Masanori Ohya and Dénes Petz

Quantum Entropy and Its Use

Monograph – Physics

Springer
Berlin Heidelberg New York
Barcelona Hong Kong
London Milan Paris
Tokyo

Preface

Entropy is a concept which appears in several fields and it is in the center of interest both in mathematical and physical subjects, sometimes even at other places, for example in communication engineering. The birthplace of quantum entropy was in quantum statistical mechanics. Quantum entropy is not a single concept but rather a family of notions which started with the von Neumann entropy of a density matrix and has developed in a number of directions. The heritage of quantum entropies from quantum mechanics is their strong relation to Hilbert space.

This book begins with the entropy of a state and through the entropy of coarse graining it proceeds to the quantum dynamical entropy. A bunch of topics shows the direct physical relevance of the entropy methods. The mathematical formalism is based on operator theory and mostly the language of operator algebras is used. The use of the concepts and methods of functional analysis not only makes the theory of quantum entropy more uniform and lucid, but it essentially simplifies and extends it. A comprehensive presentation without these elements would not be possible.

Chapters 1–3 are elementary and require of the reader only a basic knowledge of linear algebra and function theory. (Some series expansions may be beyond this level but they are standard in mathematical physics and and exact source indication provides assistance in the few cases they show up.) Most of the results of Chapters 1 and 3 are repeated in Part 2 in a more general setting. Part 1 of the book intends to demonstrate that although the quantum entropy is a more technical subject than the classical one, the basic concepts and properties are accessible without heavy mathematics.

The theory of operator algebras becomes more necessary in Parts 2–4. The very essentials are contained in any standard book on operator algebras but we suggest "Operator Algebras and Quantum Statistical Mechanics 1" by Olla Bratteli and Derek W. Robinson if a reader really would like to see a systematic treatment of this subject. What is badly needed for quantum entropy theory is concisely summarized in Chapter 4. It might be informative to emphasize that the bulk of entropy results are fully understandable without a deep knowledge of functional analysis but the proofs at a general level require sometimes sophisticated tools.

Part 5 is a collection of independent subjects which provide wide space for application of the abstract theory. In our opinion one can get a real insight into these topics without a thorough reading of all the previous chapters but with the understanding of Part 1. Precise references help those readers who have not gone through the preceding chapters.

Almost all the results in this book are accompanied with detailed rigorous proofs. Our attitude is contained in the following motto.

> "Even things that are true can be proved."
> (O. Wilde)

The end of a proof is indicated by the symbol □. Each chapter is closed by a section "Notes and Remarks". Here one can find the main sources used by the authors in the presentation and references to related results if the authors were aware of those. The literature on quantum entropy is huge and on entropy it is even more tremendous. Our selection must be very subjective and it mostly indicates our sources rather than evaluation of the literature. To find the balance between classical entropy and quantum entropy theory was a delicate point for the authors. Typically the classical theory is neither assumed nor explained in detail in the book. Of course, any slight knowledge of probabilistic entropy concepts is very helpful for the reader and yields a deeper understanding.

We want to thank many colleagues for their kind help in the preparation of the manuscript. We cannot name all of them but we ought to mention Luigi Accardi, Imre Csiszár, András Dévényi, Matthew J. Donald, Mark Fannes, Fumio Hiai, Heide Narnhofer, Walter Thirring and André Verbeure.

Masanori Ohya
Dénes Petz

Contents

Part IV Perturbation Theory

Part V Miscellanea

Introduction

The task of statistical mechanics is to derive macroscopic or bulk properties of matter from the laws governing a huge number of individual particles. (The word particle is used here for the sake of simplicity, one can think of molecules in a gas, electrons in a plasma, etc.) Entropy relates macroscopic and microscopic aspects and the concept is as old as modern thermodynamics. The name "entropy" is due to Clausius and the main steps towards the concept were taken by Boltzmann and Gibbs. Any study leading to the understanding of entropy must go through probability theory. The Boltzmann-Gibbs-Shannon entropy of a finite probability distribution $(p_1, p_2, \ldots, p_k)$ is the quantity

$$H(p_1, p_2, \ldots, p_k) = -\sum_{i=1}^{k} p_i \log p_i \,, \tag{1}$$

which is strongly related to the asymptotics of certain probabilities. This is the very reason why the same quantity appears in different subjects, statistical mechanics, information theory, ergodic theory, for example.

$$W_N(p_1, p_2, \ldots, p_k) = \frac{N!}{(p_1 N)!(p_2 N)! \ldots (p_k N)!} \tag{2}$$

is the number of ways one can put N objects (say, particles) into k boxes (say, cells of the phase space) so that the different boxes contain objects with proportions p_1, $p_2, \ldots, p_k$. Suppose we are interested in the behavior of W_N for large N. Using Stirling bounds for the factorials one can see that

$$\left| \tfrac{1}{N} \log W_N(p_1, p_2, \ldots, p_k) - H(p_1, p_2, \ldots, p_k) \right| = O(n^{-1} \log n) \,. \tag{3}$$

Boltzmann interpreted the number $\log W_N(p_1, p_2, \ldots, p_k)$ as the "thermodynamical probability" of the macro-state $(p_1, p_2, \ldots p_k)$. He had to divide the phase space into cells in order to speak of the number of "micro-states". In quantum mechanics such discretization is not necessary because contained in the theory.

Before we turn to entropy in quantum systems we summarize the abstract logical foundations of quantum theory. The primary mathematical structure associated with the set $\mathcal{P}$ of all propositions on a physical system is an ordered structure. (In place of proposition "question", "event" and "yes-no experiment" are used in the literature.) On the set $\mathcal{P}$ a reflexive antisymmetric transitive relation (that is, an order relation) is given, which is customarily denoted by $\leq$. The relation $p \leq q$ is interpreted as stating that q is true whenever p is proven to be true. Any (finite) collection $(p_i)_i$ of elements of $\mathcal{P}$ admits a greatest lower bound $\wedge_i p_i$ and a least upper bound $\vee_i p_i$. We write $p \wedge q$ and $p \vee q$ if we have to do with two elements. $\mathcal{P}$ is called a lattice in the language of algebra and it has a least element and a greatest element, which are denoted by O and I respectively. Another algebraic operation on $\mathcal{P}$ is the orthocomplementation. This is a mapping $p \mapsto p^\perp$ of $\mathcal{P}$ onto itself such that

(i) $p^{\perp\perp} = p$,
(ii) $p \leq q$ implies $p^\perp \geq q^\perp$,
(iii) $p \wedge p^\perp = 0$ and $p \vee p^\perp = I$.

Two elements p and q of $\mathcal{P}$ are said to be disjoint or orthogonal if $p \leq q^\perp$ (or equivalently $q \leq p^\perp$). Without deeply entering quantum logics we recall that the logical differences between quantum and classical behavior concerns the distributivity of the lattice $\mathcal{P}$.

In this book we always suppose that $\mathcal{P}$ is the set of all orthoprojections (i.e., idempotent self-adjoint elements: $p = p^2 = p^*$) of a *-algebra $\mathcal{A}$. In usual quantum mechanics $\mathcal{A}$ is the set of all bounded operators on a Hilbert space. On the set of projections of $\mathcal{A}$ the ordering is defined by

$$p \leq q \iff pq = p$$

and orthocomplementation is defined by

$$p^\perp = I - p$$

where I is the unit element of $\mathcal{A}$. (I is the largest projection.) The lattice operations are not directly expressed by product and sum. In the Hilbert space formulation of quantum mechanics $\mathcal{P}$ is the set of all orthogonal projections of a Hilbert space $\mathcal{H}$. Let $p, q \in \mathcal{P}$ be projections onto the closed subspaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Then the join $p \vee q$ is the orthogonal projection onto the closed subspace generated by $\mathcal{H}_1$ and $\mathcal{H}_2$ and the meet $p \wedge q$ projects onto the intersection of $\mathcal{H}_1$ and $\mathcal{H}_2$.

Given an orthocomplemented partially ordered set $\mathcal{P}$ we say that a real-valued function μ on $\mathcal{P}$ is a probability measure on $\mathcal{P}$ if

(i) $0 \leq \mu(p) \leq 1$ for all $p \in \mathcal{P}$, $\mu(0) = 0$ and $\mu(I) = 1$,
(ii) for every sequence (p_i) of pairwise orthogonal elements in $\mathcal{P}$ the additivity $\sum_i \mu(p_i) = \mu(\vee_i p_i)$ holds.

The additivity property (ii) may be called finite or countable additivity depending on the sequence required. When $\mathcal{P}$ is the set of projections of a *-algebra and φ is a linear functional on $\mathcal{A}$ such that $0 \leq \varphi(p) \leq 1$ for every projection and $\varphi(I) = 1$, then φ restricted to $\mathcal{P}$ gives rise to a finitely additive measure on $\mathcal{P}$. This is a consequence of the simple fact that for disjoint projections p and q the equality $p \vee q = p + q$ holds. In order to obtain a measure on $\mathcal{P}$ by the restriction of a linear functional φ we need to pose some further conditions on φ. A linear functional φ will be called a state if $\varphi(I) = 1$ and $\varphi(x^*x) \geq 0$ for every $x \in \mathcal{A}$. The states of a *-algebra supply an important class of measures on the set of projections of the algebra but certainly not all the measures may be obtained in this way. A pair $(\mathcal{A}, \varphi)$ consisting of a *-algebra $\mathcal{A}$ and of a state φ will be called a probability algebra. Given a probability space $(X, \mathcal{F}, \mu)$ the set of all complex-valued bounded measurable functions form a *-algebra with pointwise conjugation as involution, on which integration with respect to μ provides a state. On the basis of this example, we are tempted to borrow terminology from classical probability theory when dealing with probability algebras.

In Hilbert space quantum mechanics states are given by statistical operators. A positive operator $D \in B(\mathcal{H})$ is called a statistical operator or density if its spectrum consists only of eigenvalues and the sum of eigenvalues is 1. When the physical system is in a state described by the density D, the expected value of the bounded (self-adjoint) observable $A \in B(\mathcal{H})$ is

$$\mathrm{Trace}\, DA = \sum_i \lambda_i \langle f_i |\, A \,| f_i \rangle \tag{4}$$

if $D = \sum_i \lambda_i |f_i\rangle\langle f_i|$ in an orthonormal basis $|f_i\rangle$. In physics the underlying Hilbert space is often infinite dimensional and so is the *-algebra $B(\mathcal{H})$. However, because of their for their technical simplicity it is useful to consider finite dimensional *-algebras which readily admit an abstract trace functional. An abstract trace functional is a linear functional Tr taking the value 1 on each minimal projection (that is, on the atoms of the corresponding logic). So $\mathrm{Tr}\, I$ measures the size of the algebra; it is the number of pairwise orthogonal minimal projections the sum of which is the unity I. By means of Tr the states are described by density matrices in the form

$$\varphi(a) = \mathrm{Tr}\, D_\varphi a \qquad (a \in \mathcal{A})\,.$$

Here the density matrix D_φ is positive (semidefinite) and $\mathrm{Tr}\, D_\varphi = 1$ comes from the normalization $\varphi(I) = 1$. In this formalism for the case of the commutative algebra $\mathcal{A} = \mathbb{C}^n$ we have

$$\mathrm{Tr}\,(c_1, c_2, \dots, c_n) = \sum_{i=1}^n c_i\,.$$

In this case the density "matrix" or "statistical operator" becomes a probability distribution on the n-point-space. It is certainly an advantage of the *-algebraic approach to the foundations of quantum theory that it includes traditional probability theory and on the other hand the "matrix mechanics" of quantum systems is covered by the same formalism.

The set of all states, the state space, is a convex set. In the example $\mathcal{A} = \mathbb{C}^n$ the state space is nothing else but the standard n-simplex

$$\mathbf{S}_n = \{(c_1, c_2, \dots, c_n) \in \mathbb{R}^n \,:\, c_i \geq 0\,, \textstyle\sum_{i=1}^n c_i = 1\}\,.$$

The extreme points of the state space are called pure states. (Pure states can not be mixtures of other states.) The simplex $\mathbf{S}_n$ has n extreme points and every element is uniquely represented by their convex combination. (A convex set is called "Choquet simplex" if all points are unique mixtures of the extremal boundary.) The state space of a typical quantum system is far from being a simplex. Consider the algebra $\mathcal{A}$ of a single particle of spin 1/2 which is simply $\mathcal{A} = M_2(\mathbb{C})$, the algebra of 2×2 complex matrices. The state space with an appropriate parameterization may be identified with the three dimensional unit ball. The pure states correspond to surface points and all the inner points are mixtures of pure states in many ways.

