UNIVERSITÀ DEGLI STUDI DI PAVIA
DIPARTIMENTO DI FISICA
CORSO DI LAUREA IN FISICA

A COMMON FRAMEWORK FOR ENTROPY IN THERMODYNAMICS AND QUANTUM INFORMATION THEORY

Relazione per la laurea di:
Luca Teodori

Relatore:
Chiar.mo Prof. Paolo
Perinotti

Anno accademico 2016/2017

Contents

Chapter 1

Introduction

Thermodynamics and information theory seem very different theories: the former deals with macroscopic physical systems and relies on concepts like heat, pressure, temperature, work, etc. that are macroscopic observables. The latter deals with information on an abstract level and uses microscopic concepts such as the concept of bit.

However, they both use the notion of entropy. Thermodynamic entropy is a function of the macroscopic state of a system and characterizes the possibility or otherwise of certain processes (only the processes that don't diminish the total entropy of an isolated system are possible). Information entropy is related, as we will see, to the concept of compressibility of a source of information and quantifies the amount of its uncertainty. Information-related entropies are equivalent to the entropy of statistical mechanics, a matter we will see by exploiting the maximum entropy principle, and this suggests that even thermodynamic entropy and information entropy are, in a certain sense, two faces of the same coin. To prove that we will exploit the axiomatic framework of thermodynamics of Lieb and Yngvason [1]. This approach has the remarkable feature that introduces all of thermodynamics without the assumption of entropy with its main properties and avoids the concept of heat differently from Callen's approach.

It also avoids the differentiability of the thermodynamic functions differently from Caratheodory's approach. The axioms of their approach to thermodynamics identify basic properties of a thermodynamic system, but the remarkable result we wish to exploit is that these axioms are also fulfilled by resource theories [2].

A resource theory deals with the "resources" needed to accomplish certain tasks or needed to change the state of a certain system, if the processes that we can use are only of a predefined class, known as allowed operations. An important resource theory is the entanglement resource theory, where the states are the entangled quantum systems and the allowed operations are the so called LOCC, Local Operations and Classical Communication. If the axioms of the Lieb and Yngvason framework are satisfied, then we can demonstrate the existence and uniqueness (up to unimportant affine scale changes) of a function that has the same properties of entropy. The fact that resource the-

ories satisfy these axioms has the consequence that even in these theories we can define entropy and we will show that these entropies from the Lieb and Yngvason framework are exactly the information-theoretic entropies of information theory.

All these facts suggest a strong connection between thermodynamics and information theory that is not only of academic interest but, as we will see in this chapter, has a great importance and we will clarify that by reviewing a classical gedankenexperiment, the Maxwell demon ([4], [7]) and his resolution via the Landauer principle, a very important link between these two theories and of practical interest in a world where computers are going closer and closer to the limit this principle imposes.

Then in chapter 2 we examine some basic properties about thermodynamics, statistical mechanics and information theory, from the latter in particular the fundamental properties of the entropy related concepts ([10]).

In chapter 3 we exploit Shannon's entropy properties to establish a connection between statistical mechanics and information theory by using the maximum entropy principle ([3]).

In chapter 4 we introduce the Lieb and Yngvason approach to thermodynamics ([1]) and finally we extend this approach to information theory by exploiting resource theory ([9], [6], [8]) in chapter 5.

1.1   Maxwell demon

Maxwell demon was a gedankenexperiment first proposed by James Clerk Maxwell. In Maxwell's mind the entropy principle is valid only in a macroscopic sense and it can be violated at a fundamental level.

To assert his point of view he imagined a box divided in two parts by a partition with a gas in it, in thermal equilibrium with the environment. There is a little door on the partition and there is a demon capable of detecting the instant velocity of the gas particles. This demon opens the door on the partition in the exact moment he sees a particle with a certain velocity and let it pass in the other side and immediately close the door. Let's say we want the mean value of the velocity of the gas' particles to be greater in the left side of the box, then the demon lets pass all the particles moving from right to left that have velocity greater than the aforementioned, $\langle v \rangle$, and lets pass all the particles moving from left to right that have velocity less than $\langle v \rangle$. In this way the demon causes a difference in temperature and an exchange of heat from the colder part to the hotter part ideally without exerting work on the gas, thus violating Clausius form of the second law that states the impossibility of such a process.

One can devise another type of Maxwell demon, that lets pass all the particles moving in one direction while stopping the particles moving in the opposite direction. In this way one can produce a difference in pressure, and the net effect is to totally transform the heat transferred from the thermal reservoir into work needed to create the pressure difference. This is a clear violation of

Kelvin's form of the second law that states the impossibility to devise a cycle whose sole net effect is the extraction of heat from a reservoir and the performance of an equal amount of work.

The solution to the Maxwell demon almost surprisingly comes from the concept of information, in particular from the Landauer's principle, a matter we shall discuss in the next section.

1.2   Landauer's principle

A possible solution to Maxwell demon was proposed by Szilard, it involves the entropy increase due to the measurement process ([7]). As a matter of fact the measurement process is common to all possible Maxwell demon experiments, so its solution was believed for more than a decade to reside there. But to witness the real death of Maxwell demon paradox we must wait for the work of Landauer. In his work Landauer reconsidered the role of memory, in particular the process of erasure of information, that is the solution to this paradox.

Since information processing must be carried out by physical systems, there should be a one-to-one correspondence between logical and physical states. Logical states are described by an abstract set of variables on which some information processing can be performed. Then, a reversible logical process, i.e. an one-to-one mapping for logical states, corresponds to a reversible physical process. A logically irreversible process is instead noninjective. Memory erasure is a logically irreversible process because many possible states of memory should be set to a single fixed state after an erasing procedure. We cannot determine the state prior to erasure without further information. Landauer understood that this logical irreversibility must involve dissipation in the environment, because for example memory erasure decreases the degrees of freedom of a certain system, thus its entropy decreases, and in order not to violate the second law the system must exchange a certain amount of heat with the environment. This is the core of the Landauer's principle, that states the following:

Landauer's principle: the minimum amount of work W needed to erase the state of a bit in a thermal reservoir at temperature T is

$$W = K_B T \ln 2 \tag{1.1}$$

where K_B is the Boltzmann constant.

The validity of this principle resolves the Maxwell demon paradox since the demon must store the information he gets from the measurement in a certain memory, and the demon cyclic process must contain an erasure process since we are dealing with a cyclic process, and at the end of it the demon along with his memory must return to the initial state (the necessity of an erasure process can be seen also as a consequence of the fact that there can't exist a memory

device with infinite store capability). The paradox has then a solution, since every possible gain the demon gets is lost in the erasure of the information he has previously stored.

The derivations of Landauer's principle all involve an idealization of a memory source often described by a double-well potential. Indeed with this or similar ideas one can derive Landauer's principle in many frameworks ([4])

In the following section we show a derivation of Landauer's principle in the quantum case from microscopic considerations, obviously without direct reference to the second law. We will follow [4].

1.3 Landauer in quantum case

To derive Landauer's principle in quantum case we need a way to represent the memory of the demon and an heat reservoir that represents the environment. We also need an external parameter that allows us to change the state of the memory (in order to erase its content). We will treat this system in an abstract way using density matrices and such but in order to have a clearer picture one could think that the memory consists in a spin-$\frac{1}{2}$ particle and the external parameter being a magnetic field that can flip the particle spin. The erasure process consists then in changing the state of the demon memory to a standard state. If we think this particle to be represented by a bit, we may consider the erasure process as changing the bit to a standard state, let's say the state corresponding to 1.

We now describe the derivation of Landauer's principle, in order to represent the demon's memory we use a 2×2 density matrix that always admits a decomposition of the form

$$\rho_D = p_a|a\rangle\langle a| + p_b|b\rangle\langle b| \tag{1.2}$$

in a certain basis (not necessarily the energy eigenstate basis). Thus ρ_D is diagonal with p_a the probability that ρ_D is in the state $|a\rangle$ and p_b the probability that it is in the state $|b\rangle$.

The heat reservoir is initially in thermal equilibrium with ρ_D so we can write its density matrix ρ_H as

$$\rho_H = \frac{e^{-\beta\mathcal{H}}}{tr(e^{-\beta\mathcal{H}})} \tag{1.3}$$

where $\mathcal{H}$ is the hamiltonian of the heat reservoir. The probability of finding the heat reservoir in $|E_n\rangle$ is then

$$P(E_n) = \frac{e^{-\beta E_n}}{\sum_m e^{-\beta E_m}}. \tag{1.4}$$

We finally indicate the external parameter as $\lambda(t)$.

The erasure process consists in the following steps:

1. at $t = 0$ the memory state begins in the state described by

$$\rho_D(t = 0) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \tag{1.5}$$

   in the energy eigenstate basis; we will refer to the first state as the state representing 0 and to the second state as the one representing 1 and we assume that they are degenerate.

2. at $t = 0^+$ we couple the bit to the reservoir, we then change λ in order to produce the erasure process, and at a certain $t = \tau$ we want $\lambda(0) = \lambda(\tau)$ (the external parameter returns to the initial state).

3. we finally decouple the bit from the reservoir.

As a consequence of this result the heat reservoir begins in a certain state, let us say $|n\rangle$ and ends in a final state denoted by $|m\rangle$. We define

$$Q = E_n - E_m \tag{1.6}$$

as the heat lost by the heat reservoir (we assumed that also $|m\rangle$ is diagonal in the energy eigenbasis, so the probability distribution of E_m is analogous to the one in (1.4)). We denote the probability distribution of the initial state of the bit with $P_i(k)$ with $k = 0, 1$ and the probability distribution of the final state of the bit with $P_f(j)$ with $j = 0, 1$. $P_f(1)$ will represent for example the probability that the final state of the bit is 1. Then following our scheme we have

$$P_i(k) = \frac{1}{2} \text{ for } k = 0, 1 \tag{1.7}$$

and

$$P_f(j) = \begin{cases} 1 & \text{for } j = 1 \text{ (the standard state)} \\ 0 & \text{for } j = 0 \end{cases} \tag{1.8}$$

Let us define the following observable

$$\Gamma = \ln P_i - \ln P_f - \beta(E_n - E_m) \tag{1.9}$$

then we can write the average value of $\exp(-\Gamma)$

$$\langle e^{-\Gamma} \rangle = \sum_{n,m,k,j} P(E_n) P_i(k) e^{-\ln P_i(k) + \ln P_f(j) + \beta(E_n - E_m)}$$

$$= \sum_{n,m,k,j} P_i(k) \frac{e^{-\beta E_n}}{\sum_{n'} e^{-\beta E_{n'}}} |U_{n,m,k,j}|^2 \frac{P_f(j)}{P_i(k)} e^{\beta(E_n - E_m)} \tag{1.10}$$

$$= \frac{1}{\sum_{n'} e^{-\beta E_{n'}}} \sum_{j,m,n,k} P_f(j) e^{-\beta E_m} |U_{n,m,k,j}|^2$$

where $U_{n,m,k,j} = \langle m, j | U(\tau) | n, i \rangle$ with U the time evolution operator. We see then, exploiting this last equation and the fact that U is a unitary matrix, that the following equality holds

$$\langle e^{-\Gamma} \rangle = 1. \tag{1.11}$$

Now we want to prove that

$$e^{-\langle \Gamma \rangle_\rho} \leq \langle e^{-\Gamma} \rangle_\rho \tag{1.12}$$

for a generic state ρ. We will exploit Jensen's inequality: for $f : [a, b] \rightarrow \mathbb{R}$ convex, given $x = \sum_i p_i x_i$, with $\{p_i\}$ a probability distribution ($\sum_i p_i = 1$), we have

$$f(x) \leq \sum_i p_i f(x_i) \tag{1.13}$$

so, with the orthonormal decompositions $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, $\Gamma = \sum_i q_i |\varphi_i\rangle\langle\varphi_i|$ and exploiting the convexity of the exponential function,

$$
\begin{aligned}
e^{-\langle \Gamma \rangle_\rho} = e^{-\sum_i p_i \langle\psi_i|\Gamma|\psi_i\rangle} &\leq \sum_i e^{-\langle\psi_i|\Gamma|\psi_i\rangle} \\
= \sum_i p_i e^{-\sum_j |\langle\psi_i|\varphi_j\rangle|^2 q_j} &\leq \sum_i p_i \sum_j q_j |\langle\psi_i|\varphi_j\rangle|^2 e^{-q_j} \\
= \sum_i p_i \langle\psi_i|e^{-\Gamma}|\psi_i\rangle &= \langle e^{-\Gamma} \rangle_\rho.
\end{aligned}
\tag{1.14}
$$

