CN

Chapter 2

CT

The road to equilibrium

A

2.1 The time evolution

We found in the previous section, that a general quantum mechanical system is described by its density operator. Physical systems evolve in time, and this evolution should be contained in their density or state operator. We have arrived here to the central problem in Statistical Mechanics, since this time evolution should describe the irreversibility that we observed in macroscopic phenomena. In particular, if we prepare a system in an arbitrary initial state, the time evolution will drive the system to equilibrium. Since relaxation is an irreversible process, this poses the conundrum of how to describe this phenomenon through a dynamics which is time reversible. Boltzmann solved this problem using his kinetic equation (Boltzmann equation), but his procedure had a hidden breaking of reversibility., which was named 'molecular chaos' (*Stosszahlansatz).* Boltzmann interpreted his results in terms of most probable distributions, arguing that an actual system consists of a large number of molecules, and in the thermodynamic limit (infinite size), reversible paths are vanishingly rare. This intuition seems to be correct, but it has not been fully elucidated and still continues to be considered as an open problem, not solved by the advent of Quantum Mechanics. In what follows, we discuss the dynamics of a quantum system through the time evolution of its density operator. It was not explicitly stated, but the definition of $\boldsymbol{\rho}$ was implied to be formulated in the Schödinger's picture at a fixed time t_0. Within the above framework, dynamics is contained in state kets, which evolve following the Schrödinger equation:

$$i\hbar \frac{\partial |\psi^{(i)}(t, t_0)>}{\partial t} = \mathbf{H}|\psi(t, t_0)> \ , \tag{2.1}$$

where $\mathbf{H}$ is the Hamiltonian, and the evolving kets are written as $|\psi(t, t_0)>$. This results in a time evolution for $\boldsymbol{\rho}$ which is closely related to the Schrödinger equation, since the weighs of the mixture $\{w_i\}$ are fixed and defined at the initial time t_0. It follows that $\boldsymbol{\rho}$ is written as:

$$\boldsymbol{\rho}(t, t_0) \equiv \sum_i w_i \ |\psi^{(i)}(t, t_0)>< \psi^{(i)}(t, t_0)| \ \ ,$$

and taking the time derivative we get:

$$i\hbar \frac{\partial \boldsymbol{\rho}(t, t_0)}{\partial t} = \sum_i w_i \left\{ i\hbar \frac{\partial |\psi^{(i)}(t, t_0)>}{\partial t} < \psi^{(i)}(t, t_0)| \ + \ |\psi^{(i)}(t, t_0)> i\hbar \frac{\partial < \psi^{(i)}(t, t_0)|}{\partial t} \right\} .$$

Using (2.1), it can be cast to the form:

$$\begin{aligned} i\hbar\frac{\partial\boldsymbol{\rho}(t,t_0)}{\partial t} &= \sum_i w_i \left\{ \mathbf{H}|\psi^{(i)}(t,t_0)><\psi^{(i)}(t,t_0)| \ - |\psi^{(i)}(t,t_0)><\psi^{(i)}(t,t_0)| \ \mathbf{H} \right\} = \\ &= \mathbf{H}\boldsymbol{\rho} - \boldsymbol{\rho}\mathbf{H} \ = [\mathbf{H},\boldsymbol{\rho}] = -\left[\boldsymbol{\rho},\mathbf{H}\right] \ , \end{aligned} \tag{2.2}$$

where the symbol $[\mathbf{A},\mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA}$ is the commutator of $\mathbf{A}$ and $\mathbf{B}$. Equation (2.2) is the quantum analog of Liouville theorem. It was found by several authors, but we call it as the Liouville-von Neumann equation. It can be rewritten in the more suggestive form:

$$\frac{\partial\boldsymbol{\rho}(t,t_0)}{\partial t} + \frac{1}{i\hbar}\left[\boldsymbol{\rho},\mathbf{H}\right] = 0 \ , \tag{2.3}$$

which closely resembles the classical Liouville theorem, with the correspondences:

$$\begin{aligned} \boldsymbol{\rho}(t) &\to \rho(q,p,t) \ , \\ \frac{1}{i\hbar}\left[...,...\right] &\to \{...,...\} \ . \end{aligned} \tag{2.4}$$