Turning to thermodynamical entropy we repeat a gedanken experiment of John von Neumann. Let us have a gas of $N(\gg 1)$ molecules in a rectangular box K. Suppose that the gas behaves as a quantum system and it is described by a probability algebra $(\mathcal{A}, \varphi)$ in a certain physical state. If φ is a mixture $\lambda\varphi_1 + (1-\lambda)\varphi_2$, then we may think that λN molecules are in state φ_1 and $(1-\lambda)N$ ones are in the state φ_2. On the ground of phenomenological thermodymanics we assume that if φ_1 and φ_2 are disjoint, then there is a wall which is completely permeable for the φ_1-molecules and isolating for the φ_2-molecules. (If the states φ_1 and φ_2 are disjoint, then this should be demonstrated by a certain filter. Mathematically the disjointness of φ_1 and φ_2 is expressed in the orthogonality of the eigenvectors corresponding to nonzero eigenvalues of the density matrices of the states φ_1 and φ_2.) We add an equally large empty rectangular box K' to the left of the box K and we replace the common wall with two new walls. Wall (a), the one to the left is impenetrable, whereas the one to the right, wall (b), lets through the φ_1-molecules but keeps back the φ_2-molecules. We add a third wall (c) opposite to (b) which is semi-permeable, transparent for the φ_2-molecules and impenetrable for the φ_1-ones. Then we push slowly (a) and (c) to the left, keeping their distance. During this process the φ_1-molecules are pressed through (b) into K' and the φ_2-molecules diffuse through wall (c) and remain in K. No work is done against the gas pressure, no heat is developed. Replacing the walls (b) and (c) with a rigid absolutely impenetrable wall and removing (a) we restore the boxes K and K' and succeed in the separation of the φ_1-molecules from the φ_2-ones without any work being done, without any temperature change and without evolution of heat. The entropy of the original φ-gas (with density N/V) must be the sum of the entropies of the φ_1- and φ_2-gases (with densities $\lambda\, N/V$ and $(1-\lambda)N/V$, respectively.) If we compress the gases in K and K' to the volumes λV and $(1-\lambda)V$, respectively, keeping the temperature constant by means of a heat reservoir, the entropy change amounts to $\kappa_{\mathrm{B}}\lambda N \log \lambda$ and $\kappa_{\mathrm{B}}(1-\lambda)N\log(1-\lambda)$, respectively. Finally, mixing the φ_1- and φ_2-gases of identical density we obtain a φ-gas of N molecules in a volume V at the original temperature. If $S_0(\psi, N)$ denotes the entropy of a ψ-gas of N molecules (in a volume V and at the given temperature), we conclude that

$$
\begin{aligned}
&S_0(\varphi_1, \lambda N) + S_0(\varphi_2, (1-\lambda)N) \\
&\quad = S_0(\varphi, N) + \kappa_{\mathrm{B}} \lambda N \log \lambda + \kappa_{\mathrm{B}}(1-\lambda) N \log(1-\lambda)
\end{aligned}
$$

must hold, where κ_{B} is Boltzmann's constant. Assuming that $S_0(\psi, N)$ is proportional to N and dividing by N we have

$$
\begin{aligned}
&\lambda S(\varphi_1) + (1-\lambda) S(\varphi_2) \\
&\quad = S(\varphi) + \kappa_{\mathrm{B}} \lambda \log \lambda + \kappa_{\mathrm{B}}(1-\lambda) \log(1-\lambda)\,,
\end{aligned}
\tag{5}
$$

where S is certain thermodynamical entropy quantity (relative to the fixed temperature and molecule density). We arrived at the mixing property of entropy but we should not forget about the starting assumption: φ_1 and φ_2 are supposed to be disjoint. In order to get the entropy of a φ-gas with statistical operator D_φ we need one more postulate: We shall assume that the entropy of pure states vanishes. If $D_\varphi = \sum_i \lambda_i p_i$ is the spectral decomposition of the statistical operator (p_i's are pairwise orthogonal minimal projections), then

$$
S(\varphi) = -\kappa_{\mathrm{B}} \sum \lambda_i \log \lambda_i
\tag{6}
$$

is easily inferred by repeated application of the mixing equation (5). In this book Bolzmann's constant κ_{B} will be taken 1. This choice makes the entropy dimensionless (and correspondingly, the temperature should be measured in erg instead of Kelvin). The above gedanken experiment is a heuristical motivation for von Neumann's entropy formula (6). During the deduction we overlooked a problem arising from the ambiguity of the decomposition $\varphi = \lambda\varphi_1 + (1-\lambda)\varphi_2$. Nevertheless, a closer inspection shows that a different decomposition $\varphi = \mu\psi_1 + (1-\mu)\psi_2$ would not yield a contradicting conclusion.

The statistical operator is fundamental in quantum theory, but one should be aware of the fact that not all states of any quantum system can be described by a statistical operator. For example, type III von Neumann algebras appearing typically in quantum field theory do not have pure normal states. Notwithstanding this, the von Neumann entropy of a density matrix is right way to understand quantum entropy. Entropy is not an observable like position or angular momentum. There does not exist self-adjoint operator such that its expectation value in a state would yield the entropy of this state. The von Neumann entropy is a functional on the states of the quantum system which can not be represented by an operator.

The von Neumann entropy of a state expresses the degree of "mixing". Pure states have vanishing entropy. Since the state space of a quantum system is never a Choquet simplex, a state φ may have several extremal decompositions like

$$
\varphi = \textstyle\sum_i \lambda_i \varphi_i \qquad \left(\sum_i \lambda_i = 1, \quad \lambda_i \geq 0 \right),
\tag{7}
$$

where every φ_i is a pure state by definition. Then

$$
S(\varphi) = \inf\{H(\lambda_1, \lambda_2, \ldots) : \varphi = \textstyle\sum_i \lambda_i \varphi_i\}
$$

where the inf is over all (countable) extremal decompositions and H is given by (1). The infimum is attained on orthogonal decompositions (that is, if φ_i and φ_j are orthogonal whenever $i \neq j$). These are decompositions without any interference effects. When φ_i and φ_j are not orthogonal, λ_i and λ_j represent probabilities of non-disjoint events. As we mentioned above, it may happen that φ does not admit an extremal decomposition (7). Then one assigns the value $+\infty$ to the von Neumann entropy. The von Neumann entropy is often infinite and this fact creates a need for other entropy concepts, relative entropy, Rényi entropy, entropy density, etc.

It was Shannon who interpreted the quantity (1) as information. He showed that, up to a constant factor, (1) is the only function of the probabilities $p_1, p_2, \dots, p_k$ that satifies some postulates natural for an information quantity. Information and uncertainty are very close concepts. The quantity (1) may be thought of as uncertainty as well because we are unable to predict exactly the outcome of an experiment having different outcomes with probabilities $p_1, p_2, \dots, p_k$. For example, when one of the probabilities is near to 1, our uncertainty in the prediction is small, and so is (1). On the other hand, having performed the experiment, the amount of information we have gained (or more exactly, what we are expected to gain) is expressed by (1). Hence we shall regard information and uncertainty as complementary notions. Uncertainty is simply considered negative information or the lack of information.

A possible reason for uncertainty is the large number of alternatives. Suppose that a source produces N long messages from an "alphabet" $\{1, 2, \dots, k\}$ and $1 \leq i \leq k$ appears with probability p_i. The number of all possible messages is k^N, but for large N most of them will show up with very small probability. (This is a consequence of the law of large numbers.) If we consider a set L of messages with total probability $1 - \varepsilon$, then this set contains at least $\exp(N\,H(p_1, p_2, \dots, p_k))$ messages,

$$\#(L) \geq \exp(N\,H(p_1, p_2, \dots, p_k))\,. \tag{8}$$

Thus, asymptotically (as $N \to \infty$, ε fixed) there is a simple exponential relation between the Shannon entropy $H(p_1, p_2, \dots, p_k)$ and the number of alternatives in a set of large probability. A similar result is true in quantum probability. Assume that we are in a position to investigate a quantum system $\mathcal{A}$ composed of N independent and identical subsystems $\mathcal{B}$. Imagine, that a stistical operator D is given on the Hilbert space $\mathcal{H}$ associated with the individual systems. The composite system can be in the state described on $\mathcal{H}_N = \mathcal{H} \otimes \mathcal{H} \otimes \dots \otimes \mathcal{H}$ by the density $D_N = D \otimes D \otimes \dots \otimes D$ (both tensor products are N-fold). A subspace $\mathcal{K}$ of the tensor product $\mathcal{H}_N$ carries most of the probability if for the corresponding orthogonal projection P

$$\mathrm{Tr}\, DP \geq 1 - \varepsilon$$

holds. The number of alternatives within $\mathcal{K}$ is the dimension of $\mathcal{K}$, or equivalently, the rank of P. Corresponding to the above probabilistic result we have

$$\operatorname{Tr} P = \dim \mathcal{K} \geq \exp(N\, S(D))\,. \tag{9}$$

Although the Hilbert space probability is very different from classical probability, the analogy between (8) and (9) is striking.

The Shannon entropy (1) measures the amount of information carried by the probability distribution $(p_1, p_2, \dots, p_k)$. Suppose again that p_i is the probabilty of an outcome i in an experiment. If the experiment is observed roughly, for example, if we do not distinguish between the outcomes 1 and 2, then the observation provides less information. In this example, the relevant probability distribution is $(p_1 + p_2, p_3, \dots, p_k)$, and correspondingly

$$H(p_1 + p_2, p_3, \dots, p_k) < H(p_1, p_2, p_3, \dots, p_k)\,.$$

The partition $\{\{1,2\},\{3\},\dots\{k\}\}$ is coarser than $\{\{1\},\{2\},\{3\},\dots\{k\}\}$, and indeed a coarser partition always has smaller Shannon entropy. In statistical language, grouped data yields less information. The following analysis of the quantum mechanical analogue leads to a delicate point in the theory of quantum entropy.

Let D_{12} be a density matrix of a composite system $\mathcal{H}_1 \otimes \mathcal{H}_2$. The measurement of observables of the form like $A \otimes I$ gives only partial information on the state D_{12} of the composite system. The observer does not get complete information about the linear functional

$$\varphi_{12} : X \mapsto \operatorname{Tr} D_{12} X \qquad (X \in B(\mathcal{H}_1 \otimes \mathcal{H}_2) \equiv B(\mathcal{H}_1) \otimes B(\mathcal{H}_2))\,,$$

instead, his information is confined to the restriction φ_1 of φ_{12} to the first subsystem $B(\mathcal{H}_1)$. (This is what we shall call coarse graining.) The restriction φ_1 possesses a density matrix D_1 and according to our intuition D_1 carries less information then D_{12}. However, it is by no means necessarily true that the von Neumann entropy of D_1 is less than that of D_{12}. For example, if

$$D_{12} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{then} \quad D_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and $S(D_1) = \log 2$ while $S(D_{12}) = 0$ because D_{12} corresponds to a pure state. In fact, much more is known. Any state of $B(\mathcal{H})$ is the restriction of a pure state of $B(\mathcal{H}) \otimes B(\mathcal{H})$. This is a pure quantum phenomena, and this problem with density D_{12} has already caused a lot of problems to philosophers of quantum theory. If given a larger total system $\mathcal{A}$ (the composite system $B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ in the above situation) and a part of the observables is neglected for certain reasons and only a subalgebra $\mathcal{B} \subset \mathcal{A}$ is considered, we shall speak of a coarse graining. When coarse graining is performed, our uncertainty is decreased since prediction is easier if a part of the complicated circumstances is neglected. Let the total system $\mathcal{A}$ be in a state φ. Then the coarse grained state is $\varphi|\mathcal{B}$. Since there is no relation between the von Neumann entropies $S(\varphi)$ and $S(\varphi|\mathcal{B})$, this concept is not suitable to measure the information contained in the restriction to the subalgebra $\mathcal{B}$ in the state φ of $\mathcal{A}$. Hence there is a need for the entropy of a coarse graining, say $H_\varphi(\mathcal{B})$, which must possess the following obvious properties.

(i) $H_\varphi(\mathcal{C}) \le H_\varphi(\mathcal{B})$ whenever $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$.
(ii) $H_\varphi(\mathcal{A}) = S(\varphi)$.

From the viewpoint of communication theory any coarse graining gives rise to a channeling transformation. The state φ of $\mathcal{A}$ is transformed into $\varphi|\mathcal{B}$ and this application is an affine mapping from the state space of $\mathcal{A}$ into that of $\mathcal{B}$. (In communication theory one would speak of mutual information instead of the entropy of coarse graining.) The entropy of coarse graining is a very important concept. An observation on a system is a coarse graining in a wide sense, and our knowledge on a quantum system is based on observations.

In this book the relative entropy of two states is regarded as the fundamental notion and the different other kinds of entropies will be deduced from the relative entropy (like the entropy of coarse graining discussed above). This approach does not follow the path of history, since quantum relative entropy appeared in the work of Umegaki only in 1962. In the simplest case the relative entropy of ω with respect to φ will be defined as

$$S(\omega,\varphi) = \mathrm{Tr}\, D_\omega(\log D_\omega - \log D_\varphi)\,, \tag{10}$$

provided that the densities D_ω and D_φ have positive eigenvalues. The origin of probabilistic relative entropy is in mathematical statistics, where it is used to measure how much the state φ differs from ω in the sense of statistical distinguishability. Briefly, the larger $S(\omega,\varphi)$ is, the more information for discriminating between the hypotheses ω and φ can be obtained from an observation. There $S(\omega,\varphi)$ is called information for discrimination. The same aspect of relative entropy is present in the quantum statistical setting as well. Perhaps relative entropy is not the most appropriate terminology, but it is rather widespread. For a state ω of the matrix algebra $M_n(\mathbb{C})$ one has

$$S(\omega) = \log n - S(\omega,\tau)\,, \tag{11}$$

where τ is the tracial state Tr/n. The negative sign in (11) implies that the von Neumann entropy and the relative entropy are rather different, for example concerning convexity. The von Neumann entropy as a real uncertainty quantity is a concave functional of the state, while the relative entropy is (jointly) convex (as it is an information quantity, in fact). The von Neumann entropy may be expressed in terms of the relative entropy without the use of a trace functional. Namely,

$$S(\varphi) = \sup\left\{\textstyle\sum_i \lambda_i S(\varphi_i,\varphi) : \varphi = \sum_i \lambda_i \varphi_i\right\}, \tag{12}$$

where the greatest lower bound is over all finite convex decompositions in the state space. Formula (12) is the key to the definition of the entropy of coarse graining (satisfying the desiderata (i) and (ii)).