We have then $-\langle \Gamma \rangle \leq 0$, i.e.

$$-\langle \ln P_i \rangle + \langle \ln P_f \rangle + \beta \langle Q \rangle \leq 0 \tag{1.15}$$

now exploiting (1.7) and (1.8) we have $\langle \ln P_i \rangle = -\ln 2$ and $\langle \ln P_f \rangle = 0$ so we have

$$K_B T \ln 2 \leq -\langle Q \rangle \tag{1.16}$$

We define the work W as

$$W = \Delta E_H + \Delta E_D \tag{1.17}$$

Now since the two states representing the bit are degenerate we have $\Delta E_D = 0$ and using (1.6) we have $\langle Q \rangle = -\langle W \rangle$ so from (1.16)

$$K_B T \ln 2 \leq \langle W \rangle. \tag{1.18}$$

We can finally conclude that the work we have to spend for the erasure process is at least $K_B T \ln 2$ so the Landauer's principle has been proved in this framework.

Chapter 2

Preliminaries

In this chapter we shall review some basic topics about classical thermodynamics, statistical mechanics and information theory (both the quantum and classical version)

2.1 Classical thermodynamics

In classical thermodynamics the entropy is a state function, defined for equilibrium states as

$$dS = \frac{\delta Q}{T} \tag{2.1}$$

where we made clear that unlike entropy the differential of heat is not an exact differential. The second law states that for every process, $dS \geq 0$, and this places a strong constraint on the possible transformations between equilibrium states. The existence of such a function is what characterizes thermodynamic processes, so every formulation of thermodynamics must contain it.

Callen's approach to thermodynamics has the advantage of clarity and it rapidly brings one to the application of thermodynamics, but it doesn't clarify the minimal assumptions that are needed in order to ensure the existence of entropy.

Other formulations of the second law can be paraphrased as follows:

1. Clausius: there cannot be a process whose sole net effect is the transmission of heat from a colder body to an hotter one;

2. Kelvin-Planck: there cannot be a process whose sole net effect is that a body is cooled and work is done;

3. Caratheodory: in any neighborhood of any state there are states that cannot be reached from it by an adiabatic process.

However, all these statements have some problems. One problem is to give a precise meaning to concepts like heat, hot and cold, neighborhood etc. Another problem involves the minimal assumptions one has to make in order to achieve the derivation of entropy. For example we know that phase transitions

are often characterized by discontinuities, but almost every presentations of classical thermodynamics is based on the differential calculus, which presupposed continuous derivatives, especially the Caratheodory one.

Lieb and Yngvason approach, that we review in chapter 4, makes it clear what are the minimal requests needed to grant the existence of a function with the properties of entropy and without resorting to unrealistic devices like Carnot machines and such.

2.2   Statistical mechanics

The declared goal of statistical mechanics is to derive classical thermodynamics from the behavior of the microscopic constituents of a macroscopic system. Here we consider a system in a macrostate characterized by fixed macroscopic quantities. Every macrostate corresponds to several microstates, i.e. configurations of the constituent particles that give the same macrostate. So we consider an ensemble of ideal copies of the system, all in the same macrostate and all in a different microstate. For statistical mechanics, the entropy function according to the famous Boltzmann's formula is

$$S = K_B \ln W \tag{2.2}$$

where K_B is the Boltzmann's constant and W represents the number of all the possible microstates of a system, in general W will be a function of quantities like volume, energy, number of particles etc. The purpose of statistical mechanics is then to find expressions for the quantity W under various conditions and then identifying the thermodynamic quantities with respect to W by analogy of their behavior such as common behavior of the statistical entropy function (2.2) and the thermodynamic one (2.1) in specific models like an ideal gas. However, this link between thermodynamics and statistical mechanics by mere analogy is not a satisfactory one; the link between thermodynamics and information theory using the Lieb and Yngvason approach and then the link of information theory and statistical mechanics using the maximum entropy principle is a much more satisfactory solution.

In statistical mechanics we can consider three different types of ensembles. An ensemble is a set of all the ideal copies of a system with fixed values for the macroscopic quantities; different choices of macroscopic fixed quantities correspond to different types of ensemble.

In the microcanonical ensemble we have that the energy E, the volume V and the number of particles N are fixed. In this ensemble one can find an expression for the entropy $S = S(E, V, N)$ with (2.2) and then derive all the other thermodynamic quantities with classical thermodynamic expressions, for example with

$$dS = \frac{1}{T}dE + \frac{P}{T}dV. \tag{2.3}$$

In the canonical ensemble the fixed quantities are V, N and the temperature T. In this ensemble one can find the expression for the partition function $\mathcal{Z}$ that is given by

$$\mathcal{Z} = \sum_E g(E) e^{-\frac{E}{K_B T}} \tag{2.4}$$

where $g(E)$ is the density of states at energy E. Then one can find an expression for the free Helmoltz energy A with comparison with the relation (as one can easily derive from (2.4))

$$U = N K_B T^2 \frac{\partial}{\partial T} \ln \mathcal{Z} \tag{2.5}$$

and

$$U = \frac{\partial(\frac{A}{T})}{\partial(\frac{1}{T})} \tag{2.6}$$

derived from classical thermodynamics. The expression for A is then

$$A = -N K_B T \ln \mathcal{Z}. \tag{2.7}$$

In analogy with the microcanonical ensemble, one can derive the other thermodynamic quantities by using classical thermodynamic relations like the following

$$dA = -PdV - TdS. \tag{2.8}$$

Finally in the grand-canonical ensemble the fixed quantities are V, T and the chemical potential μ. In this ensemble one can find an expression for the grand-canonical partition function $\mathcal{Q}$, given by

$$\mathcal{Q} = \sum_{n,l} e^{\frac{\mu N_n}{K_B T} - \frac{E_l}{K_B T}}. \tag{2.9}$$

Then one can identify $\mathcal{Q}$ as (see any text on statistical mechanics)

$$P = \frac{K_B T}{V} \ln \mathcal{Q} \tag{2.10}$$

and one can easily derive from (2.9) the following two relations

$$\langle N \rangle = K_B T \frac{\partial}{\partial \mu} \ln \mathcal{Q}, \tag{2.11}$$

$$U = K_B T^2 \frac{\partial}{\partial T} \ln \mathcal{Q}. \tag{2.12}$$

A remarkable result in ensemble theory is that in the thermodynamic limit (i.e N and V tend to infinity) the three ensembles are equivalent.

To resume we can say that statistical mechanics finds specific expressions for quantities such W or $\mathcal{Z}$ related to the microscopic model under analysis and then it identifies them with respect to thermodynamic quantities by analogy with classical thermodynamic expressions, connection, as previously stated, that is not completely satisfactory.

2.3 Information theory

We shall resume here some important results in information theory regarding Shannon entropy and his quantum version, the Von Neumann entropy. In the following all logarithms are intended to be in base two unless otherwise specified ($\log_2 x \equiv \log x$)

Definition 2.1 (Shannon entropy). *Let X be a random variable with possible outcomes*
$x_1, x_2, ..., x_n$ respectively with associated probabilities $p_1, p_2, ..., p_n$; the Shannon entropy of the random variable X is defined as (we assume $0 \log 0 \equiv 0$):

$$H(X) = -\sum_x p_x \log p_x. \qquad (2.13)$$

We now resume some basic properties and definitions about Shannon entropy, in the following $p(x)$ and $q(x)$ are two generic probability distributions.

1. Relative entropy $H(p(x)\|q(x)) = \sum p(x) \log \frac{p(x)}{q(x)}$; it can be viewed as a measure of the distance of the two distributions (the following result holds: $H(p(x)\|q(x)) \geq 0$ with the equality if and only if $p(x) = q(x)$);

2. If X is a variable with d possible values, then $H(X) \leq \log d$, the equivalence holds for the uniform distribution (we can view this as a synonymous of maximum disorder for a uniform distribution);

3. Conditional entropy $H(X|Y) = H(X,Y) - H(Y)$, a measure of how uncertain we are, on average, about the value of X, given that we know the value of Y;

4. Strong subadditivity $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$;

5. Mutual information $H(X:Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y)$, a measure of the amount of information X and Y have in common;

Definition 2.2 (Von Neumann entropy). *Let ρ be a quantum state, we define the Von Neumann entropy of ρ to be*

$$S(\rho) = -tr(\rho \log \rho) = -\sum_x \lambda_x \log \lambda_x \qquad (2.14)$$

where λ_x are the eigenvalues of ρ (we assume $0 \log 0 \equiv 0$).

We now resume some definition and properties about the Von Neumann entropy that have many similarities with Shannon entropy.

1. Relative entropy $S(\rho\|\sigma) = -S(\rho) - tr(\rho \log \sigma)$; Klein's inequality holds, i.e. $S(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$;

2. The entropy is non negative and is zero if and only if the state is pure (if the state is pure it has one eigenvalue equal to one and all the others equal to zero, so the conclusion is straightforward using (2.2));

3. In a d dimensional Hilbert space the entropy is at most $\log d$, the equality holds if and only if the system is in the completely mixed state $\frac{\mathcal{I}}{d}$ (again we can view this as a synonymous of maximum disorder of the completely mixed state, as its name suggests);

4. $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$;

5. Concavity of entropy $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$;

6. A pure state of a composite system $|AB\rangle$ is entangled if and only if $S(A, B) < S(B)$;

7. Strong subadditivity $S(A, B, C) + S(B) \leq S(A, B) + S(B, C)$.

We want now to demonstrate two important results of information theory: Shannon's noiseless channel coding theorem and Schumacher's noiseless channel coding theorem, the quantum version of the former. They are about the compressibility of a source of information. In classical information theory a source consists of a sequence of random variables $X_1, X_2, ...$, whose values represent the output of the source. We assume these variables are independent and identically distributed (i.i.d.). The idea behind the theorem is that we can use the so called typical sequences, i.e. outputs of the source that are most likely to occur, in order to find a scheme for data compression.

To understand this point, suppose our source produces the sequence of bits $x_1, ..., x_n$ relatively to the random variables $X_1, ..., X_n$; each bit is equal to zero with probability p and equal to one with probability $1 - p$; then as n becomes large we expect that a fraction p of sequences' value $x_1, ..., x_n$ will be equal to zero and a fraction $1 - p$ will be equal to one. The sequences that follow this assumption are called typical sequences. We have

$$p(x_1, ..., x_n) = p(x_1)...p(x_n) \approx p^{np}(1 - p)^{(1-p)n}. \qquad (2.15)$$

We used the assumption of independence and that we have a typical sequence. So by taking the logarithms

$$-\log p(x_1, ..., x_n) \approx -np \log p - n(1 - p) \log(1 - p) = nH(X) \qquad (2.16)$$

where we have recognized the term involving the entropy of the source $H(X) = -p \log p - (1-p) \log(1-p)$ (remember, the variables are identically distributed). So $p(x_1, ..., x_n) \approx 2^{-nH(X)}$ so there are at most $2^{nH(X)}$ sequences (the sum of all probabilities cannot be greater than one, and $2^{-nH(X)}2^{nH(X)} = 1$). We are ready to give the following definition.

