J. OPERATOR THEORY
16(1986), 335–348

RELATIVE ENTROPY OF STATES: A VARIATIONAL EXPRESSION

HIDEKI KOSAKI

0. INTRODUCTION

In [4], Araki introduced the notion of relative entropy $S(\varphi, \psi)$ of states φ, ψ on a (not necessarily semi-finite) von Neumann algebra. (See [12] for the semi-finite case.) In recent applications of the theory of operator algebras to quantum statistical mechanics, Araki's relative entropy has been playing an important role. It is also crucially used in Connes' work [5] on entropy of automorphisms.

Araki defined $S(\varphi, \psi)$ by using a standard form of the algebra in question, [2], [7], and other relevant objects. (See Definition 3.1.) Due to uniqueness of a standard form (up to a spatial isomorphism), $S(\varphi, \psi)$ depends only on the states φ, ψ. Yet, if so, it is certainly more desirable that one can express $S(\varphi, \psi)$ in terms of just φ, ψ.

The main purpose of the article is to obtain certain expressions for $S(\varphi, \psi)$ and other related quantities in terms of states themselves. Partly from physical consideration, these quantities are expected to satisfy certain properties. Proving them is not so easy, and effort has been made by several authors ([4], [9], [10], [11], [13]). Our expression is variational, and all non-trivial properties are "built-in" in the expression. In fact, all of them can be immediately derived from our expression.

1. NOTATIONS AND PRELIMINARIES

Throughout, let M be a von Neumann algebra with a standard form $(M, \mathscr{H}, J, \mathscr{P}^\natural)$, [2], [7]. For each $\varphi \in M_*^+$, its unique implementing vector in the natural cone $\mathscr{P}^\natural$ is denoted by ξ_φ. ($\varphi(x) = (x\xi_\varphi|\xi_\varphi)$, $x \in M$.) Let p_φ (resp. p'_φ) be the projection onto the closure of $M'\xi_\varphi$ (resp. $M\xi_\varphi$). Thus, $p_\varphi \in M$, $p'_\varphi \in M'$, and p_φ is exactly the support projection of φ.

Following [4], we introduce a relative modular operator $\Delta_{\varphi\psi}$ for (not necessarily faithful) functionals $\varphi, \psi \in M_*^+$. Namely, $\Delta_{\varphi\psi}$ is the positive self-adjoint operator on the standard Hilbert space $\mathscr{H}$ characterized by

(1)
$$M\xi_\psi \oplus (1-p'_\psi)\mathscr{H} \text{ is a core for } \Delta_{\varphi\psi}^{1/2}, \text{ the support of } \Delta_{\varphi\psi} \text{ is } p'_\psi p_\varphi \mathscr{H},$$
$$J\Delta_{\varphi\psi}^{1/2}(x\xi_\psi + \zeta) = p_\psi x^* \xi_\varphi, \quad x \in M, \ p'_\psi \zeta = 0.$$

LEMMA 1.1. *For each* $t > 0$, *we have*

$$(\Delta_{\varphi\psi}(t+\Delta_{\varphi\psi})^{-1}\xi_\psi \mid \xi_\psi) = \inf\{\psi(x^*x) + t^{-1}\varphi(yy^*);\ x+y=1, x, y \in M\}.$$

Proof. Since $M\xi_\psi \oplus (1-p'_\psi)\mathscr{H}$ is a core for $\Delta_{\varphi\psi}^{1/2}$, it follows from Lemma 2.1, [8], that the above left side is equal to

(2)
$$\inf\{\|x\xi_\psi + \eta\|^2 + t^{-1}\|\Delta_{\varphi\psi}^{1/2}(y\xi_\psi + \zeta)\|^2;$$
$$x\xi_\psi + \eta + y\xi_\psi + \zeta = \xi_\psi, \quad x, y \in M, \ \ \eta, \zeta \in (1-p'_\psi)\mathscr{H}\}.$$

Notice that

$$\|\Delta_{\varphi\psi}^{1/2}(y\xi_\psi + \zeta)\|^2 = \|Jp_\psi y^*\xi_\varphi\|^2 = \|p_\psi y^* \xi_\varphi\|^2, \qquad \text{by (1)}$$
$$\|x\xi_\psi + \eta\|^2 = \|x\xi_\psi\|^2 + \|\eta\|^2,$$
$$x\xi_\psi + \eta + y\xi_\psi + \zeta = \xi_\psi \Leftrightarrow (x+y)p_\psi = p_\psi \text{ and } \zeta + \eta = 0.$$

