Izvestiya: *Mathematics* **76**:4 840–856

Izvestiya RAN: *Ser. Mat.* **76**:4 207–224 DOI 10.1070/IM2012v076n04ABEH002606

Stability of convex sets and applications

M. E. Shirokov

Abstract. We briefly review the results related to the notion of stability of convex sets and consider some of their applications. We prove a corollary of the stability property which enables us to develop an approximation technique for concave functions on a wide class of convex sets. This technique yields necessary and sufficient conditions for the local continuity of concave functions. We describe some examples of their applications.

Keywords: stable convex set, concave function, weak convergence of probability measures, barycentric map.

§ 1. Introduction

The notion of stability of convex sets arose in the late 1970s out of the study of questions of the continuity of convex hulls[1] of continuous functions and concave continuous functions defined on convex compact sets. The key results of that study were due to Vesterstrøm and O'Brien. In [3], these questions were related to the openness of the barycentric map, and the equivalence of the continuity of the convex hull of any continuous concave function (called the CE-*property* in [4]) and that of the convex hull of any continuous function was conjectured. This conjecture was proved in [5], where these properties were shown to be equivalent to the openness of the convex mixture map $(x,y)\mapsto \lambda x+(1-\lambda)y$. The last property of (not necessarily compact) convex sets has been studied in many papers under the name of *stability*, and convex sets possessing this property were called *stable convex sets* [6]. Relations have been established between the stability and other properties of convex sets (see [7]–[9] and the references therein).

The main result of Vesterstrøm and O'Brien does not hold for arbitrary non-compact convex sets. Nevertheless it extends to the class of so-called μ-compact convex sets, which includes many of the non-compact convex sets arising in applications [10].

An important example of a convex μ-compact set is the set of quantum states (density operators on a separable Hilbert space) [11]. The stability of this set is an efficient tool for studying the analytic properties of various characteristics of quantum systems (see § 4 in [10] and the references therein).

[1] The convex hull of a function f is the maximal convex function not exceeding f [1], [2].

This work was partially supported by the programme 'Mathematical control theory and dynamical systems' of the Russian Academy of Sciences and by RFBR (grants nos. 10-01-00139-a, 12-01-00319-a).

AMS 2010 *Mathematics Subject Classification.* 46A55, 52A07, 81P16.

It was shown in [12] that the set of quantum states possesses a property which is formally stronger than stability. This property is called *strong stability* and enables us to develop an approximation technique for concave lower semicontinuous functions on the set of quantum states and obtain a criterion for the local continuity of the von Neumann entropy (one of the basic characteristics of a quantum state). Since the proof of strong stability uses a specific structure of the set of quantum states (the purifying procedure), it is not clear how to prove this property for general stable convex sets. This is an obstacle to the direct generalization of the results in [12].

We shall prove a general property of stable convex sets (Theorem 1) which, in particular, enables us to overcome that obstacle and use the approximation technique for concave functions on μ-compact convex sets to derive necessary and sufficient conditions for the local continuity of such functions (Theorem 2).

§ 2. Preliminaries

In what follows we assume that $\mathcal{A}$ is a closed bounded convex subset of a locally convex space which is a complete separable metric space.[2] We shall write $\operatorname{extr}\mathcal{A}$ for the set of extreme points of $\mathcal{A}$, $C(\mathcal{A})$ for the set of continuous bounded functions on $\mathcal{A}$, and $\mathrm{cl}(\mathcal{B})$, $\mathrm{co}(\mathcal{B})$, $\sigma\text{-co}(\mathcal{B})$ and $\overline{\mathrm{co}}(\mathcal{B})$ for the closure, convex hull, σ-convex hull[3] and convex closure respectively of a subset $\mathcal{B}\subset\mathcal{A}$ ([1], [2]).

Let $M(\mathcal{B})$ be the set of all Borel probability measures on a closed subset $\mathcal{B}\subseteq\mathcal{A}$ endowed with the topology of weak convergence ([13], [14]). With an arbitrary measure $\mu\in M(\mathcal{B})$ we associate its barycentre (average) $\mathbf{b}(\mu)\in\overline{\mathrm{co}}(\mathcal{B})$, which is defined by the Pettis integral (see [15], [16]):

$$\mathbf{b}(\mu)=\int_{\mathcal{B}} x\,\mu(dx).$$

The barycentric map

$$M(\mathcal{B})\ni\mu\mapsto\mathbf{b}(\mu)\in\overline{\mathrm{co}}(\mathcal{B})$$

is continuous (this can easily be proved using Prokhorov's theorem [13]).

Let $M_x(\mathcal{B})$ be the convex closed subset of $M(\mathcal{B})$ consisting of all measures μ such that $\mathbf{b}(\mu)=x\in\overline{\mathrm{co}}(\mathcal{B})$.

We write $\delta(x)$ for the measure in $M(\mathcal{A})$ concentrated at a point $x\in\mathcal{A}$. The discrete measure with finitely or countably many atoms $\{x_i\}$ of weights $\{\pi_i\}$ will correspondingly be denoted by $\sum_i\pi_i\delta(x_i)$ or, briefly, $\{\pi_i,x_i\}$. For such a measure, $\mathbf{b}(\mu)=\sum_i\pi_ix_i$. Given an arbitrary (not necessarily closed) subset $\mathcal{B}\subseteq\mathcal{A}$, we write $M^a(\mathcal{B})$ (resp. $M^a_x(\mathcal{B})$) for the set of all discrete measures with atoms in $\mathcal{B}$ (resp. the subset consisting of all such measures with barycentre x).

[2]This means that the topology on $\mathcal{A}$ is defined by a countable subset of the family of seminorms defining the topology of the whole locally convex space, and $\mathcal{A}$ is separable and complete in the metric defined by this subset of seminorms.

[3]$\sigma\text{-co}(\mathcal{B})$ is the set of all countable convex combinations of points in $\mathcal{B}$.

Following [6], we introduce the main definition.

Definition 1. A convex set $\mathcal{A}$ is said to be *stable* if the map $\mathcal{A} \times \mathcal{A} \ni (x, y) \mapsto \frac{x+y}{2} \in \mathcal{A}$ is open.

This property is easily seen to be equivalent to the openness of the map $\mathcal{A} \times \mathcal{A} \ni (x, y) \mapsto \lambda x + (1 - \lambda)y \in \mathcal{A}$ for every $\lambda \in (0, 1)$ [7].

Every convex compact set in $\mathbb{R}^2$ is stable. The stability of a convex compact subset of $\mathbb{R}^3$ is equivalent to the closedness of the set of its extreme points, while in $\mathbb{R}^n$, $n > 3$, stability is stronger than the latter property [5]. A complete characterization of the stability property in $\mathbb{R}^n$ is obtained in [6]. In infinite dimensions, stability is proved for the unit balls in some Banach spaces and for the positive part of the unit ball in Banach lattices whose unit ball is stable [8]. From a physical point of view, it is essential that the set of quantum states (density operators on a separable Hilbert space) is stable [12].

A basic result about stability says that the following properties of a convex compact set $\mathcal{A}$ are equivalent.

(i) $\mathcal{A}$ is stable.

(ii) The map $M(\mathcal{A}) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ is open.

(iii) The map $M\big(\mathrm{cl}(\mathrm{extr}\, \mathcal{A})\big) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ is open.[4]

(iv) The convex hull of any continuous function on $\mathcal{A}$ is continuous.

(v) The convex hull of any concave continuous function on $\mathcal{A}$ is continuous.

An essential part of this assertion was obtained by Vesterstrøm [3], and the complete version was proved by O'Brien [5]. In what follows this result is referred to as the Vesterstrøm–O'Brien theorem. It does not hold for general non-compact convex sets, but it does hold for convex μ-compact sets ([10], Theorem 1), which are defined as follows.

Definition 2. A convex set $\mathcal{A}$ is said to be *μ-compact* if the pre-image of any compact subset of $\mathcal{A}$ under the map $\mu \mapsto \mathbf{b}(\mu)$ is a compact subset of $M(\mathcal{A})$.

Every compact set is μ-compact since the compactness of $\mathcal{A}$ implies that of $M(\mathcal{A})$ [14]. The property of μ-compactness is not purely topological: it reflects a particular relation between the topology and linear structure of a convex set. See [10] for a simple criterion for μ-compactness and further properties and examples of μ-compact sets.