From the point of view of physics the most transparent path to relative entropy starts from a simple thermodynamical model. Consider a finite quantum system $\mathcal{A}$ of spins that is described by a Hamiltonian $H = H^* \in \mathcal{A}$. The free energy functional (also called Helmholtz free energy) is defined on the state space of $\mathcal{A}$ as

$$F : \omega \mapsto \omega(H) - \frac{1}{\beta} S(\omega) \quad \text{where} \quad \beta = \frac{1}{\kappa_{\mathrm{B}} T}$$

is the inverse temperature and $\omega(H)$ is the internal energy term. This functional is minimized by the canonical state (also called Gibbs state)

$$\varphi^c(a) = \frac{\mathrm{Tr}\, a\, e^{-\beta H}}{\mathrm{Tr}\, e^{-\beta H}} \qquad (a \in \mathcal{A}). \tag{13}$$

The minimal value of the free energy functional is given by

$$F_0 = -\frac{1}{\beta} \log \mathrm{Tr}\, e^{-\beta H} \tag{14}$$

and is often called Gibbs free energy. In this example the relative entropy $S(\omega, \varphi^c)$ is the increment of free energy, $\beta(F(\omega) - F_0)$ and in the thermodynamical limit (as as the size of the system tends to infinity) one meets the most important use of quantum entropy in statistical thermodynamics. The gap between finite systems and infinite systems is very serious in the behavior of entropy. To indicate what we have in mind, take von Neumann entropy. On the state space of a finite quantum system this is a continuous real functional. In infinite dimension, or in the thermodynamical limit discontinuities occur (and become dominant) which are frequently related to interesting physical phenomenon and it is by no means an irritating pathology.

In a common quantum statistical model the states of a quantum mechanical system correspond to positive normalized functionals of an algebra of operators. The observer of the physical system obtains information by measurements. In the mathematical model measurements are measures taking positive operator values in the operator algebra. If the system is in a state φ and $E(H)$ is the measure of the set $H \subset \mathbb{R}^n$ then $\varphi(E(H))$ is interpreted as the probability that the measured quantity has its value in the set H. Suppose that we have to decide between two states φ and ω on the algebra. Performing a measurement we receive two probability distributions μ_φ and μ_ω. The decision between φ and ω must be done on the basis of μ_φ and μ_ω. The result may be satisfactory only in the case in which the measurement performed has extracted all the information on the mutual relation of the two states. The information contained in the measures μ_φ and μ_ω is typically less than the information carried by the couple φ and ω. The formulation in terms of relative entropy is the following.

$$S(\mu_\omega, \mu_\varphi) \le S(\omega, \varphi) \tag{15}$$

In the fortunate case the two information quantities coincide and we shall call such a measurement sufficient. Sufficient measurements do not exist for "non-commuting" states and this is nothing else but a form of the quantum mechanical uncertainty. Although the uncertainty principle is conventionally expressed by non-commuting observables, something similar appears in connection with two non-commuting states as well. Loss of relative entropy of two states is a form of uncertainty during a measurement. A large number of identical and independent systems makes the mean uncertainty small when the measurement is chosen properly.

Due to its importance we devote some more discussion to the position-momentum uncertainty. Recall that P and Q are self-adjoint operators on $\mathcal{H} = L^2(\mathbb{R})$ defined by

$$(e^{\mathrm{i}xP} f)(t) = f(t+x) \quad \text{and} \quad (e^{\mathrm{i}yQ} f)(t) = e^{\mathrm{i}yt} f(t) \tag{16}$$

for every $f \in L^2(\mathbb{R})$ and $x, y, t \in \mathbb{R}$. Let ξ be a vector of norm 1 (in other words, ξ is a wave function) and set

$$\xi_{xy} = e^{\mathrm{i}yQ} e^{-\mathrm{i}xP} \xi \qquad (x, y \in \mathbb{R}).$$

Then one obtains a probability density

$$\varrho(x,y) = \frac{1}{2\pi} \langle \xi_{xy}, D\xi_{xy} \rangle$$

on the phase space $\mathbb{R}^2$ for any statistical operator D on $\mathcal{H}$. It is an uncertainty relation in the phase space of quantum mechanics that the density ϱ can not be arbitrarily localized. The entropic uncertainty relation states that

$$H(D, \xi) = -\int_{\mathbb{R}^2} \varrho(x,y) \log \varrho(x,y)\, dx\, dy \ge \log 2\pi e\,. \tag{17}$$

In the most interesting applications the entropies are connected with a limiting procedure. This can be seen already in formula (3), which was the origin of the Boltzmann-Gibbs entropy $H(p_1, p_2, \dots, p_k)$. In thermodynamics the limiting procedure is the thermodynamical limit, and the emerging entropy density is a macroscopic variable, moreover, it is one of the ingredients of the Gibbs variational principle. Statistical mechanics provides much room for entropic variational expressions. To take an axample, in a d dimensional quantum lattice gas model, a copy $M_n(\mathbb{C})$ of the $n \times n$ matrix algebra is associated to each $i \in \mathbb{Z}^d$. For a finite subset $\Lambda \subset \mathbb{Z}^d$ a mean-field local hamiltonian is given by

$$H_\Lambda = \sum_{i \in \Lambda} h_i + \sum_{i \in \Lambda} x_i \times \frac{1}{|\Lambda| - 1} \sum_{\substack{j \in \Lambda \\ i \ne j}} x_j \tag{18}$$

where h_i and x_i are copies of some self-adjoint matrices at the corresponding lattice site. The thermodynamical limit of the local free energy density

$$F_\Lambda(\beta) = -\frac{1}{|\Lambda|\beta} \log \mathrm{Tr}\, \exp\big(\beta\, H_\Lambda\big)$$

resembles probabilistic large deviation theory and it is the sample of perturbational limit theorems. There entropies appear in the role of a rate function like in the case of classical theory.

In the course of testing a quantum system the upper bound of the available average information is a characteristic of the evolution of the system. This is the content of dynamical entropy providing a quantum extension of the Kolmogorov-Sinai entropy of classical dynamical systems. The quantum dynamical entropy will be obtained by abelian modelling of the dynamical evolution and become a conjugation invariant quantity.

Entropy shows up very naturally in the central limit theorems: Normal fluctuations manifest themselves in the form of a maximum entropy principle. In the course of the central limit the relative entropy with respect to the limiting state converges to 0.

Notes and Remarks. The quantum logic approach is summarized here in order to show similarities and differences in the stochastic description of classical and quantum behavior. There are several books that treat the quantum logical approach in more detail, we may mention [Emch 1972], [Gudder 1979], [Beltrametti and Cassinelli 1981]. We have to emphasize that in books on quantum mechanics states are often understood as pure states in our sense.

The expression (1) is due to [Boltzmann 1877]. In the derivation of the entropy formula (6) von Neumann benefited from earlier arguments by Einstein and Szilárd. The gedanken experiment is analyzed in more detail in Chapter 5 of [Neumann 1932]. The excellent survey [Wehrl 1978] is a suggested overview on entropy and it contains a number of references on the early developments. Since the publication of that paper several "rather tricky and involved" proofs in quantum entropy theory have become natural and standard.

Part I

Entropies for Finite Quantum Systems

1 Fundamental Concepts

The notion of entropy was introduced by Clausius in order to discuss the thermal behaviour of physical systems. After the work of Bolztmann, Gibbs and others, von Neumann and Shannon contributed to the theory of entropy. Their starting points were very different, von Neumann was motivated by quantum mechanics and Shannon founded communication theory.

The entropy of a state describing a physical system is a quantity expressing the uncertainty or randomness of the system. Shannon regarded this uncertainty attached to a physical system as the amount of information carried by the system, so that the entropy of a state can be read as the information carried by the state. His idea comes from the following consideration : If a physical system has a large uncertainty and one receives information on the system by some procedure, then the so-obtained information is more valuable than that received from a system having less uncertainty.

The *relative entropy* is an information measure representing the uncertainty of a state with respect to another state. Hence it indicates a kind of distance between the two states. In information theory the relative entropy $S(\nu, \mu)$ of two finite probability distributions $\nu = (\kappa_1, \kappa_2, \ldots, \kappa_n)$ and $\mu = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ on an n-point space is usually defined by

$$S(\nu, \mu) = \begin{cases} \sum_i \kappa_i (\log \kappa_i - \log \lambda_i) & \text{if } \lambda_i = 0 \text{ implies } \kappa_i = 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{1.1}$$

We shall regard relative entropy as the most fundamental concept and take it as a starting point for introducing several other kinds of entropy-like quantities. First we shall discuss the generalisation of the above concept of relative entropy to finite quantum systems. In this deductive treatment the basic properties of entropy will be obtained from those of the relative entropy.

We shall follow an algebraic approach and by a finite quantum system we shall mean a finite dimensional C*-algebra. (This does not mean that in this chapter any knowledge of operator algebra theory would be required.) The algebraic viewpoint does not distinguish between isomorphic algebras. For example, the algebra of $n \times n$ complex diagonal matrices is obviously identified with the algebra of n-tuples of complex numbers. To work in the framework of finite quantum systems is advantageous didactically, because all linear operators are bounded in this case, so that complicated domain problems are avoided. Another advantage is that a number of problems can be treated explicitly in finite dimension and some of these techniques and results can be carried over by an approximation procedure to the infinite dimension, which is more interesting for applications in physics.

One knows that a finite dimensional C*-algebra is a direct sum of full matrix algebras. On such an algebra $\mathcal{A}$ there exists a unique linear functional Tr, which takes the value 1 at each minimal projection. It is "tracial" in the sense that

$$\mathrm{Tr}\, ab = \mathrm{Tr}\, ba \qquad (a, b \in \mathcal{A})$$

and the integer $\mathrm{Tr}\, I$ is called the *tracial dimension* of $\mathcal{A}$. (Readers feeling uneasy with this C*-algebraic language may simply think of elements of algebras as finite square matrices.) To explain the relation of Tr and the matrix trace, assume that $\mathcal{A}$ is a direct sum

$$\oplus_{j=1}^{k} M_{n_j}(\mathbb{C})$$

where $M_{n_j}(\mathbb{C})$ is the full matrix algebra of $n_j \times n_j$ complex matrices. If tr_j is the usual matrix trace on $M_{n_j}(\mathbb{C})$ then

$$\mathrm{Tr}\,(a_1 \oplus a_2 \oplus \ldots \oplus a_k) = \sum_{j=1}^{k} \mathrm{tr}_j\, a_j\,.$$

In particular, $\mathrm{Tr}\, I = \sum_{j=1}^{k} n_j$.

Every functional ω on $\mathcal{A}$ is uniquely represented by a *density operator* $D_\omega \in \mathcal{A}$ in the form

$$\omega(a) = \mathrm{Tr}\, D_\omega\, a \qquad (a \in \mathcal{A})\,. \tag{1.2}$$

If ω is positive then $D_\omega \geq 0$ and the normalisation $\omega(I) = 1$ implies $\mathrm{Tr}\, D_\omega = 1$.

The *relative entropy* of ω with respect to φ is defined by

$$S(\omega, \varphi) = \begin{cases} \mathrm{Tr}\, D_\omega(\log D_\omega - \log D_\varphi) & \text{if } \mathrm{supp}\, D_\varphi \geq \mathrm{supp}\, D_\omega \\ +\infty & \text{otherwise.} \end{cases} \tag{1.3}$$

Here $\mathrm{supp}\, D_\psi$ denotes the smallest projection p such that $\psi(p) = \psi(I)$. In particular, $S(\omega, \varphi)$ is always finite if the density of φ has strictly positive eigenvalues. (Such a φ is called faithful.) When D_φ commutes with D_ω and their eigenvalue lists are $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $(\kappa_1, \ldots, \kappa_n)$ respectively, then $S(\,.\,,\,.\,)$ reduces to the classical expression (1.1). Although we mostly speak of the relative entropy of states it is convenient to allow ω and φ in the definition of $S(\omega, \varphi)$ to be arbitrary positive functionals. It is straightforward that the relative entropy of arbitrary positive functionals obeys the transformation rule

$$S(\lambda\omega, \mu\varphi) = \lambda S(\omega, \varphi) - \lambda \log \mu - \eta(\lambda)\omega(I), \qquad (\lambda, \mu \in \mathbb{R}^+) \tag{1.4}$$

which is called *scaling property* in the sequel. The notation $\eta(t)$ for the continuous function $-t \log t$ is standard in information theory.

The entropy $S(\omega)$ of a functional ω is defined by means of its density operator as

$$S(\omega) = \mathrm{Tr}\, \eta(D_\omega)\,. \tag{1.5}$$

This notion was introduced by von Neumann in 1927 and we shall term it *von Neumann entropy* or shortly entropy. Up to a sign the entropy is the relative entropy with respect to the functional Tr, that is,

$$S(\omega) = -S(\omega, \mathrm{Tr}\,) = -\sum_i \kappa_i \log \kappa_i = \sum_i \eta(\kappa_i)\,.$$

The unnormalised Tr plays the role of a uniform distribution on $\mathcal{A}$ since it associates the same value to each minimal projection. In terms of the normalised trace $\tau = \mathrm{Tr}\,/\mathrm{Tr}\,(I)$ we have $S(\omega) = -S(\omega, \tau) + \omega(I) \log \mathrm{Tr}\,(I)$. Observe that our definition (1.5) includes the so-called Shannon entropy of a finite probability distribution. If $\mathcal{A} = \mathbb{C}^n$ and $\omega(c_1, c_2, \ldots, c_n) = \sum_i p_i c_i$ with a probability distribution $(p_1, p_2, \ldots, p_n)$ then $S(\omega) = \sum_i \eta(p_i)$. The Shannon entropy $\sum_i \eta(p_i)$ represents the information carried by the random variable which takes the value x_i with probability p_i $(i = 1, 2, \ldots, n)$.