Therefore, (2) is equal to

$$\inf\{\|x\xi_\psi\|^2 + t^{-1}\|p_\psi y^*\xi_\varphi\|^2;\ (x+y)p_\psi = p_\psi, \quad x, y \in M\},$$

which is obviously majorized by

$$\inf\{\|x\xi_\psi\|^2 + t^{-1}\|y^*\xi_\varphi\|^2; \quad x+y=1, \ x, y \in M\}$$

since $\|p_\psi y^*\xi_\varphi\|^2 \leqslant \|y^*\xi_\varphi\|^2$. Actually, the reversed majorization is also valid by the following fact: If $x, y \in M$ satisfy $(x+y)p_\psi = p_\psi$, then $\tilde{x} = xp_\psi + (1-p_\psi)$ and $\tilde{y} = yp_\psi$ satisfy

$$\tilde{x} + \tilde{y} = 1,$$
$$\|\tilde{x}\xi_\psi\|^2 + t^{-1}\|\tilde{y}^*\xi_\varphi\|^2 = \|x\xi_\psi\|^2 + t^{-1}\|p_\psi y^*\xi_\varphi\|^2.$$

Q.E.D.

Throughout the article, $g(\lambda)$ will be used to denote an operator monotone function on the closed half-interval $[0, \infty)$. ($g(a) \leqslant g(b)$ as operators whenever operators a, b satisfy $0 \leqslant a \leqslant b$.) Our standard reference on this subject is [6]. The next result is classical.

LEMMA 1.2. *For such a function $g(\lambda)$, there exist (unique) real numbers $\alpha, \beta, \beta \geqslant 0$, and a finite Radon measure μ on $(0,\infty)$ such that*

$$g(\lambda) = \alpha + \beta\lambda + \int_0^\infty \lambda(t+\lambda)^{-1}(1+t)\mathrm{d}\mu(t), \quad \lambda \geqslant 0.$$

Conversely, such an integral expression always gives rise to an operator monotone function on $[0,\infty)$.

2. VARIATIONAL EXPRESSIONS

Throughout the section, we assume that an operator monotone function $g(\lambda)$ on $[0,\infty)$ has the integral expression described in Lemma 1.2. At first we obtain a certain variational expression for $(g(\Delta_{\varphi\psi})\xi_\psi \mid \xi_\psi)$. (The integral expression of $g(\lambda)$ guarantees that $g(\lambda)$ is concave so that $g(0) \leqslant g(\lambda) \leqslant a\lambda + b$ for some $a, b \geqslant 0$. The vector ξ_ψ being in the domain of $\Delta_{\varphi\psi}^{1/2}$, the expression $(g(\Delta_{\phi\psi})\xi_\psi|\xi_\psi)$ always makes sense as a form.)

LEMMA 2.1. *Let N be a subspace of M containing 1. If N is dense in M with respect to the strong* operator topology, then for any $\varphi, \psi \in M_*^+$ we have*

$$(g(\Delta_{\varphi\psi})\xi_\psi \mid \xi_\psi) = \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) +$$

$$+ \int_0^\infty \inf\{\psi(x^*x) + t^{-1}\varphi(yy^*);\ x + y = 1,\ x, y \in M\}(1+t)\mathrm{d}\mu(t).$$

Proof. Let $\Delta_{\varphi\psi} = \int_0^\infty \lambda \mathrm{d}e_\lambda$ be the spectral decomposition. Since the support of $\Delta_{\varphi\psi}$ is $p'_\psi p_\varphi \mathscr{H}$ ((1)), we get

$$\int_0^\infty \alpha\,\mathrm{d}\|e_\lambda\xi_\psi\|^2 = \alpha(p'_\psi p_\varphi \xi_\psi \mid \xi_\psi) = \alpha\psi(p_\varphi).$$

Also, because of $\Delta_{\varphi\psi}^{1/2}\xi_\psi = Jp_\psi\xi_\varphi$ ((1)), we get

$$\int_0^\infty \beta\lambda\,\mathrm{d}\|e_\lambda\xi_\psi\|^2 = \beta\|\Delta_{\varphi\psi}^{1/2}\xi_\psi\|^2 = \beta\|Jp_\psi\xi_\varphi\| = \beta\varphi(p_\psi).$$

