Vol. 1 (1970) REPORTS ON MATHEMATICAL PHYSICS No. 2

ON THE SHANNON ENTROPY AND RELATED FUNCTIONALS ON CONVEX SETS

A. UHLMANN

Physics Section, Karl Marx University, Leipzig

We give the definition of functionals $r(K, x)$ and $r(K, S, x)$ defined on convex sets K without or with respect to locally convex topology with the help of a strongly convex function $r(p)$ on a unit interval. If $r = -p \ln p$ we refer $r(K, x)$ to be the Shannon entropy of x relative to the convex set K. In the case of the convex set Z_n of density matrices this definition gives the usual Shannon–Gibbs entropy and yields a new defining inequality for the entropy which is independent of the representation of the algebra of $n \times n$-matrices.

1. Introduction

In 1957 E. T. Jaynes [1, 2] (see also [3, 4]) gave strong arguments in favour of using a Shannon like entropy [5] in thermodynamics, developing ideas which can be traced back to Gibbs. This concept, regardless its considering *the* entropy concept or not, is a very interesting one because of the possibility to consider it as a function on the set of *all* states of some not necessarily commuting *-algebras. Thus the definition of a Shannon entropy does not depend on the possible interpretation of states as being probability measures (a case covered by the Kolmogorov–Sinai approach). The very aim of the paper is to define Shannon entropies without referring to spectral decompositions of density matrices by the introduction of the concept of "the entropy of an element of a convex set" (for simplicity we exclude conditional entropies and entropy densities). It turns out that important properties depend on the convexity of $-p \ln p$ only, and therefore we use an arbitrary but fixed strongly convex and continuous function throughout. Heuristically, the functionals we are constructing may be considered as estimates of the "degree of mixture" of an element of a convex set or, in case of *-algebras, as estimates for the "degree of reducibility" of the associated GNS-construction.

For convenience we give some elementary definitions. Let us denote by L a real linear space and let M be a set of elements of L. An element $x \in L$ is said to be *convexly dependent on* M, iff there is a finite subset $x_1, \ldots, x_n$ of M and real numbers $p_1, \ldots, p_n$ with

$$p_i \geqslant 0; \quad \sum p_i = 1 \tag{1.1}$$

and

$$x = \sum p_i x_i \,. \tag{1.2}$$

For shortness, a linear combination (1.2), the coefficients of which satisfy (1.1), is called a "convex sum". Iff for all $p_i \neq 0$ we have $x = x_i$, the convex sum is called "trivial". The set

$$[M] = \{x \in L : x \text{ depends convex on } M\} \tag{1.3}$$

is called the "convex hull" of M. We have always

$$[[M]] = [M] \,. \tag{1.4}$$

A set K is called *convex*, iff it equals its convex hull. If $x \in K$ and if there is no non-trivial convex sum (1.2) with $x_i \in K$, x is said to be an *extremal element* of K. The set of all extremal elements of the convex set K is denoted by ex$K$ and we have

$$\mathrm{ex}\, K = \bigcap N \quad \text{with} \quad [N] = K. \tag{1.5}$$

2. Some Convex Functionals on the Set of Density Matrices

We shall start with the set Z_n of all $n \times n$-density matrices, i.e. of all matrices d with

$$d = d^*; \quad d \geqslant 0; \quad \mathrm{tr}\, d = 1. \tag{2.1}$$

As is well known, Z_n is isomorphic by

$$d \to f_d(a) = \mathrm{tr}\,(ad) \tag{2.2}$$

to the set of states (positive linear forms f with $f(e) = 1$) of the *-algebra of all $n \times n$-matrices.

Now on the interval [0, 1] we consider a continuous, strongly convex function r

$$2r\left(\frac{p_1+p_2}{2}\right) > r(p_1)+r(p_2); \quad p_1 \neq p_2; \quad p_i \in [0,1] \tag{2.3}$$

(i. e. convex in the sense of *convex from the above*) satisfying

$$r(0) = r(1) = 0, \tag{2.4}$$

and use the definition

$$r(d) = \sum r(\lambda_i) \,. \tag{2.5}$$

Here $\lambda_1, \ldots, \lambda_n$ denotes the complete set of eigenvalues of d.

