

# Contents

4	Elements of quantum statistical physics	3
4.1	Review of the structural elements of quantum mechanics . . . . .	3
4.1.1	Hilbert spaces . . . . .	3
4.1.2	Observables . . . . .	4
4.1.3	Density operators and expectation values . . . . .	7
4.1.4	Time evolution . . . . .	12
4.1.5	Composite systems . . . . .	13
4.1.6	A translation table . . . . .	15
4.2	Quantum mechanical ensembles . . . . .	16
4.2.1	Microcanonical ensemble . . . . .	16
4.2.2	Canonical ensemble . . . . .	16
4.2.3	Gibbs variational principle revisited . . . . .	17
4.2.4	Von Neumann entropy . . . . .	18
4.3	Third law of thermodynamics . . . . .	19



## Chapter 4

# Elements of quantum statistical physics

### 4.1 Review of the structural elements of quantum mechanics

We have seen that the classical description reaches its limits every once in a while and that the encountered paradoxes that a classical description could not resolve. This is no surprise, given that the fundamental theory that captures the dynamics of microscopic particles is quantum mechanics, not classical mechanics. Any serious attempt to derive the laws of equilibrium thermodynamics from microscopic laws hence must in one way or the other resort to a quantum mechanical description. In this chapter, we lay out the basics of such a quantum mechanical description. We begin with a review of the structural elements of quantum mechanics as such, and then turn to the specific topic of quantum statistical mechanics.

#### 4.1.1 Hilbert spaces

In classical mechanics, each type of particles was associated a phase space  $\Gamma$ . In quantum theory, we associate with systems a Hilbert space  $\mathcal{H}$ . This is a vector space over  $\mathbb{C}$ , that is, for  $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ , and  $\lambda \in \mathbb{C}$ , we also have that

$$|\psi\rangle + |\phi\rangle \in \mathcal{H} \quad \lambda|\psi\rangle \in \mathcal{H}. \quad (4.1)$$

Hilbert space elements are also referred to as “kets” in the Dirac notation. There also is a scalar product  $\langle\phi|\psi\rangle \in \mathbb{C}$  which is additive in each component and which satisfies

$$\lambda\langle\phi|\psi\rangle = \langle\lambda^*\phi|\psi\rangle = \langle\phi|\lambda\psi\rangle. \quad (4.2)$$

The scalar product also induces a norm

$$\|\psi\|^2 = \langle\psi|\psi\rangle \geq 0. \quad (4.3)$$

We find

$$\langle \phi | \psi \rangle^* = \langle \psi | \phi \rangle, \quad (4.4)$$

and hence the Cauchy-Schwarz inequality

$$|\langle \phi | \psi \rangle| \leq \| \phi \| \| \psi \| \quad (4.5)$$

is satisfied. Cauchy sequences with respect to this norm in  $\mathcal{H}$  are convergent.

A sequence  $\{|\psi_i\rangle\}_{i=1}^d$  is called an orthonormal basis, if

$$\langle \psi_i | \psi_j \rangle = \delta_{i,j} \quad (4.6)$$

and if every vector  $|\psi\rangle$  can be written as a norm convergence sequence

$$|\psi\rangle = \sum_i \alpha_i |\psi_i\rangle. \quad (4.7)$$

The components in turn are given by

$$\alpha_j = \langle \psi_j | \psi \rangle. \quad (4.8)$$

The number  $d$  of elements of this basis is the same for each basis and is called the dimension of the Hilbert space. This can be finite or infinite. Examples of finite-dimensional Hilbert spaces are the Hilbert spaces of a single spin-1/2 degree of freedom,

$$\mathcal{H} = \text{span}\{|0\rangle, |1\rangle\} \simeq \mathbb{C}^2, \quad (4.9)$$

where  $|0\rangle$  is associated with “spin down” and  $|1\rangle$  with “spin up”. Other finite-dimensional Hilbert spaces are

$$\mathcal{H} = \text{span}\{|0\rangle, \dots, |d-1\rangle\}, \quad (4.10)$$

of  $d$ -level systems. Position degrees of freedom are not associated with finite-dimensional Hilbert spaces, but rather infinite-dimensional ones. The familiar Hilbert spaces of a position degree of freedom of a spinless particle is

$$\mathcal{H} = \mathcal{L}^2(\mathbb{R}) = \left\{ \psi : \mathbb{R} \rightarrow \mathbb{C} : \|\psi\|^2 := \int dx |\psi(x)|^2 < \infty \right\}, \quad (4.11)$$

the Lebesgues square integrable functions.

