, p_3\right)$$

and we obtain

$$\begin{aligned}S'((q_1,q_2,q_3),(p_1,p_2,p_3)) &= S'((q_1+q_2,q_3),(p_1+p_2,p_3))+\\ &+S'\left(\left(\frac{p_1+p_2}{q_1+q_2}q_1,\frac{p_1+p_2}{q_1+q_2}q_2,p_3\right),(p_1,p_2,p_3)\right).\end{aligned}$$

Due to the direct sum condition the last term here equals the last term of (4) and the lemma follows.

Interchanging in $\mathbf{C}^3$ the second and third coordinates by means of the invariance condition we conclude the equation

$$\begin{aligned}(5)\qquad & S'((q_1+q_2,q_3),(p_1+p_2,p_3)) + (p_1+p_2)\cdot\\ &\cdot S'\left(\left(\frac{q_1}{q_1+q_2},\frac{q_2}{q_1+q_2}\right),\left(\frac{p_1}{p_1+p_2},\frac{p_2}{p_1+p_2}\right)\right) =\\ &= S'((q_1+q_3,q_2),(p_1+p_3,p_2)) + (p_1+p_3)\cdot\\ &\cdot S'\left(\left(\frac{q_1}{q_1+q_3},\frac{q_3}{q_1+q_3}\right),\left(\frac{p_1}{p_1+p_3},\frac{p_3}{p_1+p_3}\right)\right).\end{aligned}$$

With the notation

$$F(x,y) = S'((1-y,y),(1-x;x))$$

equation (5) is of the following form:

(6)

$$F(x,y) + (1-x)F\left(\frac{u}{1-x}, \frac{v}{1-y}\right) = F(u,v) + (1-u)F\left(\frac{x}{1-u}, \frac{y}{1-v}\right).$$

The functional equation (6) has been solved under the special nilpotence condition $F(1/2,1/2) = 0$ in [4]. A lengthy but elementary analysis yields that the only measurable solution of (6) is

$$F(x,y) = C\left(x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}\right).$$

(See also pp. 204-207 of [1].)

The recursion (4) remains true if p_3 and q_3 are replaced by $p_3, p_4, \dots, p_n$ and $q_3, q_4, \dots, q_n$ respectively. In this way we obtain that

$$S'(\varphi,\omega) = C\,S(\varphi,\omega)$$

whenever φ and ω are states of commutative finite dimensional C^*-algebras.

Now let φ and ω be states of an algebra $\mathcal{B}$ such that the densities D_φ and D_ω commute. Let $\mathcal{A}$ be the maximal abelian subalgebra generated by D_φ and D_ω. If E is the conditional expectation of $\mathcal{B}$ onto $\mathcal{A}$ which preserves Tr then $\varphi\circ E = \varphi$ and $\omega\circ E = \omega$. The conditional expectation property tells us that

$$S'(\varphi,\omega) = S'(\varphi|\mathcal{A},\omega|\mathcal{A}) + S'(\omega\circ E,\omega).$$

By nilpotence the second term vanishes and we arrive at

$$S'(\varphi,\omega) = C\,S(\varphi,\omega) \tag{7}$$

for commuting states.

The next step is $\mathcal{B} = M_n(\mathbf{C})$. Our aim is to show that (7) holds for arbitrary states on $\mathcal{B}$. (As always, φ is supposed to be faithful.)

LEMMA 2. *If* $\sigma = \lambda\sigma_1 + (1-\lambda)\sigma_2$ $(0<\lambda<1)$ *then*

$$\lambda S'(\varphi,\sigma_1) + (1-\lambda)S'(\varphi,\sigma_2) = S'(\varphi,\sigma) + \lambda S'(\sigma,\sigma_1) + (1-\lambda)S'(\sigma,\sigma_2). \tag{8}$$

The proof of (8) is quite similar to that of Lemma 1. The conditional expectation property should be applied to $\mathcal{B}\oplus\mathcal{B}$ and its subalgebra $\{b\oplus b : b\in\mathcal{B}\}$. The mapping

$$E(a\oplus b) = (\lambda a + (1-\lambda)b)\oplus(\lambda a + (1-\lambda)b)$$

is a conditional expectation leaving the state

$$\varphi_{12}(a \oplus b) = \lambda\varphi(a) + (1-\lambda)\varphi(b)$$

invariant. The argument is completed by referring to the invariance and direct sum properties.

Now we resume the determination of $S(\varphi,\omega)$ for states of $\mathcal{B} = M_n(\mathbf{C})$. We choose a basis such that the density of φ is diagonal. Then the density of ω is of the form

$$\begin{pmatrix} A_k & 0 \\ 0 & D_{n-k} \end{pmatrix} = D_\omega$$

where $A_k \in M_k(\mathbf{C})$ and $D_{n-k} \in M_{n-k}(\mathbf{C})$ is a diagonal matrix. We are going to prove (7) by mathematical induction on k. If $k=0$ then D_φ and D_ω commute and (7) holds.