In fact, Dirac [1] studying the classical limit of commutators, noted that they satisfy the same algebra as Poisson brackets. Indeed, one alternative way of stating the correspondence principle is written as:

$$\lim_{\hbar\to 0}\ \frac{1}{i\hbar}\left[\mathbf{A},\mathbf{B}\right] = \{A,B\} \ ,$$

where A and B are the classical functions associated with the quantum operators $\mathbf{A}$ and $\mathbf{B}$. The correspondence can be made more precise, if we pass to the Heisenberg picture, with the unitary transformation

$$\boldsymbol{\rho}_H(t) = \exp\left(\frac{i}{\hbar}\mathbf{H}t\right)\boldsymbol{\rho}(t)\exp\left(-\frac{i}{\hbar}\mathbf{H}t\right) \ , \tag{2.5}$$

with $t_0 = 0$, for simplicity. Note that $\boldsymbol{\rho}_H(t)$ has now two types of time dependences, one is the explicit dependence and the other is due to the dynamical variables, that now depend on time. Then, in the Heisenberg picture we get:

$$\frac{d\boldsymbol{\rho}_H(t)}{dt} = \frac{\partial \boldsymbol{\rho}_H(t)}{\partial t} + \frac{1}{i\hbar}\left[\boldsymbol{\rho}_H(t), \mathbf{H}\right] = 0 \ ,$$

i.e. the total time derivative of $\boldsymbol{\rho}$ in the Heisenberg picture vanishes due to relations (2.3) and (2.5). That means that $\boldsymbol{\rho}_H(t)$ is a constant, which is the exact quantum analog of the classical Liouville theorem of relation (1.30). Since averages of observables are traces, they are invariant under unitary transformations, and we have the useful identity:

$$\begin{aligned}
\left[\mathbf{A}\right](t) &= \mathrm{Tr}\left[\boldsymbol{\rho}(t)\mathbf{A}\right] = \mathrm{Tr}\left[\exp\left(-\frac{i}{\hbar}\mathbf{H}t\right)\boldsymbol{\rho}_H(t)\exp\left(\frac{i}{\hbar}\mathbf{H}t\right)\mathbf{A}\right] = \\
&= \mathrm{Tr}\left[\boldsymbol{\rho}_H(t)\exp\left(\frac{i}{\hbar}\mathbf{H}t\right)\mathbf{A}\exp\left(-\frac{i}{\hbar}\mathbf{H}t\right)\right] = \mathrm{Tr}\left[\boldsymbol{\rho}_H(0)\mathbf{A}_H(t)\right] \ ,
\end{aligned}$$

where, in Heisenberg picture, the time dependence is contained in the observable $\mathbf{A}_H(t)$, and since the state operator is constant, we can refer it to the initial condition $\boldsymbol{\rho}_H(0)$. Going back to the Shrödinger picture, we observed that the time evolution of $\boldsymbol{\rho}$ is unitary. In fact, if we write

$$|\psi^{(i)}(t,t_0)>= \exp\left[-\frac{1}{i\hbar}\mathbf{H}\left(t-t_0\right)\right]|\psi^{(i)}(t_0,t_0)>\equiv \mathbf{U}\left(t,t_0\right)|\psi^{(i)}(t_0,t_0)> \ ,$$

with $\mathbf{U}\left(t,t_0\right)$ being the time evolution operator, we get

$$\begin{aligned}\boldsymbol{\rho}(t,t_0) &= \sum_i w_i\ |\psi^{(i)}(t,t_0) >< \psi^{(i)}(t,t_0)| \ = \\ &= \sum_i w_i \mathbf{U}\left(t,t_0\right)\ |\psi^{(i)}(t_0,t_0) >< \psi^{(i)}(t,t_0)|\mathbf{U}^{\dagger}\left(t,t_0\right)\ = \\ &= \mathbf{U}\left(t,t_0\right)\boldsymbol{\rho}(t_0,t_0)\mathbf{U}^{\dagger}\left(t,t_0\right)\ . \qquad (2.6)\end{aligned}$$

