K. YANAGI
KODAI MATH. J.
5 (1982), 434—445

# COVARIANCE OPERATORS AND VON NEUMANN'S THEORY OF MEASUREMENTS

Dedicated to Prof. Kentaro Murata on his sixtieth birthday

BY KENJIRO YANAGI

## 1. Introduction.

The Gaussian channels are defined in the following; Let $H_1, H_2$ be a pair of real separable Hilbert spaces and let $\mathfrak{B}_1, \mathfrak{B}_2$ be Borel fields of $H_1, H_2$, respectively. For this, let $\nu(\cdot, \cdot)$ be a real valued function defined on $H_1 \times \mathfrak{B}_2$ such as

(1) for each $x \in H_1$, $\nu(x, \cdot) \equiv \nu_x$ is a Gaussian measure on $\mathfrak{B}_2$ with mean vector $m_x \in H_2$ and covariance operator $\rho_x$ on $H_2$,

(2) for each $B \in \mathfrak{B}_2$, $\nu(\cdot, B)$ is a measurable function on $H_1$.

Then the triple $[H_1, \nu, H_2]$ is said to be a Gaussian channel. In this paper, we consider the Gaussian channels constructed by the covariance operator $\rho_x$ which is constant or not constant with respect to $x$ and obtain the average mutual information of the compound source.

In particular in the case of the covariance operator $\rho_x$ which is not constant with respect to $x$, we can give von Neumann's theory of measurements as the models. In 1962, Nakamura-Umegaki proved that the statistical development $\rho_1 \rightarrow \rho_2$ by the measurements is nothing but the conditional expectation in the sense of Umegaki and developed the theory of noncommutative integration. In this paper, restricting the case of real separable Hilbert spaces, we tryto obtain the average mutual information of the statistical development by identifying density operators with covariance operators of probability measures on real separable Hilbert spaces. And in the last section we define the relative entropy among density operators and study the properties of them. We remark that the properties of our defined relative entropy are similar to the properties of von Neumann's relative entropy in some sense. Though we use almost known results relative to probability measures on Hilbert spaces, the obtained results are based on the essence of the theory of noncommutative integration with respect to the special von Neumann algebra $L(H)$.

The author would like to express his hearty thanks to Professor Hisaharu Umegaki for his valuable suggestions and encouragement in the course of preparing this paper.

Received October 21, 1981

## 2. Gaussian measures on Hilbert spaces.

In this section we shall describe about several useful results known relative to Gaussian measures on Hilbert spaces. Let $H$ be a real separable Hilbert space with inner product $\langle\cdot, \cdot\rangle$ and associated norm $\|\cdot\|$ and $\mathfrak{B}$ be the Borel field of $H$. A Borel probability measure $\mu$ on $\mathfrak{B}$ that satisfies

$$\int_H \|x\|^2 d\mu(x) < \infty$$

defines a vector $m$ of $H$ and an operator $R$ such that

$$\langle m, x\rangle = \int_H \langle y, x\rangle d\mu(y)$$

and

$$\langle Rx, y\rangle = \int_H \langle z-m, x\rangle\langle z-m, y\rangle d\mu(z).$$

The $m$ is said to be mean vector of the measure $\mu$. The operator $R$ is

(*) "linear, bounded, nonnegative, self-adjoint and of trace-class of $H$",

and we know

$$\operatorname{trace}(R) = \int_H \|x-m\|^2 d\mu(x).$$

In general, we call operators having the property (*) to be covariance operators. If $\mu$ is a Gaussian, then its characteristic functional $FT(\mu)$ is given by

$$FT(\mu)(x) = \exp\{i\langle m, x\rangle - \langle Rx, x\rangle/2\},$$

where $m$ is the mean vector of $\mu$ and $R$ the covariance operator of $\mu$. Conversely, if $m \in H$ and $R$ is a covariance operator, then $\exp\{i\langle m, x\rangle - \langle Rx, x\rangle/2\}$ is the characteristic functional of a Gaussian measure on $H$. For convenience, we use the notation $\mu = N(m, R)$ to denote that $\mu$ is a Gaussian measure on $H$ with mean vector $m$ and covariance operator $R$, and $\mu_1 \ll \mu_2$, $\mu_1 \sim \mu_2$ and $\mu_1 \perp \mu_2$ to denote that $\mu_1$ is absolutely continuous with respect to $\mu_2$, $\mu_1$ and $\mu_2$ are equivalent and $\mu_1$ and $\mu_2$ are orthogonal, respectively. Also we use the notations $(\sigma c)$ and $(\tau c)$ to denote that the space of all Hilbert-Schmidt operators and the space of all trace-class operators.