Definition 2.3. *A sequence is called ϵ-typical if, given $\epsilon > 0$,*

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, ..., x_n) \leq 2^{-n(H(X)-\epsilon)}. \qquad (2.17)$$

We denote the set of all ϵ-typical sequences of length n by $T(n, \epsilon)$ and the number of all ϵ-typical sequences with $|T(n, \epsilon)|$.

Theorem 2.4 (typical sequence theorem).

1. *Let $\epsilon > 0$, $\forall \delta > 0$ there exists n such that the probability that a sequence is ϵ-typical is at least $1 - \delta$*

2. *$\forall \epsilon > 0$, $\forall \delta > 0$ there exists n such that the number of ϵ-typical sequences $|T(n, \epsilon)|$ is*

$$(1 - \delta)2^{n(H(X)-\epsilon)} \leq |T(n, \epsilon)| \leq 2^{n(H(X)+\epsilon)} \tag{2.18}$$

3. *Let $S(n)$ be a collection of sequences with at most 2^{nR} sequences, where n is the dimension of a sequence and $R < H(X)$ (the n X_i variables are independent and identically distributed as usual). Then $\forall \delta > 0$ there exists n such that (we indicate x as the "vector" sequence)*

$$\sum_{x \in S(n)} p(x) \leq \delta \tag{2.19}$$

Proof.   1. We can view the thesis as the following (by taking the logarithms of (2.17)): there exists n such that

$$p(|\sum_{i=1}^{n} \frac{-\log p(X_i)}{n} - H(X)| \leq \epsilon) \geq 1 - \delta \tag{2.20}$$

we can apply the law of large numbers, i.e., given the estimator $E_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ of the average $\langle Y \rangle$, we have that, $\forall \epsilon > 0$,

$$\lim_{n \to \infty} p(|E_n - \langle Y \rangle| \geq \epsilon) = 0 \tag{2.21}$$

so we have the thesis with the substitutions $Y_i = -\log p(X_i)$ (in this case $\langle X \rangle = -\sum_{i=1}^{n} p(X_i) \log p(X_i) = H(X)$)

2. By part 1 and by (2.17) we have that the probabilities of typical sequences must lie in the range $1 - \delta$ to 1, so

$$1 \geq \sum_{x \in T(n,\epsilon)} p(x) \geq \sum_{x \in T(n,\epsilon)} 2^{-n(H(X)+\epsilon)} = |T(n, \epsilon)|2^{-n(H(X)+\epsilon)}$$
$$\implies |T(n, \epsilon)| \leq 2^{n(H(X)+\epsilon)} \tag{2.22}$$

and

$$1 - \delta \leq \sum_{x \in T(n,\epsilon)} p(x) \leq \sum_{x \in T(n,\epsilon)} 2^{-n(H(X)-\epsilon)} = |T(n, \epsilon)|2^{-n(H(X)-\epsilon)}$$
$$\implies |T(n, \epsilon)| \geq 2^{n(H(X)-\epsilon)}(1 - \delta) \tag{2.23}$$

(we used $\sum_{x \in T(n,\epsilon)} 1 = |T(n, \epsilon)|$.)

3. Choose ϵ so that $R < H(X) - \delta$ and $0 < \epsilon < \delta/2$, the idea is to split $S(n)$ into the ϵ-typical sequences and the non-typical sequences; the total probability of the non-typical sequences thanks to part 1 can be made less than $\delta/2$; but there are at most 2^{nR} typical sequences each with probability at most $2^{-n(H(X)-\epsilon)}$, so the probability of typical sequences is at most $2^{-n(H(X)-\epsilon-R)} \to 0$, so for a sufficient large n, $\sum_{x \in S(n)} p(x) \le \delta$.

$\square$

We are ready to state Shannon's noiseless channel coding theorem, we only need to specify the idea of a compression scheme of rate R: it is a scheme that allows us to store the output of a source with at most 2^{nR} possible outcomes.

Theorem 2.5 (Shannon's noiseless channel coding theorem). *Suppose $\{X_i\}$ is an i.i.d. (independent and identically distributed) source with entropy $H(X)$; then there exists a reliable compression scheme of rate R if $R > H(X)$; otherwise if $R < H(X)$ a reliable compression rate doesn't exist.*

Shannon's entropy, due to theorem 2.5, has a very interesting interpretation. Since a source can be compressed at a rate R if and only if $H(x) < R$, that means that $H(X)$ represents the minimal amount of resources needed to memorize the output from the source X. So $H(X)$ quantifies in a non ambiguous way the amount of uncertainty about the random variable X (the more we know about a source, the less are the physical resources needed to memorize the output). Among the consequences of this remarkable result there is the possibility to establish a connection between information theory and statistical mechanics, as we will see in chapter 3.

The quantum version of Shannon's theorem is similar to the classical one, and the ideas behind the proofs are the same (see [10]), here we only write the fundamental concepts and theorems.

In quantum information theory the state itself can be viewed as a source. More formally, an i.i.d. quantum source will be described by a finite dimensional Hilbert space H and a density operator ρ on that Hilbert space. A compression scheme of rate R is a quantum operation that takes state in $H^{\otimes n}$ to state in a compressed space of dimension 2^{nR}, whereas the decompression quantum operation does the inverse process.

Definition 2.6. *Suppose the density operator ρ associated with a quantum source has orthonormal decomposition*

$$\rho = \sum_x p(x)|x\rangle\langle x| \tag{2.24}$$

with $p(x)$ the eigenvalues of ρ. We have $S(\rho) = H(p(x))$ so in analogy with the classical case we define ϵ-typical sequences $x_1, ..., x_n$ as the sequences that satisfy

$$2^{-n(S(\rho)+\epsilon)} \le p(x_1, ..., x_n) \le 2^{-n(S(\rho)-\epsilon)}. \tag{2.25}$$

We can denote the subspace $|x_1\rangle...|x_n\rangle$ of ϵ-typical states of n dimension as $T(n, \epsilon)$. The projector onto this subspace will be

$$P(n, \epsilon) = \sum_{x \in T(n,\epsilon)} |x_1\rangle\langle x_1| \otimes ... \otimes |x_n\rangle\langle x_n|. \tag{2.26}$$

Theorem 2.7 (Typical subspace theorem)**.**

1. *Let $\epsilon > 0$, $\forall \delta > 0$ there exists n such that*

$$tr(P(n, \epsilon)\rho) \geq 1 - \delta \tag{2.27}$$

2. *$\forall \epsilon$, $\forall \delta$ there exists n such that*

$$(1 - \delta)2^{n(S(\rho)-\epsilon)} \leq tr(P(n, \epsilon)) \leq 2^{n(S(\rho)+\epsilon)} \tag{2.28}$$

3. *Let $S'(n)$ be a projector onto any subspace of H of dimension at most 2^{nR} sequences, where n is the dimension of the Hilbert space and $R < S(X)$. Then $\forall \delta > 0$ there exists n such that*

$$tr(S'(n)\rho) \leq \delta \tag{2.29}$$

Theorem 2.8 (Shumacher's noiseless channel coding theorem)**.** *Let $\{H, \rho\}$ be a i.i.d. quantum source and $R \in \mathbb{N}$; if $R > S(\rho)$ then there exist a reliable compression scheme of rate R, otherwise if $R < S(\rho)$ there is no reliable compression scheme of rate R.*

Chapter 3

Maximum entropy principle

In this chapter we shall see how statistical mechanics and information theory can be unified using the maximum entropy principle [3]. Suppose we know the average value $\langle f(x) \rangle = \sum_{i=1}^{n} p_i f(x_i)$, the spectrum of f but not the probability distribution p_i; then how we can evaluate the average value of another quantity $g(x)$? We could use Laplace's principle of insufficient reason that states that we must assume equal probabilities if there is no reason to think otherwise. However, this criteria is not always the most efficient. Moreover it is presented with mathematical problems e.g. for distributions over countably infinite sets. The maximum entropy principle states that in assigning a probability distribution we have to maximize the Shannon entropy of the distribution, obviously with respect to the constraints given by our prior knowledge about the system. We use the fact that Shannon entropy is a well defined function to quantify the amount of uncertainty about a physical system, as we have already stated in section 2.3 due to theorem 2.5, so the probability distribution we choose by following the maximum entropy principle is the one that maximize the uncertainty about the system given the information we have, so we do not introduce any bias in our inference from partial information.

By retaking the previous example, to find the probabilities p_i we have to maximize the quantity

$$H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \ln p_i \tag{3.1}$$

with the constraints

$$\sum_{i=1}^{n} p_i = 1, \tag{3.2}$$

$$\langle f(x) \rangle = \sum_{i=1}^{n} p_i f(x_i). \tag{3.3}$$

We use the Lagrange multipliers theorem, we find the multipliers λ (constraint (3.2)) and μ (constraint (3.3)) and we obtain

$$p_i = e^{-\lambda - \mu f(x_i)} \tag{3.4}$$

We explicitly note that we can write the average value of f as

$$\langle f(x) \rangle = -\frac{\partial}{\partial \mu} \ln \mathcal{Z}, \tag{3.5}$$

with

$$\mathcal{Z}(\mu) = \sum_{i=1}^{n} e^{-\mu f(x_i)}. \tag{3.6}$$

We can generalize this procedure to a generic number of constraints (let's say we know now the average value of m functions f_j, $j = 1, ..., m$), we can write a function totally analogous to the partition function of statistical mechanics:

$$\mathcal{Z}(\lambda_1, ..., \lambda_m) = \sum_{i=1}^{n} e^{-(\lambda_0 + \sum_{j=1}^{m} \lambda_j f_j(x_i))} \tag{3.7}$$

(now the Lagrange multiplier for the constraint (3.2) is λ_0) and we can write the average value of a function $f_j(x)$ as

$$\langle f_j(x) \rangle = -\frac{\partial}{\partial \lambda_j} \ln \mathcal{Z} \tag{3.8}$$

and

$$\lambda_0 = \ln \mathcal{Z} \tag{3.9}$$

whereas for the probabilities

$$p_i = \exp(-\lambda_0 - \sum_j \lambda_j f_j(x_i)). \tag{3.10}$$

The entropy becomes

$$S = -\sum_{i=1}^{n} p_i \ln p_i = \lambda_0 \overbrace{\sum_{i=1}^{n} e^{-(\lambda_0 + \sum_{j=1}^{m} \lambda_j f_j(x_i))}}^{1} +$$
$$+ \sum_{ij} \lambda_j \overbrace{e^{-(\lambda_0 + \lambda_j f_j(x_i))}}^{p_i} f_j(x_i) = \lambda_0 + \sum_j \lambda_j \langle f_j(x) \rangle \tag{3.11}$$

and the variance of the distribution $f_j(x)$ is (remember the definition of $\mathcal{Z}$ in (3.7))

$$\langle f_j^2(x) \rangle - \langle f_j(x) \rangle^2 = \frac{\partial^2}{\partial \lambda_j^2} \ln \mathcal{Z} \tag{3.12}$$

So we can conclude that the maximum entropy principle gives positive weights to all the possibilities that are not excluded by the constraints, and the partition function is formally equal to the statistical mechanic one. To prove that the two partition functions are actually the same, we will show that in the framework of statistical mechanics we obtain the same rules of calculations we've obtained with maximum entropy principle. Suppose we have a system in a

volume V and with other known parameters such as applied magnetic field or electric field, that will determine the possible energy levels E_i; if we now the average value $\langle E \rangle$, in the context of maximum entropy principle by reviewing our previous example we obtain exactly the Boltzmann distribution, i.e.

$$p_i = \mathcal{Z}^{-1} e^{-\beta E_i} \tag{3.13}$$

whereas by resorting to (3.10) and reminding that we are dealing with only one constraint

$$p_i = \exp(-\lambda_0 - \lambda_1 E_i) \tag{3.14}$$

but by resorting to (3.9) we can see that $\exp(-\lambda_0) = \mathcal{Z}^{-1}$ and with the identification $\lambda_0 = 1/K_B T$ we have basically the same expressions of the canonical ensemble partition function and probabilities. It is important to notice that the knowledge of $\langle E \rangle$ determines the value of temperature, that could be considered only an equilibrium parameter. We should also remember that the value of K_B is determined only by our arbitrary criterion regarding the scale of temperature, in a pure theoretic area we could just assume $K_B = 1$ and thus establish that energy and temperature have the same dimension and in this way the thermodynamic entropy becomes dimensionless as it should be (and as it is the Shannon one).