The aim of this chapter is to obtain the most important properties of the von Neumann entropy and of the relative entropy by elementary methods in the frame of finite quantum systems. Later on these properties will be reproved in more general situations but by means of heavier mathematical tools.

Proposition 1.1 If φ and ω are states on a finite quantum system then the inequality

$$S(\omega, \varphi) \ge \tfrac{1}{2} \mathrm{Tr}\,(D_\omega - D_\varphi)^2$$

holds.

Proof. The inequality is trivial if $\mathrm{supp}\, D_\varphi$ is strictly smaller than $\mathrm{supp}\, D_\omega$. Hence we assume that $\mathrm{supp}\, D_\varphi = I$. Let

$$D_\varphi = \sum_i \lambda_i p_i \qquad \text{and} \qquad D_\omega = \sum_j \kappa_j q_j$$

be the spectral decompositions. Since for some $\xi \in (x, y)$

$$-\eta(x) + \eta(y) + (x-y)\eta'(y) = -\tfrac{1}{2}(x-y)^2 \eta''(\xi) \ge \tfrac{1}{2}(x-y)^2 \tag{1.6}$$

we have

$$\begin{aligned} &\mathrm{Tr}\, p_i q_j \left(-\eta(D_\omega) + \eta(D_\varphi) + (D_\omega - D_\varphi)\eta'(D_\varphi) - \tfrac{1}{2}(D_\omega - D_\varphi)^2\right) \\ &\quad = \left(-\eta(\kappa_j) + \eta(\lambda_i) + (\kappa_j - \lambda_i)\eta'(\lambda_i) - \tfrac{1}{2}(\kappa_j - \lambda_i)^2\right) \mathrm{Tr}\, p_i q_j \ge 0\,. \end{aligned}$$

Summation over i and j gives

$$\mathrm{Tr}\,\big(-\eta(D_\omega)+\eta(D_\varphi)+(D_\omega-D_\varphi)(I-\log D_\varphi)-\tfrac{1}{2}(D_\omega-D_\varphi)^2\big)\geq 0$$

which was to be proven. □

It follows that for states ω and φ the quantity $S(\omega,\varphi)$ is always nonnegative and $S(\omega,\varphi)=0$ occurs only in the case $\omega=\varphi$. To prove convexity, a key property of the relative entropy, we need some preparation.

Lemma 1.2 Let A be a positive operator and V a contraction. Then

$$(V^*AV)^t \geq V^*A^tV$$

holds for every $0<t<1$.

Proof. For a projection P and an invertible positive operator T one obtains from the decomposition

$$PT^{-1}P = P(PTP)^{-1}P + PT^{-1}P^\perp(P^\perp T^{-1}P^\perp)^{-1}P^\perp T^{-1}P \tag{1.7}$$

the inequality

$$PT^{-1}P \geq P(PTP)^{-1}P\,. \tag{1.8}$$

Therefore, we have

$$(\lambda+PSP)^{-1} \leq P(\lambda+S)^{-1}P+\lambda^{-1}P^\perp \tag{1.9}$$

for $\lambda>0$ and $S\geq 0$. In order to estimate $(V^*AV)^t$ we use the formula

$$T^t = \frac{\sin t\pi}{\pi}\int_0^\infty \lambda^{t-1}-\lambda^t(\lambda+T)^{-1}\,d\lambda\,. \tag{1.10}$$

The contraction V may be written in the form PUP where U is a unitary in a larger Hilbert space and P is the orthogonal projection onto the domain of V. Hence (1.9) yields

$$\begin{aligned}(\lambda+V^*AV)^{-1} &= P(\lambda P+PU^*PAPUP)^{-1}P \leq P(\lambda+U^*PAPU)^{-1}P\\ &= PU^*(\lambda+PAP)^{-1}UP\\ &\leq PU^*P(\lambda+PAP)^{-1}PUP+\lambda^{-1}PU^*P^\perp UP\\ &= V^*(\lambda+A)^{-1}V+\lambda^{-1}(I-V^*V)\,.\end{aligned}$$

Finally, we infer

$$\begin{aligned}(V^*AV)^t &= \frac{\sin t\pi}{\pi}\int_0^\infty \lambda^{t-1}-\lambda^t(\lambda+V^*AV)^{-1}\,d\lambda\\ &\geq \frac{\sin t\pi}{\pi}V^*\left(\int_0^\infty \lambda^{t-1}-\lambda^t(\lambda+A)^{-1}\,d\lambda\right)V = V^*A^tV\,,\end{aligned}$$

and the proof is complete. □

Lemma 1.3 Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be finite quantum systems, $\alpha:\mathcal{A}_1\to\mathcal{A}_2$ a linear mapping and $S_1,T_1\in\mathcal{A}_1$, $S_2,T_2\in\mathcal{A}_2$ positive operators. Assume that T_1 and T_2 are invertible and the following conditions hold.

(i) $\alpha(a^*a) \geq \alpha(a)^*\alpha(a) \quad (a \in \mathcal{A}_1)$,
(ii) $\mathrm{Tr}\, S_2\, \alpha(a) \leq \mathrm{Tr}\, S_1\, a \quad (a \in \mathcal{A}_1^+)$,
(iii) $\mathrm{Tr}\, T_2\, \alpha(a) \leq \mathrm{Tr}\, T_1\, a \quad (a \in \mathcal{A}_1^+)$.

Then

$$\mathrm{Tr}\, \alpha(x)^* S_2^t \alpha(x) T_2^{1-t} \leq \mathrm{Tr}\, x^* S_1^t x T_1^{1-t}$$

for every $0 < t < 1$ and $x \in \mathcal{A}_1$.

Proof. $\mathcal{A}_1$ and $\mathcal{A}_2$ become Hilbert spaces with the Hilbert-Schmidt inner product:

$$\langle a, b \rangle = \mathrm{Tr}\, a^* b\,.$$

Set a linear operator $V : \mathcal{A}_1 \to \mathcal{A}_2$ by

$$V\, a\, T_1^{1/2} = \alpha(a)\, T_2^{1/2} \qquad (a \in \mathcal{A}_1).$$

The simple calculation

$$\|\alpha(a) T_2^{1/2}\|^2 = \mathrm{Tr}\, T_2\, \alpha(a)^*\alpha(a) \leq \mathrm{Tr}\, T_2\, \alpha(a^*a) \leq \mathrm{Tr}\, T_1\, a^*a = \|a\, T_1^{1/2}\|^2$$

shows that V is a contraction.

We introduce two other operators Δ and D by

$$\Delta(a\, T_1^{1/2}) = S_1\, a\, T_1^{-1/2} \qquad (a \in \mathcal{A}_1)$$
$$D(b\, T_2^{1/2}) = S_2\, b\, T_2^{-1/2} \qquad (b \in \mathcal{A}_2)$$

Δ and D are positive and

$$\Delta^t(a\, T_1^{1/2}) = S_1^t\, a\, T_1^{1/2-t} \qquad (a \in \mathcal{A}_1)$$
$$D^t(b\, T_2^{1/2}) = S_2^t\, a\, T_2^{1/2-t} \qquad (b \in \mathcal{A}_2)$$

hold for any $t > 0$.

The chain of inequalities

$$\begin{aligned}\langle V^* D V a T_1^{1/2}, a T_1^{1/2} \rangle &= \langle D\alpha(a) T_2^{1/2}, \alpha(a) T_2^{1/2} \rangle = \langle S_2 \alpha(a) T_2^{1/2}, \alpha(a) T_2^{-1/2} \rangle \\ &= \mathrm{Tr}\, S_2 \alpha(a) \alpha(a)^* \leq \mathrm{Tr}\, S_2 \alpha(aa^*) \leq \mathrm{Tr}\, S_1\, aa^* = \langle \Delta a T_1^{1/2}, a T_1^{1/2} \rangle\end{aligned}$$

tells us that

$$V^* D V \leq \Delta\,. \tag{1.11}$$

According to Lemma 1.2 we have

$$V^* D^t V \leq (V^* D V)^t$$

and one obtains from (1.10) and (1.11)

$$(V^* D V)^t \leq \Delta^t$$

for $0 < t < 1$. Hence

$$\langle V^* D^t V x T_1^{1/2}, x T_1^{1/2}\rangle \le \langle \Delta^t x T_1^{1/2}, x T_1^{1/2}\rangle$$

which is identical to the stated inequality. □

Lemma 1.3 is a version of the celebrated *Lieb concavity* result. The linear mapping $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ is called *Schwarz mapping* when the inequality

$$\alpha(a^* a) \ge \alpha(a)^* \alpha(a)$$

holds for every $a \in \mathcal{A}_1$. If T and S are densities of the linear functionals ω and φ, their relative entropy can be obtained by differentiation of

$$\mathrm{Tr}\, S^t T^{1-t}$$

at $t = 1$.

Let $\omega_1, \omega_2, \omega, \varphi_1, \varphi_2, \varphi$ be states of a finite quantum system $\mathcal{A}$ such that

$$\omega = \lambda\omega_1 + (1-\lambda)\omega_2 \qquad \varphi = \lambda\varphi_1 + (1-\lambda)\varphi_2$$

with some $0 < \lambda < 1$. We choose $\mathcal{A}_1 = \mathcal{A}$ and $\mathcal{A}_2 = \mathcal{A} \oplus \mathcal{A}$. Let $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ be the homomorphism defined by

$$a \mapsto a \oplus a \qquad (a \in \mathcal{A}).$$

Set the states ω_{12} and φ_{12} of $\mathcal{A}_2$ as

$$\omega_{12}(a \oplus b) = \lambda\omega_1(a) + (1-\lambda)\omega_2(b) \quad (a, b \in \mathcal{A})$$
$$\varphi_{12}(a \oplus b) = \lambda\varphi_1(a) + (1-\lambda)\varphi_2(b) \quad (a, b \in \mathcal{A}).$$

If S_1, T_1, S_2, T_2 are the densities of $\varphi, \omega, \varphi_{12}, \omega_{12}$ respectively, then the hypothesis of Lemma 1.3 holds. We have

$$(1-t)^{-1}(1 - \mathrm{Tr}\, S_2^t T_2^{1-t}) \ge (1-t)^{-1}(1 - \mathrm{Tr}\, S_1^t T_2^{1-t})$$

and letting $t \nearrow 1$ yields

$$S(\omega, \varphi) \le S(\omega_{12}, \varphi_{12}).$$

Since

$$S(\omega_{12}, \varphi_{12}) = \lambda S(\omega_1, \varphi_1) + (1-\lambda) S(\omega_2, \varphi_2),$$

what we have proven is the joint convexity of the relative entropy.

Theorem 1.4 If $\omega_1, \omega_2, \varphi_1, \varphi_2$ are states of a finite quantum system and $0 < \lambda < 1$ then

$$S(\lambda\omega_1 + (1-\lambda)\omega_2, \lambda\varphi_1 + (1-\lambda)\varphi_2) \le \lambda S(\omega_1, \varphi_1) + (1-\lambda) S(\omega_2, \varphi_2).$$

Proof. Strictly speaking, in the above proof we have used the invertibility of the densities of φ and φ_{12}. If $\mathrm{supp}\, \omega_1 \le \mathrm{supp}\, \varphi_1$ and $\mathrm{supp}\, \omega_2 \le \mathrm{supp}\, \varphi_2$ do not hold then the stated inequality is trivial. Therefore one may assume that

$$I = \mathrm{supp}\, \varphi_1 \ge \mathrm{supp}\, \varphi_2 \ge \mathrm{supp}\, \omega_2.$$

The density of φ is then invertible and a small alteration of the above proof will help. (With $p = \mathrm{supp}\, \varphi_2$, take $\mathcal{A}_2 = \mathcal{A} \oplus p\mathcal{A}p$ and $\alpha(a) = a \oplus pap$.) □

Lemma 1.3 includes also the monotonicity of the relative entropy under Schwarz mappings. The next result is *Uhlmann's monotonicity theorem.*

Theorem 1.5 Let $\alpha : \mathcal{A}_1 \to \mathcal{A}_2$ be a unital Schwarz mapping between the finite quantum systems $\mathcal{A}_1$ and $\mathcal{A}_2$. Then for states φ and ω of $\mathcal{A}_2$ the inequality

$$S(\omega \circ \alpha, \varphi \circ \alpha) \le S(\omega, \varphi)$$

holds.

Proof. The proof proceeds by differentiation from Lemma 1.3. □

Let $\mathcal{A}$ be a finite quantum system and $\mathcal{A}_0$ a subalgebra of $\mathcal{A}$. Theorem 1.5 tells us that

$$S(\omega|\mathcal{A}_0\, ,\ \varphi|\mathcal{A}_0) \le S(\omega, \varphi) \tag{1.12}$$

for arbitrary states ω and φ. Note that equality must hold in (1.12) if there exists a Schwarz mapping $\beta : \mathcal{A} \to \mathcal{A}_0$ that leaves both φ and ω invariant. Later on we shall discuss this possibility in detail. Choosing the trivial subalgebra we can conclude from (1.12) that the relative entropy of states is nonnegative.

Some properties of the entropy can be deduced from those of the relative entropy.

Proposition 1.6 Let $\mathcal{A}$ be a finite quantum system with states φ and ω. Then for every $0 < \lambda < 1$ the following inequality holds

$$\begin{aligned}\lambda\, S(\varphi) + (1-\lambda) S(\omega) &\le S(\lambda\varphi + (1-\lambda)\omega) \\ &\le \lambda\, S(\varphi) + (1-\lambda) S(\omega) + H(\lambda, 1-\lambda)\end{aligned}$$

where $H(\lambda, 1-\lambda) = \eta(\lambda) + \eta(1-\lambda)$.