Let U be a diagonal unitary matrix such that

$$U_{ii} = \begin{cases} 1 & i \neq k \\ -1 & i = k. \end{cases}$$

Then

$$D_{\sigma_2} = U\, D_\omega\, U$$

differs from $D_{\sigma_1} \equiv D_\omega$ only in the sign of the entries in the kth row and kth column but they are the same along the diagonal. The density

$$D_\sigma = \frac{1}{2}(D_{\sigma_1} + D_{\sigma_2})$$

is of the form

$$\begin{pmatrix} A_{k-1} & 0 \\ 0 & D_{n-k+1} \end{pmatrix}$$

where D_{n-k+1} is an $(n-k+1)\times(n-k+1)$ diagonal matrix. From the induction hypothesis we have

$$S'(\varphi,\sigma) = C\, S(\varphi,\sigma). \tag{9}$$

Write (8) with $\tau = \mathrm{Tr}/n$ and $\lambda = \frac{1}{2}$. Then

$$\frac{1}{2}S'(\tau,\sigma_1) + \frac{1}{2}S'(\tau,\sigma_2) = S'(\tau,\sigma) + \frac{1}{2}S'(\sigma,\sigma_1) + \frac{1}{2}S'(\sigma,\sigma_2). \tag{10}$$

The invariance, more precisely the relations $\tau(U\cdot U) = \tau(\cdot)$, $\sigma_1(U\cdot U) = \sigma_2(\cdot)$, $\sigma(U\cdot U) = \sigma(\cdot)$, ensures

$$S'(\tau,\sigma_1) = S'(\tau,\sigma_2) \quad \text{and} \quad S'(\sigma,\sigma_1) = S'(\sigma,\sigma_2).$$

Therefore

$$S'(\tau,\sigma_1) = S'(\tau,\sigma) + S'(\sigma,\sigma_1)$$

and (7) yields

$$S'(\sigma,\sigma_1) = C\, S(\sigma,\sigma_1). \tag{11}$$

In our special case

$$S'(\varphi,\sigma_1) = S'(\varphi,\sigma_2) \quad \text{and} \quad S'(\sigma,\sigma_1) = S'(\sigma,\sigma_2)$$

hold due to the invariance under the automorphism Ad U. Hence (8) reads as

$$S'(\varphi,\sigma_1) = S'(\varphi,\sigma) + S'(\sigma,\sigma_1)$$

where both terms on the right hand side have been compared with the relative entropy ((9) and (11)). So (7) holds for all states φ and ω on $M_n(\mathbf{C})$. Since a finite dimensional C^*-algebra is the direct sum of full matrix algebras, the direct sum property extends (7) to all C^*-algebras of finite dimension.

We note that (8) is Donald's mixing axiom and the subsequent argument is due to him ([3]).

By more complicated mathematical tools the concept of relative entropy may be extended to states of infinite dimensional C^*-algebras ([2]). If we restrict ourselves to C^*-algebras with good approximation property (nuclear algebras) then the characterization in the present paper may be modified easily. It is sufficient to choose instead of (ii) the monotonicity and instead of measurability the lower semicontinuity property given below.

(ii)' Monotonicity property: For every completely positive unital mapping α we have

$$S(\varphi,\omega) \geqq S(\varphi\circ\alpha, \omega\circ\alpha).$$

(v)' Lower semicontinuity: The functional

$$(\varphi,\omega) \mapsto S(\varphi,\omega)$$

is lower semicontinuous with respect to the weak topology.

These postulates (i), (ii)', (iii), (iv) and (v)' constitute the definition of the relative entropy functional up to a constant factor. Details of this remark are out of the scope of the present paper and they will be discussed in forthcoming publications.

References

[1] J. Aczél and Z. Daróczy, *On Measures of Information and their Characterizations*, Academic Press (New York, 1975).

[2] H. Araki, *Recent Progress on Entropy and Relative Entropy*, VIIIth International Congress on Mathematical Physics (Eds: M. Mebkhout, R. Sénéor), pp. 354–365, World Scientific, 1987.

[3] M. J. Donald, Further results on the relative entropy, *Math. Proc. Camb. Phil. Soc.*, **101** (1987), 363–373.
[4] Pl. Kannappan and C. T. Ng, Measurable solutions of functional equations related to information theory, *Proc. Amer. Math. Soc.*, **38** (1973), 303–310.
[5] S. Kullback and R. Leibler, On information and sufficiency, *Ann. Math. Stat.*, **22** (1951), 79–86.
[6] S. Kullback, *Information Theory and Statistics*, Wiley (New York, 1959).
[7] E. H. Lieb, Some convexity and subadditivity properties of entropy, *Bull. Amer. Math. Soc.*, **81** (1975), 1–14.
[8] G. Lindblad, Expectations and entropy inequalities for finite quantum systems, *Commun. Math. Phys.*, **39** (1974), 111–119.
[9] D. Petz, Properties of quantum entropy, *Quantum Probability and Applications II* (Eds: L. Accardi, W. von Waldenfels), 428–441, Lecture Notes in Math., 1136, Springer, 1985.
[10] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.*, **23** (1986), 57–65.
[11] D. Petz, Properties of the relative entropy of states of a von Neumann algebra, *Acta Math. Hung.*, **47** (1986), 65–72.
[12] R. Rényi, *Wahrscheinlichkeitsrechnung*, Deutscher Verlag der Wissenschaften (Berlin, 1962).
[13] M. Takesaki, *Operator Algebras I*, Springer, 1979.
[14] H. Umegaki, Conditional expectations in an operator algebra IV (entropy and information), *Kodai Math. Sem. Rep.*, **14** (1962), 59–85.

(Received February 27, 1990)

MATHEMATICAL INSTITUTE
OF THE HUNGARIAN ACADEMY OF SCIENCES
H–1364 BUDAPEST, P.O.B. 127.
HUNGARY