Izvestiya: *Mathematics* **70**:6 1265–1292

Izvestiya RAN: *Ser. Mat.* **70**:6 193–222
DOI 10.1070/IM2006v070n06ABEH002346

Entropy characteristics of subsets of states. I

M. E. Shirokov

Abstract. We study the properties of quantum entropy and χ-capacity (regarded as a function of sets of quantum states) in the infinite-dimensional case. We obtain conditions for the boundedness and continuity of the restriction of the entropy to a subset of quantum states, as well as conditions for the existence of the state with maximal entropy in certain subsets. The notion of χ-capacity is considered for an arbitrary subset of states. The existence of an optimal average is proved for an arbitrary subset with finite χ-capacity. We obtain a sufficient condition for the existence of an optimal measure and prove a generalized maximal distance property.

§ 1. Introduction

This paper is devoted to the study of the properties of quantum entropy and χ-capacity [1] (regarded as a function of sets of quantum states) in the infinite-dimensional case.

Quantum entropy is a concave lower semicontinuous function with range $[0, +\infty]$ on the set of all quantum states. It has bounded and even continuous restrictions to some non-trivial closed subsets of states [10], [20]. The problem of characterizing such subsets arises in many applications, in particular, in the condition for the existence of an optimal measure for constrained quantum channels [19]. Another interesting question is that of finding conditions for the existence of the state with maximal entropy in a given set of quantum states with bounded entropy. In this paper we consider these and other problems related to quantum entropy.

By the Holevo–Schumacher–Westmoreland theorem [13], [17], the χ-capacity of a set of states determines the maximal rate of transmission of classical information that can be achieved by using this set as an alphabet and applying a non-entangled encoding in the transmitter followed by an entangled measurement-decoding procedure in the receiver. The notion of χ-capacity is usually related to that of quantum channel. But it is easy to see that the χ-capacity of a channel is uniquely determined by its output set. So we may regard χ-capacity as a function of sets of states [14]. This approach is convenient because of its flexibility: we may speak about the χ-capacity of an *arbitrary* set of states, which need not be the output set

[1] This notion is referred to as the *Holevo capacity* in the Western literature and is usually associated with the notion of a quantum channel (see, for example, [17]). In this paper we regard it as a function of sets of quantum states and use the term χ-*capacity*.

This work was done with the support of the programme "Modern problems of theoretical mathematics" of the Russian Academy of Sciences and of the Russian Foundation for Basic Research (grant no. 06-01-00164-a).

AMS 2000 *Mathematics Subject Classification.* 62B10, 81P68, 94A40.

of a channel. From this point of view, χ-capacity is a non-additive function of sets ("non-additive measure"). It possesses many interesting properties, whose detailed investigation seems to be useful in the development of quantum information theory.

In § 3 we consider conditions for the boundedness and continuity of the restriction of the quantum entropy to sets of quantum states as well as conditions for the existence of the state with maximal entropy in such sets (Propositions 1, 3, 4, 6 and Corollaries 1–3). It is shown that the quantum entropy is continuous at some state (with respect to the convergence defined by the relative entropy) if and only if the eigenvalues of this state have a sufficiently large rate of decay (Proposition 2). We consider the relations between some properties of sets of states and the corresponding properties of their so-called "classical projections" (Proposition 5). In particular, the results obtained show that discontinuity and unboundedness of the quantum entropy are of a purely classical nature (see Remark 5).

In § 4 we consider the definition of χ-capacity for any set of quantum states. In § 4.1 we introduce the notion of an optimal average state as a unique state inheriting the most important properties of the average state of an optimal ensemble for a set of states in a finite-dimensional Hilbert space (Theorem 1 and Corollary 4). Properties of the optimal average state enable us to show that every set of finite χ-capacity is relatively compact (Corollary 5) and contained in the maximal set of the same χ-capacity. This observation yields the following result related to quantum channels. If the χ-capacity of an infinite-dimensional channel constrained by some set is finite, then the image of this set under this channel is relatively compact. In particular, every unconstrained channel of finite χ-capacity has a relatively compact output (Corollary 6). Our results on χ-capacity also enable us to make an important observation concerning general properties of quantum entropy (Corollary 7). In § 4.2 we introduce the notion of an optimal measure of a set of states and generalize the "maximal distance property" [14] to the infinite-dimensional case (Proposition 7). This yields a necessary condition for the existence of an optimal measure (Corollary 8). We also obtain a sufficient condition for the existence of an optimal measure (Theorem 2).

§ 2. Preliminaries

Let $\mathcal{H}$ be a separable Hilbert space, $\mathfrak{B}(\mathcal{H})$ the algebra of all bounded operators on $\mathcal{H}$ and $\mathfrak{T}(\mathcal{H})$ the Banach space of all trace-class operators with the trace norm $\|\cdot\|_1$. We will use the term *state* for a positive trace-class operator ρ on $\mathcal{H}$ with trace 1: $\rho \geqslant 0$, $\operatorname{Tr}\rho = 1$. $\mathfrak{B}(\mathcal{H})$ is often called the *algebra of observables of a quantum system*. Every state ρ determines an expectation functional $A \mapsto \operatorname{Tr}\rho A$, $A \in \mathfrak{B}(\mathcal{H})$, which is a normal state in the language of the theory of operator algebras [1]. The set $\mathfrak{S}(\mathcal{H})$ of all states is a closed convex subset of $\mathfrak{T}(\mathcal{H})$ and is a complete separable metric space with metric defined by the trace norm. We note that the convergence of a sequence of states to a *state* in the weak operator topology is equivalent to the convergence of that sequence to this state in the trace norm [2]. We will use the following compactness criterion for sets of states: *a closed set $\mathcal{K}$ of states is compact if and only if for every $\varepsilon > 0$ there is a finite-dimensional projector P_ε such that* $\operatorname{Tr}\rho P_\varepsilon \geqslant 1-\varepsilon$ *for all* $\rho \in \mathcal{K}$ [12], [19].

Let A and B be positive operators in $\mathfrak{T}(\mathcal{H})$. The von Neumann entropy of A and the relative entropy of A and B are defined by the formulae

$$H(A) = -\sum_i \langle i|A\log A|i\rangle, \qquad H(A\,\|\,B) = \sum_i \langle i|A\log A - A\log B + B - A|i\rangle,$$

where $\{|i\rangle\}$ is a basis of eigenvectors of A (see [8], [20]). The entropy (resp. relative entropy) is a concave (resp. convex) lower semicontinuous function of its arguments with range $[0,+\infty]$ [8], [20]. We shall use the inequality

$$H(\rho\,\|\,\sigma) \geqslant \frac{1}{2}\,\|\rho-\sigma\|_1^2, \tag{1}$$

which holds for arbitrary states ρ and σ in $\mathfrak{S}(\mathcal{H})$ [10].

The relative entropy $H(\rho\,\|\,\sigma)$ of states ρ and σ can be regarded as a measure of divergence of these states. (Its classical analogue is called the *Kullback–Leibler distance*.) Although this measure is not a metric (it is not symmetric and does not satisfy the triangle inequality), one can introduce a notion of convergence of a sequence $\{\rho_n\}$ of states to a state ρ_*. This is defined by the condition $\lim_{n\to+\infty} H(\rho_n\,\|\,\rho_*) = 0$. The topology on the set of states associated with this convergence is studied in the classical case in [5], where it is called the *strong information topology*. This type of convergence plays an important role in this paper; it will be called *H-convergence*.

By inequality (1), H-convergence is stronger than the convergence defined by the trace norm.

For an arbitrary set $\mathcal{A}$, let $\operatorname{co}(\mathcal{A})$ and $\overline{\operatorname{co}}(\mathcal{A})$ be its convex hull and convex closure respectively. Let $\operatorname{Ext}(\mathcal{A})$ be the set of all extreme points of $\mathcal{A}$ [6].

When speaking about the continuity of a particular function on some set of states, we mean the continuity of the restriction of this function to this set.

A finite set $\{\rho_i\}$ of states with a corresponding set of probabilities $\{\pi_i\}$ is called a (finite) *ensemble* and is denoted by $\{\pi_i,\rho_i\}$. The state $\bar\rho = \sum_i \pi_i\rho_i$ is called the *average* of this ensemble. In [19], the notion of a *generalized ensemble* is introduced as an arbitrary Borel probability measure μ on $\mathfrak{S}(\mathcal{H})$. The *average* of a generalized ensemble (measure) μ is the state [2] defined by the Bochner integral

$$\bar\rho(\mu) = \int_{\mathfrak{S}(\mathcal{H})} \rho\mu(d\rho).$$

Ordinary ensembles correspond to finitely supported measures.

Given an arbitrary closed subset $\mathcal{A}$ of $\mathfrak{S}(\mathcal{H})$, we denote by $\mathcal{P}(\mathcal{A})$ the set of all probability measures supported by $\mathcal{A}$ [11].

In what follows we regard all ensembles $\{\pi_i,\rho_i\}$ as particular cases of probability measures. In particular, a convex combination of ensembles is defined as the convex combination of the corresponding probability measures.

Consider the functionals

$$\chi(\mu) = \int_{\mathfrak{S}(\mathcal{H})} H\big(\rho\,\|\,\bar\rho(\mu)\big)\mu(d\rho), \qquad \widehat{H}(\mu) = \int_{\mathfrak{S}(\mathcal{H})} H(\rho)\mu(d\rho).$$

[2] This state is also called the *barycentre* of the measure μ.

As shown in [19] (Proposition 1 and proof of the theorem), these functionals are well defined and lower semicontinuous on $\mathcal{P}(\mathfrak{S}(\mathcal{H}))$, and we have

$$\chi(\mu) = H(\bar{\rho}(\mu)) - \widehat{H}(\mu) \tag{2}$$

for every measure μ with $H(\bar{\rho}(\mu)) < +\infty$.

If $\mu = \{\pi_i, \rho_i\}$, then

$$\chi(\{\pi_i, \rho_i\}) = \sum_{i=1}^{n} \pi_i H(\rho_i \,\|\, \bar{\rho}), \qquad \widehat{H}(\{\pi_i, \rho_i\}) = \sum_{i=1}^{n} \pi_i H(\rho_i).$$

We will use Donald's identity [3], [10],

$$\sum_{i=1}^{n} \pi_i H(\rho_i \,\|\, \hat{\rho}) = \sum_{i=1}^{n} \pi_i H(\rho_i \,\|\, \bar{\rho}) + H(\bar{\rho} \,\|\, \hat{\rho}), \tag{3}$$

which holds for any ensemble $\{\pi_i, \rho_i\}$ of n states with average $\bar{\rho}$ and for any state $\hat{\rho}$.

We will also use the generalized integral version of Donald's identity [19],

$$\int_{\mathfrak{S}(\mathcal{H})} H(\rho \,\|\, \hat{\rho}) \mu(d\rho) = \int_{\mathfrak{S}(\mathcal{H})} H(\rho \,\|\, \bar{\rho}(\mu)) \mu(d\rho) + H(\bar{\rho}(\mu) \,\|\, \hat{\rho}), \tag{4}$$

which holds for any probability measure μ with barycentre $\bar{\rho}(\mu)$ and for any state $\hat{\rho}$ and yields the following property of the functional $\chi(\mu)$.

Lemma 1. *Let $\{\mu_k\}_{k=1}^{m}$ be a finite set of probability measures on $\mathfrak{S}(\mathcal{H})$ and let $\{\lambda_k\}_{k=1}^{m}$ be a probability distribution. Then*

$$\chi\left(\sum_{k=1}^{m} \lambda_k \mu_k\right) = \sum_{k=1}^{m} \lambda_k \chi(\mu_k) + \chi(\{\lambda_k, \bar{\rho}(\mu_k)\}_{k=1}^{m}).$$

If $m = 2$, then the following inequality holds for every $\lambda \in [0,1]$:

$$\chi(\lambda\mu_1 + (1-\lambda)\mu_2) \geqslant \lambda\chi(\mu_1) + (1-\lambda)\chi(\mu_2) + \frac{\lambda(1-\lambda)}{2} \left\|\bar{\rho}(\mu_2) - \bar{\rho}(\mu_1)\right\|_1^2.$$

Proof. Let $\mu = \sum_{k=1}^{m} \lambda_k \mu_k$. By definition,

$$\chi(\mu) = \sum_{k=1}^{m} \lambda_k \int_{\mathfrak{S}(\mathcal{H})} H(\rho \,\|\, \bar{\rho}(\mu)) \mu_k(d\rho).$$

Applying (4) to every integral on the right-hand side, we obtain the identity in the lemma.

To prove the inequality for $m = 2$, we estimate the last term in the identity in the lemma by applying inequality (1):

$$\begin{aligned}
&\lambda H(\bar{\rho}(\mu_1) \,\|\, \lambda\bar{\rho}(\mu_1) + (1-\lambda)\bar{\rho}(\mu_2)) + (1-\lambda) H(\bar{\rho}(\mu_2) \,\|\, \lambda\bar{\rho}(\mu_1) + (1-\lambda)\bar{\rho}(\mu_2)) \\
&\qquad \geqslant \frac{1}{2}\,\lambda \left\|(1-\lambda)(\bar{\rho}(\mu_2) - \bar{\rho}(\mu_1))\right\|_1^2 + \frac{1}{2}\,(1-\lambda)\left\|\lambda(\bar{\rho}(\mu_2) - \bar{\rho}(\mu_1))\right\|_1^2 \\
&\qquad = \frac{1}{2}\,\lambda(1-\lambda)\left\|\bar{\rho}(\mu_2) - \bar{\rho}(\mu_1)\right\|_1^2.
\end{aligned}$$

Note that Lemma 1 yields the inequality

$$H\bigl(\lambda\rho_1+(1-\lambda)\rho_2\bigr)\geqslant\lambda H(\rho_1)+(1-\lambda)H(\rho_2)+\frac{\lambda(1-\lambda)}{2}\,\|\rho_2-\rho_1\|_1^2, \qquad (5)$$

which holds for all states ρ_1 and ρ_2. To prove this, it suffices to regard the spectral decompositions of these states as probability measures on $\mathfrak{S}(\mathcal{H})$ and apply Lemma 1.

§ 3. Properties of the quantum entropy

In this section we study the properties of the restrictions of the quantum entropy to sets of quantum states.