PROPOSITION 2.1 (Rao-Varadarajan [6]). *If $\mu_1 = N(m_1, R_1)$ and $\mu_2 = N(m_2, R_2)$, then $\mu_1 \sim \mu_2$ or $\mu_1 \perp \mu_2$. Also $\mu_1 \sim \mu_2$ if and only if*

(N1) $$m_1 - m_2 \in \operatorname{range}(R_1^{1/2}) = \operatorname{range}(R_2^{1/2}) \text{ and}$$

(N2) $$R_1 = R_2^{1/2}(I+T)R_2^{1/2},$$

*where $T \in (\sigma c)$ and $T$ is zero on null $(R_2)$.*

The following proposition is a small modification of Skorohod [8], where the cases of $m_1=m_2$ or $R_1=R_2$ are stated. The proof is ommited.

PROPOSITION 2.2. *Let* $\mu_1=N(m_1, R_1)$ *and* $\mu_2=N(m_2, R_2)$. *If* $\mu_1\sim\mu_2$, *then*

$$\frac{d\mu_1}{d\mu_2}(x)=\exp\left\{\frac{1}{2}\sum_{k,j}\lambda_k^{-1/2}\lambda_j^{-1/2}\langle T(I+T)^{-1}e_k, e_j\rangle\langle x-m_1, e_k\rangle\langle x-m_1, e_j\rangle\right.$$

$$\left.-\frac{1}{2}\sum_k \log(1+t_k)+\sum_k \lambda_k^{-1}\langle x-m_2, e_k\rangle\langle m_1-m_2, e_k\rangle-\frac{1}{2}\sum_k \lambda_k^{-1}\langle m_1-m_2, e_k\rangle^2\right\},$$

*where* $\{\lambda_k\}$ *are nonzero eigenvalues of* $R_2$, $\{e_k\}$ *are corresponding orthonormal eigenvalues of* $R_2$ *and* $\{t_k\}$ *are eigenvalues of* $T$. *Also we obtain*

$$\int_H \log\frac{d\mu_1}{d\mu_2}(x)d\mu_1(x)=\frac{1}{2}\sum_n\{t_n-\log(1+t_n)\}+\frac{1}{2}\sum_n \lambda_n^{-1}\langle m_1-m_2, e_n\rangle^2.$$

Let $H_1$, $H_2$ be real separable Hilbert spaces with inner products $\langle\cdot,\cdot\rangle_1$, $\langle\cdot,\cdot\rangle_2$ and associated norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\mathfrak{B}_1=\mathfrak{B}(H_1)$, $\mathfrak{B}_2=\mathfrak{B}(H_2)$ the Borel fields of $H_1$, $H_2$, respectively. Denote $H_1\times H_2$ the real separable Hilbert space under the inner product $[(u, v), (x, y)]=\langle u, x\rangle_1+\langle v, y\rangle_2$ and associated norm $|||(x, y)|||^2=[(x, y), (x, y)]$. Moreover, the norm-open sets obtained by this inner product generate the Borel field $\mathfrak{B}_1\times\mathfrak{B}_2=\mathfrak{B}(H_1\times H_2)$. Let $\mu_1$, $\mu_2$ be Borel probability measures on $\mathfrak{B}_1$, $\mathfrak{B}_2$ and $\mu_{12}$ be a joint probability measure on $\mathfrak{B}_1\times\mathfrak{B}_2$ such that $\mu_{12}$ has $\mu_1$, $\mu_2$ as projections on $H_1$, $H_2$, respectively. When $\mu_1\otimes\mu_2$ is the usual product measure on $\mathfrak{B}_1\times\mathfrak{B}_2$ of $\mu_1$ and $\mu_2$, the average mutual information $I(\mu_{12})$ of the measure $\mu_{12}$ with respect to $\mu_1\otimes\mu_2$ is defined as follows: If $\mu_{12}\ll\mu_1\otimes\mu_2$,

$$I(\mu_{12})=\int_{H_1\times H_2}\log\frac{d\mu_{12}}{d\mu_1\otimes\mu_2}(x, y)d\mu_{12}(x, y),$$

and otherwise, $I(\mu_{12})=\infty$.

## 3. Gaussian channels.

We shall define the Gaussian channels in the following.

DEFINITION. Let $H_1$, $H_2$ be a pair of real separable Hilbert spaces with inner products $\langle\cdot,\cdot\rangle_1$, $\langle\cdot,\cdot\rangle_2$ and associated norms $\|\cdot\|_1$, $\|\cdot\|_2$, and let $\mathfrak{B}_1$, $\mathfrak{B}_2$ be Borel fields of $H_1$, $H_2$, respectively. Let $\nu(\cdot,\cdot)$ be a real valued function defined on $H_1\times\mathfrak{B}_2$ such as

(C1) for each $x\in H_1$, $\nu(x,\cdot)\equiv\nu_x$ is a Gaussian measure on $\mathfrak{B}_2$ with a mean vector $m_x\in H_2$ and a covariance operator $\rho_x$ on $H_2$

(C2) for each $B\in\mathfrak{B}_2$, $\nu(\cdot, B)$ is a measurable function. Then the triple $[H_1, \nu, H_2]$ is said to be a Gaussian channel.