We thus get

$$(g(\Delta_{\varphi\psi})\xi_\psi \mid \xi_\psi) = \int_0^\infty g(\lambda)\mathrm{d}\|e_\lambda\xi_\psi\|^2 =$$

$$= \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty \left(\int_0^\infty \lambda(t+\lambda)^{-1}(1+t)\mathrm{d}\mu(t)\right)\mathrm{d}\|e_\lambda\xi_\psi\|^2 =$$

$$= \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty \left(\int_0^\infty \lambda(t+\lambda)^{-1}\mathrm{d}\|e_\lambda\xi_\psi\|^2\right)(1+t)\,\mathrm{d}\mu(t) =$$

(by the Fubini-Tonelli theorem)

$$= \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty (\Delta_{\varphi\psi}(t+\Delta_{\varphi\psi})^{-1}\xi_\psi \mid \xi_\psi)(1+t)\,\mathrm{d}\mu(t).$$

Now the result follows from Lemma 1.1 and the density of N. Q.E.D.

The above result is implicit in [10], and closely related to [1]. But now we go further. Namely, we will try to switch the order of the integral sign and the inf sign.

Following an idea in [10], (fixing $\varphi, \psi \in M_*^+$) we introduce the following two positive sesquilinear forms on the algebra M:

$$\varphi_{\mathrm{R}}: (x, y) \in M\times M \to \varphi_{\mathrm{R}}(x, y) = \varphi(xy^*) \in \mathbf{C},$$

$$\psi_{\mathrm{L}}: (x, y) \in M\times M \to \varphi_{\mathrm{L}}(x, y) = \psi(y^*x) \in \mathbf{C}.$$

Setting $L = \{x \in M; \varphi_{\mathrm{R}}(x, x) + \psi_{\mathrm{L}}(x, x) = 0\}$, we consider the pre-Hilbert space M/L equipped with the inner product

$$\langle x + L \mid y + L\rangle = \varphi_{\mathrm{R}}(x, y) + \psi_{\mathrm{L}}(x, y).$$

(L is a subspace, and $\langle\cdot|\cdot\rangle$ is well-defined due to the Cauchy-Schwarz inequality.) Let K be the Hilbert space completion (the inner product is still denoted by $\langle\cdot|\cdot\rangle$), and $i: M \to K$ is the canonical map. If a subspace N (containing 1) of M is s*-dense, then

(3) $i(N)$ is dense in K

from the construction. Since $\psi_{\mathrm{L}}(x, x) \leqslant \langle i(x) \mid i(x)\rangle$, $x \in M$, there exists a unique positive operator k, $0 \leqslant k \leqslant 1$, on K satisfying

$$\psi(y^*x) = \langle k\,i(x) \mid i(y)\rangle, \quad x, y \in M. \tag{4}$$

Obviously we get

$$\varphi(xy^*) = \langle (1-k)i(x) \mid i(y)\rangle, \quad x, y \in M. \tag{5}$$

The right side of the equality in Lemma 2.1 is equal to

$$\alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty \inf\{\langle k\,i(x) \mid i(x)\rangle + $$

$$+ t^{-1}\langle (1-k)i(y) \mid i(y)\rangle;\ x + y = 1, x, y \in N\}(1+t)\mathrm{d}\mu(t) =$$

$$= \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty \inf\{\langle tk\,i(x) \mid i(x)\rangle +$$

$$+ \langle (1-k)i(y) \mid i(y)\rangle;\ x + y = 1, x, y \in N\}(1+t)t^{-1}\mathrm{d}\mu(t).$$

Due to (3) and Lemma 3.2, [8], this is equal to

$$\alpha\psi(p_\varphi) + \beta\varphi(p_\psi) + \int_0^\infty \langle tk(1-k)\{tk + (1-k)\}^{-1}i(1) \mid i(1)\rangle(1+t)t^{-1}\mathrm{d}\mu(t). \tag{6}$$

The following two remarks are in order:

$$t \in (0, \infty) \to tk\{tk + (1-k)\}^{-1}i(1) \in K \tag{7}$$

is continuous,

$$\|tk + (1-k)\| \leqslant \max(1, t), \quad t > 0. \tag{8}$$

In fact, (7) is obvious while (8) follows from

$$tk + (1-k) \leqslant \begin{cases} k + (1-k) = 1 & \text{if } 0 < t \leqslant 1, \\ tk + t(1-k) = t1 & \text{if } t \geqslant 1. \end{cases}$$

Now we can prove:

THEOREM 2.2. *Let N be a subspace of M containing 1 which is dense in M*

with respect to the strong-operator topology. For any* $\varphi, \psi \in M_*^+$, *we have*

$$(g(\Delta_{\varphi\psi})\xi_\psi \mid \xi_\psi) = \alpha\psi(p_\varphi) + \beta\varphi(p_\psi) +$$

$$+ \inf \int_0^\infty \{\psi((1 - x(t))^*(1 - x(t))) + t^{-1}\varphi(x(t)x(t)^*)\} (1 + t)\, d\mu(t),$$

where the infimum is taken over all N-valued step functions x on $(0, \infty)$.