In particular, we observed that the normalization property is preserved:

$$\begin{aligned}\mathrm{Tr}\left[\boldsymbol{\rho}(t,t_0)\right] &= \mathrm{Tr}\left[\mathbf{U}\left(t,t_0\right)\boldsymbol{\rho}(t_0,t_0)\mathbf{U}^{\dagger}\left(t,t_0\right)\right] = \mathrm{Tr}\left[\boldsymbol{\rho}(t_0,t_0)\mathbf{U}^{\dagger}\left(t,t_0\right)\mathbf{U}\left(t,t_0\right)\right] = \\ &= \mathrm{Tr}\left[\boldsymbol{\rho}(t_0,t_0)\right] = 1\ .\end{aligned}$$

We can also demonstrate the following theorem:

Theorem 8 *The time evolution of $\boldsymbol{\rho}$ does not change the type of mixture of the ensemble*

In fact, we easily calculate this property:

$$\begin{aligned}\mathrm{Tr}\left[\boldsymbol{\rho}^2(t,t_0)\right] &= \mathrm{Tr}\left[\mathbf{U}\left(t,t_0\right)\boldsymbol{\rho}^2(t_0,t_0)\mathbf{U}^{\dagger}\left(t,t_0\right)\right] = \mathrm{Tr}\left[\boldsymbol{\rho}^2(t_0,t_0)\mathbf{U}^{\dagger}\left(t,t_0\right)\mathbf{U}\left(t,t_0\right)\right] = \\ &= \mathrm{Tr}\left[\boldsymbol{\rho}^2(t_0,t_0)\right]\ ,\end{aligned}$$

showing that the type of mixture is also preserved by the dynamics. This result is a big obstacle to our program, for it means that quantum dynamics cannot describe a relaxation process: *unitary evolution is time reversible.* This is the quantum version of the same problem encountered by Boltzmann in his theory of gases. The equilibrium situation is characterized by the condition

$$\frac{\partial \boldsymbol{\rho}(t,t_0)}{\partial t} = \frac{1}{i\hbar}\left[\mathbf{H},\boldsymbol{\rho}\right] = 0 \ , \tag{2.7}$$

which can be fulfilled by specific ensembles, but we do not know in general how the system approaches equilibrium for an arbitrary initial condition. Condition (2.7) is satisfied if $\boldsymbol{\rho}$ is a function of the Hamiltonian operator,

$$\boldsymbol{\rho} = \boldsymbol{\rho}\left(\mathbf{H}\right) \ ,$$

which is the case of several equilibrium distributions, but those solutions are frozen in time (they represent *eternal equilibrium*).

Let us analyze equation (2.2) in more detail. We work in a given representation with the density matrix $\rho_{nm} \equiv < n|\boldsymbol{\rho}|m>$. Assume that initially, the density matrix is brought to the diagonal form:

$$\rho_{nm}\left(0\right) = P_n \ \delta_{mn} \ , \tag{2.8}$$

and we analyze locally the behavior of the Liouville-von Neumann equation:

$$\begin{aligned} i\hbar \left.\frac{\partial \rho_{nm}}{\partial t}\right|_{t=0} &= \ < n|\left[\mathbf{H},\boldsymbol{\rho}(\mathbf{0})\right]|m> = \sum_j \left\{H_{nj}\rho_{jm}(0) - \rho_{nj}(0)H_{jm}\right\} = \\ &= \ H_{nm}\left(P_m - P_n\right) \ . \end{aligned} \tag{2.9}$$

i) For diagonal terms, $m = n$, we get:

$$i\hbar \left.\frac{\partial \rho_{nn}}{\partial t}\right|_{t=0} = 0 \ ,$$

that is, diagonal terms do not vary in first order, which is what we want near equilibrium;

ii) unfortunately, non diagonal terms ($m \neq n$) yield in general

$$i\hbar \left.\frac{\partial \rho_{nm}}{\partial t}\right|_{t=0} \neq 0 \ ,$$

which means that they will vary in first order, rendering the matrix non-diagonal. An exceptional instance is the so called *equiprobable* case, with equal probabilities:

$$P_m = P_n = ... = P_j = ... = P \ , \tag{2.10}$$

for all j, which makes (2.9) to vanish for all matrix elements, and represents a stationary case.