#### 4.1.2 Observables

Quantities that are captured by functions in phase space in classical mechanics become linear operators on  $\mathcal{H}$  in quantum mechanics. That is to say, they become operators  $A : \mathcal{H} \rightarrow \mathcal{H}$  with

$$A(\lambda|\phi\rangle + \mu|\psi\rangle) = \lambda A|\phi\rangle + \mu A|\psi\rangle, \quad (4.12)$$

#### 4.1. REVIEW OF THE STRUCTURAL ELEMENTS OF QUANTUM MECHANICS 5

for  $|\psi\rangle, |\phi\rangle \in \mathcal{H}$  and  $\lambda, \mu \in \mathbb{C}$ . An operator is called bounded if there exists a constant  $\infty > c > 0$  such that

$$\|A|\phi\rangle\| \leq c\|\phi\rangle\|. \quad (4.13)$$

This set of bounded operators is usually called  $\mathcal{B}(\mathcal{H})$ . For operators  $A \in \mathcal{B}(\mathcal{H})$  one can define the adjoint operator  $A^\dagger$  with

$$\langle\phi|A^\dagger|\psi\rangle = \langle A\phi|\psi\rangle. \quad (4.14)$$

An operator is called Hermitian, if

$$A = A^\dagger. \quad (4.15)$$

Observables: The observables in quantum theory correspond to Hermitian linear operators  $A = A^\dagger$ . If a system is prepared in a pure state with state vector  $|\psi\rangle$ , then the expectation value of  $A$  is given by

$$\langle A \rangle = \langle\psi|A|\psi\rangle. \quad (4.16)$$

Examples for spin-1/2 systems of observables are the Pauli operators

$$\sigma_X = |0\rangle\langle 1| + |1\rangle\langle 0|, \quad (4.17)$$

$$\sigma_Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0|, \quad (4.18)$$

$$\sigma_Z = |0\rangle\langle 0| - |1\rangle\langle 1|. \quad (4.19)$$

It is easy to see that they are Hermitian. In Stern-Gerlach-type devices, they can be measured. An operator is called positive (or positive semi-definite), if

$$\langle\psi|A|\psi\rangle \geq 0 \quad (4.20)$$

for all  $|\psi\rangle \in \mathcal{H}$ . This is equivalent with stating that all eigenvalues of  $A$  are non-negative. Positive operators are by definition Hermitian, so the concept of positivity does not make any sense for non-Hermitian operators. Hermitian operators have real eigenvalues. Observable quantities are associated with Hermitian operators, which are also called observables.

It is often to represent operators (and vectors and dual vectors) in their matrix form. One picks a basis and represents the operators in this basis. This is strictly speaking an isomorphism. But often, the operators and their matrix forms are identified. For a basis  $\{|\psi_k\rangle\}$ , one gets the components

$$A_{j,k} = \langle\psi_j|A|\psi_k\rangle. \quad (4.21)$$

For example, for the spin-1/2 system one can take the basis  $\{|0\rangle, |1\rangle\}$ , and then represent the Pauli operators as Pauli matrices

$$\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (4.22)$$

$$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad (4.23)$$

$$\sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (4.24)$$

In general, the identity operator becomes

$$\mathbb{1} = \sum_j |\psi_j\rangle\langle\psi_j|. \quad (4.25)$$

The operations on matrices are defined in terms of the components. So

$$(AB)_{j,k} = \sum_l A_{j,l}B_{l,k}, \quad (4.26)$$

so one encounters the familiar matrix product. Also, one obviously has that

$$\text{tr}(AB) = \sum_{j,k} A_{j,k}B_{k,j} = \text{tr}(BA), \quad (4.27)$$

for the trace

$$\text{tr}(A) = \sum_j A_{j,j} = \langle\psi_j|A|\psi_j\rangle. \quad (4.28)$$

In fact, the trace is cyclic, and for arbitrary  $A, B, C$ , one has that

$$\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA). \quad (4.29)$$

In fact, we will see that the trace plays an important role when computing expectation values. The analogue of expectation values

$$\langle f \rangle_\rho = \int d\gamma \rho(\gamma) f(\gamma) \quad (4.30)$$

in classical physics becomes

$$\langle A \rangle_\rho = \text{tr}(\rho A) \quad (4.31)$$

for density operators  $\rho$  in quantum physics. What does that mean? For that purpose, we will revisit the concept of a density operator, so a general quantum state, in the next subsection.