The simplest examples of non-compact μ-compact stable convex sets are given by the simplex of all probability distributions with a countable set of outcomes (regarded as a subset of the Banach space ℓ_1) and its non-commutative analogue, the set of quantum states in a separable Hilbert space [11]. A more general example is the convex set of all Borel probability measures on an arbitrary complete separable metric space endowed with the topology of weak convergence (the μ-compactness and stability of this set are proved in [10], Corollary 4 and [17], Theorem 2.4 respectively).

[4]It follows from (i)–(v) that $\mathrm{cl}(\mathrm{extr}\, \mathcal{A}) = \mathrm{extr}\, \mathcal{A}$.

We also use a weakened version of μ-compactness.

Definition 3. A convex set $\mathcal{A}$ is said to be *pointwise μ-compact* if the pre-image of any point in $\mathcal{A}$ under the map $\mu \mapsto \mathbf{b}(\mu)$ is a compact subset of $M(\mathcal{A})$.

The simplest example of a pointwise μ-compact set which is not μ-compact is the positive part of the unit ball in ℓ_1 endowed with the topology of ℓ_p, $p>1$ [10]. In contrast to μ-compactness, pointwise μ-compactness is preserved under the weakening of the topology (if the set remains closed). Although some results about the properties of convex compact sets can also be proved for convex pointwise μ-compact sets (for example, Choquet's theorem; see [10], Proposition 5), the Vesterstrøm–O'Brien theorem holds for μ-compact sets but not for pointwise μ-compact sets ([10], Proposition 18).

Remark 1. Defining a metric $\mathbf{d}(\,\cdot\,,\cdot\,)$ on $\mathcal{A}$ by the formula

$$\mathbf{d}(x,y)=\sum_{k=1}^{\infty}2^{-k}\frac{\|x-y\|_k}{1+\|x-y\|_k}\,, \qquad x,y\in\mathcal{A},$$

where $\{\|\cdot\|_k\}_{k=1}^{\infty}$ is a countable family of seminorms defining the topology on $\mathcal{A}$, we easily obtain the estimate

$$\mathbf{d}(\alpha x+(1-\alpha)y,\alpha' x'+(1-\alpha')y')\leqslant 2\delta+C_{x,y}(\varepsilon)$$

for all x, y, x', y' in $\mathcal{A}$ and α, α' in $[0,1]$ such that

$$\mathbf{d}(x,x')<\delta, \quad \mathbf{d}(y,y')<\delta, \quad |\alpha-\alpha'|<\varepsilon,$$

where $C_{x,y}(\varepsilon)=\sum_{k=1}^{\infty}2^{-k}\frac{\varepsilon\|x-y\|_k}{1+\varepsilon\|x-y\|_k}$ is a function with $\lim_{\varepsilon\to+0}C_{x,y}(\varepsilon)=0$.

Remark 2. In what follows, the continuity of a function f on a subset $\mathcal{B}\subset\mathcal{A}$ means that the restriction $f\big|_{\mathcal{B}}$ of f to $\mathcal{B}$ is continuous and, in particular, finite (in contrast to the case of lower or upper semicontinuity).

§ 3. Some auxiliary results

In this section we list some results concerning μ-compact and pointwise μ-compact sets which will be used below.

Lemma 1. *Let $\mathcal{B}$ be a closed subset of a convex μ-compact set $\mathcal{A}$. Then for every x_0 in $\overline{\mathrm{co}}(\mathcal{B})$ there is a measure μ_0 in $M(\mathcal{B})$ such that $x_0=\mathbf{b}(\mu_0)$.*

Proof. Take $x_0\in\overline{\mathrm{co}}(\mathcal{B})$ and let $\{x_n\}\subset\mathrm{co}(\mathcal{B})$ be a sequence converging to x_0. For every $n\in\mathbb{N}$ there is a measure $\mu_n\in M(\mathcal{B})$ with finite support such that $x_n=\mathbf{b}(\mu_n)$. By the μ-compactness of $\mathcal{A}$, the sequence $\{\mu_n\}$ has a limit point $\mu_0\in M(\mathcal{B})$. Since the map $\mu\mapsto\mathbf{b}(\mu)$ is continuous, we have $\mathbf{b}(\mu_0)=x_0$. □

Lemma 2. *Let $\mathcal{A}$ be a convex pointwise μ-compact set such that $\mathrm{cl}(\mathrm{extr}\mathcal{A})=\mathrm{extr}\mathcal{A}$ and $\mathcal{A}=\sigma\text{-}\mathrm{co}(\mathrm{extr}\mathcal{A})$. Then any measure μ_0 in $M(\mathrm{extr}\mathcal{A})$ can be approximated by a sequence $\{\mu_n\}$ of measures in $M^a(\mathrm{extr}\mathcal{A})$ such that $\mathbf{b}(\mu_n)=\mathbf{b}(\mu_0)$ for all n.*

Proof. Consider the Choquet ordering on $M(\mathcal{A})$: $\mu \succ \nu$ means that

$$\int_{\mathcal{A}} f(x)\,\mu(dx) \geqslant \int_{\mathcal{A}} f(x)\,\nu(dx)$$

for all convex continuous bounded functions f on $\mathcal{A}$ (see [18]).

For any given measure μ_0 in $M(\mathrm{extr}\mathcal{A})$ it is easy to construct a sequence $\{\mu_n\}$ of finitely supported measures in $M(\mathcal{A})$ converging to μ_0 and satisfying $\mathbf{b}(\mu_n) = \mathbf{b}(\mu_0)$ for all n. Decomposing every atom of μ_n into a convex combination of extreme points, we obtain a measure $\widehat{\mu}_n$ in $M^a(\mathrm{extr}\mathcal{A})$ with the same barycentre. Clearly, $\widehat{\mu}_n \succ \mu_n$. By the pointwise μ-compactness of $\mathcal{A}$, the sequence $\{\widehat{\mu}_n\}_{n>0}$ is relatively compact. Hence there is a subsequence $\{\widehat{\mu}_{n_k}\}$ converging to a measure $\widehat{\mu}_0$ in $M(\mathrm{extr}\mathcal{A})$. Since $\widehat{\mu}_{n_k} \succ \mu_{n_k}$ for all k, it follows from the definition of weak convergence that $\widehat{\mu}_0 \succ \mu_0$. Hence $\widehat{\mu}_0 = \mu_0$ by the maximality of μ_0 with respect to the Choquet ordering, which follows from the coincidence of this ordering with the dilation ordering [18]. □

Lemma 3. *Let $\mathcal{A}$ be a convex μ-compact set, and let $\{\{\pi_i^n, x_i^n\}_{i=1}^m\}_n$ be a sequence of measures in $M^a(\mathcal{A})$ having no more than $m < \infty$ atoms such that the sequence $\bigl\{\sum_{i=1}^m \pi_i^n x_i^n\bigr\}_n$ of their barycentres converges to $x_0 \in \mathcal{A}$. Then there is a subsequence $\{\{\pi_i^{n_k}, x_i^{n_k}\}_{i=1}^m\}_k$ converging to some measure*[5] *$\{\pi_i^0, x_i^0\}_{i=1}^m$ with barycentre x_0 in the following sense*:

$$\lim_{k\to+\infty} \pi_i^{n_k} = \pi_i^0, \quad \pi_i^0 > 0 \quad \Rightarrow \quad \lim_{k\to+\infty} x_i^{n_k} = x_i^0, \quad i = 1,\dots,m.$$

Proof. It suffices to note that the μ-compactness of $\mathcal{A}$ implies the relative compactness of the sequence $\{\{\pi_i^n, x_i^n\}_{i=1}^m\}_n$ and that the set of measures with no more than m atoms is a closed subset of $M(\mathcal{A})$. □

The following proposition describes an important property of μ-compact sets.

Proposition 1. *Let $\mathcal{A}$ be a convex μ-compact set, and let f be an upper-semicontinuous upper-bounded function on a closed subset $\mathcal{B} \subset \mathcal{A}$. Then the function*

$$\hat{f}^{\mu}_{\mathcal{B}}(x) = \sup_{\mu \in M_x(\mathcal{B})} \int_{\mathcal{B}} f(y)\,\mu(dy) \tag{1}$$

is upper-semicontinuous on $\overline{\mathrm{co}}(\mathcal{B})$. For every $x \in \overline{\mathrm{co}}(\mathcal{B})$ the supremum in the definition of $\hat{f}^{\mu}_{\mathcal{B}}(x)$ is achieved at some measure in $M_x(\mathcal{B})$.