Let $\mathcal{A}$ be a set of quantum states with $\sup_{\rho\in\mathcal{A}}H(\rho)<+\infty$. If this supremum is attained at some state, then that state is called the *maximal entropy state of* $\mathcal{A}$, and we denote it by $\Gamma(\mathcal{A})$. Using inequality (5), we make the following simple observation.

Lemma 2. *Let $\mathcal{A}$ be a closed convex set of states with $\sup_{\rho\in\mathcal{A}}H(\rho)<+\infty$ and let $\{\rho_n\}$ be any sequence of states in $\mathcal{A}$ such that*

$$\lim_{n\to+\infty}H(\rho_n)=\sup_{\rho\in\mathcal{A}}H(\rho).$$

Then $\{\rho_n\}$ converges[3] *to a uniquely determined state $\rho_*(\mathcal{A})$ in $\mathcal{A}$.*

If the maximal entropy state $\Gamma(\mathcal{A})$ exists, then it coincides with $\rho_*(\mathcal{A})$, and the restriction of the entropy to $\mathcal{A}$ is continuous at $\Gamma(\mathcal{A})$.

Proof. By hypothesis, for every $\varepsilon>0$ there is an N_ε such that $H(\rho_n)>\sup_{\rho\in\mathcal{A}}H(\rho)-\varepsilon$ for all $n\geqslant N_\varepsilon$. Using inequality (5) with $\lambda=1/2$, we get

$$\begin{aligned}\sup_{\rho\in\mathcal{A}}H(\rho)-\varepsilon&\leqslant\frac12\,H(\rho_{n_1})+\frac12\,H(\rho_{n_2})\\&\leqslant H\biggl(\frac12\,\rho_{n_1}+\frac12\,\rho_{n_2}\biggr)-\frac18\,\|\rho_{n_2}-\rho_{n_1}\|_1^2\leqslant\sup_{\rho\in\mathcal{A}}H(\rho)-\frac18\,\|\rho_{n_2}-\rho_{n_1}\|_1^2.\end{aligned}$$

Hence $\|\rho_{n_2}-\rho_{n_1}\|_1<\sqrt{8\varepsilon}$ for all $n_1\geqslant N_\varepsilon$ and $n_2\geqslant N_\varepsilon$. Thus $\{\rho_n\}$ is a Cauchy sequence and so converges to some state ρ_* in $\mathcal{A}$. It is easy to see that ρ_* does not depend on the choice of the sequence $\{\rho_n\}$, and so it is determined by $\mathcal{A}$ alone. We denote this state by $\rho_*(\mathcal{A})$.

If the maximal entropy state $\Gamma(\mathcal{A})$ exists, then it coincides with $\rho_*(\mathcal{A})$ by the observation above. The assertion on continuity follows from the definition of $\Gamma(\mathcal{A})$ and the lower semicontinuity of the entropy. The lemma is proved.

Since the entropy is lower semicontinuous, we have

$$H\bigl(\rho_*(\mathcal{A})\bigr)\leqslant\sup_{\rho\in\mathcal{A}}H(\rho),$$

[3] Using Proposition 1 of [16] and Lemma 1 of [16] for the identical channel Φ, one can obtain a stronger version of Lemma 2, which asserts the H-convergence of $\{\rho_n\}$ to $\rho_*(\mathcal{A})=\Omega(\Phi,\mathcal{A})$. This observation and Proposition 2 below imply that $\rho_*(\mathcal{A})=\Gamma(\mathcal{A})$ if there is a $\lambda<1$ such that $\operatorname{Tr}\bigl(\rho_*(\mathcal{A})\bigr)^\lambda<+\infty$.

and the maximal entropy state exists if and only if this is an equation. Propositions 1 and 3 give examples of sets for which equality does not hold. The possible breach of equality and its corollaries in the classical case are considered in [4], where it is called the effect of "entropy loss".

Following [19], we use the term $\mathfrak{H}$-*operator* for an unbounded positive operator H on $\mathcal{H}$ with discrete spectrum of finite multiplicity. Let Q_n be the spectral projector of H corresponding to the lowest n eigenvalues. In accordance with [18], we put

$$\operatorname{Tr}\rho H = \lim_{n\to\infty} \operatorname{Tr}\rho Q_n H, \tag{6}$$

where the sequence on the right-hand side is non-decreasing. As shown in [18], [19], any compact set $\mathcal{K}$ of states is contained in the convex compact set $\mathcal{K}_{H,h} = \{\rho \in \mathfrak{S}(\mathcal{H}) \mid \operatorname{Tr}\rho H \leqslant h\}$ determined by an $\mathfrak{H}$-operator H and a positive number h. Let $h_{\mathrm{m}}(H)$ be the minimal eigenvalue of H and let $\mathcal{H}_{\mathrm{m}}(H)$ be the corresponding (finite-dimensional) eigenspace.

Note that $\mathcal{K}_{H,h}$ is empty if $h < h_{\mathrm{m}}(H)$, $\mathcal{K}_{H,h} = \mathfrak{S}(\mathcal{H}_{\mathrm{m}}(H))$ if $h = h_{\mathrm{m}}(H)$, and $\mathcal{K}_{H,h}$ contains infinite-dimensional states if $h > h_{\mathrm{m}}(H)$.

The following proposition shows that the properties of the restriction of the quantum entropy to $\mathcal{K}_{H,h}$ are determined by the *growth coefficient* $\mathrm{g}(H)$ *of the* $\mathfrak{H}$-*operator* H, which is defined by

$$\mathrm{g}(H) = \inf\{\lambda > 0 \mid \operatorname{Tr}\exp(-\lambda H) < +\infty\}.$$

Here we assume that $\mathrm{g}(H) = +\infty$ if $\operatorname{Tr}\exp(-\lambda H) = +\infty$ for all $\lambda > 0$.

It is known [10], [20] that if $\mathrm{g}(H) = 0$, then the entropy is continuous on the compact set $\mathcal{K}_{H,h}$ and attains its (finite) maximum at the state $\Gamma(\mathcal{K}_{H,h})$ of the form $\bigl(\operatorname{Tr}\exp(-\lambda H)\bigr)^{-1}\exp(-\lambda H)$. The following proposition generalizes this observation. It also provides a necessary and sufficient condition for the existence of the maximal entropy state of $\mathcal{K}_{H,h}$. Put $h_*(H) = \frac{\operatorname{Tr} H\exp(-\mathrm{g}(H)H)}{\operatorname{Tr}\exp(-\mathrm{g}(H)H)}$ if $\operatorname{Tr}\exp\bigl(-\mathrm{g}(H)H\bigr) < +\infty$ and $h_*(H) = +\infty$ otherwise.

Proposition 1. *Let H be an $\mathfrak{H}$-operator on a Hilbert space $\mathcal{H}$, and let h be a positive number such that $h > h_{\mathrm{m}}(H)$.*

1) *The entropy is bounded on $\mathcal{K}_{H,h}$ if and only if $\mathrm{g}(H) < +\infty$.*

2) *The entropy is continuous on $\mathcal{K}_{H,h}$ if and only if $\mathrm{g}(H) = 0$.*

3) *If $h \leqslant h_*(H)$, then $\sup_{\rho\in\mathcal{K}_{H,h}} H(\rho) = \lambda^* h + \log\operatorname{Tr}\exp(-\lambda^* H)$, where $\lambda^* = \lambda^*(H,h) \geqslant \mathrm{g}(H)$ is uniquely determined by the equation*

$$\operatorname{Tr} H\exp(-\lambda H) = h\operatorname{Tr}\exp(-\lambda H),$$

and the maximal entropy state

$$\Gamma(\mathcal{K}_{H,h}) = \bigl(\operatorname{Tr}\exp(-\lambda^* H)\bigr)^{-1}\exp(-\lambda^* H)$$

exists. If $h > h_*(H)$, then $\sup_{\rho\in\mathcal{K}_{H,h}} H(\rho) = \mathrm{g}(H)h + \log\operatorname{Tr}\exp\bigl(-\mathrm{g}(H)H\bigr)$, and the maximal entropy state of $\mathcal{K}_{H,h}$ does not exist. In both cases, $\sup_{\rho\in\mathcal{K}_{H,h}} H(\rho) = \inf_{\lambda\in(\mathrm{g}(H),+\infty)}\bigl(\lambda h + \log\operatorname{Tr}\exp(-\lambda H)\bigr)$.

The function $F_H(h)=\sup_{\rho\in\mathcal{K}_{H,h}}H(\rho)$ *has the following properties.*

i) *It is a continuous increasing function on* $[h_{\mathrm{m}},+\infty)$ *such that* $F_H(h_{\mathrm{m}})=\log\dim\mathcal{H}_{\mathrm{m}}(H)$ *and* $\lim_{h\to+\infty}F_H(h)=+\infty$.

ii) *It has a continuous derivative on* $(h_{\mathrm{m}},+\infty)$:

$$\frac{dF_H(h)}{dh}=\begin{cases}\lambda^*(H,h), & h\in\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr),\\ \mathrm{g}(H), & h\in\bigl[h_*(H),+\infty\bigr),\end{cases}$$

$$\frac{dF_H(h)}{dh}\bigg|_{h=h_{\mathrm{m}}+0}=\lim_{h\to h_{\mathrm{m}}(H)+0}\frac{dF_H(h)}{dh}=+\infty,\qquad \lim_{h\to+\infty}\frac{dF_H(h)}{dh}=\mathrm{g}(H).$$

iii) *It is strictly concave on* $\bigl[h_{\mathrm{m}}(H),h_*(H)\bigr)$ *and linear on* $\bigl[h_*(H),+\infty\bigr)$ *if* $h_*(H)<+\infty$.

The results of numerical calculations of $\sup_{\rho\in\mathcal{K}_{H,h}}H(\rho)$ as a function of $h=c$ for the $\mathfrak{H}$-operator $H=-\log\sigma$ with finite $h_*(H)$ are shown among other characteristics in Fig. 2 of part II of this paper, which will be published in the next issue of this journal.

Proof. Write $H=\sum_{k=1}^{+\infty}h_k|k\rangle\langle k|$, where $\bigl\{|k\rangle\bigr\}_{k\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$ and $\{h_k\}$ is a non-decreasing sequence of positive numbers converging to $+\infty$. Let $d=\dim\mathcal{H}_{\mathrm{m}}(H)$. Then $h_k=h_{\mathrm{m}}$ for $k=1,\dots,d$, and $\bigl\{|k\rangle\bigr\}_{k=1}^{d}$ is a basis of the subspace $\mathcal{H}_{\mathrm{m}}(H)$.

Let us prove assertion 1) of the proposition.

Suppose that $\mathrm{g}(H)<+\infty$. Then there is $\lambda>0$ such that

$$\sigma=\bigl(\operatorname{Tr}\exp(-\lambda H)\bigr)^{-1}\exp(-\lambda H)$$

is a state. Since the relative entropy is non-negative, we see from the definition of $\mathcal{K}_{H,h}$ that

$$H(\rho)=\lambda\operatorname{Tr}\rho H+\log\operatorname{Tr}\exp(-\lambda H)-H(\rho\,\|\,\sigma)\leqslant\lambda h+\log\operatorname{Tr}\exp(-\lambda H)<+\infty$$

for all ρ in $\mathcal{K}_{H,h}$. Hence the entropy is bounded on $\mathcal{K}_{H,h}$.

Suppose that $\sup_{\rho\in\mathcal{K}_{H,h}}H(\rho)<+\infty$. We claim that the equation

$$\sum_{k=1}^{n}h_k\exp(-\lambda h_k)=h\sum_{k=1}^{n}\exp(-\lambda h_k)\tag{7}$$

has a unique positive solution λ_n for all sufficiently large n and that the sequence $\{\lambda_n\}$ is increasing. Indeed, (7) is equivalent to the equation $f_n(\lambda)=0$, where $f_n(\lambda)=\sum_{k=1}^{n}(h_k-h)\exp\bigl(-\lambda(h_k-h)\bigr)$. Since

$$f_n'(\lambda)=-\sum_{k=1}^{n}(h_k-h)^2\exp\bigl(-\lambda(h_k-h)\bigr)<0,$$

the function $f_n(\lambda)$ is strictly decreasing on $[0,+\infty)$. It is easy to see that

$$f_n(0)=\sum_{k=1}^{n}h_k-nh,\qquad \lim_{\lambda\to+\infty}f_n(\lambda)=-\infty,\qquad h>h_{\mathrm{m}}.$$

Since the sequence $\{h_k\}$ is non-decreasing and unbounded, we have $\sum_{k=1}^{n} h_k > nh$ for all sufficiently large n. Thus the observation above yields the existence of a unique positive solution λ_n of the equation $f_n(\lambda) = 0$. To prove that $\lambda_{n+1} > \lambda_n$, it suffices to note that $f_{n+1}(\lambda) > f_n(\lambda)$ for all λ in $[0, +\infty)$ and all n such that $h_n > h$.

For every sufficiently large n, we consider the state

$$\rho_n = \left(\sum_{k=1}^{n} \exp(-\lambda_n h_k)\right)^{-1} \sum_{k=1}^{n} \exp(-\lambda_n h_k)|k\rangle\langle k| \tag{8}$$

in $\mathcal{K}_{H,h}$. This state is the maximum point of the function $H(\rho)$ on the subset $\mathcal{K}_{H,h}^n$ of $\mathcal{K}_{H,h}$. (This subset consists of states supported by the linear hull of the vectors $\{|k\rangle\}_{k=1}^{n}$.) Indeed, since the relative entropy is non-negative, we easily see from the definition of ρ_n that

$$H(\rho) = \lambda_n \operatorname{Tr} \rho H + \log \sum_{k=1}^{n} \exp(-\lambda_n h_k) - H(\rho \,\|\, \rho_n) \leqslant \lambda_n h + \log \sum_{k=1}^{n} \exp(-\lambda_n h_k)$$

for all ρ in $\mathcal{K}_{H,h}^n$. Moreover, this inequality is an equation if and only if $\rho = \rho_n$. Using this observation and the monotonicity of the logarithm, we obtain

$$H(\rho_n) = \lambda_n h + \log \sum_{k=1}^{n} \exp(-\lambda_n h_k) \geqslant \lambda_n (h - h_{\mathrm{m}}). \tag{9}$$

Since $h > h_{\mathrm{m}}$, the assumption $\sup_{\rho \in \mathcal{K}_{H,h}} H(\rho) < +\infty$ guarantees that the sequence $\{\lambda_n\}$ is bounded. As mentioned above, this sequence is also monotone. Hence, $\lim_{n\to+\infty} \lambda_n = \lambda^* < +\infty$ exists. Since $\lambda_n \leqslant \lambda^*$ for all n, equality in (9) implies that

$$\sum_{k=1}^{n} \exp(-\lambda^* h_k) \leqslant \sum_{k=1}^{n} \exp(-\lambda_n h_k) < \exp\Bigl(\sup_{\rho \in \mathcal{K}_{H,h}} H(\rho)\Bigr) < +\infty \tag{10}$$

for all n and, therefore,

$$\sum_{k=1}^{+\infty} \exp(-\lambda^* h_k) < +\infty. \tag{11}$$

This proves that $\mathrm{g}(H) \leqslant \lambda^* < +\infty$.