LEMMA 1: If $x_1, \ldots, x_n$ denotes a complete orthonormal set of vectors, then

$$r(d) \leqslant \sum r(\langle x_i, d x_i \rangle) \tag{2.6}$$

and the equality sign holds if and only if

$$\{x_i\} \text{ is a set of eigenvectors of } d.$$

The main tool for the proof is Jensen's inequality [5]:
If $\lambda_i \in [0, 1]$ and the convex sum $\lambda = \Sigma p_i \lambda_i$ is not a trivial one, then

$$r(\lambda) > \sum p_i r(\lambda_i).$$

PROOF OF LEMMA 1: Let us denote by $y_1, \ldots, y_n$ a (necessarily complete) orthonormal system of eigenvectors of d. It follows from Jensen's inequality that

$$\begin{aligned} r(\langle x_i, d x_i \rangle) &= \sum_i r\Big(\sum_k |\langle x_i, y_k \rangle|^2 \lambda_k\Big) \\ &\geqslant \sum |\langle x_i, y_k \rangle|^2 r(\lambda_k) = \sum r(\lambda_k). \end{aligned}$$

Now the equality sign can hold iff for every given i all such numbers λ_k for which $\langle x_i, y_k \rangle$ is different from zero, are equal one to another, i.e. iff all the x_i are eigenvectors of d.

LEMMA 2: $r(d)$ is strongly convex on Z_n: For every non-trivial convex sum

$$d = \sum p_i d_i, \qquad d_i \in Z_n, \tag{2.7}$$

we have the inequality

$$r(d) > \sum p_i r(d_i). \tag{2.8}$$

PROOF: Consider an orthonormal system $y_1, \ldots, y_n$ of eigenvectors of d. By Jensen's inequality, we have

$$r(d) \geqslant \sum p_i r(\langle y_k, d_i y_k \rangle) \geqslant \sum p_i r(d_i).$$

According to Lemma 1, the second inequality sign becomes an equality sign iff $y_1, \ldots, y_n$ is an eigenvector system for every d_i.

However, the first inequality becomes an equality iff $\langle y_k, d_i y_k \rangle$ does not depend on i and hence the equality sign holds in (2.8) iff $d_i = d$ for all i.

Remark 1: The definition (2.5) may be extended to hermitian matrices, the eigenvalues of which are in $[0, 1]$. Then Lemmas 1 and 2 remain valid. For another proof of Lemma 2 see [9].

Remark 2: $r(d)$ is continuous with respect to the matrix elements of d, for $r(d)$ is a symmetric continuous function of the eigenvalues of d [10].

Before we begin the main object of this section, we note the following:

Lemma 3: ex Z_n consists of all projectors of trace one, i.e. all projection operators on one-dimensional subspaces.

Namely, we can write the spectral decomposition of $d \in Z_n$ in form of a convex sum of projection operators on one-dimensional subspaces, the coefficients of which are the eigenvalues of d. If therefore $d \in \text{ex } Z_n$, this spectral decomposition has to be trivial and d has to be a projector. On the other hand, for every projector $r(d) = 0$, and because of Lemma 2 this equation implies $d \in \text{ex } Z_n$. Lemma 3 is the well-known statement that exZ_n represent the pure states of the physical system.

Theorem 1: Let be $d \in Z_n$ and consider any convex sum

$$d = \sum p_i d_i \quad \text{with} \quad d_j \in \text{ex} Z_n, \quad p_j \neq 0. \tag{2.9}$$

Then we have

$$r(d) \leqslant \sum r(p_i) \tag{2.10}$$

and the equality sign holds if and only if (2.9) is a spectral decomposition of d.