This property yields μ-compact generalizations of several results concerning convex compact sets ([10], Proposition 6, Corollary 2). It does not hold if we relax the μ-compactness to pointwise μ-compactness ([10], Proposition 7). One can show that the upper boundedness of f is essential in Proposition 1. A proof of Proposition 1 is given in the Appendix.

Another important technical result is described in the following proposition.

[5] It is not asserted that $x_i^0 \neq x_j^0$ for all $i \neq j$.

Proposition 2. *Let $\mathcal{A}$ be a convex μ-compact set,*[6] *and let f be a lower-semicontinuous lower-bounded function on a closed subset $\mathcal{B}\subseteq\mathcal{A}$.*

A) *If the map $M(\mathcal{B})\ni\mu\mapsto\mathbf{b}(\mu)\in\overline{\mathrm{co}}(\mathcal{B})$ is open, then the function $\hat{f}^{\mu}_{\mathcal{B}}$ defined in (1) is lower-semicontinuous on $\overline{\mathrm{co}}(\mathcal{B})$.*

B) *If the map $M^a(\mathcal{B})\ni\mu\mapsto\mathbf{b}(\mu)\in\sigma\text{-}\mathrm{co}(\mathcal{B})$ is open, then the function*

$$\hat{f}^{\sigma}_{\mathcal{B}}(x)=\sup_{\mu\in M^a_x(\mathcal{B})}\int_{\mathcal{B}}f(y)\,\mu(dy)=\sup_{\{\pi_i,x_i\}\in M^a_x(\mathcal{B})}\sum_i\pi_i f(x_i)$$

is lower-semicontinuous on $\sigma\text{-}\mathrm{co}(\mathcal{B})$. If, in addition, $\sigma\text{-}\mathrm{co}(\mathcal{B})=\overline{\mathrm{co}}(\mathcal{B})$, then $\hat{f}^{\sigma}_{\mathcal{B}}$ coincides with the function $\hat{f}^{\mu}_{\mathcal{B}}$ defined in (1).

A proof of Proposition 2 is given in the Appendix.

Remark 3. One can show that $\hat{f}^{\sigma}_{\mathcal{B}}$ and $\hat{f}^{\mu}_{\mathcal{B}}$ may not coincide for a bounded upper-semicontinuous function f on a closed subset $\mathcal{B}$ of a convex μ-compact set $\mathcal{A}$ with $\sigma\text{-}\mathrm{co}(\mathcal{B})=\overline{\mathrm{co}}(\mathcal{B})$, and that Proposition 1 does not hold for the function $\hat{f}^{\sigma}_{\mathcal{B}}$.

Propositions 1, 2 have an obvious corollary.

Corollary 1. *Let $\mathcal{B}$ be a closed subset of a convex μ-compact set $\mathcal{A}$.*

A) *If $\mathcal{A}=\overline{\mathrm{co}}(\mathcal{B})$ and the map $M(\mathcal{B})\ni\mu\mapsto\mathbf{b}(\mu)\in\mathcal{A}$ is open, then $\hat{f}^{\mu}_{\mathcal{B}}\in C(\mathcal{A})$ for any function $f\in C(\mathcal{B})$.*

B) *If $\mathcal{A}=\sigma\text{-}\mathrm{co}(\mathcal{B})$ and the map $M^a(\mathcal{B})\ni\mu\mapsto\mathbf{b}(\mu)\in\mathcal{A}$ is open, then $\hat{f}^{\sigma}_{\mathcal{B}}=\hat{f}^{\mu}_{\mathcal{B}}\in C(\mathcal{A})$ for any function $f\in C(\mathcal{B})$.*

If $\mathcal{A}$ is a convex stable μ-compact set, then the set $\mathrm{extr}\mathcal{A}$ is closed and the surjection $M(\mathrm{extr}\mathcal{A})\ni\mu\mapsto\mathbf{b}(\mu)\in\mathcal{A}$ is open by the generalized Vesterstrøm–O'Brien theorem (Theorem 1 in [10]). Hence Corollary 1, A) shows that an arbitrary function f in $C(\mathrm{extr}\mathcal{A})$ has a continuous bounded concave extension $\hat{f}^{\mu}_{\mathrm{extr}\mathcal{A}}$ to the set $\mathcal{A}$. This property does not hold if we relax the μ-compactness to pointwise μ-compactness (this can be shown by modifying Example 1 in [10]).

§ 4. On a property of stable sets

Given an arbitrary subset $\mathcal{A}_1$ of a convex set $\mathcal{A}$, we consider a monotone family of subsets

$$\mathcal{A}_k=\biggl\{\sum_{i=1}^{k}\pi_i x_i\mid\{\pi_i\}\in\mathfrak{P}_k,\ \{x_i\}\subset\mathcal{A}_1\biggr\},\qquad k\in\mathbb{N},\tag{2}$$

where $\mathfrak{P}_k$ is the simplex of all probability distributions with k outcomes.

The following property of stable sets plays an essential role in this paper.

Theorem 1. *Let $\mathcal{A}_1$ be a subset of a stable convex set $\mathcal{A}$ such that $\mathcal{A}=\sigma\text{-}\mathrm{co}(\mathcal{A}_1)$, and let $\{\mathcal{A}_k\}$ be the family of subsets defined in (2). If the map*

$$M^a(\mathcal{A}_k)\ni\mu\mapsto\mathbf{b}(\mu)\in\mathcal{A}\tag{3}$$

is open for $k=1$, then it is open for every $k\in\mathbb{N}$.

[6] The μ-compactness is used only to guarantee that $\mathbf{b}(M(\mathcal{B}))=\overline{\mathrm{co}}(\mathcal{B})$ (by Lemma 1).

Note that the condition $\mathcal{A}=\sigma\text{-co}(\mathcal{A}_1)$ means that (3) is surjective for every k.

Proof of Theorem 1. The proof consists of two steps.

1. We fix any k, an arbitrary finitely supported measure $\mu_0=\sum_{i=1}^m\pi_i\delta(x_i)$, where $\{x_i\}_{i=1}^m\subset\mathcal{A}_k$, $\{\pi_i\}_{i=1}^m\in\mathfrak{P}_m$, $m\in\mathbb{N}$, and any sequence $\{x^n\}\subset\mathcal{A}$ converging to $x^0=\sum_{i=1}^m\pi_ix_i$. We claim that one can find a subsequence $\{x^{n_l}\}$ and a sequence $\{\mu_l\}\subset M^a(\mathcal{A}_k)$ such that $\lim_{l\to+\infty}\mu_l=\mu_0$ and $\mathbf{b}(\mu_l)=x^{n_l}$ for all l.

For $k=1$ this follows from the openness of the map $M^a(\mathcal{A}_1)\ni\mu\mapsto\mathbf{b}(\mu)$. We assume that the assertion holds for some k and show that it holds for $k+1$.

Write $\mu_0=\sum_{i=1}^m\pi_i\delta(x_i)\in M^a(\mathcal{A}_{k+1})$, where $\pi_i>0$ for all i and $\{x_i\}_{i=1}^m\not\subseteq\mathcal{A}_k$, and let $\{x^n\}$ be a sequence converging to $x^0=\sum_{i=1}^m\pi_ix_i$. There is no loss of generality in assuming that $x_i=\alpha_iy_i+(1-\alpha_i)z_i$ for every $i=1,\dots,m$, where $y_i\in\mathcal{A}_k$, $z_i\in\mathcal{A}_1$ and $\alpha_i\in(0,1)$. Hence $x^0=\eta y^0+(1-\eta)z^0$, where

$$\eta=\sum_{i=1}^m\alpha_i\pi_i\in(0,1),\qquad y^0=\eta^{-1}\sum_{i=1}^m\alpha_i\pi_iy_i\in\mathcal{A},$$
$$z^0=(1-\eta)^{-1}\sum_{i=1}^m(1-\alpha_i)\pi_iz_i\in\mathcal{A}.$$

Since $\mathcal{A}$ is stable, the sequence $\{x^n\}$ can be replaced by a subsequence with the property that $x^n=\eta y^n+(1-\eta)z^n$ for some sequences $\{y^n\}\subset\mathcal{A}$ and $\{z^n\}\subset\mathcal{A}$ converging to y^0 and z^0 respectively. By the inductive hypothesis we can assume (again passing to subsequences) that there are sequences $\{\nu_n\}\subset M^a(\mathcal{A}_k)$ and $\{\zeta_n\}\subset M^a(\mathcal{A}_1)$ converging respectively to the measures

$$\nu_0\doteq\eta^{-1}\sum_{i=1}^m\alpha_i\pi_i\delta(y_i),\qquad\zeta_0\doteq(1-\eta)^{-1}\sum_{i=1}^m(1-\alpha_i)\pi_i\delta(z_i)$$

such that $\mathbf{b}(\nu_n)=y^n$ and $\mathbf{b}(\zeta_n)=z^n$ for all n.