Since $\mathcal{K}_{H,h} = \overline{\bigcup_n \mathcal{K}_{H,h}^n}$ and $\sup_{\rho \in \mathcal{K}_{H,h}^n} H(\rho) = H(\rho_n)$, the lower semicontinuity of the entropy yields that

$$\sup_{\rho \in \mathcal{K}_{H,h}} H(\rho) = \lim_{n\to+\infty} H(\rho_n).$$

By Lemma 2, the sequence $\{\rho_n\}$ of states converges to the state $\rho_*(\mathcal{K}_{H,h})$. Since $\lim_{n\to+\infty} \lambda_n = \lambda^*$, the sequence

$$\left\{A_n = \sum_{k=1}^{n} \exp(-\lambda_n h_k)|k\rangle\langle k|\right\}_n$$

of operators in $\mathfrak{T}(\mathcal{H})$ converges in the weak operator topology to the operator $A_* = \sum_{k=1}^{\infty} \exp(-\lambda^* h_k)|k\rangle\langle k|$ in $\mathfrak{T}(\mathcal{H})$. Combining these observations, we easily see that

$$\lim_{n\to+\infty} \operatorname{Tr} A_n = \lim_{n\to+\infty} \sum_{k=1}^{n} \exp(-\lambda_n h_k) = \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k) = \operatorname{Tr} A_*, \tag{12}$$

$$\rho_*(\mathcal{K}_{H,h}) = \lim_{n\to+\infty} \rho_n = \biggl(\sum_{k=1}^{+\infty} \exp(-\lambda^* h_k)\biggr)^{-1} \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k)|k\rangle\langle k|. \tag{13}$$

Using (9) and (12), we obtain

$$\sup_{\rho\in\mathcal{K}_{H,h}} H(\rho) = \lim_{n\to+\infty} H(\rho_n) = h\lambda^* + \log \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k). \tag{14}$$

The lower semicontinuity of the entropy implies that

$$H\bigl(\rho_*(\mathcal{K}_{H,h})\bigr) = \lambda^* \frac{\sum_{k=1}^{+\infty} h_k \exp(-\lambda^* h_k)}{\sum_{k=1}^{+\infty} \exp(-\lambda^* h_k)} + \log \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k) \leqslant \lim_{n\to+\infty} H(\rho_n).$$

It follows from (14) that this inequality is equivalent to the inequality

$$\sum_{k=1}^{+\infty} h_k \exp(-\lambda^* h_k) \leqslant h \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k). \tag{15}$$

If these inequalities become equations, then $\rho_*(\mathcal{K}_{H,h}) = \Gamma(\mathcal{K}_{H,h})$. Conversely, if the maximal entropy state $\Gamma(\mathcal{K}_{H,h})$ exists, then it coincides with $\rho_*(\mathcal{K}_{H,h})$ by Lemma 2, and hence equality holds in (15). Thus the maximal entropy state $\Gamma(\mathcal{K}_{H,h})$ exists if and only if we have equality in (15). To complete the proof of assertion 1) of the proposition, it suffices to show that the inequality $h \leqslant h_*(H)$ is equivalent to equality in (15).

First, we claim that the inequality $\lambda^* > \mathrm{g}(H)$ yields equality in (15). Indeed, consider the function

$$f(\lambda) = \lim_{n\to+\infty} f_n(\lambda) = \sum_{k=1}^{+\infty} (h_k - h) \exp\bigl(-\lambda(h_k - h)\bigr).$$

Since the series $\sum_{k=1}^{+\infty} h_k^p \exp(-\lambda h_k)$ converges uniformly on $\bigl[\mathrm{g}(H) + \varepsilon, +\infty\bigr)$ for any $p \in \mathbb{N}$ and $\varepsilon > 0$, the function $f(\lambda)$ has a continuous derivative $f'(\lambda) = -\sum_{k=1}^{+\infty} (h_k - h)^2 \exp\bigl(-\lambda(h_k - h)\bigr) < 0$ on the interval $\bigl(\mathrm{g}(H), +\infty\bigr)$. By construction, we have $f(\lambda_n) > f_n(\lambda_n) = 0$ for all sufficiently large n. Thus the continuity of $f(\lambda)$ at the point $\lambda^* \in \bigl(\mathrm{g}(H), +\infty\bigr)$ implies that $f(\lambda^*) \geqslant 0$. Since (15) is equivalent to the reverse inequality, we get $f(\lambda^*) = 0$, which means that equality holds in (15).

If $h < h_*(H)$, then $f\bigl(\mathrm{g}(H)\bigr) > 0$ (admitting the case $f\bigl(\mathrm{g}(H)\bigr) = +\infty$). Since (15) implies that $f(\lambda^*) \leqslant 0$, we have $\lambda^* > \mathrm{g}(H)$ and, by the observation above, $f(\lambda^*) = 0$.

If $h = h_*(H)$, then $f\big(\mathrm{g}(H)\big) = 0$ and, therefore, $\lambda^* = \mathrm{g}(H)$. Indeed, if $\lambda^* > \mathrm{g}(H)$, then the observation above shows that $f(\lambda^*) = 0 = f\big(\mathrm{g}(H)\big)$, contrary to the fact that $f(\lambda)$ is strictly decreasing.

If $h > h_*(H)$, then $f\big(\mathrm{g}(H)\big) < 0$. Since the function $f(\lambda)$ is decreasing, we have $f(\lambda^*) < 0$ and, therefore, strict inequality holds in (15).

Let us prove assertion 2) of the proposition. If $\mathrm{g}(H) = 0$, then the entropy is continuous on $\mathcal{K}_{H,h}$ by an observation in [20]. This also follows from the implication (i) $\Rightarrow$ (ii) in Proposition 4 below.

To prove the converse assertion, consider the sequence of states

$$\sigma_n = (1 - q_n)|1\rangle\langle 1| + \frac{q_n}{n}\sum_{k=2}^{n+1}|k\rangle\langle k|,$$

where $\{q_n\}$ is the sequence of positive numbers

$$q_n = (h - h_{\mathrm{m}})\biggl(n^{-1}\sum_{k=2}^{n+1} h_k - h_{\mathrm{m}}\biggr)^{-1},$$

which obviously converges to zero. Here we assume that n is so large that $q_n \leqslant 1$. Since the sequence $\{\sigma_n\}$ is contained in $\mathcal{K}_{H,h}$ and converges to the pure state $|1\rangle\langle 1|$, continuity of the entropy on $\mathcal{K}_{H,h}$ implies that the following sequence of positive numbers converges to zero:

$$H(\sigma_n) = h_2(q_n) + q_n \log n = h_2(q_n) + \frac{(h - h_{\mathrm{m}})\log n}{n^{-1}\sum_{k=2}^{n+1} h_k - h_{\mathrm{m}}}\,.$$

The obvious estimate $n^{-1}\sum_{k=2}^{n+1} h_k \leqslant h_{n+1}$ shows that the sequence $\{\nu_n = h_{n+1}^{-1}\log n\}$ converges to zero. Therefore, for arbitrary $\lambda > 0$, we have

$$\operatorname{Tr}\exp(-\lambda H) = \sum_{n=0}^{+\infty}\exp(-\lambda h_{n+1}) = \sum_{n=1}^{+\infty} n^{-\frac{\lambda}{\nu_n}} < +\infty.$$

Hence $\mathrm{g}(H) = 0$.

The general expression for $\sup_{\rho\in\mathcal{K}_{H,h}} H(\rho)$ can be deduced from the previous observation by noting that the infimum in this expression is attained at $\lambda = \lambda^*$ if $h \leqslant h_*(H)$, and at $\lambda = \mathrm{g}(H)$ if $h \geqslant h_*(H)$.

The proofs of the properties of the function $F_H(\rho)$ are based on the implicit function theorem and are given in § 5.

Let σ be an arbitrary state. In what follows we use the *coefficient* $\mathrm{d}(\sigma)$ *of decrease of the state* σ, which is defined by

$$\mathrm{d}(\sigma) = \inf\bigl\{\lambda > 0 \mid \operatorname{Tr}\sigma^{\lambda} < +\infty\bigr\} \in [0, 1].$$

If σ is a full rank state, then $-\log\sigma$ is an $\mathfrak{H}$-operator and $\mathrm{d}(\sigma) = \mathrm{g}(-\log\sigma)$.

It is easy to see that if $\mathrm{d}(\sigma) < 1$, then $H(\sigma) < +\infty$. However, there are states σ with finite entropy such that $\mathrm{d}(\sigma) = 1$ (for example, the state with spectrum

$\{a((k+1)\log^3(k+1))^{-1}\}$, where a is a coefficient). The special role of these states is shown by the following proposition, which is a non-commutative generalization of Theorem 21 in [5], where a classical state (a probability distribution) σ is said to be *hyperbolic* if $\mathrm{d}(\sigma)=1$, and *power dominated* if $\mathrm{d}(\sigma)<1$.

Proposition 2. *Let σ be a state with finite entropy.*

1) *If* $\mathrm{d}(\sigma)<1$, *then*

$$\lim_{n\to+\infty} H(\rho_n)=H(\sigma)$$

for every sequence of states $\{\rho_n\}$ that H-converges[4] *to σ.*

2) *If* $\mathrm{d}(\sigma)=1$, *then for every $h\geqslant H(\sigma)$ there is a sequence $\{\rho_n\}$ of finite-rank states that H-converges to σ and satisfies*

$$\lim_{n\to+\infty} H(\rho_n)=h.$$

Remark 1. Proposition 2 shows that the convex set $\{\sigma\in\mathfrak{S}(\mathcal{H})\mid \mathrm{d}(\sigma)<1\}$ is the maximal set of continuity of the entropy with respect to H-convergence.

The proof of Proposition 2 is based on the following lemma.

Lemma 3. *Let σ be a state with* $\mathrm{d}(\sigma)<1$. *Then the entropy $H(\rho)$ is finite for every state ρ with $H(\rho\,\|\,\sigma)<+\infty$, and the following identity holds for every $\lambda>\mathrm{d}(\sigma)$:*

$$H\bigl(\rho\,\|\,(\operatorname{Tr}\sigma^\lambda)^{-1}\sigma^\lambda\bigr)=\lambda H(\rho\,\|\,\sigma)+\log\operatorname{Tr}\sigma^\lambda-(1-\lambda)H(\rho).$$

If $\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}<+\infty$, *then this identity also holds for $\lambda=\mathrm{d}(\sigma)$.*

Proof. Let $\{P_n\}$ be the increasing sequence of spectral projectors of the state σ. For every n, the positive trace-class operators $A_n=P_n\rho P_n$ and $B_n=P_n\sigma$ satisfy

$$\begin{aligned}H(A_n\,\|\,B_n^\lambda)&=\operatorname{Tr}(A_n\log A_n-A_n\log B_n^\lambda+B_n^\lambda-A_n)\\&=\operatorname{Tr}\bigl((\lambda+(1-\lambda))A_n\log A_n-\lambda A_n\log B_n+B_n^\lambda-A_n\bigr)\\&=\lambda H(A_n\,\|\,B_n)+\operatorname{Tr}B_n^\lambda-\lambda\operatorname{Tr}B_n-(1-\lambda)\operatorname{Tr}A_n-(1-\lambda)\operatorname{Tr}A_n(-\log A_n).\end{aligned}$$

Since $B_n^\lambda=P_n\sigma^\lambda$, Lemma 4 of [8] implies that

$$\lim_{n\to+\infty}\operatorname{Tr}A_n(-\log A_n)=H(\rho),\qquad \lim_{n\to+\infty}H(A_n\,\|\,B_n^\lambda)=H(\rho\,\|\,\sigma^\lambda)$$

for all $\lambda>\mathrm{d}(\sigma)$. Passing to the limit as $n\to+\infty$ in the previous equality, we get

$$H(\rho\,\|\,\sigma^\lambda)=\lambda H(\rho\,\|\,\sigma)+\operatorname{Tr}\sigma^\lambda-1-(1-\lambda)H(\rho).$$

Thus the finiteness of $H(\rho\,\|\,\sigma)$ implies that $H(\rho)$ and $H(\rho\,\|\,\sigma^\lambda)$ are finite for all $\lambda>\mathrm{d}(\sigma)$. Noting that

$$H\bigl(\rho\,\|\,(\operatorname{Tr}\sigma^\lambda)^{-1}\sigma^\lambda\bigr)=H(\rho\,\|\,\sigma^\lambda)+\log\operatorname{Tr}\sigma^\lambda-\operatorname{Tr}\sigma^\lambda+1,$$

we obtain the identity in the lemma.

[4] This means that $\lim_{n\to+\infty}H(\rho_n\,\|\,\sigma)=0$.