The above discussion allows us to formulate the *central problem* in a slightly different way: given an arbitrary initial condition, which may be (2.8) or any other, relaxation to equilibrium should lead us to the equiprobable case, where the density matrix is proportional to the identity:

$$\boldsymbol{\rho}(t \to \infty) = P \begin{pmatrix} 1 & 0 & ... & 0 & 0 \\ 0 & ... & 0 & 0 & 0 \\ ... & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 1 \end{pmatrix} \ . \tag{2.11}$$

Normalization implies

$$P = \frac{1}{\mathcal{D}} \; ,$$

where $\mathcal{D}$ is the dimension of the Hilbert space. Such an ensemble is called *random* or *ensemble of maximal mixture*, for reasons that will be fully explained afterwards. We have also seen that the above process is not described by the Liouville-von Neumann equation. Two important tasks are left: one is quantify the degree of mixture; the second, following Boltzmann procedure, is to realize that actual macroscopic systems are constituted by a large number of particles, implying that statistical arguments are essential to describe relaxation to equilibrium. But, in first place we have to discuss the physical meaning of equilibrium itself. This is done in the next subsection.

A

2.2 Thermal equilibrium

The notion of ‘equilibrium’ is a subtle one, which depends on time scale. It means that when we speak of equilibrium, we must explicitly specify the corresponding observational time, since a physical system often displays different kinds of thermal equilibrium at different time scales, when different couplings with the environment are active. Using Feynman’s informal style [2], we say that in thermal equilibrium *“all the fast things have happened and all the slow things not”.* The meaning of ‘*fast*’ and ‘*slow*’ implies the existence of different scales of relaxation times. A typical example is presented in

[3], as a simple quotidian experiment of pouring boiling water into a cold cup. This is an example of a non-equilibrium state. Within some seconds, the system water plus cup will acquire a common temperature and the volume of water will not change significantly at this time scale. We may say that the system cup + water is in *equilibrium*. For longer times, the temperature will change and some of the liquid water will evaporate, changing the water volume in the cup. Within times of the order of an hour, the temperature will be equal to the room temperature and the water volume will become stable. Now we say that the system is in thermal equilibrium with the local surrounding. Within a few days, the water begins to saturate the air of the room, and again the system will not be in an equilibrium state. The whole process defines a hierarchy of time scales, which describes different types of equilibrium states. Usually, in models, we do not include the slow degrees of freedom, working with the notion of an ideal equilibrium state that remains as such at infinite times, meaning that the system is completely isolated from the environment. But in real systems this is not the case, since external couplings, however small, always exist.

The discussion of the equilibrium state is also related to the concept of *ergodicity*. The term *ergodicity* was coined by Boltzmann when working out the foundation of Statistical Mechanics, meaning that all points of phase space available to the system (satisfying the macroscopic constrains), are visited after a long period of time. As a consequence, all microstates of the ensemble are treated as equally probable, which is the fundamental postulate that leads to the *microcanonical ensemble.* But depending on the observation time, different regions of phase space are visited, and true ergodicity is effectively attained at $t \to \infty$. The ergodic hypothesis is not able to establish typical relaxation times for equilibrium states and most systems are known to be non-ergodic. In a typical phase transition, some symmetry is spontaneously broken and this is accompanied by a breaking of ergodicity, *i.e.*, the macroscopic system is stuck within a region which is part of the whole available phase space [3].

In this monograph, we will not pursue the approach of deriving Statistical Mechanics from the ergodic theorem. Instead, following Pauli [4], we will try to establish a Master Equation to describe the time evolution of a system approaching equilibrium [5]. In principle, this approach allows for the calculation of relaxation times and holds greater promise for a satisfactory derivation of Statistical Mechanics. More on this later.