Proof. The first inequality is an immediate consequence of the convexity of the relative entropy. In order to obtain the second inequality we benefit from the formula

$$\begin{aligned}&\operatorname{Tr} A(\log(A+B) - \log A) \\ &\qquad = \int_0^\infty \operatorname{Tr} A(A+t)^{-1} B (A+B+t)^{-1}\, dt \ge 0 \qquad (A, B \ge 0)\end{aligned}$$

and infer

$$\operatorname{Tr} \lambda D_\varphi\, \log(\lambda\, D_\varphi + (1-\lambda) D_\omega) \ge \operatorname{Tr} \lambda D_\varphi\, \log \lambda D_\varphi$$

and

$$\operatorname{Tr}\,(1-\lambda) D_\omega\, \log(\lambda D_\varphi + (1-\lambda) D_\omega) \ge \operatorname{Tr}\,(1-\lambda) D_\omega\, \log(1-\lambda) D_\omega\, .$$

Adding the latter two inequalities we obtain the second inequality of the proposition. □

Since η is continuous on $\mathbb{R}^+$, the entropy is a continuous functional on the states. In some cases, a more concrete estimate for the continuity might be useful. We recall that if φ and ω are states with densities D_φ and D_ω, respectively, then

$$\|\varphi - \omega\| = \mathrm{Tr}\,|D_\varphi - D_\omega|\,,$$

where $|A|$ stands for $(A^*A)^{1/2}$. Separating the negative and nonnegative eigenvalues of $D_\varphi - D_\omega$ we may write

$$D_\varphi - D_\omega = x - y \tag{1.13}$$

where x and y are nonnegative operators with disjoint supports. So

$$|D_\varphi - D_\omega| = x + y\,.$$

(Let us note that (1.13) is called Jordan decomposition.)

Lemma 1.7 Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ be the eigenvalues of the self-adjoint matrices A and B, respectively. Then

$$\mathrm{Tr}\,|A - B| \geq \sum_{i=1}^{n} |\lambda_i - \mu_i|\,.$$

Proof. Let $x - y$ be the Jordan decomposition of $A - B$. So $\mathrm{Tr}\,|A - B| = \mathrm{Tr}\,x + \mathrm{Tr}\,y$. Consider $C = B + x = A + y$ with eigenvalues $\nu_1 \geq \nu_2 \geq \ldots \geq \nu_n$. Since $C \geq A, B$ we have

$$\nu_i \geq \lambda_i, \mu_i \qquad (i = 1, 2, \ldots, n)$$

which implies

$$2\nu_i - \lambda_i - \mu_i \geq |\lambda_i - \nu_i|\,.$$

Summing up we obtain

$$\sum_{i=1}^{n} |\lambda_i - \mu_i| \leq \sum_{i=1}^{n} (2\nu_i - \lambda_i - \mu_i) = \mathrm{Tr}\,(2C - A - B) = \mathrm{Tr}\,x + \mathrm{Tr}\,y$$

and the lemma is proven. □

The von Neumann entropy is the trace of a continuous function of the density matrix, hence it is an obviously continuous functional on the states. However, a more precise estimate for the continuity will be required in approximations. Such an estimate is due to *Fannes*.

Proposition 1.8 Let φ and ω be states of a finite quantum system $\mathcal{A}$ with tracial dimension $d = \mathrm{Tr}\,I$. If $\|\varphi - \omega\| < \frac{1}{3}$, the inequality

$$|S(\varphi) - S(\omega)| \leq \|\varphi - \omega\| \log d + \eta(\|\varphi - \omega\|)$$

holds.

Proof. Let $\lambda_1 \geq \ldots \geq \lambda_d$ and $\mu_1 \geq \ldots \geq \mu_d$ be the eigenvalues of the densities D_φ and D_ω. We set $\varepsilon_i = |\lambda_i - \mu_i|$ and $\varepsilon = \sum_i \varepsilon_i$. From the elementary inequality

$$|\eta(s) - \eta(t)| \leq \eta(s-t) \qquad 0 \leq s-t < \tfrac{1}{3}$$

we have

$$|S(\varphi) - S(\omega)| \leq \sum_{i=1}^{d} |\eta(\lambda_i) - \eta(\mu_i)| \leq \sum_{i=1}^{d} \eta(\varepsilon_i)$$

and

$$\sum_{i=1}^{d} \eta(\varepsilon_i) = \varepsilon \sum_{i=1}^{d} \eta\left(\frac{\varepsilon_i}{\varepsilon}\right) - \varepsilon \log \varepsilon \leq \varepsilon \log d + \eta(\varepsilon)\,.$$

Here the following facts were used. By the previous lemma $\varepsilon \leq \|\varphi - \omega\|$ and η is monotone on the interval $[0, \frac{1}{3}]$. □

The *strong subadditivity* of the entropy follows easily from the monotonicity of the relative entropy (Theorem 1.5). The strong subadditivity property is related to the composition of three different systems and it is essential in the proof of the existence of the global entropy density functional of lattice systems. Let $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ be finite quantum systems and set $\mathcal{A}_{123} = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3$, $\mathcal{A}_{12} = \mathcal{A}_1 \otimes \mathcal{A}_2$ and $\mathcal{A}_{23} = \mathcal{A}_2 \otimes \mathcal{A}_3$. For a state φ_{123} of $\mathcal{A}_{123}$ we denote by $\varphi_{12}, \varphi_{23}, \varphi_2$ its restrictions to $\mathcal{A}_{12}, \mathcal{A}_{23}, \mathcal{A}_2$, respectively. The strong subadditivity asserts the following.

Proposition 1.9 $S(\varphi_{123}) + S(\varphi_2) \leq S(\varphi_{12}) + S(\varphi_{23})$.

When the algebra $\mathcal{A}_2$ is trivial we have

$$S(\varphi_{13}) \leq S(\varphi_1) + S(\varphi_3) \tag{1.14}$$

which is referred to as subadditivity.

Proof. In order to prove the strong subadditivity we start with the identities

$$S(\varphi_{123}, \mathrm{Tr}_{123}) = S(\varphi_{12}, \mathrm{Tr}_{12}) + S(\varphi_{123}, \varphi_{12} \otimes \mathrm{Tr}_3) \tag{1.15}$$

and

$$S(\varphi_2, \mathrm{Tr}_2) + S(\varphi_{23}, \varphi_2 \otimes \mathrm{Tr}_3) = S(\varphi_{23}, \mathrm{Tr}_{23}) \tag{1.16}$$

where Tr with a subscript denotes the trace functional in the corresponding algebra. (Note that Tr_{123} restricted to $\mathcal{A}_i$ differs from Tr_i by a factor of $\mathrm{Tr}_i(I)/\mathrm{Tr}_{123}(I)$.) Due to the monotonicity we have

$$S(\varphi_{123}, \varphi_{12} \otimes \mathrm{Tr}_3) \geq S(\varphi_{23}, \varphi_2 \otimes \mathrm{Tr}_3)\,. \tag{1.17}$$

So that addition of (1.15)–(1.17) yields

$$S(\varphi_{123}, \mathrm{Tr}_{123}) + S(\varphi_2, \mathrm{Tr}_2) \geq S(\varphi_{12}, \mathrm{Tr}_{12}) + S(\varphi_{23}, \mathrm{Tr}_{23})$$

which is the assertion, up to sign. □

Let $\mathcal{A}$ be a finite quantum system with Hamiltonian $H \in \mathcal{A}^{sa}$. The *free energy functional* at the inverse temperature $\beta = (\kappa_{\mathrm{B}} T)^{-1}$

$$F(\varphi) = \varphi(H) - \frac{1}{\beta} S(\varphi)$$

is defined for the states of the system. (κ_{B} is Boltzmann's constant.) Next we show that there is a unique canonical state φ_c determined by the thermodynamic stability condition, namely, by the condition that φ_c should minimize the free energy functional at temperature T.

Proposition 1.10 If $\Phi(\beta) = -\frac{1}{\beta} \log \operatorname{Tr} e^{-\beta H}$, then

$$F(\varphi) \begin{cases} = \Phi(\beta) & \text{for } \varphi = \varphi_c \\ > \Phi(\beta) & \text{otherwise,} \end{cases}$$

where φ_c is the canonical state given by the density $\exp(-\beta(H - \Phi(\beta)))$.

Proof. The proof follows from the observation

$$\frac{1}{\beta} S(\varphi, \varphi_c) = F(\varphi) - \Phi(\beta).$$

Since $S(\varphi, \varphi_c) \geq 0$, with equality if and only if $\varphi_c = \varphi$ (see Proposition 1.1), we have

$$F(\varphi) \geq \Phi(\beta)$$

and equality holds only in the case $\varphi = \varphi_c$. □

In order to treat a natural generalisation of Proposition 1.10 to relative entropy we introduce *perturbed functionals*. Let φ be a state (or a positive functional) on a finite quantum system and let $h \in \mathcal{A}$ be self-adjoint. Set D_φ for the density and p for the support projection of φ. The the perturbed functional φ^h is given by the density matrix

$$p \exp(\log D_\varphi + php)$$

where $\log D_\varphi$ is understood as $p(\log(D_\varphi + p^\perp))$. Note that the functional φ^h is not normalized and $\operatorname{supp} \varphi^h = \operatorname{supp} \varphi$.

Proposition 1.11 For a state φ of $\mathcal{A}$ with support projection $p \in \mathcal{A}$ and $h = h^* \in \mathcal{A}$ the following variational formula holds

$$\log \varphi^h(I) = \max\{\omega(php) - S(\omega, \varphi) : \omega \quad \text{is a state of } \mathcal{A}\}.$$

The maximum is attained at the state $\omega = \varphi^h / \varphi^h(I)$.

Proof. The quantity $S(\omega,\varphi)-\omega(php)$ is exactly $S(\omega,\varphi^h)$. According to the monotonicity of the relative entropy

$$S(\omega,\varphi^h)\geq\omega(I)(\log\omega(I)-\log\varphi^h(I)),$$

and equality holds if and only if $\omega=\varphi^h/\varphi^h(I)$. □

The frame of finite dimensional C*-algebras contains two extremes. One extreme is the case when the centre of the algebra $\mathcal{A}$ is $\mathcal{A}$ itself, i.e. $\mathcal{A}$ is commutative and the other extreme is when the centre reduces to scalar multiples of the identity. (The latter algebras, i.e. algebras with trivial centre are called factors.) When $\mathcal{A}$ is written in the form

$$\mathcal{A}=\oplus_{j=1}^{k}M_{n_j}(\mathbb{C}),$$

then the identity of the summand $M_{n_j}(\mathbb{C})$ is a central projection of $\mathcal{A}$, and $\mathcal{A}$ is a factor if and only if $k=1$. On the other hand $\mathcal{A}$ is commutative if and only if $n_j=1$ for every $j=1,2,\dots,k$. Let us consider inclusions $\mathcal{B}\subset\mathcal{A}$ with finite dimensional C*-algebras $\mathcal{A}$ and $\mathcal{B}$. When $\mathcal{A}$ is commutative it is represented by complex functions on a finite space and to each subalgebra $\mathcal{B}$ there corresponds a partition of the space such that $\mathcal{B}$ consists of functions taking a constant value on each set of the partition. If $\mathcal{A}$ and $\mathcal{B}$ are factors, for example if $\mathcal{A}=M_n(\mathbb{C})$ and $\mathcal{B}=M_m(\mathbb{C})$ then, one can see that the inclusion $\mathcal{B}\subset\mathcal{A}$ forces $n=tm$ for an integer t, and up to an isomorphism $\mathcal{A}=\mathcal{B}\otimes M_t(\mathbb{C})$. The integer t is termed the multiplicity of $\mathcal{B}$ in $\mathcal{A}$ and

$$t=\frac{\mathrm{Tr}_{\mathcal{A}}b}{\mathrm{Tr}_{\mathcal{B}}b}\qquad(b\in\mathcal{B}).$$

(Note that in case of a factor our steadily used functional Tr coincides with the common matrix trace.) Turning to the general case, we set

$$\mathcal{B}=\oplus_{i=1}^{l}M_{m_i}(\mathbb{C})$$

and assume $\mathcal{B}\subset\mathcal{A}$. The inclusion representation $\pi:\mathcal{B}\to\mathcal{A}$ restricted to $M_{m_i}(\mathbb{C})$ is decomposed into k summands and $\pi(M_{m_i}(\mathbb{C}))I_{n_j}$ has a certain multiplicity $\mu(i,j)$ in $M_{n_j}(\mathbb{C})$. (Here I_{n_j} is written for the identity of $M_{n_j}(\mathbb{C})$ and it is a central projection of $\mathcal{A}$). The matrix $\mu(i,j)$ with l rows and k columns describes the inclusion $\mathcal{B}\subset\mathcal{A}$, and it is termed inclusion matrix.

Proposition 1.12 Let $\mathcal{B}\subset\mathcal{A}$ be finite quantum systems and $\tau=\mathrm{Tr}_{\mathcal{A}}/\mathrm{Tr}_{\mathcal{A}}I$ the normalized trace functional on $\mathcal{A}$. Then there exists a unique linear mapping $E:\mathcal{A}\to\mathcal{B}$ such that

(i) if $a\in\mathcal{A}^+$ then $E(a)\in\mathcal{B}^+$,
(ii) $E(b)=b$ for every $b\in\mathcal{B}$,
(iii) $E(ab)=E(a)b$ holds for every $a\in\mathcal{A}$ and $b\in\mathcal{B}$,
(iv) $\tau(E(a))=\tau(a)$ for every $a\in\mathcal{A}$.