## 4.1. REVIEW OF THE STRUCTURAL ELEMENTS OF QUANTUM MECHANICS 7

### 4.1.3 Density operators and expectation values

Let us imagine we have a single spin, associated with a Hilbert space  $\mathcal{H} \simeq \mathbb{C}^2$ . We now throw a coin. In case of heads, we prepare the spin in  $|0\rangle$ , in case of tails, we prepare it in  $|1\rangle$ . That is to say, with the classical probability  $1/2$  we have  $|0\rangle$ , and with classical probability  $1/2$  we get  $|1\rangle$ . How do we capture this situation? Can we describe the system by a state vector

$$|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}? \quad (4.32)$$

Not quite. This is easy to see: In case of a  $\sigma_x$  measurement, we would always get the same outcome. But this is different from the situation we encounter here. In fact, when we make a measurement of  $\sigma_x$ , we would get both outcomes with equal probability. Or

$$|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}? \quad (4.33)$$

Again, this will not work, for the same reason. In fact, no state vector is associated with such a situation, and for that, we need to generalize our concept of a quantum state slightly: to density operators. This is, however, the most general quantum state in standard quantum mechanics, and we will not have to generalize it any further.

In fact, the above situation is an instance of the situation where we prepare with probability  $p_j$ ,  $j = 1, \dots, n$ , a system in a state vector  $|\psi_j\rangle$ . Since we encounter a probability distribution, we have

$$\sum_{j=1}^n p_j = 1. \quad (4.34)$$

Such a situation is sometimes referred to as a mixed ensemble. How do we incorporate that?

Density operator of a pure state: A pure state associated with a state vector  $|\psi\rangle \in \mathcal{H}$  from some Hilbert space  $\mathcal{H}$  is given by the density operator

$$\rho = |\psi\rangle\langle\psi|. \quad (4.35)$$

We immediately find some properties of such an operator: We obviously have that

$$\rho = \rho^\dagger. \quad (4.36)$$

Then,

$$\text{tr}(\rho) = 1. \quad (4.37)$$

Finally, we have that

$$\rho \geq 0, \quad (4.38)$$

which means that all of its eigenvalues are non-negative, which is clearly the case, as all the eigenvalues are given by 0 or 1, clearly non-negative numbers. We also have the property that

$$\text{tr}(\rho^2) = \text{tr}(|\psi\rangle\langle\psi|\psi\rangle\langle\psi|) = \text{tr}(|\psi\rangle\langle\psi|) = \text{tr}(\rho) = 1. \quad (4.39)$$

How do we compute expectation values from such a density operator? Well, we know that for an observable

$$\langle A \rangle = \langle \psi | A | \psi \rangle. \quad (4.40)$$

This we can equally well (although at this instance in time this may look unnecessarily complicated, but we will see the point in a second) write as

$$\langle A \rangle = \langle \psi | A | \psi \rangle = \text{tr}(A | \psi \rangle \langle \psi |) = \text{tr}(A \rho). \quad (4.41)$$

We have hence made use of one of the above rules, and have written expectation values as a trace of the observable, multiplied with the density operator.

A general density operator is just extended by linearity from this definition.

Density operator of a mixed ensemble: Consider the situation of preparing  $|\psi_j\rangle$ ,  $j = 1, \dots, n$  with probability  $p_j$ . This is associated with a density operator

$$\rho = \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j|. \quad (4.42)$$

How do we compute expectation values from that? We have for observables  $A$  that

$$\langle A \rangle = \sum_{j=1}^n p_j \langle \psi_j | A | \psi_j \rangle, \quad (4.43)$$

from the very definition of a mixed ensemble. This we can, however, also write as

$$\begin{aligned} \langle A \rangle &= \sum_{j=1}^n p_j \text{tr}(A |\psi_j\rangle\langle\psi_j|) \\ &= \text{tr} \left( A \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j| \right) \\ &= \text{tr}(A \rho). \end{aligned} \quad (4.44)$$

So again, expectation values are just computable as the trace of the density operator multiplied with the observable.