Theorem 1 provides us with an intrinsic characterization of $r(d)$ as an infimum. Let us note that the trace condition implies

$$d_j d_k = 0 \quad \text{for} \quad j \neq k \tag{2.11}$$

as a necessary and sufficient condition for (2.9) to be a spectral decomposition. Before proving Theorem 1 let us go a step further. We denote by A the set of all sequences

$$\{p_i\}, \quad i = 1, 2, \ldots \quad \text{with} \quad p_1 \geqslant p_2 \geqslant p_3 \geqslant \ldots \geqslant 0 \tag{2.12}$$

and

$$\sum p_i = 1. \tag{2.13}$$

The set A is partially ordered by the relation

$$\{p_i\} \succ \{p_i'\} \quad \text{iff} \quad \sum_{i=1}^{j} p_i \geqslant \sum_{i=1}^{j} p_i' \quad \text{for all} \quad j. \tag{2.14}$$

Consider next the subset A_m of A defined by

$$\{p_i\} \in A_m \quad \text{iff} \quad \{p_i\} \in A \quad \text{and} \quad p_j = 0 \quad \text{for} \quad j > m. \tag{2.14}$$

A_m may be considered as a compact subset of the m-dimensional number space.

The union of all sets A_m, called A^0, consists of all $\{p_i\} \in A$ having an almost finite number of components p_j different from zero.

Definition: For $d \in Z_n$ we denote by $\{d\}$ the sequence

$$\{d\} = \{\lambda_1, \lambda_2, \ldots, \lambda_n, \ 0, 0, 0, \ldots\} \tag{2.15}$$

constructed with the help of the ordered set $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ of the eigenvalues of d.

Now we formulate

THEOREM 2: Let be $d \in Z_n$ and $\{p_i\} \in A^0$. The elements $d_i \in \mathrm{ex}\, Z_n$ satisfying

$$d = \sum p_j d_j, \text{ convex sum} \tag{2.16}$$

exist iff

$$\{d\} \succ \{p_i\}. \tag{2.17}$$

Theorem 2 may be looked at as an example of what Kato [9] calls "perturbation theory in the large".

Now Theorem 1 is quickly deduced from Theorem 2, because (2.17) is a necessary and sufficient condition for the existence of a bistochastic matrix (a_{ik})

$$a_{ik} \geqslant 0, \quad \sum_i a_{ik} = \sum_j a_{kj} = 1, \tag{2.18}$$

such that [10, 11]

$$p_i = \sum_k a_{ik} \lambda_k. \tag{2.19}$$

Therefore Theorem 1 follows from Theorem 2 with the help of Jensen's inequality.

PROOF OF THEOREM 1: Let us denote by B_m the set of all sequences $\{p_i\} \in A_m$ for which there exist $d_i \in \mathrm{ex} Z_n$ satisfying (2.16). Clearly B_m is a compact set of the m-dimensional number space (remind the compactness of $\mathrm{ex} Z_n$) and therefore there exists for every $\{p_i\} \in B_m$ a sequence $\{p'_i\} \in B_m$ which is with respect to the order relation $\succ$ a maximal element of B_m and which majorizes $\{p_i\}$, i.e. $\{p_i\} \prec \{p'_i\}$. Thus we have to prove that there is only one maximal element in B_m, namely the element $\{d\}$ (or B_m is empty, in which case the number m was chosen not large enough). To do this, let us consider a representation (2.16) and choose two numbers i, k. The matrix

$$p_i d_i + p_k d_k = d_{ik}, \quad i \geqslant k, \tag{2.20}$$

has at most two eigenvalues different from zero and we may call them $\mu_1 \geqslant \mu_2$. Now it is a matter of elementary calculations with (2×2)-matrices to prove that d_{ik} can be represented in the form (2.20) if and only if

$$\mu_1 + \mu_2 = p_i + p_k \quad \text{and} \quad \mu_1 \geqslant p_i \geqslant p_k \geqslant \mu_2. \tag{2.21}$$