Now it is clear how we can derive the grand-canonical partition function: suppose now that in our system we have r chemical species and that we know, in addition to the usual $\langle E \rangle$, the mean value of the number of particles of each chemical species $\langle n_1 \rangle$, $\langle n_2 \rangle$, ..., $\langle n_r \rangle$. So from (3.7) and by reminding the value of λ_0 and of the Lagrange multiplier associated to $\langle E \rangle$

$$\mathcal{Z} = \sum_{i,n_1,n_2,...} e^{-\beta E_i - \sum_{j=1}^{r} \lambda_j n_j} \tag{3.15}$$

and we can immediately view the following identification $\lambda_j = -\mu_j / K_B T$, where the μ_j are the chemical potentials. Up to this point we can say that the formal equivalence between statistical mechanics approach and the information theory one has been proven.

In conclusion we can say that we can derive the relations of statistical mechanics in a very simple way if we reinterpret its prediction problem in the subjective sense, i.e. by focusing on the amount of uncertainty the observer has with respect to a given physical system. Thus we can avoid concepts like ensembles, or arguments regarding ergodicity or equal a priori probabilities. In the prediction problem, the entropy maximization is merely a method of reasoning which ensures us that no arbitrary assumption other than the ones we can make exploiting our prior knowledge have been made.

We have proved the formal equivalence between entropy in statistical mechanics and in information theory, now it's time to move on the second and much more interesting connection; we will introduce the Lieb and Yngvason axiomatic framework in the next chapter.

Chapter 4

Lieb and Yngvason approach

In this chapter we illustrate the Lieb and Yngvason approach to classical thermodynamics. The idea is to define a set Γ whose elements (X, Y, etc.) are called equilibrium states. In fact physically we may consider this set as the set of all equilibrium states of a thermodynamic system, but we should think of it for now in a more general aspect (later we will include in this axiomatic framework concepts related to information theory).

In this set we can define an operation of scaling, i.e. if an element $X \in \Gamma$, then $\lambda X \in \Gamma$ with $\lambda \geq 0$, and an operation of composition, i.e given two sets Γ, Γ' then we may consider the element $(X, X') \in \Gamma \times \Gamma$. The physical interpretations of these operations in the thermodynamic framework are respectively scaling the system substance with a factor of λ and the possibility of considering a system as composed by two subsystems.

In these sets we define the order relation $\prec$ that satisfies the following axioms:

A1 Reflexivity: $X \sim X$ (we write $X \sim Y$ if $X \prec Y$ and $Y \prec X$);

A2 Transitivity: $X \prec Y$ and $Y \prec Z \implies X \prec Z$;

A3 Consistent composition: $X \prec Y$ and $X' \prec Y' \implies (X, X') \prec (Y, Y')$;

A4 Scaling invariance: $X \prec Y \implies \lambda X \prec \lambda Y \ \forall \lambda > 0$;

A5 Splitting recombination: for $0 \leq \lambda \leq 1, \ X \sim (\lambda X, (1 - \lambda)X)$;

A6 Stability: $(X, \epsilon Z) \prec (Y, \epsilon Z')$ for $\epsilon \to 0 \implies X \prec Y$.

There is another property the relation $\prec$ must have in order to obtain our results

Comparison hypothesis: Any two elements X, Y in a set $(1 - \lambda)\Gamma \times \Gamma$ with $0 \leq \lambda \leq 1$ are comparable, i.e. related by $\prec$.

The comparison hypothesis, unlike the previous six properties, isn't an obvious request. In [1] the comparison hypothesis is demonstrated using other simpler postulates, but here we shall assume it.

We are ready to prove the main result of the Lieb and Yngvason approach

Theorem 4.1. *If the order relation $\prec$ satisfies the six axioms previously stated and the comparison hypothesis, then there exists an unique (up to an affine change of scale) real valued function S such as:*

1. *Characterization of adiabatic accessibility (monotonicity): for two states X and Y,*

$$X \prec Y \iff S(X) \le S(Y). \tag{4.1}$$

2. *Additivity and extensivity*

$$S(X, X') = S(X) + S(X') \ \text{and} \ S(\lambda X) = \lambda S(X). \tag{4.2}$$

Proof. Let's take two reference systems $X_0 \prec\prec X_1$ in Γ, where we write $X \prec\prec Y$ if $X \prec Y$ but $Y \not\prec X$ (if such systems don't exist, it implies that all systems are adiabatically equivalent, i.e they can all be related by $\prec$ and we can define S to be constant, so (4.1) is satisfied); let's pick another state X such that $X_0 \prec X \prec X_1$, so for an entropy function satisfying (4.1) and (4.2) it must be $S(X_0) < S(X_1)$ and $S(X_0) \le S(X) \le S(X_1)$ and there is an unique real number λ between 0 and 1 such that

$$S(X) = (1 - \lambda)S(X_0) + \lambda S(X_1). \tag{4.3}$$

To prove that, let's take a such that $S(X) = S(X_0) + a, \ a \ge 0$ and b such that $S(X_1) = S(X) + b, \ b \ge 0$ but a and b cannot be both equal to zero; obviously a and b are unique, and if we want to find an expression for λ we have

$$S(X) = (1 - \lambda)S(X_0) + \lambda S(X_1) = S(X) - a + \lambda(a + b) \implies \lambda = \frac{a}{a + b} \tag{4.4}$$

so we have the uniqueness of λ; we can write (remember that $S(X, X') = S(X) + S(X')$)

$$X \sim ((1 - \lambda)X_0, \lambda X_1). \tag{4.5}$$

Now another entropy function with the same values on X_0 and X_1 brings to equation (4.5) with a certain λ' but exploiting axioms A1-A6 it is easy to prove that $\lambda = \lambda'$ (see [1]). So entropy is unique up to an affine change of scale (the choice of its value on the two reference systems X_0 and X_1). For the existence of entropy, from A1-A6 one shows that (see [1])

$$\sup\{\lambda' : ((1 - \lambda')X_0, \lambda'X_1) \prec X\} = \inf\{\lambda' : X \prec ((1 - \lambda')X_0, \lambda'X_1)\} \tag{4.6}$$

thus for the λ' that satisfies the previous equation (we denote it by λ) we have

$$X \sim ((1 - \lambda)X_0, \lambda X_1) \tag{4.7}$$

so if we fixed the value of our reference systems as $S(X_0) = 0$ and $S(X_1) = 1$ we have an explicit formula for the entropy we'll write on the following definition. $\square$

Definition 4.2. *The unique (up to an affine change of scale) entropy function is*

$$S(X) = \sup\{\lambda : ((1-\lambda)X_0, \lambda X_1) \prec X\} =$$
$$= \inf\{\lambda : ((1-\lambda)X_0, \lambda X_1) \prec X\}. \tag{4.8}$$

First of all we shall spend a few words on the physical meaning of the relation $\prec$. In the thermodynamic framework the relation $\prec$ corresponds to the concept of adiabatic accessibility, i.e. a state Y may be reached from a state X by means of an adiabatic process if $X \prec Y$ holds. An adiabatic process is a process that leaves no trace on the environment except that a weight may have changed its relative position (this definition of an adiabatic process avoids the concept of heat). With this identification the S function defined in (4.8) is the classical entropy function. In fact we may view the entropy as a state function that allows us to determine precisely which changes are possible, and which are not, under well-defined conditions, and state the second law of thermodynamics as the existence of such a function. We may conclude that any formulation of thermodynamics must include axioms A1-A6 and the comparison hypothesis since they are equivalent to the second law (with those axioms we can demonstrate the existence of the entropy function whereas if we assume the existence of the entropy function it is straightforward to derive the aforementioned axioms, see [1] for further details).

Now we discuss a possible extension of this framework related to non-equilibrium systems. For these systems we may expect that axioms A4 and A5 (scaling and splitting) are not satisfied. So we introduce the set Γ_{ext} for which we require:

N1 Axioms A1, A2, A3, A6 hold;

N2 $\forall X \in \Gamma_{ext} \exists X', X'' \in \Gamma$ with $X' \prec X \prec X''$, i.e. every non-equilibrium state can be obtained by means of an adiabatic process from an equilibrium state and can be brought back to an equilibrium state.

For such non-equilibrium states the following holds.

Theorem 4.3. *On condition that N1 and N2 hold in Γ_{ext} then the functions*

$$S_-(X) = \inf\{\lambda : ((1-\lambda)X_0, \lambda X_1) \prec X\} \tag{4.9}$$

$$S_+(X) = \sup\{\lambda : ((1-\lambda)X_0, \lambda X_1) \prec X\} \tag{4.10}$$

bound all possible extensions S_{ext} of S to the set Γ_{ext} that are monotonic with respect to $\prec$.

Now we have all the concepts to move on resource theory and exploit this formalism that was originally meant to give an axiomatic framework to thermodynamics only (for the curious reader see [1]) to reach the goal of finding the information-theoretic entropies always with formula (4.8) but in the appropriate context.

Chapter 5

Quantum resource theory

In this chapter we will describe some fundamental topics about quantum resource theories and how they fit in the Lieb and Yngvason framework. To demonstrate that we need the concept of an adiabatic-like process and an order relation that satisfies axioms A1-A6 in certain states. It turns out that the order relation we seek lies in the concept of majorization and the adiabatic-like processes lie in the concept of quantum operations, two topics we shall briefly review.

5.1 Majorization

In this section we shall review the theory of majorization. For now for majorization we use the same notation $\prec$ we've already used for the order relation of the Lieb and Yngvason approach. Let's say $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are two n-dimensional vectors. We use $x^\downarrow$ to denote the vector x with the components ordered in decreasing order (for example $x_1^\downarrow$ is the largest component of x).

Definition 5.1. x *is majorized by* y, $x \prec y$, *if* $\sum_{i=1}^{k} x_i^\downarrow \leq \sum_{i=1}^{k} y_i^\downarrow$ *for* $k = 1, ..., n$ *with equality when* $k = n$.

The following theorems clarify why majorization is a candidate to represent the Lieb and Yngvason order relation in quantum information theory

Theorem 5.2. $x \prec y \iff x = \sum_j p_j P_j y$ *for some probability distribution* p_j *and some permutation matrices* $\{P_j\}$.