#### 4.1. REVIEW OF THE STRUCTURAL ELEMENTS OF QUANTUM MECHANICS 9

We now once more investigate properties of such a density operator: We find that again,

$$\rho = \rho^\dagger. \quad (4.45)$$

In the same fashion as before, since now

$$\sum_{j=1}^n p_j = 1, \quad (4.46)$$

we also have that

$$\text{tr}(\rho) = 1. \quad (4.47)$$

Finally, we have that

$$\rho \geq 0, \quad (4.48)$$

since all of the probabilities are positive, and a sum of positive operators is positive. These are exactly the same properties as above, except from one that is now missing. We no longer have

$$\text{tr}(\rho^2) = 1. \quad (4.49)$$

In fact, this property is replaced by

$$\begin{aligned} \text{tr}(\rho^2) &= \text{tr} \left( \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j| \right) \left( \sum_{k=1}^n p_k |\psi_k\rangle\langle\psi_k| \right) \\ &= \sum_{j,k=1}^n p_j p_k \text{tr}(|\psi_j\rangle\langle\psi_j| \psi_k\rangle\langle\psi_k|) \\ &\leq 1, \end{aligned} \quad (4.50)$$

where we have bounded the scalar products between two arbitrary state vectors. We have now arrived at the most general concept of a state in (standard) quantum mechanics. This is surely worth a box:

Density operators: General states of quantum systems with Hilbert space  $\mathcal{H}$  are given by density operators  $\rho$ . Their properties are

$$\rho = \rho^\dagger \text{ (Hermicity),} \quad (4.51)$$

$$\rho \geq 0 \text{ (Positivity),} \quad (4.52)$$

$$\text{tr}(\rho) = 1 \text{ (Normalization).} \quad (4.53)$$

Pure states are those density operators for which

$$\text{tr}(\rho) = 1, \quad (4.54)$$

those can be represented by state vectors  $|\psi\rangle \in \mathcal{H}$  as

$$\rho = |\psi\rangle\langle\psi|. \quad (4.55)$$

Otherwise, if  $\text{tr}(\rho^2) < 1$ , the state is called mixed. For observables, expectation values are computed as

$$\langle A \rangle = \text{tr}(A\rho). \quad (4.56)$$

This is a good moment to discuss a number of examples. Let us go back to our initial situation discussed at the beginning of the chapter, of preparing  $|0\rangle$  or  $|1\rangle$  with equal probability. We can now easily associate this with a density operator

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \quad (4.57)$$

We can write this in matrix form – remember that operators and their matrix representation are identified with each other throughout the script

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \quad (4.58)$$

We have that

$$\text{tr}(\rho^2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1. \quad (4.59)$$

This in fact the minimum value  $\text{tr}(\rho^2)$  can take for a system with  $\mathcal{H} \simeq \mathbb{C}^2$ . The pure state  $\rho = |0\rangle\langle 0|$  in turn is represented as

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad (4.60)$$

obviously satisfying  $\text{tr}(\rho) = 1$ . Generally, if we have probabilities  $p_0$  and  $p_1$  to prepare  $|0\rangle$  and  $|1\rangle$ , we have the density operator

$$\rho = \begin{bmatrix} p_0 & 0 \\ 0 & p_1 \end{bmatrix}. \quad (4.61)$$

#### 4.1. REVIEW OF THE STRUCTURAL ELEMENTS OF QUANTUM MECHANICS 11

But of course, we are not forced to take the standard basis. The situation of having prepared  $|+\rangle$  and  $|-\rangle$  with equal probabilities is captured as

$$\rho = \frac{1}{2}|+\rangle\langle+| + \frac{1}{2}|-\rangle\langle-|. \quad (4.62)$$