Unless otherwise is stated, by a step function we will always mean a step function with *finitely* many values.

Proof. Thanks to Lemma 2.1 and (6), it suffices to show

$$\int_0^\infty \langle tk(1 - k)\{tk + (1 - k)\}^{-1} i(1) \mid i(1)\rangle (1 + t)t^{-1} d\mu(t) =$$

(9)

$$= \inf \int_0^\infty \{\varphi(x(t)x(t)^*) + t\psi((1 - x(t))^*(1 - x(t)))\}(1 + t)t^{-1} d\mu(t).$$

Elementary computation shows that, for any $x(t) \in M$, $t \in (0, \infty)$, one gets

$$\langle tk(1 - k)\{tk + (1 - k)\}^{-1} i(1) \mid i(1)\rangle + \langle \{tk + (1 - k)\}\xi(t) \mid \xi(t)\rangle =$$

$$= \langle (1 - k)\, i(x(t)) \mid i(x(t))\rangle + \langle tk\, i(1 - x(t)) \mid i(1 - x(t))\rangle$$

with $\xi(t) = tk\{tk + (1 - k)\}^{-1} i(1) - i(x(t))$.

By (4) and (5), in (9) the second term is equal or greater than the first. To show (9), we have to show that the (positive) error term $\langle \{tk + (1 - k)\}\xi(t) \mid \xi(t)\rangle$ (after integrated with respect to the measure $(1 + t)t^{-1} d\mu(t)$) can be arbitrarily small for an appropriate choice of a step function $x(t)$.

Choose and fix $\varepsilon > 0$. For $t > n_0 = n_0(\varepsilon)$ large enough and to be fixed later, we set $x(t) = 1$ $(\in N)$. Then

$$\xi(t) = tk\{tk + (1 - k)\}^{-1} i(1) - i(1) =$$

$$= -(1 - k)\{tk + (1 - k)\}^{-1} i(1), \quad t > n_0.$$

We thus get

$$\int_{n_0}^{\infty} \langle \{tk + (1-k)\}\xi(t) \mid \xi(t)\rangle (1+t)t^{-1}\mathrm{d}\mu(t) =$$

$$= \int_{n_0}^{\infty} \langle -(1-k)i(1) \mid -(1-k)\{tk + (1-k)\}^{-1}i(1)\rangle (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant$$

$$\leqslant \int_{n_0}^{\infty} \|1-k\| \, \|(1-k)\{tk+(1-k)\}^{-1}\| \, \|i(1)\|^2 (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant$$

$$\leqslant \|i(1)\|^2 \int_{n_0}^{\infty} (1+t)t^{-1}\mathrm{d}\mu(t).$$

Since $(1+t)t^{-1} \sim 1$ (for t large) and $\mu((n_0, \infty)) \leqslant \mu((0,\infty)) < +\infty$, it is possible to choose n_0 such that

$$\int_{n_0}^{\infty} \langle \{tk + (1-k)\}\xi(t) \mid \xi(t)\rangle (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant \varepsilon/3.$$

Next, for $0 < t < \eta_0 = \eta_0(\varepsilon)$, small enough and to be fixed later, we set $x(t) = 0 \ (\in N)$. In this case, we get

$$\xi(t) = tk\{tk + (1-k)\}^{-1}i(1) - i(0) =$$

$$= tk\{tk + (1-k)\}^{-1}i(1), \quad 0 < t < \eta_0,$$

and estimate

$$\int_{0}^{\eta_0} \langle \{tk + (1-k)\}\xi(t) \mid \xi(t)\rangle (1+t)t^{-1}\mathrm{d}\mu(t) =$$

$$= \int_{0}^{\eta_0} \langle tk\, i(1) \mid tk\{tk + (1-k)\}^{-1}i(1)\rangle (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant$$

$$\leqslant \int_0^{\eta_0} t\|k\| \, \|(tk)\{tk + (1-k)\}^{-1}\| \, \|i(1)\|^2(1+t)t^{-1}\mathrm{d}\mu(t) \leqslant$$

$$\leqslant \|i(1)\|^2 \int_0^{\eta_0} (1+t)\mathrm{d}\mu(t).$$

Since $1 + t \sim 1$ (for t small) and $\mu((0, \eta_0)) \leqslant \mu((0, \infty)) < +\infty$, it is possible to choose η_0 such that

$$\int_0^{\eta_0} \langle \{tk + (1-k)\} \xi(t) \mid \xi(t) \rangle (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant \varepsilon/3.$$