Proof. Endow $\mathcal{A}$ with the Hilbert-Schmidt inner product $\langle a_1, a_2 \rangle = \tau(a_2^* a_1)$. One can observe readily that $E : \mathcal{A} \to \mathcal{B}$ shares the properties (i)–(iv) if and only if E is the orthogonal projection onto $\mathcal{B}$ with respect to the Hilbert-Schmidt inner product. This supplies us with both the existence and the uniqueness of E. □

A linear mapping $E : \mathcal{A} \to \mathcal{B}$ fulfilling conditions (i)–(iii) in the above proposition is usually called a *conditional expectation.* This terminology is borrowed from probability theory and motivated by the case of a commutative $\mathcal{A}$. Conditional expectations are frequently used tools.

The constant multiple of the linear mapping E of Proposition 1.12 is sometimes called partial trace. We are speaking here of the mapping

$$F(a) = \frac{\mathrm{Tr}_{\mathcal{A}}\, I}{\mathrm{Tr}_{\mathcal{B}}\, I} E(a) \qquad (a \in \mathcal{A}) \tag{1.18}$$

which possesses the property

$$\mathrm{Tr}_{\mathcal{A}}\, a = \mathrm{Tr}_{\mathcal{B}}\, F(a) \qquad (a \subset \mathcal{A})$$

and sends the density of a state φ of $\mathcal{A}$ to the density of $\varphi|\mathcal{B}$.

When $\mathcal{B} \subset \mathcal{A}$, then for arbitrary states φ and ω on $\mathcal{A}$ the inequality

$$S(\omega, \varphi) - S(\omega|\mathcal{B}, \varphi|\mathcal{B}) \geq 0 \tag{1.19}$$

holds due to the monotonicity of the relative entropy (cf. Theorem 1.5 and (1.17)). We shall see that in certain situations the difference on the left hand side of (1.19) is a relative entropy quantity itself.

Theorem 1.13 Let $\mathcal{B} \subset \mathcal{A}$ be finite quantum systems and φ a state of $\mathcal{A}$ with invertible density. Assume that there exists a conditional expectation $E : \mathcal{A} \to \mathcal{B}$ such that $\varphi \circ E = \varphi$. Then for every state ω of $\mathcal{A}$ we have

$$S(\omega, \varphi) - S(\omega|\mathcal{B}, \varphi|\mathcal{B}) = S(\omega, \omega \circ E) \tag{1.20}$$

Proof. We denote by $D_\varphi^{\mathcal{A}}$, $D_\omega^{\mathcal{A}}$, $D_\varphi^{\mathcal{B}}$, $D_\omega^{\mathcal{B}}$ and D the densities of states φ, ω, $\varphi|\mathcal{B}$, $\omega|\mathcal{B}$ and $\omega \circ E$, respectively. The existence of E implies the strong condition

$$(D_\omega^{\mathcal{B}})^t (D_\varphi^{\mathcal{B}})^{-t} = D^t (D_\varphi^{\mathcal{A}})^{-t} \qquad (t \in \mathbb{R}) \tag{1.21}$$

for these densities. Now we multiply (1.21) by $D_\omega^{\mathcal{A}}$ and take $\mathrm{Tr}_{\mathcal{A}}$. The derivative of

$$\mathrm{Tr}_{\mathcal{A}}\, D_\omega^{\mathcal{A}} (D_\omega^{\mathcal{B}})^t (D_\varphi^{\mathcal{B}})^{-t} = \mathrm{Tr}_{\mathcal{B}} (D_\omega^{\mathcal{B}})^{1+t} (D_\varphi^{\mathcal{B}})^{-t}$$

at $t = 0$ gives $S(\varphi|\mathcal{B}, \omega|\mathcal{B})$. On the other hand, the derivative of

$$\mathrm{Tr}_{\mathcal{A}}\, D_\omega^{\mathcal{A}} D^t (D_\varphi^{\mathcal{A}})^{-t}$$

equals

$$\mathrm{Tr}_{\mathcal{A}}\, D_\omega^{\mathcal{A}} (\log D - \log D_\varphi^{\mathcal{A}}) = S(\omega, \varphi) - S(\omega, \omega \circ E)\,,$$

which is the conditional expectation property (1.20). □

We shall discuss condition (1.21) in terms of the theory of operator algebras later on in more detail. The conditional expectation property (1.20) will be used in the next chapter as the crucial postulate for the relative entropy.

Let $\mathcal{A}, \mathcal{B}, \varphi, E$ be as in Theorem 1.13 and ω_0 a state on $\mathcal{B}$. It is a consequence of the last theorem that the functional $\omega \mapsto S(\omega, \varphi)$ takes its least value on the convex set $\{\omega : \omega|\mathcal{B} = \omega_0\}$ at $\omega = \omega_0 \circ E$. This minimization problem is important for statistical inference.

In order to improve the lower bound of Proposition 1.1 we need the following lemma.

Lemma 1.14 For all $0 \le u \le v \le 1$ the inequality

$$2(u-v)^2 \le u \log \frac{u}{v} + (1-u) \log \frac{1-u}{1-v}$$

holds.

Theorem 1.15 For states φ and ω of a finite quantum system $\mathcal{A}$

$$S(\omega, \varphi) \ge \tfrac{1}{2} \|\omega - \varphi\|^2 .$$

Proof. Let p be the spectral projection of $D_\varphi - D_\omega$ corresponding to the positive eigenvalues and take $q = I - p$. We may assume that

$$u \equiv \mathrm{Tr}\, D_\omega p \le \mathrm{Tr}\, D_\varphi p \equiv v .$$

Set $\mathcal{B}$ for the two dimensional commutative subalgebra of $\mathcal{A}$ generated by the projection p. We have

$$\|\varphi - \omega\| = \mathrm{Tr}\,(D_\varphi - D_\omega)p - \mathrm{Tr}\,(D_\varphi - D_\omega)q = 2(v-u)$$

and Lemma 1.14 tells us

$$2(v-u) \le \sqrt{2S(\omega|\mathcal{B}, \omega|\mathcal{B})} .$$

The rest of the proof follows from the monotonicity

$$S(\omega|\mathcal{B}, \varphi|\mathcal{B}) \le S(\omega, \varphi) ,$$

established in (1.12). □

Inspection of the proof of Theorem 1.15 reveals that the lemma contained the result in the (two dimensional) commutative case and the idea was to reduce the general case to the commutative one. To compute the norm $\|\varphi - \omega\|$ it is enough to know these states on commutative subalgebras and

$$\|\varphi - \omega\| = \max\{\|\varphi|\mathcal{A}_0 - \omega|\mathcal{A}_0\| \,:\, \mathcal{A}_0 \subset \mathcal{A} \text{ is commutative}\} .$$

In contrast to this property of the norm our next aim is to show that in general

$$S(\omega, \varphi) > \sup\{S(\omega|\mathcal{A}_0, \varphi|\mathcal{A}_0) \,:\, \mathcal{A}_0 \subset \mathcal{A} \text{ is commutative}\} . \tag{1.22}$$

Proposition 1.16 Let φ_0 and φ_1 be faithful states of the finite quantum system $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{A}$ a commutative subalgebra. If

$$S(\varphi_1, \varphi_0) = S(\varphi_1 | \mathcal{B}, \varphi_0 | \mathcal{B})$$

then D_{φ_0} commutes with D_{φ_1} and $D_{\varphi_0} D_{\varphi_1}^{-1} \in \mathcal{B}$.

Proof. For $0 \le t \le 1$ we set

$$\varphi_t = (1-t)\varphi_0 + t\varphi_1$$

and we write simply D_t for the density of φ_t. With the notation

$$\begin{aligned} K(t) &= S(\varphi_t, \varphi_0) - tS(\varphi_1, \varphi_0) \\ L(t) &= S(\varphi_t | \mathcal{B}, \varphi_0 | \mathcal{B}) - tS(\varphi_1 | \mathcal{B}, \varphi_0 | \mathcal{B}) \\ y(t) &= K(t) - L(t) \end{aligned}$$

we have

$$y(0) = y(1) = 0 \qquad \text{and} \qquad y(t) \ge 0 \qquad (0 \le t \le 1)\,.$$

We are going to show that y is convex. To this end we estimate y''.

Using the formula (3.16) we infer

$$K'(t) = \operatorname{Tr}_{\mathcal{A}} S \log(tS + D_0) - \operatorname{Tr}_{\mathcal{A}}(D_1 \log D_1 - D_0 \log D_0)$$

where $S = D_1 - D_0$. Now by means of the representation

$$\log x = \int_0^\infty (1+t)^{-1} - (x+t)^{-1}\, dt$$

we obtain

$$K''(t) = \operatorname{Tr}_{\mathcal{A}} \int_0^\infty S(D_t + u)^{-1} S(D_t + u)^{-1}\, du\,, \tag{1.23}$$

and similarly

$$L''(t) = \operatorname{Tr}_{\mathcal{B}} \int_0^\infty F(S)(F(D_t) + u)^{-1} F(S)(F(D_t) + u)^{-1}\, du\,,$$

where $F : \mathcal{A} \to \mathcal{B}$ is the partial trace. Since $\mathcal{B}$ is supposed to be commutative we have

$$L''(t) = \operatorname{Tr}_{\mathcal{B}}(F(S)^2 F(D_t)^{-1})\,. \tag{1.24}$$

If the spectral resolution of D_t is $\sum_i \lambda_i^{(t)} p_i^{(t)}$, then from (1.23)

$$K''(t) = \sum_{i,j} \operatorname{Tr} S p_i^{(t)} S p_j^{(t)} / \operatorname{Lm}(\lambda_i^{(t)}, \lambda_j^{(t)})\,,$$

where Lm stands for the logarithmic mean: $\operatorname{Lm}(\alpha, \beta) = (\alpha - \beta)/(\log\alpha - \log\beta)$ if $\alpha \ne \beta$, and $\operatorname{Lm}(\alpha, \alpha) = \alpha$. The elementary inequality $\operatorname{Lm}(\alpha, \beta) \le (\alpha + \beta)/2$

holds and if α is different from β then the logarithmic mean is strictly smaller than the arithmetic one. Hence

$$K''(t) \geq \operatorname{Tr} SW(t)\,,$$

where $W(t) = 2\sum_{i,j}(p_i^{(t)} S\, p_j^{(t)})/(\lambda_i^{(t)} + \lambda_j^{(t)})$ is a self-adjoint element of $\mathcal{A}$. It is easy to check that

$$D_t W(t) + W(t) D_t = 2S\,.$$

Hence

$$\operatorname{Tr} SW(t) = \tfrac{1}{2}\operatorname{Tr}\left(D_t W(t)^2 + W(t) D_t W(t)\right) = \operatorname{Tr} D_t W(t)^2\,.$$

With the choice of $V(t) = F(S)F(D_t)^{-1}$ the identity

$$\operatorname{Tr} SW(t) = \operatorname{Tr} D_t V(t)^2 + \operatorname{Tr} D_t(W(t) - V(t))^2 + 2\operatorname{Tr}\left(SV(t) - D_t V(t)^2\right)$$

gives

$$\operatorname{Tr} SW(t) = \operatorname{Tr} D_t V(t)^2 + \operatorname{Tr} D_t(W(t) - V(t))^2 \tag{1.25}$$

and

$$K''(t) \geq \operatorname{Tr} SW(t) \geq \operatorname{Tr} D_t V(t)^2 = \operatorname{Tr}_{\mathcal{B}} F(D_t) V(t)^2 = L''(t)\,.$$

The last inequality yields $y'' \geq 0$, establishing the convexity of y.

Since

$$y'(0) = S(\varphi_1, \varphi_0) - S(\varphi_1|\mathcal{B}, \varphi_0|\mathcal{B}) = 0,$$

y must vanish on the whole interval [0,1] due to our hypothesis. This implies that

$$K''(t) = \operatorname{Tr} SW(t)$$

and for $i \neq j$ the coefficient of $(\operatorname{Lm}(\lambda_i^{(t)}, \lambda_j^{(t)}))^{-1}$, which is

$$\operatorname{Tr} S\, p_i^{(t)} S\, p_i^{(t)}\,,$$

vanishes, too. The identity

$$\operatorname{Tr}|SD_t - D_t S|^2 = \sum_{i,j}(\lambda_i^{(t)} - \lambda_j^{(t)})^2 \operatorname{Tr} S\, p_i^{(t)} S\, p_j^{(t)}$$

shows that $SD_t = D_t S$ or equivalently, D_0 and D_1 commute. From (1.25) we conclude

$$W(t) = V(t)\,,$$

which for $t = 1$ reads as

$$F(D_0)F(D_1)^{-1} = D_0\, D_1^{-1}\,.$$

This condition implies that $D_0\, D_1^{-1} \in \mathcal{B}$. □

By some compactness arguments one can see that the supremum in (1.22) is in fact a maximum. Hence, in the light of Proposition 1.16, if the densities of φ and ω do not commute, then (1.22) holds with strict inequality.

It is understood in probability theory that the notion of (Shannon, or measure theoretic) entropy has successful applications in a variety of subjects because it determines the asymptotic behaviour of certain probabilities in the course of independent trials. Now we will discuss these phenomena for finite quantum systems and to do this a kind of law of large numbers is required.