This is

$$\begin{aligned} \rho &= \frac{1}{4}((|0\rangle + |1\rangle)(\langle 0| + \langle 1|)) + \frac{1}{4}((|0\rangle - |1\rangle)(\langle 0| - \langle 1|)) \\ &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|, \end{aligned} \quad (4.63)$$

with matrix representation

$$\rho = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \quad (4.64)$$

Eh, wait a minute? Was this not the matrix representation of the ensemble consisting of  $|0\rangle$  and  $|1\rangle$ ? So we are faced here with the irritating situation that two different mixed ensembles are described by the same density operator. Let us face it: There are many different ways of preparing the same density operator! Since all expectation values of observables are computed as

$$\langle A \rangle = \text{tr}(A\rho), \quad (4.65)$$

we get exactly same value for all observables in case of

$$\rho = \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j| = \sum_{k=1}^m q_k |\phi_k\rangle\langle\phi_k|, \quad (4.66)$$

even if all of the probabilities  $\{p_j\}$  and  $\{q_k\}$  as well as all state vectors  $\{|\psi_j\rangle\}$  and  $\{|\phi_k\rangle\}$  are different. In fact, now even  $n = m$  has to hold. What matters for all outcomes in all experiments is the density operator, not the mixed ensemble we have started with.

The situation is hence quite subtle: Yes, a density operator is a concept that allows to introduce the concept of classical probability distributions into quantum mechanics. We have to have a way of incorporating probabilistic preparation, where we do one thing with some probability and another with another. In statistical physics, we will encounter such a situation frequently, where density operators are ubiquitous, to say the least.

But no, once we arrive at a given density operator, there is no way to reconstruct the mixed ensemble that can be held responsible for the density operator. In retrospect, there always would have been infinitely many other ways of preparing the same density operator (unless it is a pure state). Sometimes, people use notions of the kind, “the system is in some pure state vector  $|\psi_j\rangle$ ,  $j = 1, \dots, n$ , we simply do not know which one”. Such reasoning is not quite precise and can be plain wrong, in which case it is referred to as preferred ensemble fallacy.

## 4.1.4 Time evolution

Now that we have understood what a density operator is, the rest will be a piece of cake. How do density operators evolve in time? Well, this equation is just inherited from the Schroedinger equation by linearity. Since it was von-Neumann who first described this situation well, it is called von-Neumann equation. But really, it is just the ordinary Schroedinger equation written for density operators. There is no new physics happening here.

Von Neumann equation: Density operators of physical systems described by Hamiltonians  $H$  evolve in time according to

$$i\hbar \frac{d}{dt} \rho(t) = [H, \rho(t)]. \quad (4.67)$$

So together with the initial condition at  $\rho(0)$ , this differential equation fully specifies the density operator at a later time, exactly as the Schroedinger equation does for state vectors. Is this equation mysterious? Not at all: Let us write

$$\rho(t) = \sum_{j=1}^n p_j |\psi_j(t)\rangle \langle \psi_j(t)|, \quad (4.68)$$

then

$$\begin{aligned} i\hbar \frac{d}{dt} \rho(t) &= i\hbar \sum_{j=1}^n p_j \frac{d}{dt} |\psi_j(t)\rangle \langle \psi_j(t)| \\ &= i\hbar \sum_{j=1}^n p_j \left( \left( \frac{d}{dt} |\psi_j(t)\rangle \right) \langle \psi_j(t)| + |\psi_j(t)\rangle \left( \frac{d}{dt} \langle \psi_j(t)| \right) \right) \\ &= \sum_{j=1}^n p_j (H |\psi_j(t)\rangle \langle \psi_j(t)| - |\psi_j(t)\rangle \langle \psi_j(t)| H) \\ &= [H, \rho(t)]. \end{aligned} \quad (4.69)$$

Of course we can again write the non-differential form of time evolution:

Time evolution in terms of the time evolution operator: We have that

$$\rho(t) = U_t \rho(0) U_t^\dagger, \quad (4.70)$$

with  $U_t = e^{-iHt}$  being the time evolution operator for times  $t \geq 0$ .