An output source $\mu_2$ derived from an input source $\mu_1$ and a channel distribution $\nu$ is defined by

$$\mu_2(B)=\int_{H_1}\nu_x(B)d\mu_1(x), \qquad B\in\mathfrak{B}_2.$$

A compound source $\mu_{12}$ derived from an input source $\mu_1$ and a channel distribution $\nu$ is defined by

$$\mu_{12}(C)=\int_{H_1}\nu_x(C_x)d\mu_1(x), \qquad C\in\mathfrak{B}_1\times\mathfrak{B}_2,$$

where $C_x=\{y\in H_2: (x, y)\in C\}$. Then

$$\mu_1\otimes\mu_2(C)=\int_{H_1}\mu_2(C_x)d\mu_1(x), \qquad C\in\mathfrak{B}_1\times\mathfrak{B}_2.$$

If there exist a probability measure $\mu$ on $(H_2, \mathfrak{B}_2)$ such that

$$\nu_x \sim \mu \text{ a.e. } d\mu_1(x), \tag{3.1}$$

then $\mu_2\sim\mu$ and $\mu_{12}\sim\mu_1\otimes\mu_2$.
And the output source $\mu_2$ has the following mean vector $m_2$ and covariance operator $\rho_2$:

$$\langle m_2, x\rangle_2=\int_{H_1}\langle m_y, x\rangle_2 d\mu_1(y) \tag{3.2}$$

and

$$\langle \rho_2 x, y\rangle_2=\int_{H_1}\{\langle \rho_z x, y\rangle_2+\langle m_z-m_2, x\rangle_2\langle m_z-m_2, y\rangle_2\}\, d\mu_1(z). \tag{3.3}$$

If the condition (3.1) is satisfied, then

$$I(\mu_{12})=\int_{H_1}\int_{H_2}\log\frac{d\nu_x}{d\mu}(y)d\nu_x(y)d\mu_1(x)-\int_{H_2}\log\frac{d\mu_2}{d\mu}(y)d\mu_2(y). \tag{3.4}$$

For simplicity, we assume that $H_1=H_2=H$. The models of concrete Gaussian channels constructed by $\nu_x=N(x, \rho)$ with the covariance operator $\rho$ which is constant with respect to $x$ is obtained by Baker ([1, 2]). When the input source $\mu_1$ is restricted to be Gaussian, the followings are equivalent by setting $\mu_1=N(0, \rho_1)$ and $\mu=N(0, \rho)$:

(A1) $\nu_x\sim\mu$ a.e. $d\mu_1(x)$,

(A2) $\mu_1[\text{range}\,(\rho^{1/2})]=1$,

(A3) $\rho_1=\rho^{1/2}T\rho^{1/2}$, where $T\in(\tau c)$ and $T$ is zero on null $(\rho)$.

Let $\{t_n\}$ be eigenvalues of $T$ in the above condition (A3). Then by (3.2) and (3.3),

$$\rho_2=\rho+\rho_1=\rho^{1/2}(I+T)\rho^{1/2}.$$

By Proposition 2.2, we have

$$\int_H \int_H \log \frac{d\nu_x}{d\mu}(y) d\nu_x(y) d\mu_1(x) = \frac{1}{2} \operatorname{trace}(T) = \frac{1}{2} \sum_n t_n$$

and

$$\int_H \log \frac{d\mu_2}{d\mu}(y) d\mu_2(y) = \frac{1}{2} \sum_n \{t_n - \log(1+t_n)\}.$$

Consequently by (3.4), the following proposition is obtained.

PROPOSITION 3.1.

$$I(\mu_{12}) = \frac{1}{2} \sum_n \log(1+t_n).$$

In the following section, we shall give the model of Gaussian channels constructed by $\nu_x = N(0, \rho_x)$ with the covariance operator $\rho_x$ which is not constant with respect to $x$.