Let $\mathcal{A}$ be a finite quantum system with a faithful state ω. The n-fold algebraic tensor product $\mathcal{A}_n = \mathcal{A} \otimes \ldots \otimes \mathcal{A}$ is again a finite quantum system, and the product functional $\omega_n = \omega \otimes \ldots \otimes \omega$ is a state of $\mathcal{A}_n$. Using the obvious identifications the inclusion $(\mathcal{A}_n, \omega_n) \subset (\mathcal{A}_m, \omega_m)$ holds for $n \le m$, and we set

$$(\mathcal{A}_\infty, \omega_\infty) = \cup\{(\mathcal{A}_n, \omega_n) \,:\, n \in \mathbb{N}\}\,.$$

On the $*$-algebra $\mathcal{A}_\infty$ the right shift endomorphism γ is defined for $a_1 \otimes a_2 \otimes \ldots \otimes a_n \in \mathcal{A}_n$ as

$$\gamma(a_1 \otimes a_2 \otimes \ldots \otimes a_n) = I \otimes a_1 \otimes a_2 \otimes \ldots \otimes a_n \in \mathcal{A}_{n+1}$$

and ω_∞ is obviously invariant under γ. Now perform the GNS-construction with the state ω_∞ and arrive at the triplet $(\pi, \mathcal{H}, \Omega)$. We identify $\mathcal{A}_\infty$ through its faithful representation π with a subalgebra of the generated von Neumann algebra $\mathcal{M} = \pi(\mathcal{A}_\infty)'' \subset \mathcal{B}(\mathcal{H})$. The normal state

$$\omega(a) = \langle \Omega, a\Omega \rangle \qquad (a \in \mathcal{M})$$

is an extension of ω_∞, and the endomorphism γ extends to $\mathcal{M}$ so that the relation $\omega \circ \gamma = \omega$ is preserved. (For the sake of simpler notation we do not use a new letter for the extension.)

The following result may be called the weak law of large numbers for independent finite quantum systems.

Proposition 1.17 In the situation described above the following statements hold.

(i) If $a \in \mathcal{M}$ and $\gamma(a) = a$ then $a \in \mathbb{C}I$.
(ii) For every $a \in \mathcal{M}$ the sequence $S_n(a) = n^{-1}(a + \gamma(a) + \ldots + \gamma^{n-1}(a))$ converges to $\omega(a)I$ in the strong operator topology.
(iii) If $a \in \mathcal{M}^{sa}$ and $I \subset \mathbb{R}$ is a closed interval such that $\omega(a) \notin I$, and p_n is the spectral projection of $S_n(a)$ corresponding to the interval I, then $p_n \to 0$ in the strong operator topology.

Proof. We apply the statistical ergodic theorem for the isometry $V : \mathcal{H} \to \mathcal{H}$ defined by the formula

$$Va\Omega = \gamma(a)\Omega \qquad (a \in \mathcal{M})\,.$$

Set E for the projection onto the fixed point space of V. In order to show condition (i) we prove that E is of rank one. It is easy to see that for $a, b \in \mathcal{A}_\infty$ the relation

$$\omega(aS_n(b)) \to \omega(a)\omega(b) \tag{1.26}$$

holds. Equivalently,

$$\langle a^* \Omega,\, n^{-1}(I + V + \ldots + V^{n-1})b\Omega\rangle \to \langle \Omega,\, a\Omega\rangle\langle \Omega,\, b\Omega\rangle\,.$$

Therefore we have

$$\langle a^* \Omega,\, Eb\Omega\rangle = \langle \Omega,\, a\Omega\rangle\langle \Omega,\, b\Omega\rangle$$

for every $a, b \in \mathcal{M}$. The vector Ω being cyclic the equality

$$Eb\Omega = \langle \Omega,\, b\Omega\rangle\Omega = \omega(b)\Omega$$

must hold and E is really of rank one.

The ergodicity condition (i) implies easily that the state ω is faithful on $\mathcal{M}$. Hence the vector Ω is cyclic for the commutant $\mathcal{M}'$ too. The statistical ergodic theorem tells us that

$$S_n(a)b'\Omega \to \omega(a)b'\Omega$$

for every $b' \in \mathcal{M}'$. Since $(S_n(a))_n$ is norm bounded and $\mathcal{M}'\Omega$ is dense, we are able to conclude that $s_n(a) \to \omega(a)I$ strongly. Statement (ii) is proved and (iii) follows. □

Let us fix a positive number $\varepsilon < 1$. For a while we shall say that a projection $Q_n \in \mathcal{A}_n$ is rather sure if $\omega_n(Q_n) \geq 1 - \varepsilon$. On the other hand, the size of Q_n, i.e. the cardinality of a maximal pairwise orthogonal family of projections contained in Q_n, is given by $\mathrm{Tr}_n\, Q_n$. (The subscript n in Tr_n indicates that the algebraic trace functional on $\mathcal{A}_n$ is meant here.) The theorem below says that it is the von Neumann's entropy of ω which governs asymptotically the size of the rather sure projections: A rather sure projection in $\mathcal{A}_n$ contains at least $\exp(nS(\omega))$ pairwise orthogonal minimal projections.

Theorem 1.18 Under the above conditions and with the above notation the limit relation

$$\lim_{n\to\infty} \tfrac{1}{n} \inf\{\log \mathrm{Tr}_n\, Q_n : Q_n \in \mathcal{A}_n \text{ is a projection, } \omega_n(Q_n) \geq 1 - \varepsilon\} = S(\omega)$$

holds.

Proof. If D_n denotes the density of ω_n then one can see easily that

$$-\log D_n = \sum_{i=0}^{n-1} \gamma^i(-\log D_1)$$

where γ stands for the right shift. The sequence $(\gamma^i(-\log D_1))_i$ behaves as independent identically distributed random variables with respect to the state

ω_∞. More precisely, the previous proposition applies for $a = -\log D_1$ and tells us that

$$-\tfrac{1}{n}\log D_n \to S(\omega) I$$

strongly. Let $P(n,\delta)$ be the spectral projection of the self-adjoint operator $-n^{-1}\log D_n$ corresponding to the interval $(S(\omega)-\delta,\, S(\omega)+\delta)$. According to (iii) of Proposition 1.17 one has

$$P(n,\delta) \to I \tag{1.27}$$

strongly for every $\delta > 0$. In particular,

$$\omega(P(n,\delta)) = \langle P(n,\delta)\Omega, \Omega\rangle \to 1$$

as $n \to \infty$, and $P(n,\delta)$ is a rather sure projection if n is large enough. It follows from the definition of $P(n,\delta)$ that

$$D_n\, P(n,\delta) \exp(n\, S(\omega) - n\delta) \le P(n,\delta) \le D_n\, \exp(n\, S(\omega) + n\delta) \tag{1.28}$$

which gives

$$\tfrac{1}{n}\log\, \mathrm{Tr}_n\, P(n,\delta) \le S(\omega) + \delta\,.$$

Since $\delta > 0$ was arbitrary we establish

$$\limsup_{n\to\infty} \tfrac{1}{n}\, \inf\{\log \mathrm{Tr}_n\, Q_n \;:\; Q_n\,\} \le S(\omega)\,. \tag{1.29}$$

To prove that $S(\omega)$ is actually the limit we shall argue by contradiction. Assume that there exist a sequence $n(1) < n(2) < \ldots$ of integers, a number $t > 0$ and projections $Q(n(k)) \in \mathcal{A}_{n(k)}$ $(k = 1, 2, \ldots)$ such that

(i) $\omega_\infty(Q(n(k)) \ge 1 - \varepsilon$,
(ii) $\log \mathrm{Tr}_{n(k)} Q(n(k)) \le n(k)(S(\omega) - t)$.

The bounded sequence $(Q(n(k)))_k$ has a weak limit point in the von Neumann algebra $\mathcal{M}$, say $T \in \mathcal{M}$. Instead of selecting a subsequence we suppose that $Q(n(k)) \to T$ weakly. It is straightforward to show that from (1.27) the weak limit

$$Q(n(k))P(n(k),\delta) \to T$$

follows. Consequently,

$$\liminf_{k\to\infty} \omega_\infty(Q(n(k))\, P(n(k),\delta)) \ge \omega(T) \ge 1 - \varepsilon\,. \tag{1.30}$$

Using the first part of (1.28) we estimate

$$\begin{aligned}
\mathrm{Tr}\, Q(n(k)) &\ge \mathrm{Tr}\, Q(n(k))\, P(n(k),\delta) \\
&\ge \mathrm{Tr}\, D_{n(k)} Q(n(k))\, P(n(k),\delta)\ \exp(n\, S(\omega) - n\delta) \\
&= \exp(n\, S(\omega) - n\delta)\omega_\infty(Q(n(k))\, P(n(k),\delta))
\end{aligned}$$

and

$$\liminf_{k\to\infty} \frac{1}{n(k)} \log \mathrm{Tr}_{n(k)} Q(n(k))$$
$$\geq S(\omega) - \delta + \lim_{k\to\infty} \frac{1}{n(k)} \log \omega_\infty (Q(n(k))\, P(n(k),\delta))\,.$$

The limit term on the right hand side vanishes due to (1.30) and we arrive at a contradiction with (ii) if $0 < \delta < t$. This proves the theorem. □

A variant of Theorem 1.18 holds for the relative entropy. Let us keep the setting of Theorem 1.18 but in addition let φ be a faithful state on the finite quantum system $\mathcal{A}$. We write φ_n for the corresponding product states of $\mathcal{A}_n$ ($n \in \mathbb{N}$). Here φ_n will measure the size of the projections in $\mathcal{A}_n$ while ω_n may be interpreted as probability. Set

$$R(n,\varepsilon) = \inf\{\varphi_n\,(Q_n) \,:\, Q_n \in \mathcal{A}_n \text{ is a projection, } \omega_n(Q_n) \geq 1-\varepsilon\}$$

for positive ε. For a tracial state φ the proof of Theorem 1.18 works and yields

$$\lim_{n\to\infty} \tfrac{1}{n} \log R(n,\varepsilon) = -S(\omega,\varphi)\,. \tag{1.31}$$

This relation holds actually for an arbitrary state φ and the relation is called the *quantum Stein lemma.*

Before closing this chapter we explain the concept of mutual entropy which plays an important role both in communication theory and in the definition of dynamical entropy.

By a decomposition of a state φ of a quantum system we mean a finite convex combination $\varphi = \sum_i \lambda_i \varphi_i$ in the state space. This decomposition is called *extremal* if every φ_i is pure and it is called *orthogonal* if $\operatorname{supp}\varphi_i$ is orthogonal to $\operatorname{supp}\varphi_j$ whenever $i \neq j$. For a state of an abelian finite dimensional algebra the extremal decomposition is unique and orthogonal. In case of a general finite quantum system the extremal decomposition is neither unique nor necessarily orthogonal. If the density matrix possesses an eigenvalue with multiplicity, then there exists an infinity of orthogonal extremal decompositions.

A positive unital mapping $\alpha : \mathcal{A}_1 \to \mathcal{A}$ between two finite quantum systems $\mathcal{A}_1$ and $\mathcal{A}$ may be interpreted as a communication channel. The mapping α associates to any state φ of $\mathcal{A}$ a state $\varphi \circ \alpha$ of $\mathcal{A}_1$. On the language of communication theory, φ is an input state and $\varphi \circ \alpha$ is the corresponding output state. The notion of *mutual entropy* represents the amount of information correctly transmitted from the input to the output system. Therefore the mutual entropy does depend on the input state and on the transmission channel. Let $\varphi = \sum_i \lambda_i \varphi_i$ be an orthogonal extremal decomposition. With respect to this decomposition the compound state φ_{co} is defined on the product algebra $\mathcal{A} \otimes \mathcal{A}_1$ as

$$\varphi_{\mathrm{co}} = \sum_i \lambda_i (\varphi_i \otimes \varphi_i \circ \alpha)\,. \tag{1.32}$$

The mutual entropy describing the amount of information transmitted through the channel should indicate some relation between elementary components (pure states) of φ and $\varphi \circ \alpha$. This relation is expressed in the compound state. Let us consider the abelian case. Then $\mathcal{A}_1 = \mathbb{C}^n$, $\mathcal{A} = \mathbb{C}^k$ and a channel is given by a stochastic matrix $(c_{j,i})$ with k rows and n coloums. The formula (1.32) gives the compound measure which contains the correlation between the input and the output states.

The mutual entropy of the input state φ and the channel α for the given decomposition is $S(\varphi_{\mathrm{co}}, \varphi \otimes (\varphi \circ \alpha))$, and the mutual entropy $I(\varphi; \alpha)$ is defined to be the supremum of all this relative entropy quantities over all possible orthogonal extremal decompositions of φ. One checks readily that

$$S(\varphi_{\mathrm{co}}, \varphi \otimes (\varphi \circ \alpha)) = \sum_i \lambda_i S(\varphi_i \circ \alpha, \varphi \circ \alpha)\,, \tag{1.33}$$

hence we may write

$$I(\varphi; \alpha) = \sup\{\textstyle\sum_i \lambda_i S(\varphi_i \circ \alpha, \varphi \circ \alpha) : \varphi = \sum_i \lambda_i \varphi_i\,\}$$

where the least upper bound is taken over all orthogonal extremal decompositions $\varphi = \sum_i \lambda_i \varphi_i$.

Theorem 1.19 $I(\varphi; \alpha) \le \min\{S(\varphi),\, S(\varphi \circ \alpha)\}$.

Proof. Since

$$\sum_i \lambda_i S(\varphi_i \circ \alpha, \varphi \circ \alpha) = S(\varphi \circ \alpha) - \sum_i \lambda_i S(\varphi_i \circ \alpha)$$

the second majorization is clear without any assumption on the decomposition. To prove the first estimate we need to assume the extremality of the decomposition $\varphi = \sum_i \lambda_i \varphi_i$ and we use the monotonicity of the relative entropy.

$$\begin{aligned}\sum_i \lambda_i S(\varphi_i \circ \alpha, \varphi \circ \alpha) &\le \sum_i \lambda_i S(\varphi_i, \varphi) \\ &= -\sum_i \lambda_i \mathrm{Tr}\, \eta(D_{\varphi_i}) + \sum_i \mathrm{Tr}\, \lambda_i (D_{\varphi_i} \log D_\varphi)\end{aligned}$$

which equals to $S(\varphi)$, due to $\eta(D_{\varphi_i}) = 0$. □

The proposition tells us that the quantity of information correctly transmitted through the channel can not exceed the information content of the input state. The difference $S(\varphi) - I(\varphi; \alpha)$ is called *entropy defect*. We note that $S(\varphi \circ \alpha) \le S(\varphi)$ is not true in general.