Proof of Proposition 2. Let $\mathrm{d}(\sigma) < 1$. Then Lemma 3 implies that

$$\frac{H\big(\rho_n \,\|\, (\operatorname{Tr}\sigma^{\lambda})^{-1}\sigma^{\lambda}\big) - \lambda H(\rho_n \,\|\, \sigma)}{1-\lambda} = \frac{\log \operatorname{Tr}\sigma^{\lambda}}{1-\lambda} - H(\rho_n) \tag{16}$$

for all $\lambda > \mathrm{d}(\sigma)$. Suppose that $\liminf_{n\to+\infty} H(\rho_n) - H(\sigma) = \Delta > 0$. Since the first term on the right-hand side of (16) tends to $H(\sigma)$ as $\lambda \to 1$, there is a $\lambda' < 1$ such that the right-hand of (16) is less than $-\Delta/2$ for $\lambda = \lambda'$ and all sufficiently large n. On the other hand, since the relative entropy is non-negative, we see that the left-hand side of (16) does not exceed $-\frac{\lambda' H(\rho_n \,\|\, \sigma)}{1-\lambda'}$, which tends to zero as $n \to +\infty$.

Suppose that $\mathrm{d}(\sigma) = 1$ and $h > H(\sigma)$. Without loss of generality, we may assume that σ is a full-rank state, $-\log\sigma$ is an $\mathfrak{H}$-operator with $\mathrm{g}(-\log\sigma) = \mathrm{d}(\sigma) = 1$ and $h_*(-\log\sigma) = H(\sigma) < +\infty$. By Proposition 1, $\sup_{\rho\in\mathcal{K}_{-\log\sigma,h}} H(\rho) = h$ for all $h > h_*(-\log\sigma)$. For any given $h > h_*(-\log\sigma)$, the proof of Proposition 1 yields a sequence $\{\rho_n\}$ of states defined by formula (8) and converging to the state $\rho_*(\mathcal{K}_{-\log\sigma,h}) = \sigma$, which is defined by (13). By construction,

$$\lim_{n\to+\infty} H(\rho_n) = \sup_{\rho\in\mathcal{K}_{-\log\sigma,h}} H(\rho) = h, \qquad \lim_{n\to+\infty} H(\rho_n \,\|\, \sigma) = 0.$$

The proposition is proved.

We consider the set $\mathcal{V}_{\sigma,c} = \big\{\rho \in \mathfrak{S}(\mathcal{H}) \mid H(\rho \,\|\, \sigma) \leqslant c\big\}$, which is determined by a state σ and a non-negative number c. The properties of the relative entropy imply that $\mathcal{V}_{\sigma,c}$ is a non-empty, closed, convex subset of $\mathfrak{S}(\mathcal{H})$ for all σ and c. We may regard $\mathcal{V}_{\sigma,c}$ as a c-pseudoneighbourhood of σ with respect to the pseudometric defined by the relative entropy. In the next section, we will see that this set plays a special role related to the notion of χ-capacity of a set of states.

We put $c_*(\sigma) = H\big(\big(\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}\big)^{-1}\sigma^{\mathrm{d}(\sigma)} \,\|\, \sigma\big)$ if $\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)} < +\infty$, and $c_*(\sigma) = +\infty$ otherwise. The following proposition describes some properties of the restriction of the entropy to $\mathcal{V}_{\sigma,c}$ and gives necessary and sufficient conditions for the existence of the maximal entropy state of $\mathcal{V}_{\sigma,c}$.

Proposition 3. *Let σ be any state in $\mathfrak{S}(\mathcal{H})$ and let c be a positive number.*

1) *The set $\mathcal{V}_{\sigma,c}$ is a compact convex subset of $\mathfrak{S}(\mathcal{H})$.*

2) *The entropy is bounded on $\mathcal{V}_{\sigma,c}$ if and only if $\mathrm{d}(\sigma) < 1$.*

3) *The entropy is continuous on $\mathcal{V}_{\sigma,c}$ if and only if $\mathrm{d}(\sigma) = 0$.*

4) *If $\mathrm{d}(\sigma) < 1$ and $c \leqslant c_*(\sigma)$, then $\sup_{\rho\in\mathcal{V}_{\sigma,c}} H(\rho) = \frac{\lambda^* c + \log\operatorname{Tr}\sigma^{\lambda^*}}{1-\lambda^*}$, where $\lambda^* = \lambda^*(\sigma,c) \geqslant \mathrm{d}(\sigma)$ is uniquely determined by the equation*[5]

$$(\lambda - 1)\operatorname{Tr}(\sigma^{\lambda}\log\sigma) = (c + \log\operatorname{Tr}\sigma^{\lambda})\operatorname{Tr}\sigma^{\lambda},$$

and the maximal entropy state $\Gamma(\mathcal{V}_{\sigma,c}) = (\operatorname{Tr}\sigma^{\lambda^*})^{-1}\sigma^{\lambda^*}$ exists.

[5]This equation means that $H\big((\operatorname{Tr}\sigma^{\lambda^*})^{-1}\sigma^{\lambda^*} \,\|\, \sigma\big) = c$.

If $\mathrm{d}(\sigma)<1$ *and* $c>c_*(\sigma)$, *then* $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)=\frac{\mathrm{d}(\sigma)c+\log\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}}{1-\mathrm{d}(\sigma)}$ *and the maximal entropy state of* $\mathcal{V}_{\sigma,c}$ *does not exist. The following equation holds in both cases if* $\mathrm{d}(\sigma)<1$:

$$\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)=\inf_{\lambda\in(\mathrm{d}(\sigma),1)}\frac{\lambda c+\log\operatorname{Tr}\sigma^{\lambda}}{1-\lambda}.$$

The results of numerical calculations of $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)$ as a function of c for states σ with $\mathrm{d}(\sigma)<1$ and $c_*(\sigma)<+\infty$ are shown among other characteristics in Fig. 2 of part II of this paper.

Proof. Without loss of generality, we may assume that σ is a full-rank state and $-\log\sigma$ is an $\mathfrak{H}$-operator.[6]

1) The proof of the assertion on compactness is based on the compactness criterion described in § 2 and the inequality

$$H(\rho\,\|\,\sigma)\geqslant H(P\rho P\,\|\,P\sigma P)\geqslant\operatorname{Tr}(P\rho)\log\frac{\operatorname{Tr}(P\rho)}{\operatorname{Tr}(P\sigma)}+\operatorname{Tr}(P\sigma)-\operatorname{Tr}(P\rho),\tag{17}$$

which holds for any states ρ,σ and any projector P. This inequality follows from Lemma 3 of [8] and the monotonicity property of the relative entropy [9] applied to the trace-preserving completely positive map $\Phi(A)=(\operatorname{Tr}A)\tau$, where τ is an arbitrary state.

For a given σ, let $\{P_n\}$ be a sequence of finite-rank projectors such that $\operatorname{Tr}P_n\sigma>1-n^{-1}$. Suppose that $\mathcal{V}_{\sigma,c}$ is non-compact. By the compactness criterion, for every n there is a state ρ_n in $\mathcal{V}_{\sigma,c}$ such that $\operatorname{Tr}(I_{\mathcal{H}}-P_n)\rho_n>\varepsilon$ for some positive ε. Using inequality (17) with $P=I_{\mathcal{H}}-P_n$, we get

$$\begin{aligned}H(\rho_n\,\|\,\sigma)\geqslant\operatorname{Tr}\bigl((I_{\mathcal{H}}-P_n)\rho_n\bigr)\log\frac{\operatorname{Tr}\bigl((I_{\mathcal{H}}-P_n)\rho_n\bigr)}{\operatorname{Tr}\bigl((I_{\mathcal{H}}-P_n)\sigma\bigr)}\\ +\operatorname{Tr}\bigl((I_{\mathcal{H}}-P_n)\sigma\bigr)-\operatorname{Tr}\bigl((I_{\mathcal{H}}-P_n)\rho_n\bigr)\geqslant\varepsilon\log(\varepsilon n)-1\end{aligned}$$

for all sufficiently large n. Hence $H(\rho_n\,\|\,\sigma)$ tends to $+\infty$ as $n\to+\infty$. This contradicts the definition of $\mathcal{V}_{\sigma,c}$.

2) If $\mathrm{d}(\sigma)=1$, then the entropy is unbounded on $\mathcal{V}_{\sigma,c}$ by assertion 2) of Proposition 2.

If $\mathrm{d}(\sigma)<1$, then Lemma 3 implies that

$$H(\rho)=\frac{\lambda H(\rho\,\|\,\sigma)+\log\operatorname{Tr}\sigma^{\lambda}-H(\rho\,\|\,\sigma_\lambda)}{1-\lambda}\leqslant\frac{c\lambda+\log\operatorname{Tr}\sigma^{\lambda}}{1-\lambda}\tag{18}$$

for all λ in $\bigl(\mathrm{d}(\sigma),1\bigr)$ and all ρ in $\mathcal{V}_{\sigma,c}$. Therefore $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)<+\infty$.

3) If $\mathrm{d}(\sigma)>0$, then Proposition 1 shows that the entropy is discontinuous on the set $\mathcal{K}_{-\log\sigma,c}$, which is contained in $\mathcal{V}_{\sigma,c}$.

If $\mathrm{d}(\sigma)=0$, then $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)=d<+\infty$ by the observation above. Hence the set $\mathcal{V}_{\sigma,c}$ is contained in $\mathcal{K}_{-\log\sigma,c+d}$. By Proposition 1, the entropy is continuous on $\mathcal{K}_{-\log\sigma,c+d}$.

[6]This assumption and the assumption that the dimension of $\mathcal{H}$ is infinite (which is used in the proof) imply that the rank of σ is infinite. However, one can show that all the assertions in Proposition 3 are valid for all states σ of finite rank.

4) We denote the state $(\operatorname{Tr}\sigma^{\lambda})^{-1}\sigma^{\lambda}$ by σ_{λ} and note that the continuous function $f(\lambda)=H(\sigma_{\lambda}\,\|\,\sigma)$ is decreasing on the interval $\big(\mathrm{d}(\sigma),1\big)$. Indeed, it is easy to see that this function has derivative

$$f'(\lambda)=-(1-\lambda)\big(\operatorname{Tr}\sigma_{\lambda}\log^{2}\sigma-(\operatorname{Tr}\sigma_{\lambda}\log\sigma)^{2}\big)<0$$

for every λ in $\big(\mathrm{d}(\sigma),1\big)$. We also note that

$$\lim_{\lambda\to\mathrm{d}(\sigma)+0}f(\lambda)=c_{*}\leqslant+\infty,\qquad f(1)=0.$$

Suppose that $c\leqslant c_{*}$. Then the observation above implies that there is a unique solution λ^{*} of the equation $f(\lambda)=c$. Thus $H(\sigma_{\lambda^{*}}\,\|\,\sigma)=c$ and, therefore,

$$H(\sigma_{\lambda^{*}})=\frac{c\lambda^{*}+\log\operatorname{Tr}\sigma^{\lambda^{*}}}{1-\lambda^{*}}\,.$$

Inequality (18) implies that $H(\rho)\leqslant H(\sigma_{\lambda^{*}})$ for all ρ in $\mathcal{V}_{\sigma,c}$.

Suppose that $c_{*}<+\infty$ and $c>c_{*}$. Then

$$h=\frac{\mathrm{d}(\sigma)c+\log\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}}{1-\mathrm{d}(\sigma)}>\frac{\mathrm{d}(\sigma)c_{*}+\log\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}}{1-\mathrm{d}(\sigma)}=H(\sigma_{\mathrm{d}(\sigma)}).$$

Since $\mathrm{d}(\sigma_{\mathrm{d}(\sigma)})=1$, Proposition 2 yields that for every sufficiently large m there is a sequence $\{\rho_n^m\}_n$ of states such that

$$\lim_{n\to+\infty}H\big(\rho_n^m\,\|\,\sigma_{\mathrm{d}(\sigma)}\big)=0,\qquad\lim_{n\to+\infty}H(\rho_n^m)=h-\frac{1}{m}\,.\tag{19}$$

Using Lemma 3, we get

$$\begin{aligned}\lim_{n\to+\infty}H(\rho_n^m\,\|\,\sigma)&=\lim_{n\to+\infty}\frac{H(\rho_n^m\,\|\,\sigma_{\mathrm{d}(\sigma)})-\log\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}+\big(1-\mathrm{d}(\sigma)\big)H(\rho_n^m)}{\mathrm{d}(\sigma)}\\&=\frac{\big(1-\mathrm{d}(\sigma)\big)h-\log\operatorname{Tr}\sigma^{\mathrm{d}(\sigma)}}{\mathrm{d}(\sigma)}-\frac{1-\mathrm{d}(\sigma)}{\mathrm{d}(\sigma)m}=c-\frac{1-\mathrm{d}(\sigma)}{\mathrm{d}(\sigma)m}\,.\end{aligned}$$

Thus for every m there is an $N(m)$ such that ρ_n^m belongs to $\mathcal{V}_{\sigma,c}$ for all $n\geqslant N(m)$. Using this and (19), we see that the family $\{\rho_n^m\}_{n,m}$ contains a sequence $\{\hat\rho_n\}_n$ of states in $\mathcal{V}_{\sigma,c}$ such that $\{\hat\rho_n\}_n$ converges to $\sigma_{\mathrm{d}(\sigma)}$ and $\lim_{n\to+\infty}H(\hat\rho_n)=h$. Therefore $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)\geqslant h$. Since the reverse inequality follows from (18), we have $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)=h>H(\sigma_{\mathrm{d}(\sigma)})$. By Lemma 2, the set $\mathcal{V}_{\sigma,c}$ contains no maximal entropy state in this case.

The general expression for $\sup_{\rho\in\mathcal{V}_{\sigma,c}}H(\rho)$ can be deduced from the previous observations by observing that the infimum in this expression is attained at $\lambda=\lambda^{*}$ if $c\leqslant c_{*}(\sigma)$, and at $\lambda=\mathrm{d}(\sigma)$ if $c\geqslant c_{*}(\sigma)$.

The following proposition concerns the question of continuity of the entropy on arbitrary sets of states.

Proposition 4. *Let $\mathcal{A}$ be an arbitrary closed subset of $\mathfrak{S}(\mathcal{H})$. Then the following properties are equivalent.*

(i) *$\mathcal{A} \subseteq \mathcal{K}_{H,h}$ for some $\mathfrak{H}$-operator H with $\mathrm{g}(H) = 0$ and some positive number h.*

(ii) *The entropy is continuous on $\mathcal{A}$ and there is a state σ in $\mathfrak{S}(\mathcal{H})$ such that the relative entropy $H(\rho \,\|\, \sigma)$ is continuous and bounded on $\mathcal{A}$.*

(iii) *There is an $\mathfrak{H}$-operator $\widetilde{H}$ with $\mathrm{g}(\widetilde{H}) < +\infty$ such that the linear function $\operatorname{Tr} \rho\widetilde{H}$ is continuous and bounded on $\mathcal{A}$.*

If $\mathcal{A}$ possesses the equivalent properties (i)–(iii), then the $\mathfrak{H}$-operators H, $\widetilde{H}$ and the state σ can be chosen in such a way that $\operatorname{Tr} \sigma H < +\infty$, $\widetilde{H} = -\log \sigma$ and $H(\sigma) < +\infty$.