Moreover, the equality sign holds if and only if $d_i d_k = 0$. Hence, if the equality sign does not hold in (2.21), we can replace the components p_i and p_k of $\{p_j\}$ by μ_1 and μ_2 and get, after reordering in the natural order, a new sequence $\{p'_j\}$ which can be shown [11] to be larger than $\{p_i\}$ with respect to our ordering. Hence $\{p_i\}$ can be a maximal sequence of B_m if and only if $d_i d_k = 0$ for all i, k with $p_i p_k \neq 0$. But this condition is satisfied by the sequence $\{d\}$ only. On the other hand, if $\{p_i\} \prec \{d\}$ there is [11] a finite sequence of steps (each step consists of replacing two components p_i, p_k of a sequence by p_{io}, p_{ko} satisfying

$$p_i + p_k = p_{io} + p_{ko}, \quad p_i \geqslant p_{io} \geqslant p_{ko} \geqslant p_k,$$

and reestablishing the natural order of the new sequence), such that starting with $\{d\}$ we arrive at $\{p_i\}$. Because of condition (2.21) for the form (2.20), we can associate with each of these steps a certain representation (2.16). Hence B_m consists only of all sequences fulfilling (2.17) q.e.d.

3. The General Definition

Theorem 1 suggests to define $r(d)$ by the infimum of the sums (2.10) and to derive formula (2.5) and other properties, because going this way we start with the structure of the convex set Z_n as a whole. Then one could try to generalize this for arbitrary convex sets. However, in general, only for a finite-dimensional compact convex set the convex hull of its extremal points covers a given convex set. So we shall not start with any assumption on extremal points.

Let us denote by M a subset of a real linear space and let us fix a certain strongly convex function $r = r(p)$ on $[0, 1]$, vanishing at the endpoints of the interval.

Definition: If x depends convexly on M, we denote by

$$r^M(x)$$

the infimum of all sums

$$\sum r(p_i), \tag{3.1}$$

where $\{p_j\}$ runs over all such sequences of A^0 for which there exist elements $x_i \in M$ with

$$x = \sum p_i x_i. \tag{3.2}$$

Remark: In the case $r(p) = -p \ln p$, the number $r^M(x)$ is called "entropy of x relative to M".

$r^M(x)$ is a non-negative finite function on the convex hull of M. Further, if x depends convexly on M

$$r^M(x) \geqslant r^N(x) \quad \text{for} \quad M \subseteq N, \tag{3.3}$$

$$x \in M \quad \text{implies} \quad r^M(x) = 0. \tag{3.4}$$

We now derive an estimate, showing how rapidly r^M increases near the points of M. We consider first an identity. If x and y are convex sums

$$x = \sum_{i=1}^{m} p_i x_i, \quad \Big(\sum_{j=1}^{m} p_j\Big) y = \sum_{k=1}^{m} p_k x_k, \tag{3.5}$$

it follows

$$x - y = \sum_{i=m+1}^{n} p_i (x_i - y). \tag{3.6}$$

Next we use an arbitrary seminorm q defined on L and form the number

$$d_q(M) = \sup_{x,\, x' \in M} q(x-x') = \sup_{x,\, x' \in [M]} q(x-x') \tag{3.7}$$

which may be called "q-diameter of M". Let us choose a natural number m and a number $0 < p_0 < 1$. r is convex, and therefore the following inequalities hold:

$$pr(p_0) \leqslant p_0 r(p), \quad \text{if} \quad 0 \leqslant p \leqslant p_0, \tag{3.8}$$

$$(1-p)r(p_0) \leqslant (1-p_0)r(p), \quad \text{if} \quad p_0 \leqslant p \leqslant 1. \tag{3.9}$$

Let us now assume $mr(p_0) > (1-p_0)r^M(x)$. Then x can be represented as a convex combination (3.5) with $x_i \in M$ and we may assume $p_1 \geqslant p_2 \geqslant \dots$ and

$$mr(p_0) > (1-p_0)\sum_i r(p_i) \geqslant (1-p_0) \sum_{p_i \geqslant p_0} r(p_i)$$

and hence

$$mr(p_0) > \sum_{p_i \geqslant p_0} (1-p_i) r(p_0) \geqslant (m'-1) r(p_0),$$

where m' is the number of coefficients p_j with $p_j \geqslant p_0$. It follows that $m' \leqslant m$ and $m' \leqslant p_0^{-1}$. Now we use (3.6) to get

$$q(x-y) \leqslant \sum_{i > m'} p_i q(x_i - y) \leqslant p_0 r(p_0)^{-1} d_q(M) \sum_{i > m'} r(p_i).$$