Proof. We suppose, without loss of generality, that $x = x^\downarrow$ and $y = y^\downarrow$. We prove the theorem by induction. For $n = 1$ the thesis is trivial, then suppose x and y are $n + 1$ dimensional vector and that $x \prec y$, for hypothesis $x_1 \leq y_1$. Now we choose j such that $y_j \leq x_1 \leq y_{j-1}$ (the components of y cannot be all greater than x_1 for the constraint $\sum_{i=1}^{n+1} x_i^\downarrow = \sum_{i=1}^{n+1} y_i^\downarrow$); let be $t \in [0, 1]$ such that $x_1 = ty_1 + (1 - t)y_j$ and define the operator $D = t\mathcal{I} + (1 - t)T$ with T the transpose matrix between element 1 and element j, so

$$Dy = (\overbrace{ty_1 + (1 - t)y_j}^{x_1}, ..., (1 - t)y_1 + ty_j, ..., y_{n+1}) \tag{5.1}$$

now with $x' = (x_2, ..., x_{n+1})$ and $y' = (y_2, ..., (1-t)y_1 + ty_j, ..., y_{n+1})$ we can use the inductive hypothesis (they both are n dimensional vectors) and write $x' = \sum_j p'_j P'_j y'$ if we show $x' \prec y'$ but we have for hypothesis $\sum_{i \neq 1,j}^{n+1} x'_i = \sum_{i \neq 1,j}^{n+1} y'_i$, it is sufficient to show that the term $(1-t)y_1 + ty_j$ doesn't give problems (it could help view $(1-t)y_1 + ty_j$ as equal to $y_1 + y_j - x_1$). So we can write $x = (\sum_j p'_j P'_j)Dy$ (the extended P'_j don't act on the first entry) and since a product of permutation matrices is a permutation matrix we obtain the result. $\qquad \square$

The previous theorem tells us that if $x \prec y$, then x is more "disordered" than y (x can be obtain by permuting the components of y and combine with some probability coefficient the resulting permuted vectors).

Definition 5.3. *A matrix D that can be viewed as a convex combination of permutation matrices is called doubly stochastic*

For a doubly stochastic matrix the following holds:

Theorem 5.4 (Birkhoff's theorem). *A matrix D is doubly stochastic if and only if has no negative entries and each row and column sums to 1.*

Now we can extend the concept of majorization to hermitian operators.

Definition 5.5. *Let H and K be hermitian operators; we say $H \prec K$ if $\lambda(H) \prec \lambda(K)$, where $\lambda(H)$ is the vector of eigenvalues of H.*

We notice that with this definition all density operators can be related by majorization (they all have unitary trace).

Theorem 5.6. *Let H and K be hermitian operators; then $H \prec K$ if and only if there is a probability distribution p_j and unitary matrices U_j such that*

$$H = \sum_j p_j U_j K U_j^\dagger. \tag{5.2}$$

Proof. Suppose $H \prec K$, then by theorem 5.2 the diagonal matrices whose entries are eigenvalues of the respective operator $\Lambda(H), \ \Lambda(K)$ can be related by

$$\Lambda(H) = \sum_j p_j P_j \Lambda(K) P_j^\dagger \implies V\Lambda(H)V^\dagger = \sum_j p_j V P_j W K W^\dagger P_j^\dagger V^\dagger \tag{5.3}$$

where P_j are permutation matrices. With the identifications $H = V\Lambda(H)V^\dagger$ and $\Lambda(K) = WKW^\dagger$ with V and W unitary matrices, we have the thesis with $U_j = VP_jW$. Suppose now $H = \sum_j p_j U_j K U_j^\dagger$, with the same identifications as before we arrive at

$$\lambda(H)_k = \sum_{jl} p_j V_{jkl} \lambda(K)_l V_{jkl}^\dagger = \sum_{jl} p_j |V_{jkl}|^2 \lambda(K)_l \tag{5.4}$$

and we can write $\lambda(H) = D\lambda(K)$ with $D_{kl} = \sum_j p_j |V_{jkl}|^2$, and by noticing that D is doubly stochastic for theorem 5.4 we conclude. $\qquad \square$

With this last theorem it becomes clear how majorization is connected to entropy and all the formalism we have previously developed. In fact, under certain restriction we should discuss later, majorization-like concepts are the Lieb and Yngvason relations for certain equilibrium states, and by theorem 4.1 we can find the unique entropy function associated, a matter we should discuss in detail in this chapter.

5.2 Quantum operations

Quantum operations are a tool very useful in describing the evolution of quantum systems in a great variety of circumstances, in particular they can be used to implement the idea of interaction of the quantum system with an environment, so they are useful to represent the evolution of an open quantum system. For example let's assume that a state ρ, the principal system, interacts with an environment represented by the state ρ_{env}, together they represent a closed quantum system. Suppose the evolution of the closed system is given by the unitary operator U, then we can say that the final state of the principal system can be written by using a quantum operation defined as a partial trace over the environment of the evolved closed state, i.e.

$$\mathcal{E}(\rho) = tr_2(U(\rho \otimes \rho_{env})U^\dagger). \tag{5.5}$$

This is quite a general formula for a quantum operation; actually there are three equivalent different approaches to quantum operation:

1. quantum operation as the result of an interaction between a system and an environment, as already seen;

2. quantum operations in the operator-sum representation;

3. quantum operations as the operations that satisfy physically motivated axioms.

The operator-sum representation consists in rewriting (5.5) exploiting the orthonormal basis of the environment $|e_k\rangle$; we assume that the environment starts in the pure state $|e_0\rangle\langle e_0|$; this assumption doesn't imply a loss of generality because, in case the environment isn't in a pure state we can consider its purification and the conclusions are the same. So we rewrite (5.5) as

$$\mathcal{E}(\rho) = \sum_k (\mathcal{I} \otimes \langle e_k|)U(\rho \otimes |e_0\rangle\langle e_0|)U^\dagger(\mathcal{I} \otimes |e_k\rangle)$$
$$= \sum_k \langle e_k|U|e_0\rangle \, \rho \, \langle e_0|U^\dagger|e_k\rangle = \tag{5.6}$$

$$= \sum_k E_k \rho E_k^\dagger \tag{5.7}$$

with $E_k = \langle e_k|U|e_0\rangle$ the so called operation elements; the (5.7) is the operator-sum representation of the quantum operation $\mathcal{E}$. Quantum operations are useful not only to represent the evolution of an open quantum system but they can

also describe the effects of measurements. In particular quantum operation of the first kind have operation elements that are trace preserving, i.e. they satisfy $\sum_k E_k E_k^\dagger = \mathcal{I}$, because the following must hold:

$$1 = tr\mathcal{E}(\rho) = \sum_k tr(E_k \rho E_k^\dagger) = \sum_k tr(E_k E_k^\dagger \rho) \ \forall \rho \implies \sum_k E_k E_k^\dagger = \mathcal{I}. \quad (5.8)$$

Otherwise quantum operations of the second kind don't have the constraint $tr\mathcal{E}(\rho) = 1$, for them $\sum_k E_k E_k^\dagger \leq \mathcal{I}$ (i.e $\mathcal{I} - \sum_k E_k E_k^\dagger$ could be positive) because the state after the measure is not $\mathcal{E}(\rho)$ but

$$\rho' = \frac{P_m \mathcal{E}(\rho) P_m^\dagger}{tr(P_m \mathcal{E}(\rho) P_m^\dagger)} \quad (5.9)$$

where P_m is the projector onto the subspace corresponding to the measurement outcome m.

In the axiomatic approach to quantum operations we can see them as the operations $\mathcal{E}$ that satisfy the following axioms:

E1 $0 \leq tr(\mathcal{E}(\rho)) \leq 1$ for any state ρ;

E2 $\mathcal{E}$ is a convex-linear map on the set of density operators, that is, for certain probabilities p_i, we have $\mathcal{E}(\sum_i p_i \rho_i) = \sum_i p_i \mathcal{E}(\rho_i)$;

E3 $\mathcal{E}$ is a positive operator for all A positive operator and $(\mathcal{I} \otimes \mathcal{E})B$ is positive for all B positive operator defined in the tensor product Hilbert space, i.e. $\mathcal{E}$ is a completely positive map.

Axiom E1 allows us to contain quantum operations that represents measurement, in particular we want $\mathcal{E}$ to be defined in such a way that $tr(\mathcal{E}(\rho))$ is equal to the probability of the measurement outcome described by $\mathcal{E}$.

From axiom E2 we want the evolution of the system to be given from a random selection from the ensemble $\{p(i/\mathcal{E}), \frac{\mathcal{E}(\rho_i)}{tr(\mathcal{E}(\rho_i))}\}$ with $p(i/\mathcal{E})$ the probability of obtaining ρ_i given the process represented by $\mathcal{E}$.

Finally axiom E3 assure us that we have density operators even if we do a partial trace over the system.

A remarkable result [10] is the following theorem, that implies as already stated the equivalence of these two last approaches.

Theorem 5.7. *$\mathcal{E}$ satisfies axioms E1-E3 $\iff$ we can write $\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger$ for all quantum state ρ for certain operation elements E_k.*

We will see different types of quantum operation in the following section in the context of resource theory.

5.3 Resource theories

A resource theory is a means to investigate which tasks can be achieved or how certain states of systems may change, if the processes affecting a system are of a restricted class, called the allowed operations. In quantum resource theory the systems are represented by density operators defined in an Hilbert space H. We will consider only finite-dimensional Hilbert spaces. The differences between the various resource theories come from the different types of allowed quantum operations one wishes to consider. In the following we will describe two resource theories, the resource theory of noisy operations and the resource theory of thermal operations.

Definition 5.8 (Noisy quantum operation). *A noisy quantum operation is a quantum operation $\mathcal{E}$ that admits an ancillary system H_a of dimension d_a and a unitary operator U in $H \otimes H_a$ such that $\forall \rho$, the following holds:*

$$\mathcal{E}(\rho) = tr_2(U(\rho \otimes \frac{1}{d_a}\mathcal{I}_a)U^\dagger) \tag{5.10}$$

where $\frac{1}{d_a}\mathcal{I}_a$ denotes the maximally mixed state.

We redefine the concept of majorization in this framework using the concept of step function. The spectrum of a state ρ with ordered eigenvalues $p_1 \geq p_2 \geq ... \geq p_d$ can be indeed represented as a step function

$$f_\rho(x) = \begin{cases} p_i, \ i-1 \leq x \leq i \\ 0, \ \text{otherwise.} \end{cases} \tag{5.11}$$

We express majorization as

$$\rho \prec \sigma \iff \int_0^k f_\rho(x)dx \geq \int_0^k f_\sigma(x)dx \ \forall k \geq 0. \tag{5.12}$$

With this redefinition the following holds:

Theorem 5.9. *For states ρ and σ of a system of dimension d there exists a noisy operation $\mathcal{E}$ such that $\mathcal{E}(\rho) = \sigma$ if and only if $\rho \prec \sigma$.*

Proof. see [6]. □

We define now the min and max-entropies for a density operator, they are of interest because as we will see in the following section they correspond to the Lieb and Yngvason S_- and S_+ in the appropriate framework regarding noisy operations.

Definition 5.10. *Let ρ be a density operator, then its min and max-entropies are defined as*

$$H_{min}(\rho) = -\log \|\rho\|_\infty \tag{5.13}$$

$$H_{max} = \log rank\rho \tag{5.14}$$

where $\|\rho\|_\infty$ denotes the maximal eigenvalue of ρ.

Now we discuss quantum resource theory of thermal operations.

Definition 5.11. *A thermal state is a state of the form*

$$\tau = \sum_i \frac{e^{-\beta E_i}}{\mathcal{Z}} |E_i\rangle\langle E_i| \tag{5.15}$$

where $\mathcal{Z}$ is the partition function and $|E_i\rangle$ are the energy eigenstates of a corresponding hamiltonian $\mathcal{H}_\tau$.

Definition 5.12 (Thermal operation). *A thermal operation is a quantum operation $\mathcal{E}$ that admits an ancillary thermal state τ with arbitrary hamiltonian $\mathcal{H}_\tau$ and with temperature T and a unitary operator U in $H \otimes H_\tau$ that commutes with the total hamiltonian (i.e. it is an energy-preserving operator) such that $\forall \rho$, the following holds:*

$$\mathcal{E}(\rho) = tr_2(U(\rho \otimes \tau)U^\dagger) \tag{5.16}$$

Note that in the extreme case of a system where all levels are degenerate, τ reduces to the maximally mixed state and a thermal operation reduces to a noisy operation.