Remark 2. The last assertion of Proposition 4 implies that if properties (i)–(iii) hold for $\mathcal{A}$, then they hold for the set $\overline{\mathrm{co}}\{\mathcal{A}, \sigma\}$.

Proof of Proposition 4. (ii) $\Rightarrow$ (iii). Since the entropy is finite on $\mathcal{A}$ by (ii), we have

$$H(\rho \,\|\, \sigma) = -H(\rho) + \operatorname{Tr} \rho(-\log \sigma) \quad \forall\, \rho \in \mathcal{A}. \tag{20}$$

By Proposition 3, $\mathcal{A}$ is compact. Therefore the entropy is bounded on $\mathcal{A}$. Hence (ii) and (20) imply that the function $\operatorname{Tr} \rho(-\log \sigma)$ is continuous and bounded on $\mathcal{A}$. Thus (iii) holds with $\widetilde{H} = -\log \sigma$.

(iii) $\Rightarrow$ (ii). Suppose that $\lambda > \mathrm{g}(\widetilde{H})$ and let $\sigma = \bigl(\operatorname{Tr} \exp(-\lambda \widetilde{H})\bigr)^{-1} \exp(-\lambda \widetilde{H})$ be a state in $\mathfrak{S}(\mathcal{H})$ with finite entropy. Then (iii) means that the function $\operatorname{Tr} \rho(-\log \sigma)$ is continuous and bounded on $\mathcal{A}$. Since the entropy and relative entropy are lower semicontinuous, it follows from (20) that the functions $H(\rho)$ and $H(\rho \,\|\, \sigma)$ are continuous and bounded on $\mathcal{A}$.

(i) $\Rightarrow$ (iii). Let $H = \sum_k h_k |k\rangle\langle k|$, where $\{|k\rangle\}$ is an orthonormal basis of $\mathcal{H}$. By assumption, we have $\sum_k \exp(-\lambda h_k) < +\infty$ for all $\lambda > 0$ and, therefore, $\sum_k h_k \exp(-\lambda h_k) < +\infty$ for all $\lambda > 0$. Hence there is a sequence $\{\lambda_k\}$ of positive numbers converging monotonically to zero and satisfying $\sum_k h_k \exp(-\lambda_k h_k) < +\infty$. This sequence can be constructed as follows. For every positive integer m, let $N(m)$ be the minimal positive integer such that $\sum_{k=N(m)}^{+\infty} h_k \exp(-h_k/m) < 2^{-m}$. It is easy to see that the sequence

$$\lambda_k = \begin{cases} 1, & k < N(2), \\ \dfrac{1}{m}, & N(m) \leqslant k < N(m+1), \ \ m \geqslant 2 \end{cases}$$

satisfies the desired condition. Since $\operatorname{Tr} \rho H = \sum_k h_k \langle k|\rho|k\rangle \leqslant h$ for all ρ in $\mathcal{A}$, the series $\sum_k \lambda_k h_k \langle k|\rho|k\rangle$ converges uniformly on $\mathcal{A}$. This yields the continuity on $\mathcal{A}$ of the function $\operatorname{Tr} \rho(-\log \sigma)$, where

$$\sigma = \biggl(\sum_k \exp(-\lambda_k h_k)\biggr)^{-1} \sum_k \exp(-\lambda_k h_k) |k\rangle\langle k|.$$

Note that the condition $\sum_k h_k \exp(-\lambda_k h_k) < +\infty$ implies that $\operatorname{Tr} \sigma H < +\infty$ and $H(\sigma) < +\infty$. Thus (iii) holds with $\widetilde{H} = -\log \sigma$.

(iii) $\Rightarrow$ (i). Let $\widetilde{H}=\sum_k \tilde{h}_k|k\rangle\langle k|$, where $\{|k\rangle\}$ is an orthonormal basis in $\mathcal{H}$. Since (iii) is equivalent to (ii), Proposition 3 shows that $\mathcal{A}$ is compact. By the assumption (iii), the series $\sum_k \tilde{h}_k\langle k|\rho|k\rangle$ converges on the compact set $\mathcal{A}$ to the continuous function $\operatorname{Tr}\rho\widetilde{H}$. By Dini's lemma, it converges uniformly on $\mathcal{A}$. This yields the existence of a sequence $\{\lambda_k\}$ of positive numbers that converges monotonically to $+\infty$ and satisfies $\sum_k \lambda_k\tilde{h}_k\langle k|\rho|k\rangle \leqslant h < +\infty$ for all ρ in $\mathcal{A}$. It is easy to see that (i) holds with $H=\sum_k \lambda_k\tilde{h}_k|k\rangle\langle k|$.

The last assertion of the proposition follows from the construction above.

The following observation is a corollary of Propositions 1 and 4.

Corollary 1. *If H is an $\mathfrak{H}$-operator with $\mathrm{g}(H)=0$, then one can find a state σ in $\mathfrak{S}(\mathcal{H})$ and an $\mathfrak{H}$-operator $\widetilde{H}$ with $\mathrm{g}(\widetilde{H})<+\infty$ such that the relative entropy $H(\rho\,\|\,\sigma)$ and the linear function $\operatorname{Tr}\rho\widetilde{H}$ are continuous on $\mathcal{K}_{H,h}$.*

Since the set $\mathcal{K}_{H,h}$ is convex, Propositions 1 and 4 yield the following result.

Corollary 2. *If the entropy is continuous on the closed set $\mathcal{A}$ and there is a state σ in $\mathfrak{S}(\mathcal{H})$ such that the relative entropy $H(\rho\,\|\,\sigma)$ is continuous and bounded on $\mathcal{A}$, then the entropy is continuous on $\overline{\mathrm{co}}(\mathcal{A})$.*

Remark 3. The assumption on the existence of σ is essential in assertion (ii) of Proposition 4 and in Corollary 2. Indeed, let $\mathcal{A}$ be the closed set of all pure states in $\mathfrak{S}(\mathcal{H})$. Then the entropy is equal to zero (and hence is continuous) on $\mathcal{A}$, but is not continuous on $\overline{\mathrm{co}}(\mathcal{A})=\mathfrak{S}(\mathcal{H})$. There is a *compact* set $\mathcal{A}$ of pure states (a convergent sequence) such that the entropy is unbounded on $\overline{\mathrm{co}}(\mathcal{A})$ (see Example 1 in part II of this paper).

Proposition 4 and Corollary 2 enable us to show that the entropy is continuous on some non-trivial sets of states. The following result will be used in part II of this paper.

Corollary 3. *Let $\{U_\lambda\}_{\lambda\in\Lambda}$ be a closed family of unitary (anti-unitary) operators on $\mathcal{H}$ and let ω be a state in $\mathfrak{S}(\mathcal{H})$ such that $U_\lambda\omega U_\lambda^*=\omega$ for all $\lambda\in\Lambda$. Then the functions $H(\rho)$ and $H(\rho\,\|\,\omega)$ are continuous on the set $\overline{\mathrm{co}}\big(\{U_\lambda\sigma U_\lambda^*\}_{\lambda\in\Lambda}\big)$ for any state σ with $\operatorname{Tr}\sigma(-\log\omega)<+\infty$.*

Given an arbitrary orthonormal basis $\{|k\rangle\}\subset\mathcal{H}$, we consider the trace-preserving completely positive map

$$\Pi_{\{|k\rangle\}}\colon \rho\mapsto\sum_k\langle k|\rho|k\rangle|k\rangle\langle k|.$$

The set of output states of the map $\Pi_{\{|k\rangle\}}$ can be regarded as the set of all classical states (probability distributions). Hence the set $\Pi_{\{|k\rangle\}}(\mathcal{A})$ may be called the *classical projection* of $\mathcal{A}$ with respect to the basis $\{|k\rangle\}$.

The following proposition shows that some properties of sets of quantum states are closely related to properties of the classical projections of these sets.

Proposition 5. *Let $\mathcal{A}$ be an arbitrary closed subset of $\mathfrak{S}(\mathcal{H})$.*

1) *The set $\mathcal{A}$ is compact if $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is compact for at least one basis $\{|k\rangle\}$.*
2) *If $\mathcal{A}$ is compact, then $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is compact for any basis $\{|k\rangle\}$.*

3) *The entropy is bounded on* $\mathcal{A}$ *if it is bounded on* $\Pi_{\{|k\rangle\}}(\mathcal{A})$ *for at least one basis* $\{|k\rangle\}$.

4) *If* $\mathcal{A}$ *is convex and the entropy is bounded on* $\mathcal{A}$, *then it is bounded on* $\Pi_{\{|k\rangle\}}(\mathcal{A})$ *for at least one basis* $\{|k\rangle\}$.

5) *The entropy is continuous on* $\mathcal{A}$ *if it is continuous on* $\Pi_{\{|k\rangle\}}(\mathcal{A})$ *for at least one basis* $\{|k\rangle\}$.

6) *If the entropy is continuous on* $\mathcal{A}$ *and there is a state* σ *in* $\mathfrak{S}(\mathcal{H})$ *such that the relative entropy* $H(\rho\,\|\,\sigma)$ *is continuous and bounded on* $\mathcal{A}$, *then the entropy is continuous on* $\Pi_{\{|k\rangle\}}(\mathcal{A})$ *for at least one basis* $\{|k\rangle\}$.

Proof. Suppose that $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is compact. By the compactness criterion for sets of classical states, for every $\varepsilon > 0$ there is an N_ε such that

$$\operatorname{Tr} P_\varepsilon \rho = \sum_{k=1}^{N_\varepsilon} \langle k|\rho|k\rangle \geqslant 1-\varepsilon \quad \forall\,\rho\in\mathcal{A},$$

where $P_\varepsilon = \sum_{k=1}^{N_\varepsilon} |k\rangle\langle k|$ is a finite-rank projector. By the compactness criterion for subsets of $\mathfrak{S}(\mathcal{H})$, it follows that $\mathcal{A}$ is compact.

If $\mathcal{A}$ is compact and $\{|k\rangle\}$ is any basis, then $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is compact, being the image of a compact set under a continuous map.

To prove the remaining assertions, we use the identity

$$H\bigl(\rho\,\|\,\Pi_{\{|k\rangle\}}(\rho)\bigr) = H\bigl(\Pi_{\{|k\rangle\}}(\rho)\bigr) - H(\rho), \tag{21}$$

which holds for any state ρ with $H\bigl(\Pi_{\{|k\rangle\}}(\rho)\bigr) < +\infty$.

If the entropy is bounded on $\Pi_{\{|k\rangle\}}(\mathcal{A})$, then it is bounded on $\mathcal{A}$. Indeed, since the relative entropy is non-negative, we see from (21) that $H(\rho) \leqslant H\bigl(\Pi_{\{|k\rangle\}}(\rho)\bigr)$ for any ρ in $\mathcal{A}$.

If the entropy is bounded on the convex set $\mathcal{A}$, then Corollary 7 below shows that $\mathcal{A}$ is contained in $\mathcal{K}_{H,h}$ for some $\mathfrak{H}$-operator H with $\mathrm{g}(H) < +\infty$. Let $\{|k\rangle\}$ be a basis of eigenvectors of the $\mathfrak{H}$-operator H. Then $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is also contained in $\mathcal{K}_{H,h}$ and, therefore, the entropy is bounded on $\Pi_{\{|k\rangle\}}(\mathcal{A})$ by Proposition 1.

Suppose that the entropy is bounded on $\Pi_{\{|k\rangle\}}(\mathcal{A})$. Then it is finite on this set. By (21), it is finite on $\mathcal{A}$. Let ρ_0 be a state in $\mathcal{A}$ and let $\{\rho_n\}$ be a sequence of states in $\mathcal{A}$ converging to ρ_0. Since the entropy is continuous on $\Pi_{\{|k\rangle\}}(\mathcal{A})$ (by assumption) and the relative entropy is lower semicontinuous, we see from (21) that

$$\begin{aligned}\limsup_{n\to+\infty} H(\rho_n) &= \lim_{n\to+\infty} H\bigl(\Pi_{\{|k\rangle\}}(\rho_n)\bigr) - \liminf_{n\to+\infty} H\bigl(\rho_n\,\|\,\Pi_{\{|k\rangle\}}(\rho_n)\bigr)\\ &\leqslant H\bigl(\Pi_{\{|k\rangle\}}(\rho_0)\bigr) - H\bigl(\rho_0\,\|\,\Pi_{\{|k\rangle\}}(\rho_0)\bigr) = H(\rho_0).\end{aligned}$$

This inequality and lower semicontinuity of the entropy imply that

$$\lim_{n\to+\infty} H(\rho_n) = H(\rho_0).$$

If the entropy is continuous on $\mathcal{A}$ and there is a state σ in $\mathfrak{S}(\mathcal{H})$ such that the relative entropy $H(\rho\,\|\,\sigma)$ is continuous and bounded on $\mathcal{A}$, then Proposition 4 shows that $\mathcal{A}$ is contained in $\mathcal{K}_{H,h}$ for some $\mathfrak{H}$-operator H with $\mathrm{g}(H) = 0$. Let $\{|k\rangle\}$ be a basis of eigenvectors of H. Then $\Pi_{\{|k\rangle\}}(\mathcal{A})$ is also contained in $\mathcal{K}_{H,h}$ and hence the entropy is continuous on $\Pi_{\{|k\rangle\}}(\mathcal{A})$ by Proposition 1.

Remark 4. In contrast to assertion 2), one cannot replace the expression "for at least one" by "for any" in assertions 4) and 6) of Proposition 5. Indeed, for every pure state ρ there is a basis $\{|k\rangle\}$ such that $H\big(\Pi_{\{|k\rangle\}}(\rho)\big) = +\infty$.