By the definition of weak convergence, for every N and any sufficiently small[7] $\varepsilon>0$, $\delta>0$ there is an $\bar n>N$ such that

$$\nu_{\bar n}=\sum_{i=1}^m\nu_{\bar n}^i+\nu_{\bar n}^r,\qquad\zeta_{\bar n}=\sum_{i=1}^m\zeta_{\bar n}^i+\zeta_{\bar n}^r,\tag{4}$$

where $\nu_{\bar n}^i$ and $\zeta_{\bar n}^i$ are non-zero measures with finite supports lying in $U_\delta(y_i)$ and $U_\delta(z_i)$ respectively such that

$$|\nu_{\bar n}^i(\mathcal{A})-\eta^{-1}\alpha_i\pi_i|<\eta^{-1}\varepsilon\pi_i,\quad|\zeta_{\bar n}^i(\mathcal{A})-(1-\eta)^{-1}(1-\alpha_i)\pi_i|<(1-\eta)^{-1}\varepsilon\pi_i,\tag{5}$$

all the atoms of the measures $\nu_{\bar n}^i$ and $\zeta_{\bar n}^i$ have rational weights, and

$$\nu_{\bar n}^r(\mathcal{A})<\eta^{-1}\varepsilon,\quad\zeta_{\bar n}^r(\mathcal{A})<(1-\eta)^{-1}\varepsilon.\tag{6}$$

[7] Here δ is assumed to be so small that the δ-neighbourhoods of distinct points of the sets $\{y_i\}_{i=1}^m$ and $\{z_i\}_{i=1}^m$ are disjoint.

The existence of a representation (4) is obvious when the sets $\{y_i\}_{i=1}^m$ and $\{z_i\}_{i=1}^m$ consist of distinct points. If these sets have coinciding points, then the existence of such a representation can be shown by 'splitting' the atoms of $\nu_{\bar{n}}$ and $\zeta_{\bar{n}}$ as follows. Assume, for example, that $y_1 = y_2 = \cdots = y_p = y$. Then the component $\sum_t \lambda_t \delta(y_t)$ of the measure $\nu_{\bar{n}}$ whose atoms lie in $U_\delta(y)$ can be 'represented' in the form

$$\sum_t \lambda_t \delta(y_t) = \sum_t \gamma_1 \lambda_t \delta(y_t) + \cdots + \sum_t \gamma_p \lambda_t \delta(y_t),$$

where $\gamma_i = \alpha_i \pi_i/(\alpha_1\pi_1 + \cdots + \alpha_p\pi_p)$ and the measure $\nu^i_{\bar{n}}$ is 'constructed' using the measure $\gamma_i \sum_t \lambda_t \delta(y_t)$.

For given i let

$$\nu^i_{\bar{n}} = \sum_{j=1}^{n_i^y} \frac{p^y_{ij}}{q_i} \delta(y_{ij}), \qquad \zeta^i_{\bar{n}} = \sum_{j=1}^{n_i^z} \frac{p^z_{ij}}{q_i} \delta(z_{ij}),$$

where p^*_* and q_* are positive integers. There are positive integers P_i, Q^y_i and Q^z_i such that

$$\nu^i_{\bar{n}}(\mathcal{A}) = \sum_{j=1}^{n_i^y} \frac{p^y_{ij}}{q_i} = \frac{P_i}{Q^y_i}\,, \qquad \zeta^i_{\bar{n}}(\mathcal{A}) = \sum_{j=1}^{n_i^z} \frac{p^z_{ij}}{q_i} = \frac{P_i}{Q^z_i}\,.$$

Let $d^y_i = (q_i Q^y_i)^{-1}$ and $d^z_i = (q_i Q^z_i)^{-1}$. Using the 'decomposition'

$$\frac{p^y_{ij}}{q_i}\delta(y_{ij}) = \underbrace{d^y_i \delta(y_{ij}) + \cdots + d^y_i \delta(y_{ij})}_{p^y_{ij} Q^y_i},$$

we obtain a representation $\nu^i_{\bar{n}} = \sum_{l=1}^{P_i q_i} d^y_i \delta(\bar{y}^l_i)$, where $\{\bar{y}^l_i\}_l$ is a set of $P_i q_i$ points (possibly, coinciding) belonging to $U_\delta(y_i) \cap \mathcal{A}_k$. We similarly obtain a representation $\zeta^i_{\bar{n}} = \sum_{l=1}^{P_i q_i} d^z_i \delta(\bar{z}^l_i)$, where $\{\bar{z}^l_i\}_l$ is a set of $P_i q_i$ points belonging to $U_\delta(z_i) \cap \mathcal{A}_1$.

Consider the measure

$$\mu_{\bar{n}} = \sum_{i=1}^m \sum_{l=1}^{P_i q_i} (\eta d^y_i + (1-\eta) d^z_i)\delta(\bar{x}^l_i) + \eta \nu^r_{\bar{n}} + (1-\eta)\zeta^r_{\bar{n}}, \qquad \bar{x}^l_i = \frac{\eta d^y_i \bar{y}^l_i + (1-\eta) d^z_i \bar{z}^l_i}{\eta d^y_i + (1-\eta) d^z_i}\,,$$

having barycentre $\eta y^{\bar{n}} + (1-\eta) z^{\bar{n}} = x^{\bar{n}}$ and belonging to $M^a(\mathcal{A}_{k+1})$. Since

$$\bar{\alpha}_i = \frac{\eta d^y_i}{\eta d^y_i + (1-\eta) d^z_i} = \frac{\frac{\eta P_i}{Q^y_i \pi_i}}{\frac{\eta P_i}{Q^y_i \pi_i} + \frac{(1-\eta) P_i}{Q^z_i \pi_i}}$$

and $\left|\frac{\eta P_i}{Q^y_i \pi_i} - \alpha_i\right| < \varepsilon$ and $\left|\frac{(1-\eta)P_i}{Q^z_i \pi_i} - (1-\alpha_i)\right| < \varepsilon$ by (5), we easily see that $|\bar{\alpha}_i - \alpha_i| < 6\varepsilon$ (if $\varepsilon < \min\left\{\frac{1}{4}, \alpha_i, 1-\alpha_i\right\}$ for all i). Hence $\bar{x}^l_i = \bar{\alpha}_i \bar{y}^l_i + (1-\bar{\alpha}_i)\bar{z}^l_i \in U_{\delta(i)}(x_i)$ for all $i = 1, \ldots, m$ and $l = 1, \ldots, P_i q_i$, where $\delta(i) = 2\delta + C_{y_i, z_i}(6\varepsilon)$ (see Remark 1). Since $P_i q_i(\eta d^y_i + (1-\eta) d^z_i) = \eta \frac{P_i}{Q^y_i} + (1-\eta)\frac{P_i}{Q^z_i}$, we easily deduce from (5) and (6) that

$$|\mu_{\bar{n}}(U_{\delta(i)}(x_i)) - \pi_i| \leqslant 4\varepsilon \tag{7}$$

provided that $U_{\delta(i)}(x_i) \cap U_{\delta(i')}(x_{i'}) = \varnothing$ for all $i \neq i'$.

For every positive integer l we put $n_l = \bar{n}$ and $\mu_l = \mu_{\bar{n}}$, where $\bar{n}$ and $\mu_{\bar{n}}$ are obtained by the above construction with $N = l$ and $\varepsilon = \delta = \frac{1}{l}$. Then $\mathbf{b}(\mu_l) = x^{n_l}$ and the weak convergence of the sequence $\{\mu_l\}$ to μ_0 follows from (7).