It remains to show that one can choose a finitely many valued step function $x(t)$ on the compact interval $[\eta_0, n_0]$. Thanks to (3) and (7), we can choose such an $x(t)$ such that

$$\|\xi(t)\|^2 \leqslant \begin{cases} \varepsilon(3\mu((0,\infty)))^{-1}t(1+t)^{-1} & \text{if } t \in (0,1] \cap [\eta_0, n_0], \\ \varepsilon(3\mu((0,\infty)))^{-1}(1+t)^{-1} & \text{if } t \in (1,\infty) \cap [\eta_0, n_0]. \end{cases}$$

Then it follows from (8) that

$$\int_{\eta_0}^{n_0} \langle \{tk + (1-k)\} \xi(t) \mid \xi(t) \rangle (1+t)t^{-1}\mathrm{d}\mu(t) \leqslant$$

$$\leqslant \int_{\eta_0}^{n_0} \max(1, t)\|\xi(t)\|^2(1+t)t^{-1}\mathrm{d}\mu(t) \leqslant \varepsilon/3.$$

Q.E.D.

From the above proof, it is obvious that a step function $x(t)$ can be further assumed $x(t) = 0$ for t small enough and $x(t) = 1$ for t large enough.

3. RELATIVE ENTROPY

We will obtain variational expressions of $S(\varphi, \psi)$, $\varphi, \psi \in M_*^+$. For a moment (see Remark 3.3), we assume $p_\varphi \geqslant p_\psi$. Hence, by (1), the unique implementing vector ξ_ψ in $\mathscr{P}^\natural$ is in the support of $\Delta_{\varphi\psi}$. Keeping this fact in mind, we introduce:

DEFINITION 3.1 ([4]). Using the spectral decomposition $\Delta_{\varphi\psi} = \int_0^\infty \lambda \mathrm{d}e_\lambda$, we set

$$S(\varphi, \psi) = -\int_0^\infty \log\lambda \,\mathrm{d}\|e_\lambda \xi_\psi\|^2,$$

the *relative entropy* of φ relative to ψ.

Notice that

$$-\infty < -\int_1^\infty \log\lambda \,\mathrm{d}\|e_\lambda \xi_\psi\|^2 \leqslant 0$$

because of $0 \leqslant \log\lambda \leqslant \lambda$, $\lambda \geqslant 1$, and the fact that ξ_ψ is in the domain of $\Delta_{\varphi\psi}^{1/2}$. Thus the integral in the above definition always make sense ($S(\varphi, \psi) \in (-\infty, \infty]$). Furthermore, when $\varphi(1) = \psi(1)$, $1 - \lambda \leqslant -\log\lambda$, $\lambda > 0$, implies $S(\varphi, \psi) \in [0, \infty)$.

Because of $\lim_{\lambda \downarrow 0} \log\lambda = -\infty$ we cannot directly apply the result in §2 to $\log\lambda$ (although $\log\lambda$ is operator monotone on $(0, \infty)$). Instead, we set

$$g_n(\lambda) = -\log(\lambda + n^{-1}), \quad \lambda \geqslant 0,$$

for $n = 1, 2, \ldots$, and observe

$$g_n(\lambda) \uparrow -\log\lambda \quad \text{as } n \uparrow \infty$$

for each $\lambda > 0$. Using the monotone convergence theorem on (0, 1) and Lebesgue's dominated convergence theorem on $[1, \infty)$, we get

$$\int_0^\infty g_n(\lambda)\mathrm{d}\|e_\lambda \xi_\psi\|^2 \uparrow S(\varphi, \psi) \quad \text{as } n \uparrow \infty. \tag{10}$$

We also notice that

$$-g_n(\lambda) = \log(\lambda + n^{-1}) = -\log n + \int_{n^{-1}}^\infty \lambda(t + \lambda)^{-1} t^{-1} \mathrm{d}t.$$