## 4. Von Neumann's theory of measurements.

To avoid the complication, let us assume that $A$, an observable corresponding to a physical quantity, is bounded and has a pure discrete simple spectrum. Let $\phi_1, \phi_2, \cdots$ be the complete orthonormal basis corresponding to the proper values $\lambda_1, \lambda_2, \cdots$ of $A$, respectively. Von Neumann [4] observed that $A$ has the value $\lambda_n$ in the fraction $\langle \rho_1 \phi_n, \phi_n \rangle$ after the measurement under density operator $\rho_1$, and that we obtain a mixture with the density operator

$$\rho_2 = \sum_{n=1}^{\infty} \langle \rho_1 \phi_n, \phi_n \rangle \phi_n \odot \phi_n \tag{4.1}$$

after the measurement, where $(\phi_n \odot \phi_n)x = \langle x, \phi_n \rangle \phi_n$ is similar with the notation of Shatten [7]. This change, given by the process

(M1) $\rho_1 \longrightarrow \rho_2$,

is the statistical development of a state by measurement, and it deffers essentially from the classical development of a state given by the process

(M2) $\rho_1 \longrightarrow \rho_t = \exp\{-(2\pi i)/ntH\} \rho_1 \exp\{(2\pi i/n)tH\}$,

where $H$ is the Hamiltonean. Just as in classical mechanics, von Neumann [4] taught us, process (M2) does not reproduce the most important property of the real world, namely its irreversibility, the fundamental difference between the time direction, "future" and "past", where (M1) is certainly not prima facie reversible. The present section is interested in the process (M1).

In 1962, Nakamura-Umegaki [3] proved that the statistical development $\rho_1 \to \rho_2$ by the measurements (with the operator $A$) is nothing but the conditional expectation $E[\rho_1 | \mathfrak{A}]$ in the sense of Umegaki [9, 10], where $\mathfrak{A}$ is the von

Neumann algebra generated by $A$.

On the other hand, statistical development $\rho_1 \to \rho_2$ by the measurement $A$ can be regarded by the Gaussian channel constructed by $\nu_x = N(0, \rho_x)$ with the covariance operator $\rho_x$ which is not constant with respect to $x$. Let us assume that input space $H_1$ and output space $H_2$ are the same real separable Hilbert space $H$. It seems that the essense of von Neumann's theory of measurements in quantum statistics is considered in the following theorem.

THEOREM 4.1. *Suppose that the ınput source* $\mu_1 = N(0, \rho_1)$ *and the channel ditribution* $\nu_x = N(0, \rho_x)$, *where*

$$\rho_x = \sum_n f_n(x) \phi_n \odot \phi_n$$

*and* $\{f_n(x)\}$ *satısfy the followıng conditions:*

(1) $\int_H f_n(x) d\mu_1(x) = \langle \rho_1 \phi_n, \phi_n \rangle$,

(2) $\mu_1\{x \in H : \nu_x \sim \nu_y\} = 1$ *for some* $y \in H$.

*Then the followings are satisfied:*

(3) *the output source* $\mu_2 = N(0, \rho_2)$, *where* $\rho_2$ *ıs given by* (4.1),

(4) $\mathrm{trace}\,(\rho_1 A) = \mathrm{trace}\,(\rho_2 A)$,

(5) $I(\mu_{12}) = \dfrac{1}{2} \sum_n \left\{ \log \langle \rho_1 \phi_n, \phi_n \rangle - \int_H \log f_n(x) d\mu_1(x) \right\}$.

*Proof.* By (2), $\nu_x \sim \nu_y$ a.e. $d\mu_1(x)$. In (3.3),

$$\begin{aligned}
\langle \rho_2 x, y \rangle &= \int_H \langle \rho_z x, y \rangle d\mu_1(z) \\
&= \sum_n \langle x, \phi_n \rangle \langle y, \phi_n \rangle \int_H f_n(z) d\mu_1(z) \\
&= \sum_n \langle x, \phi_n \rangle \langle y, \phi_n \rangle \langle \rho_1 \phi_n, \phi_n \rangle \qquad \text{(by (1))} \\
&= \langle \sum_n \langle \rho_1 \phi_n, \phi_n \rangle (\phi_n \odot \phi_n) x, y \rangle .
\end{aligned}$$

Then $\rho_2$ is given by (4.1). It is easy to obtain (4). Because

$$\begin{aligned}
\mathrm{trace}\,(\rho_2 A) &= \sum_n \lambda_n \,\mathrm{trace}\,(\rho_2(\phi_n \odot \phi_n)) \\
&= \sum_n \lambda_n \int \langle (\phi_n \odot \phi_n) y, y \rangle d\mu_2(y) \\
&= \sum_n \lambda_n \iint \langle y, \phi_n \rangle \langle y, \phi_n \rangle d\nu_x(y) d\mu_1(x)
\end{aligned}$$