The last inequality results from (3.8). Thus we have proved

LEMMA 4: If $x \in [M]$, $0 < p_0 < 1$ and one of the inequalities

$$m^{-1} \geqslant p_0 \quad \text{or} \quad m^{-1} > r^M(x)\,(1-p_0) r(p_0)^{-1} \tag{3.10}$$

is valid for the natural number m, then there exists a convex sum

$$y = p_1 x_1 + \dots + p_m x_m; \quad x_i \in M, \tag{3.11}$$

such that

$$r(p_0)\, q(x-y) \leqslant p_0 d_q(M) r^M(x). \tag{3.12}$$

This inequality is especially interesting in the case $p^{-1} r(p) \to \infty$ for $p \to 0$.

Next we consider an arbitrary convex set K in L.

Definition: $r(K, x) = \sup r^N(x)$ with $K = [N]$. (3.13)

Clearly

$$0 \leqslant r^M(x) \leqslant r(K, x) \leqslant \infty \quad \text{if} \quad [M] = K. \tag{3.14}$$

If K is the convex hull of M, we have of course $\operatorname{ex} K \subseteq M$. Inequalities (3.3) and (3.14) now tell us

$$r(K, x) \leqslant r^{\operatorname{ex} K}(x) \quad \text{for} \quad x \in [\operatorname{ex} K]. \tag{3.15}$$

Lemma 5: If $K = [\mathrm{ex}\, K]$, we have

$$r(K, x) = r^{\mathrm{ex}\, K}(x). \tag{3.16}$$

In particular,

$$r(Z_n, d) = r(d). \tag{3.17}$$

This follows from (3.14) and (3.15) and Theorem 1. For $K = Z_n$ it is just another form of Theorem 1. The inequality (3.14) provides us further with

Lemma 6: Lemma 4 remains valid if we replace $r^M(x)$ by $r([M], x)$ in (3.10) and (3.12).

A convex subset K_0 of a convex set K is called an "extremal part of K", iff for every non-trivial convex sum

$$\sum p_i x_i \in K_0 \quad \text{with} \quad x_j \in K,$$

it follows: $x_j \in K_0$ for all j with $p_j \neq 0$. The intersection of extremal parts of a convex set is an extremal part again.

Lemma 7: If the convex set K_0 is an extremal part of the convex set K, then for $x \in K_0$

$$r(K_0, x) = r(K, x). \tag{3.18}$$

For the proof we consider a set N_0 that generates K_0. The set $N = N_0 \cup (K \setminus N)$ then satisfies $K = [N]$ and because K_0 is an extremal part of K, we have $r^N(x) = r^{N_0}(x)$ for $x \in K_0$. Hence $r(K, x) \geqslant r(K_0, x)$ because N_0 was arbitrary up to the condition $[N_0] = K_0$. On the other hand, if $[N] = K$ we obviously have (again by virtue of the extremality of K_0) $[(N \cap K_0)] = K_0$ and therefore we get every set N_0 that generates K_0 by an intersection $N_0 = N \cap K_0$ with $[N] = K$ and we have always $r^N = r^{N_0}$ on K_0. Hence (3.18) is valid.

Examples: Denote by R a $*$-algebra with unit element e and by ZR the set of states of R, i.e. the set of all positive linear forms normed by the condition $f(e) = 1$.

(a) Denote by J a left ideal of R and define

$$K = \{f \in ZR : f(a) = 0 \text{ all } a \in J\}.$$

K is an extremal part of ZR.

Proof: Let be $f \in ZR$. Clearly $f \in K$ if and only if $f(a^*a) = 0$ for all $a \in J$ because f is positive. Let $f = \sum p_i f_i$ be a non-trivial convex sum. Again, by the positivity of the linear forms, $f(a^*a) = 0$ iff $f_i(a^*a) = 0$.