2. Let $\mu_0 = \sum_{i=1}^{\infty} \pi_i \delta(x_i)$ be an arbitrary measure in $M^a(\mathcal{A}_k)$ and let $\{x^n\} \subset \mathcal{A}$ be a sequence converging to $x^0 = \sum_{i=1}^{\infty} \pi_i x_i$. For every positive integer m we put $\mu_0^m = (\lambda_m)^{-1} \sum_{i=1}^{m} \pi_i \delta(x_i)$, where $\lambda_m = \sum_{i=1}^{m} \pi_i$. We also put $z_m^0 = (1-\lambda_m)^{-1} \sum_{i>m} \pi_i x_i$.

Since the sequence $\{\mu_0^m\}_m$ converges to μ_0, for every positive integer l there is[8] an m_l such that $\mu_0^{m_l} \in U_{1/l}(\mu_0)$ and $\lambda_{m_l} > 1 - \frac{1}{l}$. We have $x^0 = \lambda_{m_l} \mathbf{b}(\mu_0^{m_l}) + (1-\lambda_{m_l}) z_{m_l}^0$. Since $\mathcal{A}$ is stable, we can assume (passing from the sequence $\{x^n\}$ to a subsequence) that there are sequences $\{y^n\} \subset \mathcal{A}$ and $\{z^n\} \subset \mathcal{A}$ converging to $\mathbf{b}(\mu_0^{m_l})$ and $z_{m_l}^0$ respectively such that $x^n = \lambda_{m_l} y^n + (1-\lambda_{m_l}) z^n$.

By the first step of the proof, we can assume (again passing to subsequences) that there is a sequence $\{\mu_n\} \subset M^a(\mathcal{A}_k)$ converging to $\mu_0^{m_l}$ and satisfying $\mathbf{b}(\mu_n) = y^n$ for all n. Hence there is an $n_l > l$ such that $\mu_{n_l} \in U_{1/l}(\mu_0^{m_l}) \subset U_{2/l}(\mu_0)$. We put

$$\mu_l = \lambda_{m_l} \mu_{n_l} + (1-\lambda_{m_l}) \nu_{n_l},$$

where ν_{n_l} is any measure in $M^a(\mathcal{A}_1)$ with $\mathbf{b}(\nu_{n_l}) = z^{n_l}$.

It is easy to see that the sequence $\{\mu_l\}$ lies in $M^a(\mathcal{A}_k)$ and converges to μ_0, while $\mathbf{b}(\mu_l) = \lambda_{m_l} y^{n_l} + (1-\lambda_{m_l}) z^{n_l} = x^{n_l}$ for all l by construction. □

Theorem 1 and the μ-compact version of the Vesterstrøm–O'Brien theorem (Theorem 1 in [10]) yield the following result.

Corollary 2. *Let $\mathcal{A}$ be a convex μ-compact set such that $\mathcal{A} = \sigma\text{-co}(\text{extr}\mathcal{A})$, and let $\{\mathcal{A}_k\}$ be the family of subsets defined in (2) with $\mathcal{A}_1 = \text{extr}\mathcal{A}$. Then $\mathcal{A}$ is stable if and only if the map (3) is open for every $k \in \mathbb{N}$.*

Proof. By the Vesterstrøm–O'Brien theorem, the stability of the convex μ-compact set $\mathcal{A}$ is equivalent to the openness of the map $M(\text{extr}\mathcal{A}) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$, which is in its turn equivalent to the openness of the map $M^a(\text{extr}\mathcal{A}) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ by Lemma 2. □

Remark 4. Since the subset $\mathcal{A}_1 = \text{extr}\mathcal{A}$ is closed for every μ-compact stable set $\mathcal{A}$, the family $\{\mathcal{A}_k\}$ consists of closed subsets by Lemma 3. The map

$$M(\mathcal{A}_k) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A} \tag{8}$$

is surjective for every $k \in \mathbb{N}$ by Proposition 5 in [10] and open for $k = 1$ by the generalized Vesterstrøm–O'Brien theorem. In [12], the openness of the map (8) for every $k \in \mathbb{N}$ is called *strong stability* and is shown to hold for the set of quantum states (the proof uses special properties of the set of quantum states such as the purification procedure). Corollary 2 shows that a stable μ-compact set $\mathcal{A}$ with $\mathcal{A} = \sigma\text{-co}(\text{extr}\mathcal{A})$ is strongly stable if the set $M_x^a(\mathcal{A}_k)$ is dense in $M_x(\mathcal{A}_k)$ for every $x \in \mathcal{A}$ and all $k \in \mathbb{N}$ (for $k = 1$ this property holds by Lemma 2).

[8] The set $M(\mathcal{A})$ may be regarded as a metric space [14].

Question 1. Does the stability of an arbitrary μ-compact set $\mathcal{A}$ imply the openness of the map (8) for every $k\in\mathbb{N}$?

A positive answer would enable us to generalize the constructions and results of the next section by removing the condition $\mathcal{A}=\sigma\text{-co}(\mathcal{A}_1)$.

§ 5. The approximation of concave functions and local continuity conditions

Let $\mathcal{A}_1$ be a subset of a convex set $\mathcal{A}$ such that $\mathcal{A}=\sigma\text{-co}(\mathcal{A}_1)$ and $\{\mathcal{A}_k\}$ the family of subsets defined in (2).

For every concave non-negative function f on $\mathcal{A}$ we consider a monotone sequence of concave functions

$$\mathcal{A}\ni x\mapsto f_k(x)=\sup_{\{\pi_i,x_i\}\in M_x^a(\mathcal{A}_k)}\sum_i\pi_i f(x_i),\qquad k=1,2,\dots,$$

such that $f_k\leqslant f$ and $f_k|_{\mathcal{A}_k}=f|_{\mathcal{A}_k}$. These relations follow from the discrete Jensen inequality for f.

Clearly, $f_*\doteq\sup_k f_k$ is a concave function on $\mathcal{A}$ and we have

$$f_*\leqslant f,\qquad f_*\big|_{\mathcal{A}_*}=f\big|_{\mathcal{A}_*},\quad \mathcal{A}_*=\bigcup_{k=1}^{\infty}\mathcal{A}_k. \tag{9}$$

One can show that under certain conditions f and f_* coincide.

Lemma 4. *If a concave non-negative function f is lower semicontinuous on $\mathcal{A}$, then $f_*=f$.*

Proof. An arbitrary point $x_0\in\mathcal{A}$ can be represented in the form $x_0=\sum_{i=1}^{\infty}\pi_i y_i$, where $\{\pi_i\}\in\mathfrak{P}_{+\infty}$ and $\{y_i\}\in\mathcal{A}_1$. We put $x_n=\lambda_n^{-1}\sum_{i=1}^n\pi_i y_i$ and $y_n=(1-\lambda_n)^{-1}\sum_{i>n}\pi_i y_i$, where $\lambda_n=\sum_{i=1}^n\pi_i$. The sequence $\{x_n\}$ lies in $\mathcal{A}_*$ and converges to x_0.

For every n we have $x_0=\lambda_n x_n+(1-\lambda_n)y_n$ and hence $f_*(x_0)\geqslant\lambda_n f_*(x_n)=\lambda_n f(x_n)$ by the concavity and non-negativity of f_* and the formulae (9). Thus $\limsup_{n\to+\infty}f(x_n)\leqslant f_*(x_0)$ and hence $f(x_0)\leqslant f_*(x_0)$ by the lower semicontinuity of f. This and (9) yield that $f(x_0)=f_*(x_0)$. □

If $\mathcal{A}$ is μ-compact and $\mathcal{A}_1$ is closed, then all the subsets $\mathcal{A}_k$ are closed by Lemma 3. We make the following assumption.

($*$) *The restriction of f to $\mathcal{A}_k$ is continuous for every k.*

This assumption (with $\mathcal{A}_1=\operatorname{extr}\mathcal{A}$) is motivated by applications (see Examples 1, 2 below and § 6 in [19]).