Thus, each g_n is operator monotone on $[0, \infty)$. (In Lemma 1.2, we choose $\alpha = -\log n$, $\beta = 0$, and $(1 + t)\mathrm{d}\mu(t) = \chi_{(n^{-1}, \infty)}(t)t^{-1}\mathrm{d}t$.) Theorem 2.2 shows that

$$(g(\Delta_{\varphi\psi})\xi_\psi \mid \xi_\psi) = \sup[\psi(p_\varphi)\log n - \int_{n^{-1}}^{\infty} \{\psi((1 - x(t))^*(1 - x(t))) +$$

$$+ t^{-1}\varphi(x(t)x(t)^*)\}t^{-1}\mathrm{d}t,$$

where the sup is taken over all N-valued step functions x on (n^{-1}, ∞). Because of the assumption $p_\varphi \geqslant p_\psi$, we get

$$\psi(p_\varphi) = \psi(1)$$

(see Remark 3.3). Therefore, by (10) we have:

THEOREM 3.2. *Assume that N is a subspace of a von Neumann algebra M and that 1 is included in N. If N is dense in the strong*-operator topology, then for any $\varphi, \psi \in M_*^+$ we get*

$$S(\varphi, \psi) = \sup_{n \in \mathbf{N}^+} \sup[\psi(1)\log n - \int_{n^{-1}}^{\infty} \{\psi((1 - x(t))^*(1 - x(t))) +$$

$$+ t^{-1}\varphi(x(t)x(t)^*)\}t^{-1}\mathrm{d}t],$$

where the second supremum is taken over all (finitely many) N-valued step functions x on (n^{-1}, ∞).

REMARK 3.3. When $p_\varphi \not\geqslant p_\psi$, Araki, [4], set $S(\varphi, \psi) = +\infty$. We claim that our variational expression, that is, the right side of the equation in the above theorem, also gives $+\infty$ in this case. In fact, for example set $x(t) = 1 - p_\varphi$, $0 < t \leqslant 1$, and $x(t) = 1$, $t > 1$. Then the inside of the double sup sign is equal to

$$\psi(1 - p_\varphi)\log n - \varphi(1)$$

by straightforward computation. Since $p_\varphi \not\geqslant p_\psi$, we get $\psi(1 - p_\varphi) \gneq 0$ and

$$\sup_{n \in \mathbf{N}^+} \{\psi(1 - p_\varphi)\log n - \varphi(1)\} = +\infty.$$

In the proof of Theorem 3.2, we approximated $\log\lambda$ by operator monotone functions $\log(\lambda + n^{-1})$ on $[0, \infty)$ from above. Notice that there are many other ways to approximate $\log\lambda$. For example,

$$\log\lambda = \frac{d}{d\theta}\Big|_{\theta=0} \lambda^{\theta} = \lim_{n\to\infty} n(\lambda^{n^{-1}} - 1)$$

and

$$n(\lambda^{n^{-1}} - 1) \downarrow \log\lambda \quad \text{as } n \uparrow \infty.$$

In this case,

$$n(\lambda^{n^{-1}} - 1) = -n + \int_0^\infty \lambda(t+\lambda)^{-1} n\pi^{-1}\sin(n^{-1}\pi)t^{n^{-1}-1}\mathrm{d}t.$$

Therefore, the same argument as the proof of Theorem 2.2 gives rise to another similar expression of $S(\varphi, \psi)$.

Also it is possible to avoid a double sup sign (as shown below). Instead we have to take the sup over all countably many N-valued step functions. Recall the integral expression:

$$-\log\lambda = \int_0^\infty \{(t+1)^{-1} - \lambda(t+\lambda)^{-1}\}t^{-1}\mathrm{d}t, \quad \lambda > 0.$$

Based on this, we can prove:

$$S(\varphi, \psi) = \sup \int_0^\infty \{(t+1)^{-1}\psi(1) - \psi((1 - x(t))^*(1 - x(t)) -$$
$$- t^{-1}\varphi(x(t)x(t)^*)\}t^{-1}\mathrm{d}t,$$

where the supremum is taken over all countably many N-valued step functions x on $(0, \infty)$ which are bounded and $x(t) = 1$ for t large enough. This expression is less powerful than the previous ones, and will not be used later. Full details are left to the reader. The estimate for t small enough in the proof of Theorem 2.2 fails for the integral expression of $-\log\lambda$. This is the reason why we have to look at countably many valued step functions.