(b) Denote by q a symmetric multiplicative seminorm on R

$$q(ab) \leqslant q(a)q(b), \qquad q(b) = q(b^*).$$

The set $ZR(q)$ of states of R which are continuous with respect to q is an extremal part of ZR.

Proof: Consider a non-trivial convex sum $f = \sum p_i f_i \in ZR(q)$. Because of the positivity $f_j(b^*b) \leqslant p_j^{-1} f(b^*b)$. But $f(b^*b) \leqslant \mathfrak{L}\, q(b^*b)$ and hence with $\mathfrak{L}_j = \mathfrak{L}\, p_j^{-1}$ we get

$$(f_j(b))^2 \leqslant f_j(b^*b) \leqslant \mathfrak{L}_j\, q(b^*b) \leqslant \mathfrak{L}_j q(b)^2.$$

Our next task is to generalize an argument used in the proof of Lemma 7. We shall deal with linear mappings and therefore we explicitely note the linear space in which the convex set is imbedded: By a pair $[L, K]$ we shall denote a real linear space L and a convex set K of L.

Definition: Consider two pairs $[L, K]$ and $[L', K']$. A linear map φ is said to be a "c-linear map of $[L, K]$ into $[L', K']$" iff

(i) φ is a linear map of L into L',

(ii) φ maps K into K',

(iii) If K' is the convex hull of the set N', then

$$N = \{x \in K : \varphi x \in N'\} \tag{3.19}$$

generates K, i.e. $K = [N]$.

Remarks: It is straightforward to prove: (a) The composition of two c-linear mappings is c-linear again. (b) Every c-linear map φ can be decomposed naturally $\varphi = \varphi_1 \cdot \varphi_2$, where φ_2 is an epimorphism and φ_1 is a monomorphism. φ_1 and φ_2 are c-linear maps. By virtue of (a) we may consider the pairs $[L, K]$ to be the objects and the c-linear mappings to be the morphisms of a category. Further we note the essentiality of condition (iii): To be c-linear, it is necessary for the map φ to satisfy $\varphi(\mathrm{ex}\, K) \subseteq (\mathrm{ex}\, K')$.

Let us consider a map φ from $[K, L]$ into $[K', L']$. If K is the convex hull of M, then $\varphi K = K_0$ is the convex hull of $M_0 = \varphi M$. The set $M' = \{x' \in L' : x' \in M_0$ or $x' \in K' \setminus M_0\}$ has the property $K' = [M']$ and $M' \cap K_0 = M_0$. Therefore

$$\inf_{x' = \varphi x} r^M(x) \geqslant r^{M'}(x') \quad \text{for} \quad x' \in K_0. \tag{3.20}$$

Let us now assume the map to be c-linear. If M_1 generates K', then $M = \varphi^{-1}(K_0 \cap M_1)$ generates K convexly. The construction above shows $M_1 \subseteq M'$ with $M_0 = K_0 \cap M_1$. Therefore (3.20) remains valid if we replace M' by M_1 in this formula.

Applying definition (3.13), we arrive at

Lemma 8: Let be φ a c-linear map from $[K, L]$ into $[K', L']$. If $x' = \varphi x$ for $x \in K$ we have

$$r(K, x) \geqslant r(K', x'). \tag{3.20}$$

As an example of a c-linear map we mention the restriction on a C^*-subalgebra of the states of a C^*-algebra.

4. Topological Considerations

According to Treves [13] we call a set of seminorms S an irreducible one, iff (a) S is a convex cone and (b) with $q \in S$ and $q' \leqslant q$ also $q' \in S$.

If K is a convex set in L, then the set of all seminorms q satisfying the condition

$$d_q(K) < \infty \tag{4.1}$$

is an irreducible set of seminorms. It will be denoted by S^K. By $\ker S$ we denote the set

$$\ker S = \{x \in L : q(x) = 0 \text{ all } q \in S\}. \tag{4.2}$$

The topology defined by S is called a Hausdorff topology iff $\ker S$ consists of the zero element of L only.