A

2.3 Quantifying the degree of mixture

A pure ensemble is characterized by a single ket-state. This is the maximal information that one can have of a quantal system. Part of this information is lost when one introduces mixture of states. In the case shown in the example of 1.2 for spin 1/2, the pure ensemble displays maximal polarization. As long as one considers mixed ensembles, the polarization degrades and finally vanishes for the random ensemble (maximum mixture). We would like to measure this degree of randomness, or the loss of information caused by the mixture. This loss of information resembles instances described by the classical probability theory. Suppose that our mixed ensemble given by the density operator

$$\boldsymbol{\rho} = \sum_i w_i \; |\psi^{(i)} >< \psi^{(i)}|$$

contains a big number of similar systems, that we called $\mathbb{N}$ (for the ideal ensemble, $\mathbb{N} \to \infty$). The pure state $|\psi^{(i)} >$ is occupied by $N_i = w_i \mathbb{N}$ systems, since w_i is by definition the fraction of systems with wave function $|\psi^{(i)} >$. We would like to calculate the number of different ways that the $\mathbb{N}$ systems of the ensemble can be distributed among the pure states $(\psi^{(1)}, \psi^{(2)}, ...\psi^{(i)} ...)$, given that N_i systems occupy the state $\psi^{(i)}(i = 1, 2, ...)$, with the proviso

$$\sum_i N_i = \mathbb{N} \sum_i w_i = \mathbb{N} \ .$$

The bigger this number, the bigger the degree of randomness of the ensemble. This number correspond to a multinomial coefficient given by:

$$\binom{\mathbb{N}}{\{N_i\}} = \frac{\mathbb{N}!}{N_1! N_2! ... N_i! ...} \ ,$$

saying that from the total number of permutations, one has to discount internal permutations. Since the above is a big number, one takes the logarithm and uses the Stirling approximation $\ln N! \approx N \ln N$, for large N :

$$\ln \binom{\mathbb{N}}{\{N_i\}} \approx \mathbb{N} \ln \mathbb{N} - \sum_i N_i \ln N_i = -\mathbb{N} \sum_i w_i \ln w_i \ .$$

The quantity $\mathbb{N}^{-1}\ln\binom{\mathbb{N}}{\{N_i\}}$ is independent of $\mathbb{N}$, and can be used to measure the mixture:

Definition 9 *Information entropy, $\mathcal{E}$*

$$\mathcal{E} \equiv -\sum_i w_i \ln w_i \ . \tag{2.12}$$

1. In first place, we note that $\mathcal{E}$ vanishes for a pure ensemble.

2. Secondly, we may ask what kind of mixture yields maximum entropy. This is a constrained variational problem, since the numbers $\{w_i\}$ satisfy the condition (1.3),

   $$\sum_i w_i = 1 \ .$$

   So, we maximize the 'functional'

   $$\mathcal{E} + \lambda \sum_i w_i$$

   where λ is a Lagrange multiplier associated with (1.3). The variation process yields:

   $$0 = \delta\left(\mathcal{E} + \lambda \sum_i w_i\right) = \sum_i \delta w_i \left(\lambda - 1 - \ln w_i\right) \ ,$$

   and since variations δw_i are considered as independent, we obtain:

   $$\ln w_i = \lambda - 1 \ ,$$

independent of the subindex i, that is $w_i = w$ for all $i = 1, 2, ..., N_0$, where N_0 is the number of states of the mixture. Condition (1.3) now yields

$$\sum_i w_i = wN_0 = 1 \ ,$$

with the solution

$$w_1 = w_2 = ... = w_i = ... = \frac{1}{N_0} \ ,$$

which we have called before as the 'random case', corresponding to the maximum value of the information entropy:

$$\mathcal{E}_{\max} = \ln N_0 \ . \tag{2.13}$$

3. The value of $\mathcal{E}$ 'depends on the observer', since the decomposition (1.7) of $\boldsymbol{\rho}$ in pure states is not unique, and we may have different values of $\mathcal{E}$ associated with the same density operator.