Let σ be a state with basis of eigenvectors $\{|k\rangle\}$. We consider the set $\Pi^{-1}_{\{|k\rangle\}}(\sigma)$ of all states whose diagonal entries in the basis $\{|k\rangle\}$ coincide with those of σ. The set $\Pi^{-1}_{\{|k\rangle\}}(\sigma)$ is called the *layer*[7] *corresponding to* σ and is denoted by $\mathcal{L}(\sigma)$. In a sense, one can regard layers as the simplest purely quantum sets of states.

By (21), we have

$$H(\rho) \leqslant H(\sigma) \quad \forall\, \rho \in \mathcal{L}(\sigma). \tag{22}$$

Hence the entropy is bounded on the layer corresponding to σ if and only if $H(\sigma) < +\infty$. The following proposition shows that boundedness of the entropy on a layer guarantees its continuity.

Proposition 6. *Let σ be an arbitrary state in $\mathfrak{S}(\mathcal{H})$.*

1) *The set $\mathcal{L}(\sigma)$ is a compact convex subset of $\mathfrak{S}(\mathcal{H})$.*

2) *The entropy $H(\rho)$ is continuous on $\mathcal{L}(\sigma)$ if and only if $\sup_{\rho\in\mathcal{L}(\sigma)} H(\rho) = H(\sigma) < +\infty$.*

3) *If $H(\sigma) < +\infty$, then $H(\rho\,\|\,\sigma) = H(\sigma) - H(\rho)$ for any state ρ in $\mathcal{L}(\sigma)$.*

4) *If $H(\sigma) = +\infty$, then $H(\rho\,\|\,\sigma) = +\infty$ for every pure state ρ in $\mathcal{L}(\sigma)$.*

Proof. Assertions 1) and 2) follow from assertions 1) and 5) (respectively) of Proposition 5 because $\Pi_{\{|k\rangle\}}\big(\mathcal{L}(\sigma)\big) = \{\sigma\}$ if $\{|k\rangle\}$ is a basis of eigenvectors of σ.

The expression for the relative entropy in the case $H(\sigma) < +\infty$ is a restatement of the identity (21).

Suppose that $H(\sigma) = +\infty$ and ρ is an arbitrary pure state in $\mathcal{L}(\sigma)$. We consider the sequences of states $\{\sigma_n = (\operatorname{Tr} P_n\sigma)^{-1}P_n\sigma\}$ and $\{\rho_n = (\operatorname{Tr} P_n\rho)^{-1}P_n\rho P_n\}$, where P_n is the spectral projector of σ that corresponds to its n maximal eigenvalues.

Since the pure state ρ_n lies in $\mathcal{L}(\sigma_n)$ for every n, we see from (21) that

$$H(\rho_n\,\|\,\sigma_n) = H(\sigma_n) - H(\rho_n) = H(\sigma_n).$$

By Lemma 4 of [8], the left- and right-hand sides of this equation tend to $H(\rho\,\|\,\sigma)$ and $H(\sigma) = +\infty$ (respectively) as $n \to +\infty$.

Remark 5. Propositions 5 and 6 enable us to make the following observation: *discontinuity and unboundedness of the quantum entropy in the infinite-dimensional case are of a purely classical nature.* Indeed, the set of all quantum states may be regarded as the union of the layers corresponding to all states that are diagonalizable in some basis. The set of these states can be identified with the set of all classical states, and every layer can be identified with a set of purely quantum states. Proposition 6 shows that the entropy is continuous on the whole layer if it is finite on the corresponding classical state. By Proposition 5, the possible discontinuity of the quantum entropy is connected with transitions between layers corresponding to a set of classical states on which the entropy is discontinuous.

[7]If σ has distinct eigenvalues, then the basis $\{|k\rangle\}$ is essentially unique, and the set $\mathcal{L}(\sigma)$ depends only on σ. If σ has multiple eigenvalues, then $\mathcal{L}(\sigma)$ depends on the choice of the basis $\{|k\rangle\}$. In the latter case, all the "variants" of the set $\mathcal{L}(\sigma)$ are isomorphic to each other, and so we shall assume that one of them has been chosen.

§ 4. χ-capacity

4.1. The optimal average state. Let $\mathcal{A}$ be an arbitrary subset of $\mathfrak{S}(\mathcal{H})$. The *$\chi$-capacity* of $\mathcal{A}$ is defined by

$$\overline{C}(\mathcal{A}) = \sup_{\{\pi_i,\rho_i\}} \chi\bigl(\{\pi_i,\rho_i\}\bigr), \tag{23}$$

where the supremum is taken over all ensembles $\{\pi_i,\rho_i\}$ of states in $\mathcal{A}$.

If the entropy is bounded on the set $\overline{\mathrm{co}}(\mathcal{A})$, then

$$\overline{C}(\mathcal{A}) = \sup_{\{\pi_i,\rho_i\}} \biggl(H\biggl(\sum_i \pi_i\rho_i\biggr) - \sum_i \pi_i H(\rho_i)\biggr) \leqslant \sup_{\rho\in\overline{\mathrm{co}}(\mathcal{A})} H(\rho) < +\infty.$$

But boundedness of the entropy is not necessary for finiteness of the χ-capacity. This is shown by examples in part II of this paper.

Let $\bigl\{\{\pi_i^n,\rho_i^n\}\bigr\}_n$ be a sequence of ensembles of states in $\mathcal{A}$ such that

$$\lim_{n\to+\infty} \chi\bigl(\{\pi_i^n,\rho_i^n\}\bigr) = \overline{C}(\mathcal{A}).$$

In accordance with [15], such a sequence is called an *approximating sequence* for $\mathcal{A}$.

If $\mathcal{A}$ is a set of states (density operators) in a finite-dimensional Hilbert space, then the supremum in definition (23) of χ-capacity is attained at some ensemble $\{\pi_i,\rho_i\}$ (the optimal ensemble for $\mathcal{A}$), whose average state possesses some special properties [14]. If $\mathcal{A}$ is a set of states (density operators) in an infinite-dimensional Hilbert space, then the optimal ensemble generally does not exist. However, one can prove the existence of a unique state that possesses the properties of the average state of the optimal ensemble in the finite-dimensional case.

Theorem 1. *Let $\mathcal{A}$ be a set with finite χ-capacity $\overline{C}(\mathcal{A})$.*

1) *There is a unique state $\Omega(\mathcal{A})$ in $\mathfrak{S}(\mathcal{H})$ such that*

$$H\bigl(\rho\,\|\,\Omega(\mathcal{A})\bigr) \leqslant \overline{C}(\mathcal{A}) \quad \forall\,\rho\in\mathcal{A}.$$

The state $\Omega(\mathcal{A})$ belongs to $\overline{\mathrm{co}}(\mathcal{A})$. If $\bigl\{\{\pi_i^n,\rho_i^n\}\bigr\}_n$ is any approximating sequence for $\mathcal{A}$, then the corresponding sequence $\{\bar\rho_n\}$ of average states H-converges[8] *to the state $\Omega(\mathcal{A})$.*

2) *The χ-capacity of $\mathcal{A}$ satisfies*

$$\overline{C}(\mathcal{A}) = \inf_{\sigma\in\mathfrak{S}(\mathcal{H})} \sup_{\rho\in\mathcal{A}} H(\rho\,\|\,\sigma) = \inf_{\sigma\in\overline{\mathrm{co}}(\mathcal{A})} \sup_{\rho\in\mathcal{A}} H(\rho\,\|\,\sigma) = \sup_{\rho\in\mathcal{A}} H\bigl(\rho\,\|\,\Omega(\mathcal{A})\bigr), \tag{24}$$

where the first two equalities remain valid in the case $\overline{C}(\mathcal{A}) = +\infty$.

Proof. 1) Let $\bigl\{\mu_n = \{\pi_i^n,\rho_i^n\}_{i=1}^{N(n)}\bigr\}$ be any approximating sequence of ensembles for $\mathcal{A}$. We claim that the corresponding sequence $\{\bar\rho_n\}$ of average states converges to some state in $\mathfrak{S}(\mathcal{H})$. Indeed, by the definition of an approximating sequence,

[8] This means that $\lim_{n\to+\infty} H\bigl(\bar\rho_n\,\|\,\Omega(\mathcal{A})\bigr) = 0$.

for every $\varepsilon > 0$ there is an N_ε such that $\chi(\mu_n) > \overline{C}(\mathcal{A}) - \varepsilon$ for all $n \geqslant N_\varepsilon$. By Lemma 1 with $m = 2$ and $\lambda = 1/2$, we have

$$\begin{aligned}\overline{C}(\mathcal{A}) - \varepsilon &\leqslant \frac{1}{2}\,\chi(\mu_{n_1}) + \frac{1}{2}\,\chi(\mu_{n_2}) \\ &\leqslant \chi\biggl(\frac{1}{2}\,\mu_{n_1} + \frac{1}{2}\,\mu_{n_2}\biggr) - \frac{1}{8}\,\|\bar{\rho}_{n_2} - \bar{\rho}_{n_1}\|_1^2 \leqslant \overline{C}(\mathcal{A}) - \frac{1}{8}\,\|\bar{\rho}_{n_2} - \bar{\rho}_{n_1}\|_1^2,\end{aligned}$$

whence $\|\bar{\rho}_{n_2} - \bar{\rho}_{n_1}\|_1 < \sqrt{8\varepsilon}$ for all $n_1 \geqslant N_\varepsilon$ and $n_2 \geqslant N_\varepsilon$. Therefore $\{\bar{\rho}_n\}$ is a Cauchy sequence. Thus it converges to some state ρ_* in $\mathfrak{S}(\mathcal{H})$.

Let σ be an arbitrary state in $\mathcal{A}$. Given any positive integer n and any η in $[0,1]$, we consider the ensemble[9] μ_n^η that consists of states $\bigl\{\rho_1^n,\dots,\rho_{N(n)}^n,\sigma\bigr\}$ with corresponding probability distribution $\bigl\{(1-\eta)\pi_1^n,\dots,(1-\eta)\pi_{N(n)}^n,\eta\bigr\}$. We obtain the sequence of ensembles $\{\mu_n^\eta\}$ with corresponding sequence of average states $\bigl\{\bar{\rho}_n^\eta = (1-\eta)\bar{\rho}_n + \eta\sigma\bigr\}_n$ that converges to the state $\bar{\rho}_\eta = (1-\eta)\rho_* + \eta\sigma$ as $n \to +\infty$.

For any positive integer n, we have

$$\chi(\mu_n^\eta) = (1-\eta)\sum_i \pi_i^n H(\rho_i^n \,\|\, \bar{\rho}_n^\eta) + \eta H(\sigma \,\|\, \bar{\rho}_n^\eta). \tag{25}$$

Both summands on the right-hand side of (25) are finite because $\overline{C}(\mathcal{A})$ is finite by hypothesis. Applying Donald's identity (3) to the first of them, we get

$$\sum_i \pi_i^n H(\rho_i^n \,\|\, \bar{\rho}_n^\eta) = \chi(\mu_n^0) + H(\bar{\rho}_n \,\|\, \bar{\rho}_n^\eta).$$

Substituting this expression in (25), we obtain

$$\chi(\mu_n^\eta) = \chi(\mu_n^0) + (1-\eta)H(\bar{\rho}_n \,\|\, \bar{\rho}_n^\eta) + \eta\bigl(H(\sigma \,\|\, \bar{\rho}_n^\eta) - \chi(\mu_n^0)\bigr).$$

Since the relative entropy is non-negative, it follows that

$$H(\sigma \,\|\, \bar{\rho}_n^\eta) \leqslant \eta^{-1}\bigl(\chi(\mu_n^\eta) - \chi(\mu_n^0)\bigr) + \chi(\mu_n^0), \qquad \eta \neq 0. \tag{26}$$

By the definition of an approximating sequence, we have

$$\lim_{n\to+\infty} \chi(\mu_n^0) = \overline{C}(\mathcal{A}) \geqslant \chi(\mu_n^\eta) \tag{27}$$

for all n and $\eta > 0$. It follows that

$$\liminf_{\eta\to+0}\,\liminf_{n\to+\infty}\,\eta^{-1}\bigl[\chi(\mu_n^\eta) - \chi(\mu_n^0)\bigr] \leqslant 0. \tag{28}$$

Since the relative entropy is lower semicontinuous, inequalities (26)–(28) imply that

$$H(\sigma \,\|\, \rho_*) \leqslant \liminf_{\eta\to+0}\,\liminf_{n\to+\infty}\, H(\sigma \,\|\, \bar{\rho}_n^\eta) \leqslant \overline{C}(\mathcal{A}).$$

[9]This extension of an ensemble by “inserting” an additional state was first used in [14] in the finite-dimensional case.

Thus we have proved that

$$\sup_{\sigma\in\mathcal{A}} H(\sigma\,\|\,\rho_*)\leqslant \overline{C}(\mathcal{A}). \tag{29}$$

Let $\{\{\lambda_j^n,\sigma_j^n\}\}_n$ be an arbitrary approximating sequence of ensembles. By inequality (29), we have

$$\sum_j \lambda_j^n H(\sigma_j^n\,\|\,\rho_*)\leqslant \overline{C}(\mathcal{A}).$$

Applying Donald's identity (3), we obtain

$$\sum_j \lambda_j^n H(\sigma_j^n\,\|\,\rho_*)=\sum_j \lambda_j^n H(\sigma_j^n\,\|\,\bar\sigma_n)+H(\bar\sigma_n\,\|\,\rho_*). \tag{30}$$

The last two expressions show that

$$H(\bar\sigma_n\,\|\,\rho_*)\leqslant \overline{C}(\mathcal{A})-\sum_j \lambda_j^n H(\sigma_j^n\,\|\,\bar\sigma_n).$$

By the approximating property of the sequence $\{\{\lambda_j^n,\sigma_j^n\}\}_n$, the right-hand side of this inequality tends to zero as $n\to+\infty$. Hence the sequence $\{\bar\sigma_n\}_n$ H-converges to ρ_* and, therefore, converges to this state in the trace-norm topology. Thus the state ρ_* is independent of the choice of the approximating sequence and is determined by $\mathcal{A}$ alone. We denote this state by $\Omega(\mathcal{A})$. The observation above implies that $\rho_*=\Omega(\mathcal{A})$ is the unique state in $\mathfrak{S}(\mathcal{H})$ such that inequality (29) holds.