LEMMA 9: Let $K = [M]$. Then $r^M(x) = 0$ for $x \in K$ if and only if

$$x \in M + \ker S^K.$$

The "if-part" is trivial and the "only-part" is a straightforward application of Lemma 4. From Lemma 9 we conclude:

THEOREM 3: If S^K defines a Hausdorff topology, then

$$\operatorname{ex} K = \{x \in K : r(K, x) = 0\}. \tag{4.3}$$

Namely, in this case, $r(K, x) = 0$ iff x is contained in every set M with $[M] = K$. But the intersection of this set equals ex K.

We now introduce for a given seminorm q the notation

$$r^M(q, x) = \inf r^M(y), \qquad q(x-y) < 1. \tag{4.4}$$

If $U(q)$ denotes the set

$$U(q) = \{z \in L : q(z) < 1\}, \tag{4.5}$$

the definition (4.4) makes sense if x is contained in

$$[M + U(q)] = [M_1]; \qquad M_1 = M + U(q). \tag{4.6}$$

Here we have used the fact that U is a convex set and therefore $[M+U] = [M]+U$. Now if x is represented by a convex sum $\sum p_i x_i$ with $x_i \in M_1$, then $x_i = y_i + z_i$ with $y_i \in M$ and $z_i \in U(q)$, and therefore $x = \sum p_i y_i + z$ with $z = \sum p_i z_i \in U(q)$. Thus $y = \sum p_i y_i \in M$ and $q(x-y) < 1$. On the other hand, if $q(x-y) < 1$ and $y = \sum p_i y_i$ with $y_i \in M$, we have

$$x = \sum p_i(y_i + x - y); \qquad y_i + x - y \in M_1.$$

Therefore we can conclude

LEMMA 10: Define with a seminorm q

$$M_1 = M + U(q).$$

Then for $x \in [M_1]$ we have

$$r^M(q, x) = r^{M_1}(x). \tag{4.7}$$

From the definition (4.4) we get

$$r^M(q', x) \leqslant r^M(q, x), \quad \text{if} \quad q' \leqslant q, \tag{4.8}$$

whenever the left-hand side of this inequality is defined. Next we consider a convex set K and an irreducible set S of seminorms (defining a not necessarily Hausdorff locally convex topology).

LEMMA 11: K is contained in the S-closure of $[M]$ if and only if

$$K \subseteq [M + U(q)] \quad \text{for all} \quad q \in S. \tag{4.9}$$

PROOF: If the element x of K is a limit point of $[M]$, then for every $q \in S$ there is an element $y \in [M]$ with $q(x-y) < 1$ and hence $x \in [M] + U(q) = [M + U(q)]$. On the other hand, if

$$y \notin \bigcap [M + U(q)], \quad q \in S,$$

then y is not contained with a certain q' of S in $M + U(q')$ and hence

$$y + U(3q') \cap [M] + U(3q') = \emptyset \tag{4.10}$$

because otherwise with a certain $y_0 \in M$ and $3q'(z_s) < 1$, $s = 1,2$, it would follow

$$q'(y-x) \leqslant q'(z_1 - z_2) \leqslant 2/3$$

and that contradicts (4.10). Hence the intersection

$$\bigcap [M + U(q)], \quad q \in S, \tag{4.11}$$

is closed in the S-topology.

Definition:

$$r(K, S, x) = \sup r^M(q, x), \tag{4.12}$$

where the supremum is taken with respect of all $q \in S$ and all sets M with

$$S\text{-closure } [M] = S\text{-closure } K. \tag{4.13}$$

Remark: $r(K, S, x)$ is defined (by virtue of Lemma 11) for all x contained in the S-closure of K and because of the definition

$$r(K_1, S, x) = r(K_2, S, x), \tag{4.14}$$

if the S-closures of K_1 and K_2 coincide.

LEMMA 12: $r(K, S, x)$ is an S-subcontinuous function.