Definition 5.13. *Let ρ be a density operator that is block diagonal in the energy eigenbasis, coupled with a thermal state τ and let d denote the dimension of its Hilbert space. We define the Gibbs-rescaled step function f_ρ^T as*

$$f_\rho^T(x) = \begin{cases} p_i e^{\beta E_i}, & \sum_{k=1}^{i-1} e^{-\beta E_k} \le x \le \sum_{k=1}^i e^{-\beta E_k} \\ 0, & otherwise \end{cases} \tag{5.17}$$

with eigenvalues reordered such that $p_1 e^{\beta E_1} \ge p_2 e^{\beta E_2} \ge ... \ge p_d e^{\beta E_d}$.

We now define a majorization-like concept for thermal operations that uses the Gibbs-rescaled step function.

Definition 5.14. *Let ρ and σ be two density operators that are block diagonal in the energy eigenbasis. We define the order relation of thermo-majorization $\prec_T$ as*

$$\rho \prec_T \sigma \iff \int_0^k f_\rho^T(x)dx \ge \int_0^k f_\sigma^T(x)dx \ \ \forall k \ge 0 \tag{5.18}$$

Even for thermal states we have a theorem analogous to 5.9 that allows us to connect resource theory of thermal operations with the Lieb and Yngvason approach:

Theorem 5.15 (Horodecki and Oppenheim). *Let ρ and σ be two density operators that are block diagonal in the energy eigenbasis, then there exists a thermal operation $\mathcal{E}$ such that $\mathcal{E}(\rho) = \sigma$ if and only if $\rho \prec_T \sigma$.*

Proof. See [8]. □

Setting	Processes	Equilibrium states	S	S_-, S_+
Isolated System	Noisy operations	$\sum_i \frac{1}{rank\rho}|X_i\rangle\langle X_i|$	H	H_{min}, H_{max}
Interaction with a heat bath	Thermal operations	$\sum_i \frac{e^{-\beta E_i}}{\mathcal{Z}}|E_i\rangle\langle E_i|$	A	A_{min}, A_{max}
Interaction with a heat bath and particle reservoir	$N-T$ operations	$\sum_i \frac{e^{-(\beta E_i - \mu N_i)}}{Q} \times$ $\times |E_i, N_i\rangle\langle E_i, N_i|$	Ω	$\Omega_{min}, \Omega_{max}$
Interaction with angular momentum reservoir	J operations	$\sum_i \frac{e^{-\hbar\gamma J_i}}{\mathcal{Z}}|J_i\rangle\langle J_i|$	S_j	S_{J_-}, S_{J_+}

Table 5.1: a little resume of the connection with Lieb and Yngvason approach and the various resource theories, taken from [2].

We finally give some definitions that will turn out to correspond respectively to the Lieb and Yngvason S, S_+ and S_- in the appropriate framework regarding thermal operations.

Definition 5.16. *Let ρ be a block diagonal density operator and τ the thermal state of the system, then we define*

$$A(\tau) = -K_B T \ln \mathcal{Z}_\tau \tag{5.19}$$

$$A_{min}(\rho) = -K_B T \log tr(\Pi_\rho \tau) \ln 2 + A(\tau) \tag{5.20}$$

$$A_{max}(\rho) = K_B T \log(\min\{\lambda : \rho \le \lambda\tau\}) \ln 2 + A(\tau) \tag{5.21}$$

where Π_ρ is the projector onto the support of ρ.

5.4 Lieb and Yngvason in resource theory

Now we have to make the connection between resource theory and the Lieb and Yngvason approach, so we will continuously referring to the notation and concepts explained in chapter 4. The set of states Γ_{ext} is now the set of all density operators whereas the subset of equilibrium states Γ will be specified in the various scenarios we will encounter; the adiabatic processes consist in certain quantum operations described in the previous section and the majorization-like concepts correspond to the Lieb and Yngvason order relation. The composition of two systems ρ and σ corresponds to the tensor product $\rho \otimes \sigma$ (all systems are considered uncorrelated before any interaction); the scaling of an

equilibrium state of a factor $\lambda \in \mathbb{N}$ is defined as the composition of the state ρ with itself λ times. The step function of the scaled state $\lambda\rho \equiv \rho^{\otimes\lambda}$ is then:

$$f_{\lambda\rho}(x) = f_\rho^\lambda(x^{\frac{1}{\lambda}}) = \begin{cases} (\frac{1}{rank(\rho)})^\lambda, \ 0 \leq x \leq rank^\lambda(\rho) \\ 0 \ \text{otherwise} \end{cases} \tag{5.22}$$

(one can extend this argument to a generic $\lambda \in \mathbb{R}_{>0}$, see the supplemental material of Ref. [2]).

In the framework of noisy operations, the set of equilibrium states consists of all density operators with flat spectrum (i.e. all non zero eigenvalues are equal), the adiabatic-like processes coincide (without surprise) with noisy operations and the order relation corresponds to majorization defined in 5.5.

For equilibrium states, the step function is simply

$$f_\rho(x) = \begin{cases} \frac{1}{rank\rho}, \ 0 \leq x \leq rank\rho \\ 0, \ \text{otherwise} \end{cases} \tag{5.23}$$

so we have

$$\rho \prec \sigma \iff rank\rho \geq rank\sigma \tag{5.24}$$

Now we are able to prove the main result we were waiting for:

Theorem 5.17. *In the framework we have developed above (definition of equilibrium and non equilibrium state in quantum resource theory of noisy operations and the scaling and composition operations), for equilibrium states A1-A6 and the comparison hypothesis hold and for non-equilibrium state N1 and N2 hold.*

Proof.   A1 Reflexivity. Clearly from (5.12) we have $\rho \prec \rho$.

A2 Transitivity. Let ρ, σ, η be density operators, $\rho \prec \sigma$ and $\sigma \prec \eta$ means that

$$\int_0^k f_\rho(x)dx \geq \int_0^k f_\sigma(x)dx \geq \int_0^k f_\eta(x)dx \ \forall k \geq 0 \tag{5.25}$$

so $\rho \prec \eta$.

A3 Consistent composition. The proof is not straightforward, see the supplement material of [2].

A6 Stability. Let ρ be a density operator with eigenvalues $p_1 \geq p_2 \geq ... \geq p_d$ with d the dimension of the Hilbert space. Let ρ', σ and σ' be other density operators and assume that there exists a sequence $\{\epsilon_i\}$ that tends to 0 such that the following holds

$$(\rho, \epsilon_i\rho') \prec (\sigma, \epsilon_i\sigma') \ \forall i \tag{5.26}$$

with ρ', σ' equilibrium states (i.e. with flat spectrum).

Then using step functions and taking the limit for $i \to \infty$ we have

$$\int_0^k \lim_{i \to \infty} f_{(\rho, \epsilon_i \rho')}(x) dx \geq \int_0^k \lim_{i \to \infty} f_{(\sigma, \epsilon_i \sigma')}(x) dx \quad \forall k \geq 0. \tag{5.27}$$

However, the step function of the two composed systems can be written as (remember the definition of a scaled quantum state (5.22))

$$f_{(\rho, \epsilon_i \rho')}(x) = \begin{cases} p_k \frac{1}{rank(\rho')^{\epsilon_i}}, & (k-1)rank(\rho')^{\epsilon_i} \leq x \leq k(rank(\rho')^{\epsilon_i}) \\ 0, & \text{otherwise} \end{cases} \tag{5.28}$$

thus in the limit the term $\frac{1}{rank(\rho')^{\epsilon_i}}$ becomes closer and closer to one and we can indeed write

$$\int_0^k f_\rho(x) dx \geq \int_0^k f_\sigma(x) dx \quad \forall k \geq 0. \tag{5.29}$$

i.e $\rho \prec \sigma$.

We have proved that axiom N1 holds too (we didn't restrict to equilibrium states in the previous proofs), for N2 it is sufficient to take the maximally mixed state ρ'' and the pure state ρ' and it is straightforward to show that for any non-equilibrium state ρ, $\rho' \prec \rho \prec \rho''$ (remember that for equilibrium states the rank suffices to assure majorization). Now we have to prove the remaining axioms only for equilibrium states.

A4 Scaling invariance. By using (5.24) we have

$$\rho \prec \sigma \iff rank\rho \geq rank\sigma \iff rank^\lambda \rho \geq rank^\lambda \sigma \quad \forall \lambda > 0$$
$$\iff \lambda \rho \prec \lambda \sigma. \tag{5.30}$$

A5 Splitting and recombination. Let $0 < \lambda < 1$, then

$$f_{(\lambda \rho, (1-\lambda)\rho)}(x) = \begin{cases} \frac{1}{rank^\lambda \rho} \frac{1}{rank^{1-\lambda} \rho}, & 0 \leq x \leq rank^\lambda(\rho)rank^{1-\lambda} \rho \\ 0, & \text{otherwise} \end{cases}$$
$$= \begin{cases} \frac{1}{rank\rho}, & 0 \leq x \leq rank\rho \\ 0, & \text{otherwise} \end{cases} = f_\rho(x). \tag{5.31}$$

For the comparison hypothesis, let $0 \leq \lambda \leq 1$ and $0 \leq \mu \leq 1$ and let $\rho, \rho', \sigma, \sigma'$ be equilibrium states, then the state $(\lambda \rho, (1-\lambda)\sigma)$ can be related with $\prec$ to any $(\mu \rho', (1-\mu)\sigma')$ by exploiting the fact that the step function takes the form

$$f_{(\lambda \rho, (1-\lambda)\sigma)}(x) = \begin{cases} \frac{1}{rank^\lambda \rho} \frac{1}{rank^{1-\lambda} \sigma}, & 0 \leq x \leq rank^\lambda(\rho)rank^{1-\lambda} \sigma \\ 0, & \text{otherwise} \end{cases} \tag{5.32}$$

$\square$

As a consequence of the previous theorem, we can define in the context of quantum resource theory of noisy operations the Lieb and Yngvason thermodynamic entropy thanks to equation (4.1). We have then the following result.

Theorem 5.18. *Let quantum states be ordered by $\prec$ and let the equilibrium states be those with flat spectrum; then the unique Lieb and Yngvason thermodynamic entropy corresponds to the Von Neumann entropy H, whereas S_- and S_+ correspond respectively to H_{min} and H_{max}.*

Proof. Let ρ_0 and ρ_1 be our reference equilibrium states, and $\rho_0 \prec\prec \rho_1$; let ρ be a density operator with ordered eigenvalues $p_1 \geq p_2 \geq ... \geq p_d$ with d the dimension of the Hilbert space. Let λ be such that $((1 - \lambda)\rho_0, \lambda\rho_1) \prec \rho$, i.e. in terms of step functions:

$$\int_0^k f_{((1-\lambda)\rho_0, \lambda\rho_1)}(x)dx \geq \int_0^k f_\rho(x)dx \quad \forall k \geq 0. \tag{5.33}$$

Let $k' = rank^{1-\lambda}(\rho_0)rank^\lambda(\rho_1)$, but ρ_0 and ρ_1 are equilibrium states, so from (5.24), the inequality in (5.33) implies for $k \leq 1$ (remember that for $0 \leq x \leq 1$, $f_\rho(x) = p_1$)

$$\left(\frac{1}{rank(\rho_0)}\right)^{1-\lambda} \left(\frac{1}{rank(\rho_1)}\right)^\lambda \geq p_1 \tag{5.34}$$

which can be rewritten, with obvious notation of C_1 and C_0 as

$$C_1 \log \frac{1}{p_1} + C_0 \geq \lambda. \tag{5.35}$$

The inequality (5.34) implies also

$$\left(\frac{1}{rank(\rho_0)}\right)^{1-\lambda} \left(\frac{1}{rank(\rho_1)}\right)^\lambda \min\{k, k'\} \geq p_1 \min\left\{k, \frac{1}{p_1}\right\}$$
$$\geq \int_0^k f_\rho(x)dx \tag{5.36}$$

for all $k \geq 0$; the first inequality follows because if $k \leq k'$, then since from (5.33) we have $k' \leq \frac{1}{p_1}$, we conclude that $k \leq \frac{1}{p_1}$, whereas if $k \geq k'$ we have two cases:

1. $k \leq \frac{1}{p_1} \implies 1 \geq kp_1$, and this is exactly what the first inequality in (5.36) becomes;

2. $k \geq \frac{1}{p_1}$, and (5.36) becomes the obvious $1 \geq 1$.