$$=\sum_n \lambda_n \int \langle \rho_x \phi_n, \phi_n \rangle d\mu_1(x)$$

$$=\sum_n \lambda_n \int f_n(x) d\mu_1(x)$$

$$=\sum_n \lambda_n \langle \rho_1 \phi_n, \phi_n \rangle$$

$$=\sum_n \langle \rho_1 A \phi_n, \phi_n \rangle$$

$$=\operatorname{trace}(\rho_1 A).$$

When we set $\mu=\nu_y$ in (3.4), we have by Proposition 2.2,

$$I(\mu_{12})=\iint \log \frac{d\nu_x}{d\nu_y}(z)d\nu_x(z)d\mu_1(x)-\int \log \frac{d\mu_2}{d\nu_y}(z)d\mu_2(z)$$

$$=\int \frac{1}{2}\sum_n \left\{\frac{f_n(x)}{f_n(y)}-1-\log \frac{f_n(x)}{f_n(y)}\right\}d\mu_1(x)$$

$$-\frac{1}{2}\sum_n \left\{\frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}-1-\log \frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}\right\}$$

$$=\frac{1}{2}\sum_n \left\{\frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}-1\right\}-\frac{1}{2}\sum_n \int \log \frac{f_n(x)}{f_n(y)}d\mu_1(x)$$

$$-\frac{1}{2}\sum_n \left\{\frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}-1-\log \frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}\right\}$$

$$=\frac{1}{2}\sum_n \left\{\log \frac{\langle \rho_1 \phi_n, \phi_n \rangle}{f_n(y)}-\int \log \frac{f_n(x)}{f_n(y)}d\mu_1(x)\right\}$$

$$=\frac{1}{2}\sum_n \left\{\log < \rho_1 \phi_n, \phi_n \rangle-\int \log f_n(x) d\mu_1(x)\right\}.$$

Hence (5) is given. Q. E. D.

*Remark* 4.2. In Theorem 4.1, when $f_n(x)=\langle \rho_1 \phi_n, \phi_n \rangle$, the conditions (1) and (2) are satisfied. Then this is the simplest example. But when $f_n(x)=\langle x, \phi_n \rangle^2$, the condition (1) is satisfied but (2) is not.

## 5. Relative entropy among density operators.

In this section we shall define the relative entropy among density operators or covariance operators and study their properties. In [11], Umegaki defined a measure of Kullback-Leibler's information in a von Neumann algebra and studied the several important properties. We shall define the relative entropy in the different way. Suppose that $\rho_1$, $\rho_2$ are density operators. Let $\mu_1=N(0, \rho_1)$ and

$\mu_2=N(0, \rho_2)$. Then the equivalence of $\rho_1$ and $\rho_2$, denoted by $\rho_1\sim\rho_2$, is defined $\mu_1\sim\mu_2$. And the relative entropy $I(\rho_1|\rho_2)$ of $\rho_1$ relative to $\rho_2$ is defined in the following; If $\rho_1\sim\rho_2$, then

$$I(\rho_1|\rho_2)=\int\log\frac{d\mu_1}{d\mu_2}(x)d\mu_1(x) \quad \text{and otherwise } I(\rho_1|\rho_2)=\infty\,.$$

It is clear that $\rho_1=\rho_2$ if and only if $I(\rho_1|\rho_2)=0$. More generally, we can define the relative entropy among covariance operators in the same way. In order to give the quantum mechanical expression, we shall obtain the following relations about the relative entropy among density operators.

THEOREM 5.1. *Suppose that $\rho_1$, $\rho_2$, $\rho$, $\eta_1$, $\eta_2$ are all density operators. Then the followings hold:*

(1) *If $\rho_1$, $\rho_2$, $\rho$ are commutative, then*

$$I(\alpha\rho_1+\beta\rho_2|\rho)\leqq\alpha I(\rho_1|\rho)+\beta I(\rho_2|\rho)\,,$$

*where $\alpha+\beta=1$, $\alpha$, $\beta\geqq 0$.*

(2) $$I\Big(\frac{1}{2}(\rho_1\otimes\rho_2)\Big|\frac{1}{2}(\eta_1\otimes\eta_1)\Big)=I(\rho_1|\eta_1)+I(\rho_2|\eta_2)\,,$$

*where $\rho_1\otimes\rho_2$, $\eta_1\otimes\eta_2$ are the covariance operators of $\mu_1\otimes\mu_2$, $\nu_1\otimes\nu_2$, respectively, and $\mu_i=N(0, \rho_i)$, $\nu_i=N(0, \eta_i)$ for $i=1, 2$.*

*Proof.* (1): We may assume $\rho_1\sim\rho_2\sim\rho$. By commutativity, the following spectral decompositions hold;

$$\rho_1=\sum_n a_n\phi_n\odot\phi_n\,,\qquad \rho_2=\sum_n b_n\phi_n\odot\phi_n \quad\text{and}\quad \rho=\sum_n c_n\phi_n\odot\phi_n\,.$$