## 4.1.5 Composite systems

The last ingredient that we still need is the concept of a composite quantum system. If we have a system that is composed of two spins, we already encounter a composite system. Or a particle that has three motional degrees of freedom. Or a particle with a spin. These are simple examples. But obviously, in statistical physics we have the situation in mind where we have very many similar or identical systems at hand. The mathematical object that reflects this situation is the one of the tensor product.

Let us first have a look at the situation for two parts only, corresponding to Hilbert spaces  $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ , where  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are the Hilbert spaces of the two parts, respectively. The way to define the tensor product is as follows. We first consider the vectors, the “pure tensors”,

$$|\psi_1\rangle \otimes |\psi_2\rangle \in \mathcal{H}, \quad (4.71)$$

for which  $|\psi_1\rangle \in \mathcal{H}_1$  and  $|\psi_2\rangle \in \mathcal{H}_2$ . Since  $\mathcal{H}$  is a vector space, it of course contains all linear combinations of such vectors, and as a Hilbert space also the norm limits. The scalar product, and hence the norm, only have to be defined on these pure tensors, as

$$\langle \phi_1 | \otimes \langle \phi_2 | \psi_1 \rangle \otimes |\psi_2\rangle = \langle \phi_1 | \psi_1 \rangle \langle \phi_2 | \psi_2 \rangle. \quad (4.72)$$

This defines the tensor product. When  $A_1$  and  $A_2$  are linear operators on  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , then one has

$$(A_1 \otimes A_2) |\psi_1\rangle \otimes |\psi_2\rangle = (A_1 |\psi_1\rangle) \otimes (A_2 |\psi_2\rangle), \quad (4.73)$$

defined a tensor product of linear operators, by considering the linear extension. Of course,

$$(A_1 \otimes A_2)(B_1 \otimes B_2) = (A_1 B_1) \otimes (A_2 B_2). \quad (4.74)$$

We remember that parts were considered statistically independent, if the density was of the form

$$\rho(\gamma_1, \gamma_2) = \rho_1(\gamma_1) \rho_2(\gamma_2). \quad (4.75)$$

The quantum mechanical analogue is a state of the form

$$\rho = \rho_1 \otimes \rho_2. \quad (4.76)$$

We give this situation a box:

Product states: Product states are states of the form

$$\rho = \rho_1 \otimes \rho_2. \quad (4.77)$$

They reflect an independent preparation of two parts.

Of course, we cannot only consider the composition of parts. We always have the option to only look at a part. Classically, we know what that means. That was related to the marginal distribution, reflecting “disregarding” a part. We had that the marginal distribution was given by

$$(R_1\rho)(\gamma_1) = \int d\gamma_2 \rho(\gamma_1, \gamma_2). \quad (4.78)$$

Of course, if the distribution of the total system was pointlike and we “knew everthing of the system”, the same was true for any marginal distribution. This is actually quite intuitive.

Of course, we can also quantum mechanically define the analogue of the marginal distribution, called the reduced state: It is defined as

$$\text{tr}((R_1\rho)A) = \text{tr}(\rho(A \otimes \mathbb{1})), \quad (4.79)$$

for observables  $A$  that are only supported on the first tensor factor.  $R_1\rho$  is called the reduced state or reduced density operator. In matrix components, one has

$$(R_1\rho)_{j,k} = \sum_n \rho_{j,n;k,n}. \quad (4.80)$$

So one fixes a basis in the part one disregards, and takes the partial trace over this basis. The specific choice of basis does not matter. As such, this is a very straightforward procedure. There is only one subtlety encountered here: Even if a joint system is in a pure state, the reduced state does not have to be!

Entangled pure states: Pure states that have the property that reduced states are mixed are called entangled pure states.

This is quite remarkable. So even if we had a pure preparation and know that a joint system is in a pure state, there is no way one can consistently assign a state vector to any of the parts. For example, consider two spins, in a pure state  $\rho = |\psi\rangle\langle\psi|$  with

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle). \quad (4.81)$$

Then

$$R_1\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \quad (4.82)$$

So in fact, the reduced state is not only mixed, but in fact maximally mixed. We come back to that remarkable feature later. Entanglement manifests itself in correlations that are in a sense “stronger” than classically attainable. Also, the proof that quantum mechanics cannot be captured as a classical statistical theory is based on entangled states of this kind.