The second inequality follows since $f_\rho(x)$ is monotonously decreasing and normalized. Thus, taking the supremum over λ in (5.35) we conclude that $S_-(\rho) = H_{min}(\rho) = -\log\|\rho\|_\infty$ (in this case $\|\rho\|_\infty = p_1$) up to an affine change of scale. For S_+ one starts with λ such that $\rho \prec ((1 - \lambda)\rho_0, \lambda\rho_1)$ and with analogous reasoning as before one arrives at

$$\int_0^k f_{((1-\lambda)\rho_0, \lambda\rho_1)}(x)dx \leq \int_0^k f_\rho(x)dx \quad \forall k \geq 0. \tag{5.37}$$

We show that

$$k' \geq rank(\rho) \tag{5.38}$$

by contradiction: let $k' < rank(\rho)$, we have then

$$\int_0^{k'} f_\rho(x)dx < 1 \tag{5.39}$$

but this contradicts (5.37) since

$$\int_0^{k'} f_{((1-\lambda)\rho_0,\lambda\rho_1)}(x)dx = 1. \tag{5.40}$$

Thus we can rewrite (5.38) with obvious notation for C_1 and C_0 as

$$\lambda \geq C_1 \log(rank(\rho)) + C_0. \tag{5.41}$$

The inequality in (5.38) (with analogous reasoning already seen for the inequality (5.36)) also implies

$$\left(\frac{1}{rank(\rho_0)}\right)^{1-\lambda} \left(\frac{1}{rank(\rho_1)}\right)^{\lambda} \min\{k,k'\} \leq \frac{1}{rank(\rho)} \min\{k, rank(\rho)\}$$
$$\leq \int_0^k f_\rho(x)dx \tag{5.42}$$

i.e. implies (5.37) and taking the infimum over λ in (5.41) one concludes $S_+(\rho) = \log(rank(\rho)) = H_{max}(\rho)$. In the case of equilibrium states both $\rho \prec ((1-\lambda)\rho_0,\lambda\rho_1)$ and $((1-\lambda)\rho_0,\lambda\rho_1) \prec \rho$ hold, so form the previous results we have

$$\log(rank(\rho)) = \log \frac{1}{p_1} \tag{5.43}$$

so

$$\lambda = C_1 \log(rank(\rho)) + C_0 \tag{5.44}$$

and this coincides with the Von Neumann entropy H up to an affine change of scale (remember that we are dealing with density operators with flat spectrum). □

Due to theorem 5.18 we have established a profound connection between the concept of entropy in thermodynamics and in quantum information theory: we can indeed view quantum information theory as a particular case of thermodynamics in the sense of Lieb and Yngvason, and the corresponding thermodynamic entropy is precisely the information-theoretic entropy.

We obviously have the analogous of 5.17 and of 5.18 in the framework of thermal operations.

Theorem 5.19. *The order relation $\prec_T$ in the framework of thermal operations satisfies axioms A1-A6 and the comparison hypothesis for thermal states τ (the equilibrium states) and axioms N1-N2 for non equilibrium states that are block-diagonal in the energy eigenbasis.*

Proof. The proof is basically identical to the one in 5.17 with the Gibbs rescaled step function instead of the usual step function. The only step that is slightly different is the proof of the validity of axiom N1; let ρ be a block-diagonal state in the energy eigenbasis of the form

$$\rho = \sum_i \rho_i |E_i\rangle\langle E_i| \tag{5.45}$$

then an equilibrium state that thermomajorizes ρ is $\tau = |E_1\rangle\langle E_1|$ (τ could be viewed as a thermal state with 0 temperature) and an equilibrium state that is thermomajorized by $\prec_T$ is $\tau' = \sum_i \frac{e^{-\beta E_i}}{\mathcal{Z}}|E_i\rangle\langle E_i|$. The validity of $\rho \prec_T \tau$ is obvious from the definition of $\prec_T$, for $\tau' \prec_T \rho$ one can exploit the fact that for a thermal state the Gibbs rescaled step function takes the simple form

$$f_{\tau'}^T(x) = \begin{cases} \frac{1}{\mathcal{Z}}, & 0 \le x \le \mathcal{Z} \\ 0, & \text{otherwise} \end{cases} \tag{5.46}$$

$\square$

Theorem 5.20. *Let quantum states be ordered by $\prec_T$ and let the equilibrium states be thermal states; then the unique Lieb and Yngvason thermodynamic entropy corresponds to the Helmoltz free energy A, whereas S_- and S_+ correspond respectively to A_{max} and A_{min}.*

Proof. Let τ_0 and τ_1 be our reference equilibrium states, and $\tau_0 \prec\prec_T \tau_1$; let ρ be a density operator with rescaled eigenvalues $p_i^{(r)} = p_i e^{\beta E_i}$ ordered as $p_1^{(r)} \ge p_2^{(r)} \ge \dots \ge p_d^{(r)}$ with d the dimension of the Hilbert space. Let λ be such that $((1-\lambda)\tau_0, \lambda\tau_1) \prec_T \rho$, in terms of step functions this statement takes the following form:

$$\int_0^k f_{((1-\lambda)\tau_0, \lambda\tau_1)}^T(x)dx \ge \int_0^k f_\rho^T(x)dx \quad \forall k \ge 0. \tag{5.47}$$

Since τ_0 and τ_1 are equilibrium states, from (5.46), we have that

$$\int_0^k f_{((1-\lambda)\tau_0, \lambda\tau_1)}^T(x)dx = \int_0^k \left(\frac{1}{\mathcal{Z}_{\tau_0}}\right)^{1-\lambda}\left(\frac{1}{\mathcal{Z}_{\tau_1}}\right)^{\lambda} dx \tag{5.48}$$

so the inequality in (5.47) implies for $k \le 1$ (remember that for $0 \le x \le 1$, $f_\rho^T(x) = p_1^{(r)}$)

$$\left(\frac{1}{\mathcal{Z}_{\tau_0}}\right)^{1-\lambda}\left(\frac{1}{\mathcal{Z}_{\tau_1}}\right)^{\lambda} \ge p_1^{(r)} \tag{5.49}$$

which can be rewritten, with obvious notation of a_T and b_T as

$$a_T \log \frac{1}{p_1^{(r)}} + b_T \ge \lambda. \tag{5.50}$$

The inequality in (5.49) in a way analogous to what we have seen in (5.36) implies (we denote $k' = \mathcal{Z}_{\tau_0}^{1-\lambda} \mathcal{Z}_{\tau_1}^{\lambda}$)

$$
\begin{aligned}
\left(\frac{1}{\mathcal{Z}_{\tau_0}}\right)^{1-\lambda} \left(\frac{1}{\mathcal{Z}_{\tau_1}}\right)^{\lambda} \min\{k, k'\} &\geq p_1^{(r)} \min\left\{k, \frac{1}{p_1^{(r)}}\right\} \\
&\geq \int_0^k f_\rho^T(x)dx \quad \forall k \geq 0
\end{aligned}
\tag{5.51}
$$

and so conditions (5.49) and (5.47) are equivalent. For a state that is block-diagonal in the energy eigenbasis we have

$$
\begin{aligned}
A_{max}(\rho) &= -K_B T \ln \mathcal{Z}_\tau + K_B T \ln \min\{\lambda : \rho \leq \lambda \tau\} \\
&= K_B T \ln \min\{\mu : \rho \leq \mu \mathcal{Z}_\tau \tau\} \\
&= K_B T \ln p_{max}^{(r)}
\end{aligned}
\tag{5.52}
$$

where $p_{max}^{(r)}$ is the maximal rescaled eigenvalue of ρ, so taking the supremum over λ in (5.50) we have that $S_-(\rho) = A_{max}(\rho)$. For S_+ one again starts with λ such that $\rho \prec ((1-\lambda)\tau_0, \lambda\tau_1)$ and with analogous reasoning as before one arrives at

$$
\int_0^k f_{((1-\lambda)\tau_0,\lambda\tau_1)}^T(x)dx \leq \int_0^k f_\rho^T(x)dx.
\tag{5.53}
$$

We want to show by contradiction that

$$
\mathcal{Z}_{\tau_0}^{1-\lambda} \mathcal{Z}_{\tau_1}^{\lambda} \geq \mathcal{Z}_\rho
\tag{5.54}
$$

where $\mathcal{Z}_\rho$ is the partition function of a generic state ρ block-diagonal in the energy eigenbasis. Let $\mathcal{Z}_{\tau_0}^{1-\lambda} \mathcal{Z}_{\tau_1}^{\lambda} < \mathcal{Z}_\rho$, we have then

$$
\int_0^{k'} f_\rho^T(x)dx < 1
\tag{5.55}
$$

but this contradicts (5.53) that implies

$$
\int_0^{k'} f_{((1-\lambda)\tau_0,\lambda\tau_1)}^T(x)dx = 1.
\tag{5.56}
$$

So we can rewrite (5.54) as

$$
\lambda \geq a_T \ln \mathcal{Z}_\rho + b_T.
\tag{5.57}
$$

Again one can see that (5.54) implies

$$
\begin{aligned}
\left(\frac{1}{\mathcal{Z}_{\tau_0}}\right)^{1-\lambda} \left(\frac{1}{\mathcal{Z}_{\tau_1}}\right)^{\lambda} \min\{k, k'\} &\leq \frac{1}{\mathcal{Z}_\rho} \min\{k, \mathcal{Z}_\rho\} \\
&\leq \int_0^k f_\rho^T(x)dx \quad \forall k \geq 0
\end{aligned}
\tag{5.58}
$$

i.e. it implies (5.53) and taking the infimum over λ in (5.57) one concludes $S_+(\rho) = A_{min}(\rho)$ since for A_{min} we find that (remember that ρ is always a block-diagonal state in the energy eigenbasis)

$$
\begin{aligned}
A_{min}(\rho) &= - K_B T \ln \mathcal{Z}_\tau - K_B T \ln(tr(\Pi_\rho \tau)) \\
&= - K_B T \ln(\mathcal{Z}_\tau tr(\Pi_\rho \tau)) \\
&= - K_B T \ln \mathcal{Z}_\rho.
\end{aligned}
\tag{5.59}
$$

In the case of equilibrium state $\rho \sim_T ((1-\lambda)\tau_0, \lambda\tau_1)$ holds, so from the previous results we have

$$
- \ln p_{max}^{(r)} = \ln \mathcal{Z}_\tau
\tag{5.60}
$$

so

$$
\lambda = a_T \ln \mathcal{Z}_\tau + b_T
\tag{5.61}
$$

and this coincides with the Helmholtz free energy A up to an affine change of scale. $\qquad\square$

In table 5.4 are resumed some applications of the Lieb and Yngvason approach to various scenarios; until now we have seen the first two, now we briefly review the last two that are very similar to the case of thermal operations (that's because they consist in adding other types of reservoir to the system and that leads to mathematically equivalent situations).