Since $\rho_1\sim\rho$ and $\rho_2\sim\rho$, we obtain

$$\sum_n\Big(\frac{a_n}{c_n}-1\Big)^2<\infty \quad\text{and}\quad \sum_n\Big(\frac{b_n}{c_n}-1\Big)^2<\infty\,.$$

Then eigenvalues of the operators corresponding to $T$ in Proposition 2.2 are $a_n/c_n-1$ and $b_n/c_n-1$, respectively. Hence we can apply Proposition 2.2 and we have

$$I(\rho_1|\rho)=\frac{1}{2}\sum_n\Big\{\frac{a_n}{c_n}-1-\log\frac{a_n}{c_n}\Big\}$$

and similarly

$$I(\rho_2|\rho)=\frac{1}{2}\sum_n\Big\{\frac{b_n}{c_n}-1-\log\frac{b_n}{c_n}\Big\}\,.$$

And since

$$\alpha\rho_1+\beta\rho_2=\sum_n(\alpha a_n+\beta b_n)\phi_n\odot\phi_n$$

where

$$\sum_n \left(\frac{\alpha a_n+\beta b_n}{c_n}-1\right)^2<\infty\,,$$

we have $\alpha\rho_1+\beta\rho_2\sim\rho$, and so

$$I(\alpha\rho_1+\beta\rho_2|\rho)=\frac{1}{2}\sum_n\left\{\frac{\alpha a_n+\beta b_n}{c_n}-1-\log\frac{\alpha a_n+\beta b_n}{c_n}\right\}.$$

Consequently

$$I(\alpha\rho_1+\beta\rho_2|\rho)\leqq\alpha I(\rho_1|\rho)+\beta I(\rho_2|\rho)\,,$$

because $f(x)=x-1-\log(x)$ is convex in $(0, \infty)$.

(2): We remark that $\rho_1\otimes\rho_2\sim\eta_1\otimes\eta_2$ if and only if $\rho_1\sim\eta_1$ and $\rho_2\sim\eta_2$. By Proposition 2.1, we have $\rho_1=\eta_1^{1/2}(I+T_1)\eta_1^{1/2}$, where $T_1\in(\sigma c)$, $T_1$ is zero on null $(\eta_1)$ and range $(\rho_1^{1/2})=$range$(\eta_1^{1/2})$, and $\rho_2=\eta_2^{1/2}(I+T_2)\eta_2^{1/2}$, where $T_2\in(\sigma c)$, $T_2$ is zero on null$(\eta_2)$ and range$(\rho_2^{1/2})=$range$(\eta_2^{1/2})$. And also $\rho_1\otimes\rho_2=(\eta_1\otimes\eta_2)^{1/2}\cdot(\mathfrak{J}+\mathfrak{T})(\eta_1\otimes\eta_2)^{1/2}$, where $\mathfrak{J}$ is identity operator on $H\times H$ and $\mathfrak{T}=T_1\otimes T_2$. Here $T_1\otimes T_2(x, y)=(T_1x, T_2y)$. Let $T_1=\sum_n t_n\phi_n\odot\phi_n$ and $T_2=\sum_m s_m\psi_m\odot\psi_m$. Since

$$\mathfrak{T}(\phi_n, 0)=(T_1\phi_n, T_2 0)=(t_n\phi_n, 0)=(t_n\phi_n, t_n 0)=t_n(\phi_n, 0)$$

and

$$\mathfrak{T}(0, \psi_m)=(T_1 0, T_2\psi_m)=(0, s_m\psi_m)=(s_m 0, s_m\psi_m)=s_m(0, \psi_m)\,,$$

$\{t_n: n=1, 2, \cdots\}\cup\{s_m: m=1, 2, \cdots\}$ are all the eigenvalues of $\mathfrak{T}$. Consequently,

$$\begin{aligned}I(\rho_1\otimes\rho_2|\eta_1\otimes\eta_2)&=\frac{1}{2}\sum_n\{t_n-\log(1+t_n)\}+\frac{1}{2}\sum_m\{s_m-\log(1+s_m)\}\\&=I(\rho_1|\eta_1)+I(\rho_2|\eta_2)\,.\end{aligned}$$