PROOF: Let $r(K, S, x) > b \geqslant 0$. Then there is a set M satisfying (4.13) and a seminorm $q \in S$ with $r^M(q, x) > b$. It follows

$$r^M(y) \geqslant r^M(q, x) > b \quad \text{for all } y \text{ with} \quad q(y-x) < 1. \tag{4.15}$$

Now we may choose y_0 with $q(y_0-x) < 1/2$. Then for $q(y_0-y) < 1/2$, the inequality (4.15) remains valid. Hence $r^M(2q, y) \geqslant r^M(q, x)$ and $r^M(q, y) > \mathrm{b}$ for $2q(y-x) < 1$. Therefore the set of all x with $r(K, S, x) > b$ is an open one. This proof shows further that

$$r^M(S, x) = \sup r^M(q, x); \qquad q \in S \tag{4.16}$$

is a subcontinuous function.

Now assume $S \subseteq S^K$. Then $r^M(q, x) = 0$ implies that x is contained in $M + U(q) + \ker S^K$ but $\ker S^K \subset U(q)$ if q belongs to S^K. Hence $x \in U(q) + M$. On the other hand, $x \in M + U(q)$ implies $r^M(q, x) = 0$. From Lemma 11 follows

LEMMA 13: Under the condition $S \subseteq S^K$ we have

$$S\text{-closure } M = \{x : r^M(S, x) = 0\} \tag{4.17}$$

and under the condition (4.13)

$$\{x : r(K, S, x) = 0\} = \bigcap_M (S\text{-closure } M). \tag{4.18}$$

From this follows

THEOREM 4: Let K be a compact convex set with respect to the Hausdorff topology S. $r(K, S, x) = 0$ if and only if x is contained in the S-closure of $\operatorname{ex} K$.

PROOF: By the Krein–Millman theorem [7, 8] $\operatorname{ex} K$ fulfils condition (4.13) as well as $S \subseteq S^K$ for S-compact K. By Lemma 13 it follows for $r(K, S, x) = 0$ that x is in the S-closure of $\operatorname{ex} K$. On the other hand from $x \in \operatorname{ex} K$ it follows $r(K, S, x) = 0$. But by Lemma 12 the same is true if x is in the S-closure of $\operatorname{ex} K$.

Next, if N is the S-closure of M we have $M + U(q) = N + U(q)$ for q in S. Hence we can restrict (4.12) to S-closed sets. According to Krein–Millman, the S-closure of $\operatorname{ex} K$ is the smallest S-closed set satisfying (4.13). Hence

LEMMA 14: For S-compact convex sets K with Hausdorff topology S we have

$$r(K, S, x) = r^{\operatorname{ex} K}(S, x). \tag{4.19}$$

Aknowledgements

I wish to thank G. Lassner, G. Vojta and P. M. Alberti for valuable discussions.

REFERENCES

[1] Jaynes, E. T., Phys. Rev. **106** (1957), 620.
[2] Jaynes, E. T., Phys. Rev. **108** (1957), 171.
[3] Jaynes, E. T., Am. I. Phys. **33** (1965), 391.
[4] Robertson, B., Phys. Rev. **144** (1966), 151.
[5] Jaglom, A. M. and I. M. Jaglom, *Wahrscheinlichkeit und Information*, Berlin 1965.
[6] Berge, C. and Gihoula-Houri, *Programme, Spiele, Transportnetze*, Leipzig 1967.
[7] Day, M. M. *Normed Linear Spaces*, Berlin 1962.
[8] Neumark, M. A.; *Normierte Algebren*, Berlin 1959.
[9] Miracle-Sole, S. and B. W. Robinson, Comm. Math. Phys. **14** (1969), 235.
[10] Kato, T., *Perturbation Theory for Linear Operators*, Berlin 1966.
[11] Hardy, G. H., J. E. Littlewood and G, Polya, *Inequalities*, Cambridge 1952.
[12] Glazman, I. M. and Yu. I. Gubicz, *Konetchnomernyj lineynyi analis*, Moscow 1969.
[13] Traves, F., *Locally Convex Spaces and Linear Partial Differential Equations*, Berlin, Heidelberg, New York 1967.