Remark 5. The assumption ($*$) implies that the restriction of f to $\mathcal{A}_k$ is bounded for every k. Indeed, if $\{x_n\}\subset\mathcal{A}_k$ is a sequence with $\lim_{n\to+\infty}f(x_n)=+\infty$, then the sequence $\{\lambda_n x_n+(1-\lambda_n)y_0\}\subset\mathcal{A}_{k+1}$, where y_0 is any point in $\mathcal{A}_1$ and $\lambda_n=(f(x_n))^{-1}$, converges to y_0 (since $\mathcal{A}$ is bounded). Since f is concave, we have

$$\liminf_{n\to+\infty}f(\lambda_n x_n+(1-\lambda_n)y_0)\geqslant\liminf_{n\to+\infty}\bigl(\lambda_n f(x_n)+(1-\lambda_n)f(y_0)\bigr)=1+f(y_0),$$

contrary to the assumption ($*$).

Remark 5 enables us to deduce the following proposition (showing that the sequence $\{f_k\}$ is useful in problems of approximation) from Theorem 1 and Corollary 1, B).

Proposition 3. *Let $\mathcal{A}_1$ be a closed subset of a μ-compact convex stable set $\mathcal{A}$ such that $\mathcal{A} = \sigma\text{-co}(\mathcal{A}_1)$ and the map $M^a(\mathcal{A}_1) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ is open. For every concave non-negative function f on $\mathcal{A}$ satisfying condition $(*)$, the sequence $\{f_k\}$ consists of continuous bounded functions.*

By the generalized Vesterstrøm–O'Brien theorem (Theorem 1 in [10]) and Lemma 2, the set $\operatorname{extr}\mathcal{A}$ is closed and the map $M^a(\operatorname{extr}\mathcal{A}) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ is open for every μ-compact stable set $\mathcal{A}$ with $\mathcal{A} = \sigma\text{-co}(\operatorname{extr}\mathcal{A})$. Thus Proposition 3 yields the following result.

Corollary 3. *Let $\mathcal{A}$ be a μ-compact stable convex set such that $\mathcal{A} = \sigma\text{-co}(\operatorname{extr}\mathcal{A})$, and let $\{\mathcal{A}_k\}$ be the family of subsets defined in (2) with $\mathcal{A}_1 = \operatorname{extr}\mathcal{A}$. For every concave non-negative function f on $\mathcal{A}$ satisfying condition $(*)$, the sequence $\{f_k\}$ consists of continuous bounded functions.*

Under the hypotheses of Proposition 3 (Corollary 3), Lemma 4 shows that the monotone sequence $\{f_k\}$ of continuous bounded functions converges pointwise to f if and only if f is lower semicontinuous.

Example 1. The Shannon entropy is the concave lower-semicontinuous non-negative function

$$S(\{x^j\}_{j=1}^{\infty}) = -\sum_{j=1}^{\infty} x^j \ln x^j$$

on the set $\mathfrak{P}_{+\infty} = \{\{x^j\}_{j=1}^{\infty} \in \ell_1 \mid x^j \geqslant 0 \ \forall\, j,\ \sum_{j=1}^{\infty} x^j = 1\}$ of all probability distributions with a countable set of outcomes [20]. This function takes the value $+\infty$ on a dense subset of $\mathfrak{P}_{+\infty}$.

As described in §2, the convex set $\mathfrak{P}_{+\infty}$ is stable and μ-compact. The set $\operatorname{extr}\mathfrak{P}_{+\infty}$ consists of degenerate probability distributions (sequences having 1 at some place and zeros elsewhere). Clearly, $\mathfrak{P}_{+\infty} = \sigma\text{-co}(\operatorname{extr}\mathfrak{P}_{+\infty})$ and, for every $k \in \mathbb{N}$, the function $x \mapsto S(x)$ has a continuous restriction to the set

$$\mathfrak{P}_{+\infty}^{k} = \left\{ \sum_{i=1}^{k} \pi_i x_i \mid \{\pi_i\} \in \mathfrak{P}_k,\ \{x_i\} \subset \operatorname{extr}\mathfrak{P}_{+\infty} \right\}$$

of all probability distributions having no more than k non-zero elements. By Corollary 3, the concave function

$$\mathfrak{P}_{+\infty} \ni x \mapsto S_k(x) = \sup_{\{\pi_i, x_i\} \in M_x^a(\mathfrak{P}_{+\infty}^{k})} \sum_i \pi_i S(x_i),$$

which coincides with the Shannon entropy S on $\mathfrak{P}_{+\infty}^{k}$, is continuous for every k. By Lemma 4, the monotone sequence $\{S_k\}$ converges pointwise to the Shannon entropy S on $\mathfrak{P}_{+\infty}$.

The sequence $\{f_k\}$ can be used to derive local continuity conditions for f.

Theorem 2. *Let $\mathcal{A}$ be a convex μ-compact set, $\mathcal{A}_1$ a closed subset of $\mathcal{A}$ with $\mathcal{A} = \sigma\text{-co}(\mathcal{A}_1)$, and f a concave non-negative function on $\mathcal{A}$ satisfying condition* $(*)$. *Suppose that one of the following conditions holds.*

a) *$\mathcal{A}$ is stable and the map $M^a(\mathcal{A}_1) \ni \mu \mapsto \mathbf{b}(\mu) \in \mathcal{A}$ is open.*

b) *f is lower semicontinuous.*

Then a sufficient condition for the continuity of f on a given subset $\mathcal{B} \subseteq \mathcal{A}$ is given by the relation

$$\lim_{k\to+\infty} \sup_{x\in\mathcal{B}} \Delta_k(x|f) = 0, \quad \textit{where} \ \ \Delta_k(x|f) = \inf_{\{\pi_i,x_i\}\in M_x^a(\mathcal{A}_k)} \Big[f(x) - \sum_i \pi_i f(x_i) \Big]. \tag{10}$$

If both conditions a) *and* b) *hold and $\mathcal{B}$ is compact, then* (10) *is a necessary and sufficient condition for the continuity of f on $\mathcal{B}$.*

Condition a) *always holds if $\mathcal{A}$ is stable and $\mathcal{A}_1 = \operatorname{extr}\mathcal{A}$.*

Remark 6. Since $\Delta_k(x|f) = f(x) - f_k(x)$, the condition (10) means the uniform convergence of the sequence $\{f_k\}$ to f on $\mathcal{B}$.

Remark 7. The application of the sufficient condition for continuity in Theorem 2 is based on finding a convenient upper bound for the function in square brackets in (10) (see Example 2). The necessity of this condition enables us to obtain assertions of the form 'if f is continuous on a subset $\mathcal{B} \subseteq \mathcal{A}$, then f (or some function related to f) is continuous on any subset $\mathcal{B}' \subseteq \mathcal{A}$ obtained from $\mathcal{B}$ by operations that preserve the infinitesimality of $\Delta_k(x|f)$' (see § 5 in [12]).

Proof of Theorem 2. If condition a) holds, then $f_k \in C(\mathcal{A})$ for all k by Proposition 3. Thus, by Remark 6, the continuity of f on $\mathcal{B}$ follows from (10). If, in addition, f is lower semicontinuous and $\mathcal{B}$ is compact, then Lemma 4 and Dini's lemma show that condition (10) is equivalent to the continuity of f on $\mathcal{B}$.

If condition b) holds, then to prove the continuity of f on $\mathcal{B}$ it suffices to show that f is upper semicontinuous and bounded on $\mathcal{B}$. Since all the subsets $\mathcal{A}_k$ are closed by Lemma 3, we can consider the sequence of functions

$$\mathcal{A} \ni x \mapsto f_k^{\mu}(x) = \sup_{\mu\in M_x(\mathcal{A}_k)} \int_{\mathcal{A}_k} f(y)\,\mu(dy), \qquad k = 1, 2, \dots$$

(the lower semicontinuity of f guarantees its measurability). Since f satisfies Jensen's integral inequality (see Remark 8 in the Appendix), we have

$$f_k \leqslant f_k^{\mu} \leqslant f, \qquad f_k^{\mu}\big|_{\mathcal{A}_k} = f\big|_{\mathcal{A}_k},$$

while Remark 5 and Proposition 1 show that all the functions in this sequence are upper semicontinuous and bounded. Hence, by Remark 6, the upper semicontinuity and boundedness of f on $\mathcal{B}$ follow from condition (10).