Since

$$\begin{aligned}\mathrm{trace}\,(\rho_1\otimes\rho_2)&=\int_{H\times H}|||(x, y)|||^2 d\mu_1\otimes\mu_2(x, y)\\&=\int_{H\times H}(\|x\|^2+\|y\|^2)d\mu_1\otimes\mu_2(x, y)\\&=\int_H\|x\|^2 d\mu_1(x)+\int_H\|y\|^2 d\mu_2(y)\\&=\mathrm{trace}\,(\rho_1)+\mathrm{trace}\,(\rho_2)\\&=2\end{aligned}$$

and similarly trace$(\eta_1\otimes\eta_2)=2$, $(1/2)(\rho_1\otimes\rho_2)$ and $(1/2)(\eta_1\otimes\eta_2)$ are density operators. On the other hand,

$$I(\rho_1\otimes\rho_2|\eta_1\otimes\eta_2)=I\Big(\frac{1}{2}(\rho_1\otimes\rho_2)\Big|\frac{1}{2}(\eta_1\otimes\eta_2)\Big)^{*)}.$$

*) More general case is obtained in the following theorem

Thus we obtain

$$I\Big(\frac{1}{2}(\rho_1\otimes\rho_2)\Big|\frac{1}{2}(\eta_1\otimes\eta_2)\Big)=I(\rho_1|\eta_1)+I(\rho_2|\eta_2)\,.$$

Q. E. D.

Finally, we shall obtain the following relations about the relative entropy among general covariance operators.

THEOREM 5.2. *Suppose that $\rho_1$, $\rho_2$, $\rho_3$ are all covariance operators. Then the followings hold:*

(1) *If $\rho_1\geqq\rho_2$, then $I(\rho_1|\rho_2)\geqq I(\rho_2|\rho_1)$.*

(2) *If $k>0$, then $I(k\rho_1|k\rho_2)=I(\rho_1|\rho_2)$.*

(3) *If $\rho_1$, $\rho_2$, $\rho_3$ are commutative, $\rho_1\geqq\rho_2\geqq\rho_3$ or $\rho_1\leqq\rho_2\leqq\rho_3$ and $\rho_1\sim\rho_2$ or $\rho_2\sim\rho_3$, then*

$$I(\rho_1|\rho_3)\geqq I(\rho_1|\rho_2)+I(\rho_2|\rho_3)\,.$$

*Proof.* (1): It is clear in the case of $\rho_1\not\sim\rho_2$. If $\rho_1\sim\rho_2$, then we have $\rho_1=\rho_2^{1/2}(I+T)\rho_2^{1/2}$, where $T\in(\sigma c)$, $T$ is zero on null$(\rho_2)$ and range$(\rho_1^{1/2})=$range$(\rho_2^{1/2})$. We can obtain

$$\rho_1^{1/2}=\rho_2^{1/2}(I+T)^{1/2}W^*\,, \tag{5.1}$$

where $W$ is a partial isometry. On the other hand, we also have $\rho_2=\rho_1^{1/2}(I+S)\rho_1^{1/2}$, where $S\in(\sigma c)$, $S$ is zero on null$(\rho_1)$ and range$(\rho_2^{1/2})=$range$(\rho_1^{1/2})$. By (5.1),

$$\rho_2=\rho_2^{1/2}(I+T)^{1/2}W^*(I+S)W(I+T)^{1/2}\rho_2^{1/2}\,.$$

For simplicity we can and may assume that $\overline{\text{range}(\rho_1)}=\overline{\text{range}(\rho_2)}=H$. Hence $I=(I+T)^{1/2}W^*(I+S)W(I+T)^{1/2}$ and so $I+S=W(I+T)^{-1}W^*$. Let $\{s_n\}$, $\{t_n\}$ be eigenvalues of $S$, $T$, respectively. Then $1+s_n=(1+t_n)^{-1}$ and $s_n=-t_n/(1+t_n)$. Since $\rho_1\geqq\rho_2$, $t_n\geqq 0$. By Proposition 2.2,

$$I(\rho_1|\rho_2)=\frac{1}{2}\sum_n\{t_n-\log(1+t_n)\}$$

and

$$I(\rho_2|\rho_1)=\frac{1}{2}\sum_n\Big\{-\frac{t_n}{1+t_n}+\log(1+t_n)\Big\}\,.$$

Then we have

$$\begin{aligned}I(\rho_1|\rho_2)-I(\rho_2|\rho_1)&=\frac{1}{2}\sum_n\Big\{t_n+\frac{t_n}{1+t_n}-2\log(1+t_n)\Big\}\\&=\frac{1}{2}\sum_n\{t_n+1-(1+t_n)^{-1}-2\log(1+t_n)\}\,.\end{aligned}$$

Consequently $I(\rho_1|\rho_2)\geqq I(\rho_2|\rho_1)$, because $f(x)=x+1-(x+1)^{-1}-2\log(x+1)\geqq 0$ in $[0,\infty)$.

(2): It is clear by Proposition 2.1 and 2.2.