2) To prove (24), we start by showing that inequality (29) is actually an equation. Indeed, since (30) holds for any approximating sequence $\{\{\lambda_j^n,\sigma_j^n\}\}_n$ and the relative entropy is non-negative, we see that

$$\sum_j \lambda_j^n H(\sigma_j^n\,\|\,\bar\sigma_n)\leqslant \sum_j \lambda_j^n H(\sigma_j^n\,\|\,\rho_*)\leqslant \sup_{\sigma\in\mathcal{A}} H(\sigma\,\|\,\rho_*).$$

By the approximating property of the sequence $\{\{\lambda_j^n,\sigma_j^n\}\}_n$, the left-hand side of this inequality tends to $\overline{C}(\mathcal{A})$ as $n\to+\infty$. This proves that (29) is an equation.

Consider the function $F(\sigma)=\sup_{\rho\in\mathcal{A}} H(\rho\,\|\,\sigma)$ on $\mathfrak{S}(\mathcal{H})$. Equality in (29) means that $F\bigl(\Omega(\mathcal{A})\bigr)=\overline{C}(\mathcal{A})$. Hence the state $\Omega(\mathcal{A})$ is the unique minimum point of the function $F(\sigma)$ on $\mathfrak{S}(\mathcal{H})$. Indeed, let σ_0 be a state in $\mathfrak{S}(\mathcal{H})$ such that

$$\sup_{\rho\in\mathcal{A}} H(\rho\,\|\,\sigma_0)=F(\sigma_0)\leqslant F\bigl(\Omega(\mathcal{A})\bigr)=\overline{C}(\mathcal{A}).$$

By assertion 1) of the theorem, it follows that $\sigma_0=\Omega(\mathcal{A})$.

If $\overline{C}(\mathcal{A})=+\infty$, then the right-hand side of (24) also equals $+\infty$. Indeed, let σ_0 be a state in $\mathfrak{S}(\mathcal{H})$ such that $\sup_{\rho\in\mathcal{A}} H(\rho\,\|\,\sigma_0)=c<+\infty$. Using Donald's identity and the fact that the relative entropy is non-negative, we get

$$\sum_i \pi_i H(\rho_i\,\|\,\bar\rho)\leqslant \sum_i \pi_i H(\rho_i\,\|\,\sigma_0)-H(\bar\rho\,\|\,\sigma_0)\leqslant c$$

for an arbitrary ensemble $\{\pi_i,\rho_i\}$ of states in $\mathcal{A}$. Hence $\overline{C}(\mathcal{A})\leqslant c<+\infty$.

Definition 1. The state $\Omega(\mathcal{A})$ introduced in Theorem 1 is called the *optimal average state of the set* $\mathcal{A}$.

Using Theorem 1, Donald's identity (3) and inequality (1), we get the following useful inequality.

Corollary 4. *Let $\mathcal{A}$ be a set with finite χ-capacity. The following inequality holds for any ensemble $\{\pi_i,\rho_i\}$ of states in $\mathcal{A}$ with average state $\bar{\rho}$:*

$$\overline{C}(\mathcal{A})-\chi\bigl(\{\pi_i,\rho_i\}\bigr)\geqslant H\bigl(\bar{\rho}\,\|\,\Omega(\mathcal{A})\bigr)\geqslant\frac{1}{2}\,\bigl\|\bar{\rho}-\Omega(\mathcal{A})\bigr\|_1^2.$$

Theorem 1 and Proposition 3 yield the following important result.

Corollary 5. *Any set of states with finite χ-capacity is relatively compact.*

We note that the converse does not hold. There are compact sets of infinite χ-capacity (for example, convergent sequences of states; see § 3 in part II of this paper).

Corollary 5 enables us to make an important observation related to the χ-capacity of constrained quantum channels [18], [19], [15].

Corollary 6. *Let $\Phi\colon\mathfrak{S}(\mathcal{H})\mapsto\mathfrak{S}(\mathcal{H}')$ be an arbitrary quantum channel and let $\mathcal{A}$ be a subset of $\mathfrak{S}(\mathcal{H})$. If $\overline{C}(\Phi,\mathcal{A})<+\infty$, then $\Phi(\mathcal{A})$ is a relatively compact subset of $\mathfrak{S}(\mathcal{H}')$.*

Proof. It is easy to see from the definitions that

$$\overline{C}\bigl(\Phi(\mathcal{A})\bigr)\leqslant\overline{C}(\Phi,\mathcal{A}).$$

The corollary is proved.

This observation shows that the χ-capacity of an unconstrained quantum channel can be finite only if the output set of the channel is relatively compact.

Theorem 1 and Proposition 1 enable us to make the following observation on properties of the entropy.

Corollary 7. *The entropy is bounded on a convex set $\mathcal{A}$ if and only if $\mathcal{A}$ is relatively compact and is contained in $\mathcal{K}_{H,h}$ for some $\mathfrak{H}$-operator H with $\mathrm{g}(H)<+\infty$ and some positive h.*

Proof. If $\mathcal{A}$ is contained in $\mathcal{K}_{H,h}$ with $\mathrm{g}(H)<+\infty$, then we have $\sup_{\rho\in\mathcal{A}}H(\rho)<+\infty$ by Proposition 1.

If $\sup_{\rho\in\mathcal{A}}H(\rho)<+\infty$, then $\overline{C}(\mathcal{A})<+\infty$. By Theorem 1, we have

$$H\bigl(\rho\,\|\,\Omega(\mathcal{A})\bigr)=\operatorname{Tr}\rho\bigl(-\log\Omega(\mathcal{A})\bigr)-H(\rho)\leqslant\overline{C}(\mathcal{A})$$

for all ρ in $\mathcal{A}$. It follows that

$$\operatorname{Tr}\rho\bigl(-\log\Omega(\mathcal{A})\bigr)\leqslant\overline{C}(\mathcal{A})+\sup_{\rho\in\mathcal{A}}H(\rho)$$

for all ρ in $\mathcal{A}$. Hence we have $\mathcal{A}\subseteq\mathcal{K}_{H,h}$, where $H=-\log\Omega(\mathcal{A})$ and $h=\overline{C}(\mathcal{A})+\sup_{\rho\in\mathcal{A}}H(\rho)$.

By Corollary 7, boundedness of the entropy on a convex set $\mathcal{A}$ means that $\mathcal{A}$ is contained in $\mathcal{K}_{H,h}$ for some $\mathfrak{H}$-operator H with finite $\mathrm{g}(H)$. By Theorem 1, finiteness of the χ-capacity of an arbitrary set $\mathcal{A}$ means that $\mathcal{A}$ is contained in the set $\mathcal{V}_{\Omega(\mathcal{A}),\overline{C}(\mathcal{A})}$, which has the same χ-capacity and the same optimal average state.

4.2. Optimal measures. Let $\mathcal{A}$ be a closed set with finite χ-capacity. By Corollary 5, $\mathcal{A}$ is compact. Hence the set $\mathcal{P}(\mathcal{A})$ of all probability measures supported by $\mathcal{A}$ is compact in the topology of weak convergence (the Prokhorov topology) [11]. Since an arbitrary measure in $\mathcal{P}(\mathcal{A})$ can be weakly approximated by a sequence of finitely supported measures and the functional $\chi(\mu)$ is lower semicontinuous, it follows that

$$\overline{C}(\mathcal{A}) = \sup_{\mu\in\mathcal{P}(\mathcal{A})} \chi(\mu). \tag{31}$$

In other words, the supremum over all measures in $\mathcal{P}(\mathcal{A})$ coincides with the supremum over all finitely supported measures.

Definition 2. A measure μ_* in $\mathcal{P}(\mathcal{A})$ such that

$$\overline{C}(\mathcal{A}) = \chi(\mu_*) = \int_{\mathcal{A}} H\big(\rho \,\|\, \bar{\rho}(\mu_*)\big)\mu_*(d\rho)$$

is called an *optimal measure for the set* $\mathcal{A}$.

Using Theorem 1 and the generalized Donald's identity (4), we easily obtain the following generalization of Corollary 4. *If $\mathcal{A}$ is an arbitrary closed set with finite χ-capacity and μ is an arbitrary measure in $\mathcal{P}(\mathcal{A})$, then*

$$\overline{C}(\mathcal{A}) - \chi(\mu) \geqslant H\big(\bar{\rho}(\mu) \,\|\, \Omega(\mathcal{A})\big) \geqslant \frac{1}{2}\,\big\|\bar{\rho}(\mu) - \Omega(\mathcal{A})\big\|_1^2.$$

Using this inequality and Theorem 1, we can generalize the "maximal distance property" of optimal ensembles [14] to the infinite-dimensional case.

Proposition 7. *Let μ_* be an optimal measure for a closed set $\mathcal{A}$ with finite χ-capacity. Then its barycentre $\bar{\rho}(\mu_*)$ coincides with the optimal average state $\Omega(\mathcal{A})$, and we have $H\big(\rho \,\|\, \Omega(\mathcal{A})\big) = \overline{C}(\mathcal{A})$ for μ_*-almost all ρ.*

In particular, if the infimum in the definition (23) of χ-capacity is attained on some finite or countable ensemble $\{\pi_i, \rho_i\}$ (an optimal ensemble for $\mathcal{A}$), then its average state $\bar{\rho}$ coincides with the optimal average state $\Omega(\mathcal{A})$ and we have $H\big(\rho_i \,\|\, \Omega(\mathcal{A})\big) = \overline{C}(\mathcal{A})$ for all i such that $\pi_i > 0$.

Corollary 8. *Let $\mathcal{A}$ be a closed set with finite χ-capacity. Then the existence of an optimal measure for $\mathcal{A}$ implies that $\overline{C}(\mathcal{A}) \leqslant H\big(\Omega(\mathcal{A})\big)$.*

Proof. It suffices to consider the case when $H\big(\Omega(\mathcal{A})\big) < +\infty$. Using (2), the definition of an optimal measure μ_* and Proposition 7, we get

$$\overline{C}(\mathcal{A}) = \chi(\mu_*) = H\big(\bar{\rho}(\mu_*)\big) - \widehat{H}(\mu_*) \leqslant H\big(\bar{\rho}(\mu_*)\big) = H\big(\Omega(\mathcal{A})\big).$$

The corollary is proved.

This corollary provides a simple way of proving the non-existence of optimal measures for some sets of states. This will be used in part II of this paper.

The following theorem provides a sufficient condition for the existence of an optimal measure.

Theorem 2. *Let $\mathcal{A}$ be a closed set with finite χ-capacity. If one of the following conditions holds, then there is an optimal measure for $\mathcal{A}$.*

1) $H\big(\Omega(\mathcal{A})\big) < +\infty$ and $\lim_{n\to+\infty} H(\rho_n) = H\big(\Omega(\mathcal{A})\big)$ for any sequence $\{\rho_n\}$ of states in $\mathrm{co}(\mathcal{A})$ that H-converges[10] *to the state $\Omega(\mathcal{A})$.*

2) The function $\rho \mapsto H\big(\rho \,\|\, \Omega(\mathcal{A})\big)$ is continuous on $\mathcal{A}$.

The proof of this theorem is based on the following lemma.

Lemma 4. *Let $\mathcal{A}$ be a closed set with finite χ-capacity. Then there is a sequence $\{\mu_n\}$ of finitely supported measures in $\mathcal{P}(\mathcal{A})$ that weakly converges to some measure μ_* in $\mathcal{P}(\mathcal{A})$ with barycentre $\Omega(\mathcal{A})$ and satisfies*

$$\lim_{n\to+\infty} H\big(\bar{\rho}(\mu_n) \,\|\, \Omega(\mathcal{A})\big) = 0, \qquad \lim_{n\to+\infty} \chi(\mu_n) = \overline{C}(\mathcal{A}).$$

Proof. Let $\mu_n = \big\{\{\pi_i^n, \rho_i^n\}\big\}_n$ be an approximating sequence of ensembles for $\mathcal{A}$ and let $\big\{\bar{\rho}_n(\mu_n)\big\}$ be the corresponding sequence of average states. Theorem 1 implies that

$$\lim_{n\to+\infty} H\big(\bar{\rho}(\mu_n) \,\|\, \Omega(\mathcal{A})\big) = 0.$$

Since $\mathcal{A}$ is compact by Corollary 5, the set $\mathcal{P}(\mathcal{A})$ is weakly compact. Hence the sequence $\{\mu_n\}$ contains a subsequence weakly convergent to some measure μ_* in $\mathcal{P}(\mathcal{A})$. Since the map $\mu \mapsto \bar{\rho}(\mu)$ is continuous, we have $\bar{\rho}(\mu_*) = \Omega(\mathcal{A})$. Thus this subsequence has the required properties.

Proof of Theorem 2. The two conditions of the theorem provide two different ways of showing that the limit measure μ_* (introduced in Lemma 4) is an optimal measure for $\mathcal{A}$.

Let $\{\mu_n\}$ be a sequence given by Lemma 4.

Condition 1) implies that

$$\lim_{n\to+\infty} H\big(\bar{\rho}(\mu_n)\big) = H\big(\bar{\rho}(\mu_*)\big) = H\big(\Omega(\mathcal{A})\big) < +\infty.$$

Using (2) and the lower semicontinuity of the functional $\widehat{H}(\mu)$, we get

$$\limsup_{n\to+\infty} \chi(\mu_n) = \limsup_{n\to+\infty} \big(H\big(\bar{\rho}(\mu_n)\big) - \widehat{H}(\mu_n)\big) \leqslant H\big(\bar{\rho}(\mu_*)\big) - \widehat{H}(\mu_*) = \chi(\mu_*).$$

Since $\lim_{n\to+\infty} \chi(\mu_n) = \overline{C}(\mathcal{A})$ and $\chi(\mu_*) \leqslant \overline{C}(\mathcal{A})$, this inequality implies that $\chi(\mu_*) = \overline{C}(\mathcal{A})$. Hence the measure μ_* is optimal.