The last assertion in Theorem 2 follows from the μ-compact version of the Vesterstrøm–O'Brien theorem (Theorem 1 in [10]) and Lemma 2. □

Example 2. Returning to Example 1, we will use Theorem 2 to obtain a criterion for the local continuity of the Shannon entropy. If $f = S$, then the expression in the square brackets in (10) can be represented as

$$S(x) - \sum_i \pi_i S(x_i) = \sum_i \pi_i S(x_i \| x),$$

where $S(\cdot \| \cdot)$ is the relative entropy (the Kullback–Leibler distance [20]), which is defined for arbitrary probability distributions $x = \{x^j\}_{j=1}^{\infty}$ and $y = \{y^j\}_{j=1}^{\infty}$ in $\mathfrak{P}_{+\infty}$ by the formula

$$S(x \| y) = \begin{cases} \sum_{i=1}^{\infty} x^j \ln(x^j/y^j), & \{y^j = 0\} \Rightarrow \{x^j = 0\}, \\ +\infty & \text{otherwise.} \end{cases}$$

Theorem 2 gives the following criterion for the local continuity of the Shannon entropy. *The function $x \mapsto S(x)$ is continuous on a compact subset $\mathfrak{P} \subseteq \mathfrak{P}_{+\infty}$ if and only if*

$$\lim_{k\to+\infty} \sup_{x\in\mathfrak{P}} \Delta_k(x|S) = 0, \quad \text{where} \quad \Delta_k(x|S) = \inf_{\{\pi_i, x_i\}\in M_x^a(\mathfrak{P}_{+\infty}^k)} \sum_i \pi_i S(x_i \| x). \tag{11}$$

This criterion can be applied in a direct way using well-known properties of the relative entropy. For example, since the relative entropy is jointly convex and lower semicontinuous, the validity of (11) for convex subsets $\mathfrak{P}'$ and $\mathfrak{P}''$ of $\mathfrak{P}_{+\infty}$ implies that of (11) for the convex closure $\overline{\mathrm{co}}(\mathfrak{P}' \cup \mathfrak{P}'')$ of these subsets. Thus the criterion for continuity shows that *the continuity of the Shannon entropy on convex closed subsets $\mathfrak{P}'$ and $\mathfrak{P}''$ implies*[9] *its continuity on their convex closure $\overline{\mathrm{co}}(\mathfrak{P}' \cup \mathfrak{P}'')$.*

The continuity criterion can also be applied using the estimate

$$\Delta_k(x|S) \leqslant S(k(x)), \qquad k \in \mathbb{N}, \tag{12}$$

where $k(x)$ is the probability distribution obtained from x by the coarse-graining of order k, that is, $(k(x))^j = x^{(j-1)k+1} + \dots + x^{jk}$ for all $j = 1, 2, \dots$. This estimate is proved by using the decomposition

$$x = \sum_{i=1}^{\infty} \lambda_i^k p_i^k(x),$$

where $\lambda_i^k = (k(x))^i$ and $p_i^k(x)$ is the probability distribution with $(p_i^k(x))^j = (\lambda_i^k)^{-1} x^j$ for $j = (i-1)k + 1, \dots, ik$ and $(p_i^k(x))^j = 0$ for other values of j, since it is easy to check that

$$\sum_{i=1}^{\infty} \lambda_i^k S(p_i^k(x) \| x) = \sum_{i=1}^{\infty} \lambda_i^k (-\ln \lambda_i^k) = S(k(x)).$$

[9] One can show that if the Shannon entropy is continuous on a convex subset of $\mathfrak{P}_{+\infty}$, then this subset is relatively compact.

The following assertion is obtained from the continuity criterion and (12).

Assertion 1. *Let x_0 be a probability distribution in $\mathfrak{P}_{+\infty}$ with finite Shannon entropy. Then the Shannon entropy is continuous on the set*

$$\{x \in \mathfrak{P}_{+\infty} \mid x \prec x_0\}, \tag{13}$$

where $x \prec y$ means that the probability distribution $y = \{y^j\}_{j=1}^{\infty}$ is more chaotic than the probability distribution $x = \{x^j\}_{j=1}^{\infty}$ in the sense of Uhlmann [21], [22], that is, $\sum_{j=1}^{n} \widehat{x}^j \geqslant \sum_{j=1}^{n} \widehat{y}^j$ for every positive integer n, where the sequences $\{\widehat{x}^j\}_{j=1}^{\infty}$ and $\{\widehat{y}^j\}_{j=1}^{\infty}$ are obtained by arranging the sequences $\{x^j\}_{j=1}^{\infty}$ and $\{y^j\}_{j=1}^{\infty}$ in non-increasing order.[10]

Indeed, assuming that the elements of the probability distributions x and x_0 are arranged in non-increasing order, we have

$$x \prec x_0 \Rightarrow k(x) \prec k(x_0) \Rightarrow S(k(x)) \leqslant S(k(x_0))$$

by the Shur concavity of the Shannon entropy [22]. Hence the validity of (11) for the set (13) follows from (12) and the easily verifiable implication $S(x_0) < +\infty \Rightarrow \lim_{k\to+\infty} S(k(x_0)) = 0$. $\square$

The main area of application of the results of this section is the quantum information theory dealing with analytic properties of different entropic characteristics of quantum systems and channels. In this case $\mathcal{A}$ is the μ-compact stable set of quantum states (density operators on a separable Hilbert space). Our results apply because many important entropic characteristics, being discontinuous functions on the set of all quantum states with possibly infinite values, have continuous bounded restrictions to the sets of states of rank $\leqslant k$ (which play the role of $\mathcal{A}_k$) for every k. The simplest (and most important) example is the von Neumann entropy (a non-commutative analogue of the Shannon entropy), for which Theorem 2 (more precisely, a reduced version) gives a criterion for local continuity which leads to several useful 'convergence conditions' [12]. Applications of the results of this section to other entropic characteristics of quantum systems and channels are given in § 6 of [19].

§ 6. Appendix

For an arbitrary Borel function f on a closed subset $\mathcal{B} \subseteq \mathcal{A}$ we consider the functional

$$M(\mathcal{B}) \ni \mu \mapsto \mathbf{f}(\mu) = \int_{\mathcal{B}} f(x)\,\mu(dx). \tag{14}$$

It is easy to show that this functional is lower semicontinuous (resp. upper semicontinuous) if f is lower semicontinuous and lower bounded (resp. upper semicontinuous and upper bounded) [13].

[10] The relation $\prec$ is opposite to the majorization relation in linear algebra [23].

Proof of Proposition 1. The function $\hat{f}^{\mu}_{\mathcal{B}}$ is well-defined on $\overline{\mathrm{co}}(\mathcal{B})$ by Lemma 1. By the upper semicontinuity of the functional $\mathbf{f}$ defined in (14) and the compactness of the set $M_x(\mathcal{B})$ for every x in $\overline{\mathrm{co}}(\mathcal{B})$ (which follows from the μ-compactness of $\mathcal{A}$), the supremum in the definition of $\hat{f}^{\mu}_{\mathcal{B}}(x)$ is achieved at some measure μ_x in $M_x(\mathcal{B})$ such that $\hat{f}^{\mu}_{\mathcal{B}}(x) = \mathbf{f}(\mu_x)$.