Finally, we show an example. Let $0 < \lambda < 1$ and define a state φ_λ of $M_2(\mathbb{C})$ by its density

$$D_\lambda = \frac{1}{2}\begin{pmatrix} 1 & 1 - 2\lambda \\ 1 - 2\lambda & 1 \end{pmatrix}.$$

For $\lambda \neq 1/2$ the orthogonal decomposition of φ_λ is unique, in fact

$$D_\lambda = \frac{\lambda}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{1-\lambda}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and for the diagonal subalgebra $\mathcal{D} \subset M_2(\mathbb{C})$ we easily obtain

$$I(\varphi_\lambda; \mathcal{D}) = 0 . \qquad (\lambda \neq 1/2)$$

For $\lambda = 1/2$ the state $\varphi_{1/2}$ is the tracial one and it possesses several orthogonal decompositions. The decomposition

$$D_{1/2} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

yields the lower bound $I(\varphi_{1/2}; \mathcal{D}) \geq \log 2$. Since $I(\varphi_{1/2}; \mathcal{D}) \leq S(\varphi_{1/2}) = \log 2$, we have

$$I(\varphi_{1/2}; \mathcal{D}) = \log 2 .$$

This example shows that $I(\varphi; \mathcal{D})$ is not continuous in φ and it may vanish for a mixed state φ.

Notes and Remarks. In mathematical statistics the relative entropy of two measures was introduced in [Kullback and Leibler 1951] under the name information for discrimination. Rényi later proposed the name information gain but several other expressions are used in the literature of statistics, information and communication theory, such as for example, I-divergence, cross-information etc. In the context of operator algebras $S(\omega, \varphi)$ was defined in [Umegaki 1962] and it was called relative information. The *relative entropy* $S(\omega, \varphi)$ as we use it here is rather an information quantity than an entropy quantity. Indeed, it is convex and decreasing under positive mappings. An entropy quantity is expected to be concave and increasing. (Relying upon the slogan "entropy is the lack of information", one obtains an entropy quantity from an information quantity by changing sign and vice versa.) Concerning $S(\omega, \varphi)$ there are different sign and ordering conventions in the literature. For example, those of the book [Bratteli and Robinson 1981] are related to ours by $-S_{BR}(\omega, \varphi) = S(\omega, \varphi)$.

The quantum-mechanical entropy $S(\varphi) = -\mathrm{Tr}\, D_\varphi \log D_\varphi$ appeared in [von Neumann 1927] where the form of the density of the canonical state was also established. The information theoretical entropy originated from [Shannon 1948]. Proposition 1.1 is from [Streater 1985]. The first concavity results for trace functions were proved in [Lieb 1973a] by a method based on operator-valued analytic functions and it implies the joint convexity of the relative entropy. The present proof for of this property benefits from the operator monotonicity of the function $x \mapsto x^t \quad (0 < t < 1)$, and deduces joint convexity from monotonicity under Schwarz mappings ([Petz 1986b]). (In fact, the basic properties of entropy and relative entropy are rather equivalent, [Wehrl 1978] is a good source in this respect.) Theorem 1.5 is a special case of Theorem 5.3 which was proven in [Uhlmann 1977]. The use of relative entropy in mathematical physics was initiated in the papers [Lindblad 1973, 1974, 1975], too. The *strong subadditivity* of the entropy attracted a lot of attention at the beginning of the rigorous quantum statistical mechanics. The first proof is due to [Lieb and Ruskai 1973a] (see also the review [Lieb 1975]). The estimate of Proposition 1.8 is from [Fannes 1973b]. Lemma 1.7 is a simple consequence of standard eigenvalue inequalities, see for example II.6.5 in [Kato 1966]. Concerning conditional expectations in operator algebra theory we refer to [Strătilă 1981] or to the short survey [Petz 1988a]. The conditional expectation property (1.20) was obtained in [Hiai, Ohya and Tsukada 1981] under the strong condition that φ, restricted to the subalgebra $\mathcal{B}$, be tracial. The form presented here is from [Petz 1986a]. Lemma 1.14 is from [Csiszár 1967] and the subsequent Theorem 1.15 is due to [Hiai, Ohya and Tsukada 1981]. The paper [Petz 1985a] contains Proposition 1.16. It is worthwhile to note that no positive lower bound is known for the difference of the left and right hand sides of inequality (1.22).

Proposition 1.17 is a very special case of well-known ergodic theorems in operator algebras, our formulation is not more than a functional analytic formulation of the simplest law of large numbers. More sophisticated limit theorems may be found in [Jajte 1985]. Theorem 1.18 is the analogue of the so-called fundamental theorem of information theory. Its generalization (1.31) was discussed in [Hiai and Petz 1991].

In the quantum context the *mutual entropy* and the *compound state* was discussed in [Ohya 1983].

Comments. In the standard formalism of quantum mechanics, a quantum mechanical system is described by a Hilbert space and physical observables correspond to (self-adjoint) operators acting on the Hilbert space. When $A \in B(\mathcal{H})$ is an observable and $\Phi \in \mathcal{H}$ is a state vector, then $\langle \Phi, A\Phi \rangle$ is interpreted as the expectation value of the observable A in the quantum state Φ.

If the system of interest is a *composite system* of two subsystems described by the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, then the composite system is described by the tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. Assume that $\{\xi_j\}$ is a basis of $\mathcal{H}_1$ and $\{\eta_i\}$ is a basis of $\mathcal{H}_2$. Given a *state vector* Φ of the composite system, we have an expansion

$$\Phi = \sum_{i,j} w_{ij}\, \xi_j \otimes \eta_i$$

in the product basis. Let A be a linear operator acting on $\mathcal{H}_1$. It is easy to compute that

$$\langle \Phi,\, (A \otimes I)\Phi \rangle = \mathrm{Tr}\, AD_1$$

for an operator D_1 which has the properties $D_1 \geq 0$ and $\mathrm{Tr} D_1 = 1$. Therefore, D_1 is a *statistical operator* or *density matrix* if both Hilbert spaces are of finite dimension. (Actually $D_1 = W^*W$, when w_{ij} is the ij-element of the matrix W.) This argument shows how to arrive at a density matrix from a state vector of a composite system. Note that the statistical operator $|\Phi\rangle\langle\Phi|$ describes the original state of the composite system, and D_1 is the *reduced density matrix* on $\mathcal{H}_1$.

If D_j is a density matrix on $\mathcal{H}_j$ $(j = 1, 2)$, then $D_1 \otimes D_2$ is a density matrix on $\mathcal{H}_1 \otimes \mathcal{H}_2$. This density matrix is of product-type and the corresponding state is a product state. The convex hull of densities of product-type is an important subset of all densities acting on $\mathcal{H}_1 \otimes \mathcal{H}_2$. The elements of this set are called *disentangled* (or separable) states, and the other states are *entangled*. The phenomenon of *entanglement* is a remarkable feature of the Hilbert space probability theory.

The Hilbert space of a spin-$\frac{1}{2}$ is $\mathbb{C}^2$ which has a canonical basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. The space $\mathbb{C}^2 \otimes \mathbb{C}^2$ describes a system of two spins and has the basis $|\uparrow\rangle \otimes |\uparrow\rangle$, $|\uparrow\rangle \otimes |\downarrow\rangle$, $|\downarrow\rangle \otimes |\uparrow\rangle$ and $|\downarrow\rangle \otimes |\downarrow\rangle$. The state vector

$$\Phi = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle) \tag{1.34}$$

bears the name "singlet" and plays a leading role in a paradox due to Einstein, Podolsky and Rosen. The reduced density of $|\Phi\rangle\langle\Phi|$ is

$$\tfrac{1}{2}|\uparrow\rangle\langle\uparrow| + \tfrac{1}{2}|\downarrow\rangle\langle\downarrow|\,.$$

$|\Phi\rangle\langle\Phi|$ is an extreme point in the convex set of density matrices, hence the convex decomposition

$$|\Phi\rangle\langle\Phi| = \sum_i \lambda_i D_1^i \otimes D_2^i$$

implies $|\Phi\rangle\langle\Phi| = D_1 \otimes D_2$. From the equality of the entropies we infer that $S(D_1) = 0$. Therefore, the reduced density matrix of the right hand side has 0 entropy, while the reduced density of $|\Phi\rangle\langle\Phi|$ is the most mixed state (with entropy $\log 2$). The contradiction shows that the vector state $|\Phi\rangle\langle\Phi|$ is entangled. The argument shows that all vector states are entangled except for the product ones.

The proof of Proposition 1.9 gives that equality holds in the subadditivity (1.14) if and only if $\varphi_{13} = \varphi_1 \otimes \varphi_3$. The case of equality in the strong subadditivity is more delicate, see the comments at the end of Chapter 9.

Theorem 1.18 is an abstract form of Shannon's *noiseless coding theorem* in the quantum setting, see Theorem 1.1 in [Csiszár and Körner 1981]. Schumacher completed this result by a coding-encoding procedure and obtained his *compression theorem* for memoryless sources, see [Schumacher 1995], or [Jozsa and Schumacher 1994]. The compression theorem has many extensions, see, for example, [Winter and Massar 2001].

(1.31) is related to the *hypothesis testing* problem. The state ω is the *null hypothesis* and φ is the *alternative hypothesis.* The problem is to decide which hypothesis is true and in the asymptotic situation the n-fold product states

ω_n and φ_n are given. The decision is performed by a two-valued measurement $\{A_n, I - A_n\}$, where $0 \le A_n \le I$ is an observable and A_n corresponds to the acceptance of ω_n and $I - A_n$ corresponds to the acceptance of φ_n. A_n is called a *test*, moreover $\alpha_n := \omega_n(I - A_n)$ is the *error of the first kind* and $\beta_n := \varphi_n(A_n)$ is the *error of the second kind*. Set

$$\beta^*(n,\varepsilon) = \inf\{\varphi_n(A_n) : A_n \in \mathcal{A}_n, \quad 0 \le A_n \le I, \quad \omega_n(I - A_n) \le \varepsilon\},$$

which is the infimum of the error of the second kind when the error of the first kind is at most ε. Obviously,

$$\beta^*(n,\varepsilon) \le R(n,\varepsilon),$$

since in the later minimization is over all tests A_n which are projections. (Such a test is called sharp.) The asymptotic behaviour of the two quantities is the same and the first analysis was done in [Hiai and Petz 1991]. In this paper

$$\limsup_{n\to\infty} \tfrac{1}{n} \log R(n,\varepsilon) \le -S(\omega,\varphi) \tag{1.35}$$

was proved in full generality and

$$\liminf_{n\to\infty} \tfrac{1}{n} \log R(n,\varepsilon) \ge -S(\omega,\varphi) \tag{1.36}$$

was conjectured on the basis of a weaker form. (1.36) was obtained in [Ogawa and Nagaoka 2000] by the method of quasi-entropies and they proved a strong converse of the *quantum Stein lemma* as well.

2 Postulates for Entropy and Relative Entropy

Von Neumann used arguments of phenomenological thermodynamics to deduce certain properties of the entropy of a state described by a statistical operator. His argument (reproduced also in the Introduction) supports the *mixing condition*

$$\begin{aligned} &S(\lambda\varphi + (1-\lambda)\psi) \\ &\quad = \lambda S(\varphi) + (1-\lambda)S(\psi) - \lambda \log \lambda - (1-\lambda)\log(1-\lambda) \end{aligned} \tag{2.1}$$

(for $0 < \lambda < 1$ and disjoint states φ and ψ) and the property

$$S(\varphi) = 0 \quad \text{if} \quad \varphi \quad \text{is pure.} \tag{2.2}$$

The conditions (2.1) and (2.2) determine a unique functional

$$S(\omega) = \operatorname{Tr} \eta(D_\omega),$$

on states of a finite quantum system. (Here $\eta(t) = -t \log t$.)

Creating the foundations of information theory in 1948, Shannon derived from certain desiderata the formula

$$H(p_1, p_2, \dots, p_n) = -\sum_{k=1}^{n} p_k \log p_k = \sum_{k=1}^{n} \eta(p_k) \tag{2.3}$$

for the amount of information contained in a single observation of a random variable ξ which takes the different values $x_1, x_2, \dots, x_n$ with probabilities $p_k = \text{Prob}\,(\xi = x_k) \quad (k = 1, 2, \dots, n)$.

In the present chapter Shannon's axiomatic point of view will be adopted and we shall deduce the von Neumann entropy and the relative entropy from certain postulates. For the sake of simplicity we restrict ourselves to states of finite quantum systems (that is, finite dimensional C*-algebras).

The original set of axioms given by Shannon himself was simplified by Hinčin and Faddeev. In this latter form these axioms are as follows: Let Π denote the set of all finite probability distributions $(p_1, p_2, \dots p_n) \quad (p_i \geq 0, \sum_i p_i = 1)$. Suppose that a function $I : \Pi \to \mathbb{R}$ satisfies the following conditions.

(i) Continuity: $p \mapsto I(p, 1-p)$ is continuous on $[0, 1]$.
(ii) Normalization: $I(1/2, 1/2) = \log 2$.