(3): We may assume that $\rho_1 \sim \rho_2 \sim \rho_3$. By commutativity, we have $\rho_1 = \rho_2(I+T)$, where $T \in (\sigma c)$, $T$ is zero on null $(\rho_2)$ and range $(\rho_1^{1/2})=$ range $(\rho_2^{1/2})$, and $\rho_2=\rho_3(I+S)$, where $S \in (\sigma c)$, $S$ is zero on null $(\rho_3)$ and range $(\rho_2^{1/2})=$ range $(\rho_3^{1/2})$. Then $\rho_1=\rho_3(I+S)(I+T)$. On the other hand, we also obtain $\rho_1=\rho_3(I+U)$, where $U \in (\sigma c)$, $U$ is zero on null $(\rho_3)$ and range $(\rho_1^{1/2})=$ range $(\rho_3^{1/2})$. For simplicity, we can and may assume that $\overline{\text{range}\,(\rho_1)}=\overline{\text{range}\,(\rho_2)}=\overline{\text{range}\,(\rho_3)}=H$. Hence $U=(I+S)(I+T)-I$. Let $\{s_n\}$, $\{t_n\}$ be eigenvalues of $S$, $T$, respectively. Then $\{(1+s_n)(1+t_n)-1\}$ are eigenvalues of $U$. By Proposition 2.2,

$$I(\rho_1 | \rho_3)=\frac{1}{2} \sum_n \{(1+s_n)(1+t_n)-1-\log (1+s_n)(1+t_n)\},$$

$$I(\rho_1 | \rho_2)=\frac{1}{2} \sum_n \{t_n-\log (1+t_n)\}$$

and

$$I(\rho_2 | \rho_3)=\frac{1}{2} \sum_n \{s_n-\log (1+s_n)\}.$$

Hence $I(\rho_1|\rho_3)-I(\rho_1|\rho_2)-I(\rho_2|\rho_3)=(1/2)\sum_n s_n t_n \geqq 0$, because $s_n \geqq 0$, $t_n \geqq 0$ or $s_n \leqq 0$, $t_n \leqq 0$. Consequently

$$I(\rho_1 | \rho_3) \geqq I(\rho_1 | \rho_2)+I(\rho_2 | \rho_3).$$

Q. E. D.

## REFERENCES


[1] C.R. BAKER, Absolute continuity and applications to information theory, "Probability in Banach spaces" in Lecture Notes in Math. No. **526**, Springer, Pub., New York, 1976.

[2] C.R. BAKER, Capacity of the Gaussian channel without feedback, Information and Control, **37** (1978), 70-89.

[3] M. NAKAMURA AND H. UMEGAKI, On von Neumann's theory of measurements in quantum statistics, Math. Jap., **7** (1962), 151-157.

[4] J. VON NEUMANN, Mathematical foundations of quantum mechanics, Princeton Univ. Press, Princeton, 1955.

[5] K.R. PARTHASARATHY, Probability measures on metric spaces, Academic Press, New York, 1967.

[6] C.R. RAO AND V.S. VARADARAJAN, Discrimination of Gaussian processes, Sankhyā Ser. A. **25** (1963), 303-330.

[7] R. SCHATTEN, Norm ideals of completely continuous operators, Springer-Verlag, Berlin, 1960.

[8] R. SKOROHOD, Integration in Hilbert space, Springer-Verlag, Berlin, 1974.

[9] H. UMEGAKI, Conditional expectation in an operator algebra I, Tôhoku Math. J., **6** (1954), 608-612.

[10] H. UMEGAKI, Conditional expectation in an operator algebra II, Tôhoku Math. J., **8** (1956), 86-100.

[11] H. UMEGAKI, Conditional expectation in an operator algebra IV, Kodai Math. Sem. Rep., **14** (1962), 59-85.
[12] H. UMEGAKI, Absolute continuity of information channels, J. Multivariate Anal., **4** (1974), 382-400.
[13] K. YANAGI, On some properties of Gaussian channels, to appear in J. Math. Anal. Appl.
[14] K. YANAGI, Quantum mechanics and Gaussian channels, Physics Letters, **88A** (1982), 13-14.

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
YAMAGUCHI UNIVERSITY
YOSHIDA, YAMAGUCHI 753
JAPAN

**Added in proof.** The author was suggested by Prof. H. Araki that (1) in Theorem 5.1 also holds in the noncommutative case. Indeed, $-\log(1+x)$ is operator-convex in $|x|<1$ by the method of Nakamura-Umegaki's paper "A note on the entropy for operator algebras (Proc. Japan Acad., vol 37, no 3, pp 149-154, 1961)". Since we can reduce the case that absolute values of the eigenvalues of present operator is less than 1, it is possible to show that (1) holds by using the abore assertion.