Condition 2), the compactness of $\mathcal{A}$ and the definition of weak convergence imply that

$$\chi(\mu_*) = \int_{\mathcal{A}} H\big(\rho \,\|\, \Omega(\mathcal{A})\big)\mu_*(d\rho) = \lim_{n\to+\infty} \int_{\mathcal{A}} H\big(\rho \,\|\, \Omega(\mathcal{A})\big)\mu_n(d\rho).$$

[10] This means that $\lim_{n\to+\infty} H\big(\rho_n \,\|\, \Omega(\mathcal{A})\big) = 0$.

Since the relative entropy is non-negative, we see from the generalized Donald identity (4) that

$$\int_{\mathcal{A}} H\bigl(\rho\,\|\,\Omega(\mathcal{A})\bigr)\mu_n(d\rho)=\chi(\mu_n)+H\bigl(\bar\rho(\mu_n)\,\|\,\Omega(\mathcal{A})\bigr)\geqslant\chi(\mu_n).$$

Since $\lim_{n\to+\infty}\chi(\mu_n)=\overline{C}(\mathcal{A})$, the last two expressions imply that $\chi(\mu_*)=\overline{C}(\mathcal{A})$. Hence the measure μ_* is optimal.

Remark 6. The conditions of Theorem 2 are essential, although not necessary. There are sets with finite χ-capacity that have no optimal measures. This may happen even for a countable closed set (a converging sequence of states) with finite χ-capacity. There are also sets with finite χ-capacity that possess an optimal measure but do not satisfy the conditions of Theorem 2. All the appropriate examples will be given in § 3 of part II of this paper.

§ 5. Appendix

Here we give a detailed proof of the properties described in Proposition 1 of the function $F_H(h)=\sup_{\rho\in\mathcal{K}_{H,h}}H(\rho)$.

First of all, the lower semicontinuity of the entropy implies that

$$\lim_{h\to+\infty}F_H(h)=\sup_{\rho\in\mathfrak{S}(\mathcal{H})}H(\rho)=+\infty$$

for any value of $\mathrm{g}(H)$ since $\overline{\bigcup_{h\in\mathbb{R}_+}\mathcal{K}_{H,h}}=\mathfrak{S}(\mathcal{H})$.

We consider the function

$$g(\lambda,h)=\sum_{k=1}^{+\infty}(h_k-h)\exp(-\lambda h_k).$$

The theorem on series depending on a parameter [7] yields that this function is differentiable at every point (λ,h) with $\lambda>\mathrm{g}(H)$, and

$$\frac{\partial g(\lambda,h)}{\partial\lambda}=\sum_{k=1}^{+\infty}h_k(h-h_k)\exp(-\lambda h_k),\qquad \frac{\partial g(\lambda,h)}{\partial h}=-\sum_{k=1}^{+\infty}\exp(-\lambda h_k). \tag{32}$$

As shown in the proof of Proposition 1, for every h in $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$ there is a unique $\lambda^*=\lambda^*(h)>\mathrm{g}(H)$ such that $g\bigl(\lambda^*(h),h\bigr)=0$. It follows from (32) that

$$\frac{\partial g(\lambda,h)}{\partial\lambda}\bigg|_{\lambda=\lambda^*(h)}=-\sum_{k=1}^{+\infty}(h_k-h)^2\exp\bigl(-\lambda^*(h)h_k\bigr)<0.$$

By the implicit function theorem, the function $\lambda^*(h)$ is differentiable on the interval $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$ and

$$\begin{aligned}\frac{d\lambda^*(h)}{dh}&=-\biggl[\frac{\partial g(\lambda,h)}{\partial\lambda}\biggr]^{-1}\frac{\partial g(\lambda,h)}{\partial h}\\&=-\biggl[\sum_{k=1}^{+\infty}(h_k-h)^2\exp\bigl(-\lambda^*(h)h_k\bigr)\biggr]^{-1}\sum_{k=1}^{+\infty}\exp\bigl(-\lambda^*(h)h_k\bigr)<0.\end{aligned}\tag{33}$$

Formula (14) means that

$$F_H(h)=\lambda^*(h)h+\log\sum_{k=1}^{+\infty}\exp\bigl(-\lambda^*(h)h_k\bigr) \tag{34}$$

for all h in $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$.

A direct calculation of derivatives using the equation $g\bigl(\lambda^*(h),h\bigr)=0$ shows that

$$\frac{dF_H(h)}{dh}=\frac{d}{dh}\biggl[\lambda^*(h)h+\log\sum_{k=1}^{+\infty}\exp\bigl(-\lambda^*(h)h_k\bigr)\biggr]=\lambda^*(h). \tag{35}$$

Thus (33) implies that

$$\frac{d^2F_H(h)}{dh^2}=\frac{d\lambda^*(h)}{dh}<0.$$

Hence the function $F_H(h)$ is strictly concave on $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$.

Suppose that $h_*(H)<+\infty$. If $h>h_*(H)$, then the part of Proposition 1 proved above yields that

$$F_H(h)=\mathrm{g}(H)h+\log\sum_{k=1}^{+\infty}\exp\bigl(-\mathrm{g}(H)h_k\bigr) \tag{36}$$

is a linear function and

$$\frac{dF_H(h)}{dh}=\mathrm{g}(H). \tag{37}$$

If $h=h_*(H)$, then $\lambda^*(h)=\mathrm{g}(H)$ by the observation in the proof of Proposition 1. Hence the representations (34) and (36) coincide in this case.

To prove that $F_H(h)$ is smooth at the point $h_*(H)$, we note that $\lambda^*(h)\to\mathrm{g}(H)$ as $h\to h_*(H)-0$. Indeed, the function $\lambda^*(h)$ is decreasing on $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$ by (33), and for every $\lambda>\mathrm{g}(H)$ there is an

$$h_\lambda=\biggl[\sum_{k=1}^{+\infty}\exp(-\lambda h_k)\biggr]^{-1}\sum_{k=1}^{+\infty}h_k\exp(-\lambda h_k)$$

such that $\lambda=\lambda^*(h_\lambda)$.

Then equations (34)–(37) imply that

$$\lim_{h\to h_*(H)-0}F_H(h)=F_H\bigl(h_*(H)\bigr),\qquad \lim_{h\to h_*(H)-0}\frac{dF_H(h)}{dh}=\frac{dF_H(h)}{dh}\bigg|_{h=h_*(H)+0}.$$

Hence the function $F_H(h)$ has a continuous derivative at the point $h_*(H)$.

To prove that $F_H(h)$ is right continuous at the point $h_{\mathrm{m}}(H)$, we note that

$$\lambda^*(h)\to+\infty\qquad\text{as}\quad h\to h_{\mathrm{m}}+0. \tag{38}$$

Indeed, the function $\lambda^*(h)$ decreases on $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$ by (33). Hence $\lambda^{\mathrm{m}}=\lim_{h\to h_{\mathrm{m}}(H)+0}\lambda^*(h)$ exists. If $\lambda^{\mathrm{m}}<+\infty$, then we pass to the limit as $h\to h_{\mathrm{m}}(H)+0$ in the identity

$$\sum_{k=1}^{+\infty}h_k\exp\bigl(-\lambda^*(h)h_k\bigr)\equiv h\sum_{k=1}^{+\infty}\exp\bigl(-\lambda^*(h)h_k\bigr),$$

which holds for all h in $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$. We get an obvious contradiction.

Let $d=\dim\mathcal{H}_{\mathrm{m}}(H)$. It is easy to see that

$$P(h)=\log\sum_{k=1}^{+\infty}\exp\bigl(-\lambda^*(h)h_k\bigr)=-\lambda^*(h)h_{\mathrm{m}}(H)+Q(h), \tag{39}$$

where

$$Q(h)=\log\biggl(d+\sum_{k>d}\exp\bigl(-\lambda^*(h)\bigl(h_k-h_{\mathrm{m}}(H)\bigr)\bigr)\biggr)$$

is a non-decreasing function on $\bigl(h_{\mathrm{m}}(H),h_*(H)\bigr)$ tending to $\log d$ as $h\to h_{\mathrm{m}}(H)+0$.

Since the function $F_H(h)$ is non-decreasing on $\bigl(h_{\mathrm{m}}(H),+\infty\bigr)$, the limit

$$\lim_{h\to h_{\mathrm{m}}(H)+0}F_H(h)\geqslant F_H\bigl(h_{\mathrm{m}}(H)\bigr).$$

exists. Thus (34) and (39) imply the existence of $\lim_{h\to h_{\mathrm{m}}(H)+0}\lambda^*(h)\bigl(h-h_{\mathrm{m}}(H)\bigr)=C<+\infty$ and that

$$\lim_{h\to h_{\mathrm{m}}(H)+0}F_H(h)=C+\log d=C+F_H\bigl(h_{\mathrm{m}}(H)\bigr).$$

Thus, to prove the right continuity of $F_H(h)$ at $h_{\mathrm{m}}(H)$, it suffices to show that $C=0$. This can be done by proving that

$$\int_{h_{\mathrm{m}}(H)}^{h''}\lambda^*(h)\,dh=\lim_{h'\to h_{\mathrm{m}}(H)+0}\int_{h'}^{h''}\lambda^*(h)\,dh<+\infty \tag{40}$$

for some $h''>h_{\mathrm{m}}(H)$. Indeed, (40) and the assumption $C>0$ imply that

$$\int_{h_{\mathrm{m}}(H)}^{h''}\bigl(h-h_{\mathrm{m}}(H)\bigr)^{-1}\,dh<+\infty.$$

This is a contradiction.

It is easy to see that $\frac{dP(h)}{dh}=-h\frac{d\lambda^*(h)}{dh}$ and, therefore,

$$\frac{dQ(h)}{dh}=-\frac{d\lambda^*(h)}{dh}\bigl(h-h_{\mathrm{m}}(H)\bigr). \tag{41}$$

Integrating (41), we obtain

$$Q(h'')-Q(h')=\lambda^*(h')\bigl(h'-h_{\mathrm{m}}(H)\bigr)-\lambda^*(h'')\bigl(h''-h_{\mathrm{m}}(H)\bigr)+\int_{h'}^{h''}\lambda^*(h)\,dh.$$

Thus (40) follows from the existence of the limits $\lim_{h'\to h_{\mathrm{m}}(H)+0}Q(h')=\log d$ and $\lim_{h'\to h_{\mathrm{m}}(H)+0}\lambda^*(h')\bigl(h'-h_{\mathrm{m}}(H)\bigr)=C<+\infty$.

It follows from the previous observation that

$$\frac{F_H(h)-F_H\bigl(h_{\mathrm{m}}(H)\bigr)}{h-h_{\mathrm{m}}(H)}\geqslant\lambda^*(h)\quad\forall\,h>h_{\mathrm{m}}(H).$$

Using (38), we get $\frac{dF_H(h)}{dh}\big|_{h=h_{\mathrm{m}}(H)+0}=+\infty$.

The author is grateful to A. S. Holevo for his help in preparing this paper.

Bibliography

[1] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics. Vol. 1. C^*- and W^*-algebras, algebras, symmetry groups, decomposition of states*, Springer, New York–Heidelberg–Berlin 1979; Russian transl., Mir, Moscow 1982.

[2] D. F. Dell'Antonio, "On the limits of sequences of normal states", *Comm. Pure Appl. Math.* **20** (1967), 413–429.

[3] M. J. Donald, "Further results on the relative entropy", *Math. Proc. Cambridge Philos. Soc.* **101** (1987), 363–373.

[4] P. Harremoës and F. Topsøe, "Maximum entropy fundamentals", *Entropy* **3**:3 (2001), 191–226.

[5] P. Harremoës, "Information topologies with applications", *Entropy, search, complexity* (I. Csiszár, G. O. H. Katona, and G. Tardos, eds.), Bolyai Soc. Math. Stud., vol. 16, Springer, New York 2007, pp. 113–150.

[6] A. D. Ioffe and V. M. Tikhomirov, *Theory of extremal problems*, Nauka, Moscow 1974; English transl., North-Holland, Amsterdam–New York–Oxford 1979.

[7] L. D. Kudryavtsev, *A course of mathematical analysis*, Vysshaya Shkola, Moscow 1988. (Russian)

[8] G. Lindblad, "Expectations and entropy inequalities for finite quantum systems", *Comm. Math. Phys.* **39**:2 (1974), 111–119.

[9] G. Lindblad, "Completely positive maps and entropy inequalities", *Comm. Math. Phys.* **40**:2 (1975), 147–151.

[10] M. Ohya and D. Petz, *Quantum entropy and its use*, Texts Monogr. Phys., Springer-Verlag, Berlin 1993.

[11] K. R. Parthasarathy, *Probability measures on metric spaces*, Probab. Math. Statist., vol. 3, Academic Press, New York–London 1967.

[12] T. A. Sarymsakov, *Introduction to quantum probability theory*, Fan, Tashkent 1985. (Russian)

[13] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels", *Phys. Rev. A* **56**:1 (1997), 131–138.

[14] B. Schumacher and M. Westmoreland, *Optimal signal ensembles*, `arXiv: quant-ph/9912122`.

[15] M. E. Shirokov, "The Holevo capacity of infinite dimensional channels and the additivity problem", *Comm. Math. Phys.* **262**:1 (2006), 137–159.

[16] M. E. Shirokov, *On entropic quantities related to the classical capacity of infinite dimensional quantum channels*, `arXiv: quant-ph/0411091`.

[17] A. S. Holevo, "Quantum coding theorems", *Uspekhi Mat. Nauk* **53**:6 (1998), 193–230; English transl., *Russian Math. Surveys* **53** (1998), 1295–1331.

[18] A. S. Holevo, "Entanglement-assisted capacities of constrained quantum channels", *Teor. Veroyatn. Primen.* **48** (2004), 359–374; English transl., *Theory Probab. Appl.* **48** (2004), 243–255.

[19] A. S. Holevo and M. E. Shirokov, "Continuous ensembles and the capacity of infinite-dimensional channels", *Teor. Veroyatn. Primen.* **50** (2006), 98–114; English transl., *Theory Probab. Appl.* **50** (2006), 86–98.

[20] A. Wehrl, "General properties of entropy", *Rev. Mod. Phys.* **50**:2 (1978), 221–260.

M. E. Shirokov
Steklov Mathematical Institute
E-mail: `msh@mi.ras.ru`

Received 23/DEC/05
Translated by THE AUTHOR