If $\hat{f}^{\mu}_{\mathcal{B}}$ is not upper semicontinuous, then there is a sequence $\{x_n\} \subset \overline{\mathrm{co}}(\mathcal{B})$ converging to $x_0 \in \overline{\mathrm{co}}(\mathcal{B})$ such that

$$\exists \lim_{n\to+\infty} \hat{f}^{\mu}_{\mathcal{B}}(x_n) > \hat{f}^{\mu}_{\mathcal{B}}(x_0). \tag{15}$$

As proved above, for every n there is a measure $\mu_n \in M_{x_n}(\mathcal{B})$ such that $\hat{f}^{\mu}_{\mathcal{B}}(x_n) = \mathbf{f}(\mu_n)$. Since $\mathcal{A}$ is μ-compact, we have a subsequence $\{\mu_{n_k}\}$ converging to some measure μ_0 in $M(\mathcal{B})$. By the continuity of the map $\mu \mapsto \mathbf{b}(\mu)$, the measure μ_0 lies in $M_{x_0}(\mathcal{B})$. Since the functional $\mathbf{f}$ is upper semicontinuous, we have

$$\hat{f}^{\mu}_{\mathcal{B}}(x_0) \geqslant \mathbf{f}(\mu_0) \geqslant \limsup_{k\to+\infty} \mathbf{f}(\mu_{n_k}) = \lim_{k\to+\infty} \hat{f}^{\mu}_{\mathcal{B}}(x_{n_k}),$$

contrary to (15). □

Proof of Proposition 2. A) The function $\hat{f}^{\mu}_{\mathcal{B}}$ is well defined on $\overline{\mathrm{co}}(\mathcal{B})$ by Lemma 1. If $\hat{f}^{\mu}_{\mathcal{B}}$ is not lower semicontinuous, then there is a sequence $\{x_n\} \subset \overline{\mathrm{co}}(\mathcal{B})$ converging to $x_0 \in \overline{\mathrm{co}}(\mathcal{B})$ such that

$$\exists \lim_{n\to+\infty} \hat{f}^{\mu}_{\mathcal{B}}(x_n) < \hat{f}^{\mu}_{\mathcal{B}}(x_0). \tag{16}$$

For arbitrary $\varepsilon > 0$ let μ_0^{ε} be a measure in $M_{x_0}(\mathcal{B})$ such that $\hat{f}^{\mu}_{\mathcal{B}}(x_0) \leqslant \mathbf{f}(\mu_0^{\varepsilon}) + \varepsilon$ (where $\mathbf{f}$ is the functional defined in (14)). Since the map $M(\mathcal{B}) \ni \mu \mapsto \mathbf{b}(\mu) \in \overline{\mathrm{co}}(\mathcal{B})$ is open, one can find a subsequence $\{x_{n_k}\}$ and a sequence $\{\mu_k\} \subset M(\mathcal{B})$ converging to the measure μ_0^{ε} such that $\mathbf{b}(\mu_k) = x_{n_k}$ for every k. By the lower semicontinuity of the functional $\mathbf{f}$ we have

$$\hat{f}^{\mu}_{\mathcal{B}}(x_0) \leqslant \mathbf{f}(\mu_0^{\varepsilon}) + \varepsilon \leqslant \liminf_{k\to+\infty} \mathbf{f}(\mu_k) + \varepsilon \leqslant \lim_{k\to+\infty} \hat{f}^{\mu}_{\mathcal{B}}(x_{n_k}) + \varepsilon.$$

This contradicts (16) since ε is arbitrary.

B) The function $\hat{f}^{\sigma}_{\mathcal{B}}$ is well defined and concave on σ-co$(\mathcal{B})$. Its lower semicontinuity is proved by an obvious modification of the argument in the proof of part A).

If σ-co$(\mathcal{B}) = \overline{\mathrm{co}}(\mathcal{B})$, then the concave lower-bounded function $\hat{f}^{\sigma}_{\mathcal{B}}$, being lower semicontinuous, satisfies Jensen's integral inequality (see Remark 8 below). Since $\hat{f}^{\sigma}_{\mathcal{B}}\big|_{\mathcal{B}} \geqslant f$ by the definition of $\hat{f}^{\sigma}_{\mathcal{B}}$, we have

$$\hat{f}^{\sigma}_{\mathcal{B}}(x) \geqslant \int_{\mathcal{B}} \hat{f}^{\sigma}_{\mathcal{B}}(y)\,\mu(dy) \geqslant \int_{\mathcal{B}} f(y)\,\mu(dy)$$

for any $x \in \overline{\mathrm{co}}(\mathcal{B})$ and any measure μ in $M_x(\mathcal{B})$. Thus $\hat{f}^{\sigma}_{\mathcal{B}} \geqslant \hat{f}^{\mu}_{\mathcal{B}}$, and hence $\hat{f}^{\sigma}_{\mathcal{B}} = \hat{f}^{\mu}_{\mathcal{B}}$.

Remark 8. Jensen's integral inequality,

$$f\big(\mathbf{b}(\mu)\big) \geqslant \int_{\mathcal{A}} f(x)\,\mu(dx), \qquad \mu \in M(\mathcal{A}),$$

holds for a concave non-negative function f on a convex set $\mathcal{A}$ if this function is lower semicontinuous. This can be shown by approximating the measure μ by a sequence of finitely supported measures with the same barycentre and using the lower semicontinuity of the functional $\mathbf{f}$. Note that the condition of the lower semicontinuity of f is essential (it cannot be replaced by the condition of measurability of this function).

The author is grateful to V. M. Tikhomirov and the participants of his seminar at Moscow State University for useful remarks and discussions.

Bibliography

[1] A. D. Ioffe and V. M. Tihomirov, *Theory of extremal problems*, Nauka, Moscow 1974; English transl., Stud. Math. Appl., vol. 6, North-Holland, Amsterdam–New York 1979.

[2] E. S. Polovinkin and M. V. Balashov, *Elements of convex and strongly convex analysis*, 1st ed., Fizmatlit, Moscow 2004. (Russian)

[3] J. Vesterstrøm, "On open maps, compact convex sets, and operator algebras", *J. London Math. Soc.* (2) **6** (1973), 289–297.

[4] A. Lima, "On continuous convex functions and split faces", *Proc. London Math. Soc.* (3) **25** (1972), 27–40.

[5] R. C. O'Brien, "On the openness of the barycentre map", *Math. Ann.* **223**:3 (1976), 207–212.

[6] S. Papadopoulou, "On the geometry of stable compact convex sets", *Math. Ann.* **229**:3 (1977), 193–200.

[7] A. Clausing and S. Papadopoulou, "Stable convex sets and extremal operators", *Math. Ann.* **231**:3 (1978), 193–203.

[8] R. Grząślewicz, "Extreme continuous function property", *Acta Math. Hungar.* **74**:1–2 (1997), 93–99.

[9] J. Lukeš, J. Malý, I. Netuka, and J. Spurný, *Integral representation theory. Applications to convexity, Banach spaces and potential theory*, de Gruyter Stud. Math., vol. 35, de Gruyter, Berlin–New York 2010.

[10] V. Yu. Protasov and M. E. Shirokov, "Generalized compactness in linear spaces and its applications", *Mat. Sb.* **200**:5 (2009), 71–98; English transl., *Sb. Math.* **200**:5 (2009), 697–722.

[11] A. S. Holevo, *Statistical structure of quantum theory*, ICS, Moscow–Izhevsk 2003; English transl., Lect. Notes Phys. Monogr., vol. 67, Springer-Verlag, Berlin 2001.

[12] M. E. Shirokov, "Continuity of the von Neumann entropy", *Comm. Math. Phys.* **296**:3 (2010), 625–654.

[13] P. Billingsley, *Convergence of probability measures*, Wiley, New York–London–Sydney 1968; Russian transl., Nauka, Moscow 1977.

[14] K. R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, New York–London 1967.

[15] N. N. Vakhania and V. I. Tarieladze, "Covariance operators of probability measures in locally convex spaces", *Teor. Verojatnost. i Primenen.* **23**:1 (1978), 3–26; English transl., *Theory Probab. Appl.* **23**:1 (1978), 1–21.

[16] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, Springer-Verlag, Berlin 2006.

[17] L. Q. Eifler, "Open mapping theorems for probability measures on metric spaces", *Pacific J. Math.* **66**:1 (1976), 89–97.

[18] G. A. Edgar, "On the Radon–Nikodým property and martingale convergence", *Vector space measures and applications* (Dublin 1977), Lecture Notes in Math., vol. 645, Springer-Verlag, Berlin–New York 1978, pp. 62–76.

[19] M. E. Shirokov, *Continuity condition for concave functions on convex μ-compact sets and its applications in quantum physics*, arXiv: abs/1006.4155.

[20] S. Kullback, *Information theory and statistics*, Wiley, New York 1959.

[21] P. M. Alberti and A. Uhlmann, *Stochasticity and partial order*, Math. Monogr., vol. 18, VEB, Berlin 1981.

[22] A. Wehrl, "How chaotic is a state of a quantum system?", *Rep. Math. Phys.* **6** (1974), 15–28.

[23] R. Bhatia, *Matrix analysis*, Grad. Texts in Math., vol. 169, Springer-Verlag, New York 1997.

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

Received 15/FEB/11
Translated by THE AUTHOR