4. PROPERTIES

As mentioned in §0, all non-trivial properties of the relative entropy are "built in" in the expression of Theorem 3.2. Although these properties have been already known ([3], [4], [9], [11], [13]), for the sake of completeness we will derive them from Theorem 3.2.

THEOREM 4.1. (i) (Lower semi-continuity). *The map:* $(\varphi, \psi) \in M_*^+ \times M_*^+ \to$ $\to S(\varphi, \psi) \in (-\infty, \infty]$ *is jointly lower semi-continuous with respect to the* $\sigma(M_*, M)$-*-topology.*

(ii) (Joint convexity). *The map in* (i) *is jointly convex.*

(iii) *If* $\varphi_1 \leqslant \varphi_2$, *then* $S(\varphi_1, \psi) \geqslant S(\varphi_2, \psi)$.

(iv) *If* M_1 *is a von Neumann subalgebra of* M, *then* $S(\varphi_1, \psi_1) \leqslant S(\varphi, \psi)$, *where* φ_1, ψ_1 *are the restrictions of* φ, ψ *to* M_1. *In particular, with* $M_1 = \mathbf{C}1$, *we have the Peierls-Bogolubov inequality:*

$$-\psi(1)\log(\varphi(1)/\psi(1)) \leqslant S(\varphi, \psi).$$

(As explained in [3], *when* $l_1\varphi \leqslant \psi \leqslant l_2\varphi$ *for some* $l_1, l_2 > 0$, *this means that* $\varphi(1) \geqslant \psi(1)\exp(\psi(h)/\psi(1))$ *with the relative Hamiltonian operator* $h = h^* = i^{-1}\frac{d}{dt}\Big|_{t=0} (D\varphi; D\psi)_t$.*)*

(v) (Effect of a strongly positive map). *Let* M_1 *be another von Neumann algebra, and* $\gamma: M \to M_1$ *be a unit preserving strongly positive* $(\gamma(x^*x) \geqslant \gamma(x)^*\gamma(x))$ *normal linear map. Then for* $\varphi, \psi \in (M_1)_*^+$ *we have* $S(\varphi \circ \gamma, \psi \circ \gamma) \leqslant S(\varphi, \psi)$.

(vi) (Martingale convergence). *Let* $\{M_\iota\}_{\iota \in I}$ *be an increasing net of von Neumann subalgebras of* M *with* $(\bigcup_{\iota \in I} M_\iota)'' = M$. *Then the increasing* ((iv)) *net* $\{S(\varphi|M_\iota, \psi|M_\iota)\}_{\iota \in I}$ *converges to* $S(\varphi, \psi)$.

Proof. (i) ~ (v) immediately follow from Theorem 2.2 (with $N = M$).

(vi) We assume $S(\varphi, \psi) < +\infty$. (The case $S(\varphi, \psi) = +\infty$ can be handled similarly.) Choose $N = \bigcup_{\iota \in I} M_\iota$ in Theorem 2.2. For any $\varepsilon > 0$, there exist $n_0 \in \mathbf{N}_+$ and an N-valued step function x_0 on (n_0^{-1}, ∞) such that

$$S(\varphi, \psi) - \varepsilon \leqslant \psi(1)\log n_0 - \int_{n_0^{-1}}^{\infty} \{\psi((1 - x_0(t))^*(1 - x_0(t))) +$$

$$+ t^{-1}\varphi(x_0(t)x_0(t)^*)\}t^{-1}dt.$$

Since x_0 takes only finitely many values in $\bigcup_{\iota \in I} M_\iota$ and $\{M_\iota\}_{\iota \in I}$ is increasing, all the values of x_0 occur in some M_{ι_0}. Thus we have

$$S(\varphi, \psi) - \varepsilon \leqslant S(\varphi|M_{\iota_0}, \psi|M_{\iota_0}).$$

Q.E.D.

Now we make some comments on the C^*-algebra case. So let us assume that $\mathfrak{A}$ is a unital C^*-algebra and that φ, ψ are elements in $\mathfrak{A}_*^+$. Let $M = \mathfrak{A}^{**}$ be the universal enveloping von Neumann algebra of $\mathfrak{A}$ and $\tilde{\varphi}, \tilde{\psi} \in M_*^+$ be the normal extensions of φ, ψ. Then $S(\varphi, \psi)$ is defined as $S(\tilde{\varphi}, \tilde{\psi})$. Let $\mathfrak{B}$ be a (norm) dense subspace (containing 1) of $\mathfrak{A}$. Then $\mathfrak{B}$ is s*-dense in the von Neumann algebra M. Therefore, in the C^*-algebra set-up, Theorem 3.2 is also valid if N is replaced by $\mathfrak{B}$. Consequently, all the properties in Theorem 4.1 remain valid under suitable modifications.