A representation of $\boldsymbol{\rho}$ that is unique (except for permutations of the basis vectors), is given in terms of its eigenvalues and eigenvectors:

$$\boldsymbol{\rho}|\rho_i >= \rho_i|\rho_i > \ ,$$

with the kets $\{|\rho_i >\}$ forming and orthonormal basis. Note that some of the eigenvalues ρ_i may be zero. This mixture is written as:

$$\boldsymbol{\rho} = \sum_i \rho_i|\rho_i >< \rho_i| \ , \tag{2.14}$$

where the index i runs over the dimension os the linear space. Relation (2.14) suggests a new definition of the entropy:

Definition 10 *von Neumann entropy (or simply, entropy),* σ

$$\sigma \equiv -\sum_i \rho_i \ln \rho_i \ . \tag{2.15}$$

Note that σ is an intrinsic property of the density operator, and since $\boldsymbol{\rho}$ is in diagonal form, we may write:

$$\sigma \equiv -\sum_i \rho_i \ln \rho_i \ = -\mathrm{Tr}\,(\boldsymbol{\rho} \ln \boldsymbol{\rho}) = -[\ln \boldsymbol{\rho}] \ . \tag{2.16}$$

The von Neumann entropy shares some properties with the information entropy. A pure state has no entropy and the random mixture (2.11) has the maximum value of σ (same proof as above):

$$\sigma_{\max} = \ln \mathcal{D} \ ,$$

where $\mathcal{D}$ is the dimension of the Hilbert space. Since diagonal elements of $\boldsymbol{\rho}$ (and thus its eigenvalues) are interpreted as probabilities, the maximum value of the entropy correspond to the equiprobable case (2.10), where $\boldsymbol{\rho}$ is stationary (*equilibrium*).

The following theorem is stated here without proof [6]:

Theorem 11 *The von Neumann entropy is the smallest of all mixing entropies,*

$$\mathcal{E}\left(\{w_i\}\right) \geqslant \sigma(\boldsymbol{\rho}) \ .$$

It follows that (2.14) is the least random of all possible decomposition of the density operator.

EXP **Example 1**

Consider a random mixture of two states $|\varphi_1>$ and $|\varphi_2>$, which are normalized but are not orthogonal. Let the overlap between the two state be $<\varphi_1|\varphi_2>= m$. We calculate the information entropy $\mathcal{E}$ and the entropy σ. We get:

$$\mathcal{E} = -\frac{1}{2}\ln\frac{1}{2} \ - \ \frac{1}{2}\ln\frac{1}{2} = \ln 2 \ .$$

To calculate σ, we need to obtain the eigenvalues of $\boldsymbol{\rho}$. A simple calculation leads to

$$\rho_{1,2} = \frac{1 \pm |m|}{2} \ ,$$

so that σ is given by:

$$\sigma = -\frac{1+|m|}{2}\ln\left(\frac{1+|m|}{2}\right) - \frac{1-|m|}{2}\ln\left(\frac{1-|m|}{2}\right) \leq \ln 2 \ ,$$

which is displayed in Fig.1, below:

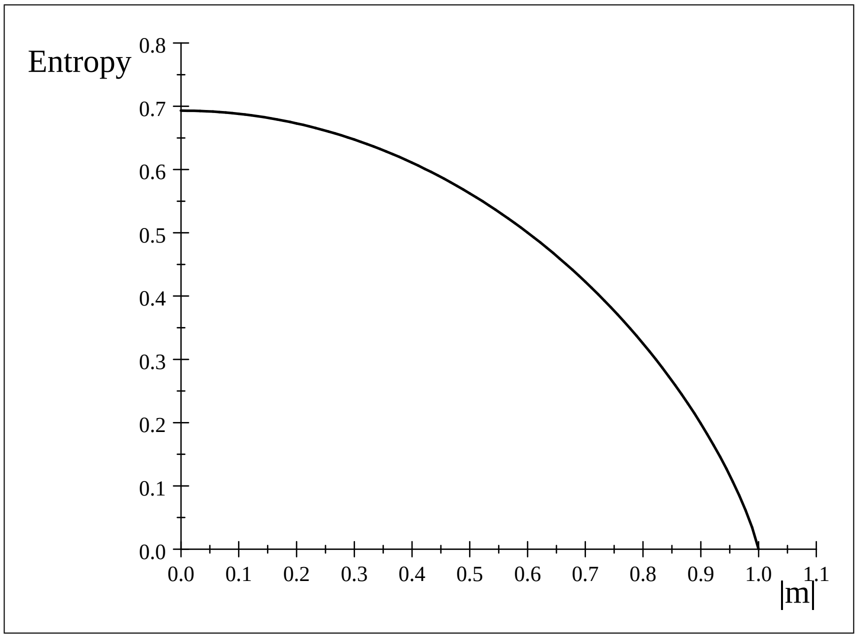

Fig.1. von Neumann entropy as a function of the overlap of quantum states.

Note the limit $\sigma \to \ln 2$, for $m \to 0$, and that $\sigma(|m|)$ is a decreasing function of $|m|$ in the interval $0 \leq |m| < 1$. ■

B

2.3.1 The extensive property of the von Neumann entropy

The entropy of a system composed of two uncorrelated parts is additive. This peculiarity is called ‘extensive property’. We proceed to its demonstration. Since the subsystems are uncorrelated, observables relative to different subsystems commute. If we assume that $\rho^{(1)}$ and $\rho^{(2)}$ are the Density Operators of the two subsystems, states of the whole system are obtained as a tensor product of both operators, as shown in

CHAPTER 2 — MANUSCRIPT

Section 1.3:

$$\rho = \rho^{(1)} \otimes \rho^{(2)} \ ,$$

and the dimension of ρ is the product of the dimensions of $\rho^{(1)}$ and $\rho^{(2)}$. Let us write as $\rho^{(1)}_{\alpha\beta}$ and $\rho^{(2)}_{\mu\nu}$ the matrix elements of the corresponding Density Matrices. Then, matrix elements of ρ are obtained with pairs of indexes in the form

$$\rho_{\alpha\mu,\beta\nu} = \rho^{(1)}_{\alpha\beta}\rho^{(2)}_{\mu\nu} \ ,$$

where (α,β) run over the dimension of $\rho^{(1)}$, and (μ,ν) over the dimension of $\rho^{(2)}$. To calculate the entropy, we need all the matrices given in diagonal form. Since the systems are uncorrelated, $\rho^{(1)}$ and $\rho^{(2)}$ can be diagonalized simultaneously, and the total ρ also results diagonal, with:

$$\rho_{\alpha\mu,\alpha\mu} = \rho^{(1)}_{\alpha\alpha}\rho^{(2)}_{\mu\mu} \ .$$

Then it follows

$$\begin{aligned} -\mathrm{Tr}\left(\boldsymbol{\rho}\ln\boldsymbol{\rho}\right) &= -\sum_{\alpha,\mu}\rho_{\alpha\mu,\alpha\mu}\ \ln\rho_{\alpha\mu,\alpha\mu} = -\sum_{\alpha,\mu}\rho^{(1)}_{\alpha\alpha}\rho^{(2)}_{\mu\mu}\ \left(\ln\rho^{(1)}_{\alpha\alpha} + \ln\rho^{(2)}_{\mu\mu}\right) = \\ &= -\sum_{\alpha}\rho^{(1)}_{\alpha\alpha}\ \ln\rho^{(1)}_{\alpha\alpha} - \sum_{\mu}\rho^{(2)}_{\mu\mu}\ \ln\rho^{(2)}_{\mu\mu} \ , \end{aligned}$$

where we have used the normalization condition $\mathrm{Tr}\boldsymbol{\rho}^{(1)} = \mathrm{Tr}\boldsymbol{\rho}^{(2)} = 1$. The above equality means that

$$\sigma = \sigma^{(1)} + \sigma^{(2)} \ ,$$

which is the additive property. ■