In addition to an heat bath, one can consider the system in contact with a particle reservoir (a situation analogous to the grand-canonical ensemble in statistical mechanics). We can then go back over the particular case of the system in contact with an heat reservoir only with minimal variations. In this new framework the equilibrium states are that of the form

$$
\xi = \sum_i \frac{e^{-(\beta E_i - \mu N_i)}}{\mathcal{Q}} |E_i, N_i\rangle\langle E_i, N_i|
\tag{5.62}
$$

with $\mathcal{Q}$ the grand-canonical partition function. The adiabatic-like processes are defined as quantum operations we shall call $N - T$ operations of the form

$$
\mathcal{E}(\rho) = tr_2(U(\rho \otimes \xi)U^\dagger)
\tag{5.63}
$$

with U an unitary operator that commutes with the total hamiltonian and the total number operator (that means U represent an unitary transformation conserving the total energy and the particle number of system and the reservoir). We have now to define the $N - T$ majorization $\prec_{N-T}$ in a way analogous to the definition of $\prec_T$.

Definition 5.21. *Let ρ be a block-diagonal state in the basis $\{|E, N\rangle\}$. The $N - T$ rescaled step function is*

$$
f_\rho^{N,T}(x) = \begin{cases} p_i e^{\beta(E_i - \mu N_i)}, & \sum_{k=1}^{i-1} e^{-\beta(E_k - \mu N_k)} \le x \le \sum_{k=1}^{i} e^{-\beta(E_k - \mu N_k)} \\ 0, & otherwise \end{cases}
\tag{5.64}
$$

Definition 5.22. *Let ρ, σ be two states which are block diagonal in the basis $\{|E, N\rangle\}$, then the $N - T$ majorization $\prec_{N-T}$ is defined as*

$$\rho \prec_{N-T} \sigma \iff \int_0^k f_\rho^{N,T}(x)dx \geq \int_0^k f_\sigma^{N,T}(x)dx \ \forall k \geq 0. \tag{5.65}$$

The order relation $\prec_{N-T}$ in a way analogous to $\prec_T$ fulfills axioms A1-A6 and the comparison hypothesis, there is thus a unique Lieb and Yngvason thermodynamic entropy function in this framework.

Theorem 5.23. *Let ξ be an equilibrium state, the unique Lieb and Yngvason entropy functions are*

$$S_{N,T}(\xi) = a_{N,T} \ln \mathcal{Q} + b_{N,T} \tag{5.66}$$

(where we made explicit the possible affine change of scale) with $a_{N,T} > 0$.

Proof. Let ξ_0 and ξ_1 be our reference equilibrium states, and $\xi_0 \prec\prec_{N-T} \xi_1$, then by exploiting 4.2 we can write

$$((1 - S_{N,T}(\xi))\xi_0, S_{N,T}(\xi)\xi_1) \sim_{N-T} \xi. \tag{5.67}$$

Now the generic $N - T$ rescaled step function for an equilibrium state ξ' is

$$f_{\xi'}^{N,T}(x) = \begin{cases} \frac{1}{\mathcal{Q}}, & 0 \leq x \leq \mathcal{Q} \\ 0, & \text{otherwise} \end{cases} \tag{5.68}$$

so for equilibrium states the following holds

$$\xi' \prec \xi'' \iff \mathcal{Q}_{\xi''} \geq \mathcal{Q}_{\xi'} \tag{5.69}$$

and returning to (5.67) we have

$$\mathcal{Q}_{\xi_0}^{1 - S_{N,T}(\xi)} \mathcal{Q}_{\xi_1}^{S_{N,T}(\xi)} = \mathcal{Q}_\xi \tag{5.70}$$

and resolving with respect to $S_{N,T}(\xi)$ we obtain the thesis. $\qquad\square$

We can then choose the affine change of scale that allows us to identify $S_{N,T}$ with the grand-canonical potential Ω

$$\Omega = -K_B T \ln \mathcal{Q}. \tag{5.71}$$

In the case of non-equilibrium states, the bounding functions we shall call Ω_{max} and Ω_{min} can be calculated analogously to the scenario including only the heat bath.

We finally treat the last case, a system in contact with an angular momentum reservoir. A concrete model of angular momentum reservoir can consists of N spin-$\frac{1}{2}$ particles. The spin states are assumed to be degenerate in energy. The equilibrium probability for the reservoir to be in a particular state with

n particles with z component of spin equal to $+\frac{1}{2}$ and $N - n$ particles with z component of spin equal to $-\frac{1}{2}$ is

$$p_n = \frac{e^{-n\hbar\gamma}}{(1 + e^{-\hbar\gamma})^N} \tag{5.72}$$

where the parameter γ is the analogue of β encountered in the context of an heat bath. The denominator in equation (5.72) has the form of an angular momentum partition function $\mathcal{Z}_J$. This allows us to consider an angular momentum reservoir of size $N \to \infty$.

The equilibrium states in this last framework are that of the form

$$\mathcal{J} = \sum_J \frac{e^{-\gamma\hbar J}}{\mathcal{Z}_J}|J\rangle\langle J| \tag{5.73}$$

where $\mathcal{Z}_J$ is the angular momentum partition function and γ is the parameter analogous to β that we have already encountered. The adiabatic-like processes are defined as quantum operations we shall call J operations of the form

$$\mathcal{E}(\rho) = tr_2(U(\rho \otimes \mathcal{J})U^\dagger) \tag{5.74}$$

with U an unitary operator that commutes with the total angular momentum operator (that means U represent an unitary transformation conserving the angular momentum of system and the reservoir). Then for the J majorization $\prec_J$ we need the usual following definition.

Definition 5.24. *Let ρ be a block-diagonal state in the eigenbasis of the z component of the angular momentum. The J rescaled step function is*

$$f_\rho^J(x) = \begin{cases} p_i e^{\gamma\hbar J_i}, & \sum_{k=1}^{i-1} e^{-\gamma\hbar J_k} \leq x \leq \sum_{k=1}^{i} e^{-\gamma\hbar J_k} \\ 0, & otherwise \end{cases} \tag{5.75}$$

Definition 5.25. *Let ρ, σ be two states which are block diagonal in the basis $\{|J\rangle\}$, then the J majorization $\prec_J$ is defined as*

$$\rho \prec_J \sigma \iff \int_0^k f_\rho^J(x)dx \geq \int_0^k f_\sigma^J(x)dx \;\; \forall k \geq 0. \tag{5.76}$$

The order relation as usual fulfills axioms A1-A6 and the comparison hypothesis for equilibrium states and N1-N2 for non-equilibrium states, there is thus a unique Lieb and Yngvason thermodynamic entropy function

$$S_J \propto \ln \mathcal{Z}_J \tag{5.77}$$

for equilibrium states and S_{J-} and S_{J+} for non-equilibrium states.

Chapter 6

Conclusions

In this thesis we showed how the entropy functions of statistical mechanics, thermodynamics and information theory may be connected and thus established a profound connection between these theories: statistical mechanics and information theory with the maximum entropy principle and information theory and thermodynamics with the Lieb and Yngvason axiomatic approach, as summarized in figure 6.

One could ask why such an effort is necessary. Viewing different fields of knowledge in the same light can greatly help one to understand the connections between them and at last it greatly enhances the understanding of these different subjects.

Here we have the concept of entropy in two main theories: thermodynamics and (quantum) information theory, and in both theories entropy is a fundamental concept. By reviewing its definition in chapter 2 we immediately notice that the thermodynamic entropy definition is very different from the information-theoretic one but they have similar properties; the most important similar properties they share are additivity and extensivity, and in the Lieb and Yngvason approach we have used this fact as a characterization of the abstract entropy as we have seen in theorem 4.1. Then we have the remarkable result that entropy-like functions in information theory could really be viewed as an analogue of the thermodynamic one: the only requirement is that the axioms of the Lieb and Yngvason approach hold in a particular framework and then all the entropy concepts follow in the same manner the thermodynamic entropy function follows from the framework of adiabatic processes and equilibrium states.

We have also seen a possible extension of the thermodynamic entropy to non-equilibrium states, but again thermodynamics and quantum information theory (in the framework of resource theories) can both be viewed in the light of Lieb and Yngvason approach, so if there is this possible extension for thermodynamics, there must be an analogous extension in quantum information theory, as can be seen in table 5.4.

This last feature might not be so obvious without the common framework for entropy in the two theories we have reviewed in this thesis.
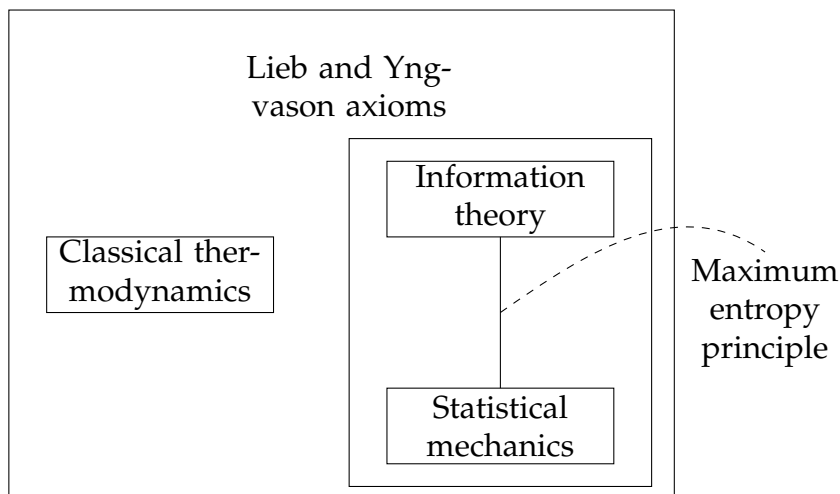
Figure 6.1: A simple scheme that summarizes the connections we have discussed in this thesis.

Another connection we have discussed in this thesis is the one between statistical mechanics and information theory with Jayne's maximum entropy principle, as seen in chapter 3. It is astonishing how we have easily derived all the most important quantities of statistical mechanics (the partition function $\mathcal{Z}$, the grand-canonical one $\mathcal{Q}$ and such) with minimal efforts only by using the maximum entropy principle. This fact emphasize the strong connection with statistical mechanics and information theory.

We must remark, however, that in chapter 5 for a system interacting with a reservoir we always assumed that the states are block-diagonal on a certain eigenbasis (for the heat bath we assumed states that are block-diagonal in the energy eigenbasis, for the heat bath and particle reservoir we assumed block-diagonal states in the basis of both the Hamiltonian and the number of particle operator and finally for the angular momentum reservoir the states were assumed to be block-diagonal in the angular momentum eigenbasis). Future works may investigate whether the results we have proved in chapter 5 also hold for generic (equilibrium) states. Maybe future works can also investigate the possibility to relate other thermodynamic and information theoretic quantities with the Lieb and Yngvason approach. Finally other future works can treat processes where quantum side information about the system is exploited.

Bibliography

[1] H. Lieb and J. Yngvason, Physics Reports 310 (1999) 1-96

[2] M. Weilenmann, L. Kraemer, P. Faist, and R. Renner, Phys Rev. Letters 117, 260601 (2016)

[3] E. T. Jaynes, Phys. Rev. 106, 620 (1957).

[4] B. Piechocinska, Phys. Rev. A, volume 61, 062314 (2000)

[5] H. Lieb, J. Yngvason 2013 The entropy concept for non-equilibrium states. Proc R Soc A 469: 20130408.

[6] M. Horodecki, P. Horodecki, and J. Oppenheim, Phys Rev A 67, 062104 (2003)

[7] K. Maruyama, F. Nori, V. Vedral, The physics of Maxwell demon and information, Rev. Of Modern Phys., Volume 81

[8] M. Horodecki, J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics. Nat. Commun. 4:2059 doi: 10.1038/ncomms3059 (2013)

[9] A mathematical theory of resources, B. Coecke, T. Fritz, and R. W. Spekkens University of Oxford, Department of Computer Science Perimeter Institute for Theoretical Physics December 1, 2014

[10] Nielsen, Michael A., and Chuang, Isaac L., Quantum Computation And Quantum Information, Cambridge University Press (2000)