Expressions such as $\|\Delta_{\varphi\psi}^{\theta/2} x\xi_\psi\|^2$, $0 < \theta < 1$, $x \in M$ (which corresponds to $\mathrm{Tr}(x^* h^\theta x k^{1-\theta})$, h, k being positive trace class operators) appear in many contexts (information, Rényi entropy, and Wigner-Yanase-Dyson-Lieb concavity). Since

$$\lambda^\theta = \int_0^\infty \lambda(t+\lambda)^{-1}\pi^{-1}\sin(\theta\pi)t^{\theta-1}\mathrm{d}t,$$

Theorem 2.2 (actually, its easy generalization to $(g(\Delta_{\varphi\psi})x\xi_\psi \mid x\xi_\psi)$) implies

$$\|\Delta_{\varphi\psi}^{\theta/2} x\xi_\psi\|^2 = \inf\int_0^\infty \{\psi((x - x(t))^*(x - x(t))) +$$

$$+ t^{-1}\varphi(x(t)x(t)^*)\}\pi^{-1}\sin(\theta\pi)t^{\theta-1}\mathrm{d}t.$$

In particular, with $\theta = 1/2$, $x = 1$, we get

$$(\sqrt{\varphi} \mid \sqrt{\psi}) (= (\xi_\varphi \mid \xi_\varphi)) =$$

$$= \inf\int_0^\infty \{\psi((1 - x(t))^*(1 - x(t))) + t^{-1}\varphi(x(t)x(t)^*)\}\pi^{-1}t^{-1/2}\mathrm{d}t.$$

Acknowledgement. Summary of the paper was circulated as "Variational Expressions of Relative Entropy of States on W^*-algebras (research announcement)". The present detailed version was written while the author stayed at MSRI, and he is grateful to MSRI for generous support.

This research is partially supported by NSF.

REFERENCES

1. Ando, T.; Kubo, F., Mean of positive linear operators, *Math. Ann.*, 246(1980), 205–224.
2. Araki, H., Some properties of modular conjugation operator of von Neumann algebras and a non-commutative Radon-Nikodym theorem with a chain rule, *Pacific J. Math.*, 50(1974), 309–354.

3. ARAKI, H., Golden-Thompson and Peierls-Bogolubov inequalities for a general von Neumann algebra, *Comm. Math. Phys.*, **34**(1973), 167--178.
4. ARAKI, H., Relative entropy of states on von Neumann algebras. I, II, *Publ. Res. Inst. Math. Sci.*, **11**(1976), 809--833, and **13**(1977), 173--192.
5. CONNES, A., Entropie de Kolmogoroff-Sinai et mécanique statistique quantique, *C. R. Acad. Sci. Paris*, **301**(1985), 1--6.
6. DONOGHUE JR., W., *Monotone matrix functions and analytic continuation*, Springer-Verlag, 1974.
7. HAAGERUP, U., The standard form of von Neumann algebras, *Math. Scand.*, **37**(1975), 271--282.
8. KOSAKI, H., Interpolation theory and the Wigner-Yanase-Dyson-Lieb concavity, *Comm. Math. Phys.*, **87**(1982), 315--329.
9. PETZ, D., Quasi-entropies for states of a von Neumann algebra, *Publ. Res. Inst. Math. Sci.*, **21**(1985), 787--800.
10. PUSZ, W.; WORONOWICZ, S., Functional calculus for sesquilinear forms and the purification map, *Rep. Math. Phys.*, **5**(1975), 159--170.
11. UHLMANN, A., Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, *Comm. Math. Phys.*, **54**(1977), 21--32.
12. UMEGAKI, H., Conditional expectation in an operator algebra. III, *Kodai Math. Sem. Rep.*, **11**(1959), 51--64.
13. WEHRL, A., General properties of entropy, *Rev. Modern Phys.*, **50**(1978), 221--260.

HIDEKI KOSAKI
Mathematical Sciences Research Institute,
Berkeley, CA 94720,
U.S.A.

Permanent address:
Department of Mathematics,
College of General Education,
Kyushu University,
Fukuoka 810,
Japan.

Received February 4, 1986.