## 4.1. REVIEW OF THE STRUCTURAL ELEMENTS OF QUANTUM MECHANICS 15

### 4.1.6 A translation table

We are now in the position to recapitulate what we have learned. In this subsection, we summarise the state concepts in classical and quantum statistical mechanics, which have been and will be used from now on. Of course, this table is not only applicable to our statistical physics concept at hand: It generally summarises the relevant concepts in classical and quantum physics.

	Classical physics	Quantum physics
Configuration space	Phase space $\Gamma$ , e.g., $\Gamma = \mathbb{R}^{6N}$	Hilbert space $\mathcal{H}$ e.g., $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^{3N})$
Coordinates	Canonical coords. $\mathbf{p}_1, \dots, \mathbf{q}_N$	Orthonormal basis
State	Density function $\rho \geq 0, \int d\gamma \rho(\gamma) = 1$	Density operator $\rho \geq 0, \text{tr}(\rho) = 1$
Property	Function $f : \Gamma \rightarrow \mathbb{R}$	Hermitian operator $A = A^\dagger$
Expectation value	Integral of a function $\langle f \rangle_\rho = \int d\gamma f(\gamma) \rho(\gamma)$	Trace of an operator $\langle A \rangle_\rho = \text{tr}(\rho A)$
Pure states	$\rho(\gamma) = \delta(\gamma - \gamma_0)$	$\rho =  \psi\rangle\langle\psi $
One particle	$\Gamma = \{\mathbf{p}, \mathbf{q} : \mathbf{p}, \mathbf{q} \in \mathbb{R}^3\}$	$\mathcal{H} = \mathcal{L}^2(\mathbb{R}^3)$
Composition	$\Gamma = \Gamma_1 \times \Gamma_2$	$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$
Observable on one part	$f(\gamma_1, \gamma_2) = f_1(\gamma_1)$	$A( \phi_1\rangle \otimes  \phi_1\rangle) = (A_1 \phi_1\rangle) \otimes  \phi_2\rangle$ $A = A_1 \otimes \mathbb{1}$
Product observable	$(f_1 \times f_2)(\gamma_1, \gamma_2) = f_1(\gamma_1)f_2(\gamma_2)$	$(A_1 \otimes A_2)( \phi_1\rangle \otimes  \phi_1\rangle)$ $= (A_1 \phi_1\rangle) \otimes A_2 \phi_2\rangle$
Independent prep.	$\rho(\gamma_1, \gamma_2) = \rho_1(\gamma_1)\rho_2(\gamma_2)$	$\rho = \rho_1 \otimes \rho_2$
Reduced description	Marginal distribution $\rho_1(\gamma_1) = \int d\gamma_2 \rho(\gamma_1, \gamma_2)$	Reduced state $\rho_1 = \text{tr}_2(\rho)$
Hamiltonian	Hamiltonian function	Hamilton operator
Dynamics	$\dot{\rho}_t = \{\rho_t, H\}$ $\dot{p}_i = -\frac{\partial H}{\partial q_i}, \dot{q}_i = \frac{\partial H}{\partial p_i}$ $\rho_t(\gamma) = \rho(\mathcal{F}_{-t}(\gamma))$	$\dot{\rho}_t = i[\rho, H]$ $i\dot{ \psi\rangle}(t) = H \psi(t)\rangle$ $U_t = e^{-iHt}$
Independent dynamics	$H(\gamma_1, \gamma_2) = H_1(\gamma_1) + H_2(\gamma_2)$	$H = H_1 \otimes \mathbb{1} + \mathbb{1} \otimes H_2$

This translation table should be quite helpful from now on, when we look at further implications and applications of quantum and classical statistical physics.

## 4.2 Quantum mechanical ensembles

### 4.2.1 Microcanonical ensemble

This will be quick: We have all the tools at hand. In fact, we will define the micro canonical ensemble just in the analogous way compared to the classical situation. We write the spectral decomposition of the Hamilton operator as

$$H = \sum_{j=1}^n E_j |E_j\rangle\langle E_j|, \quad (4.83)$$

assuming that the spectrum of the Hamiltonian has no continuous parts; otherwise, one has to replace this expression by spectral projectors.

Microcanonical ensemble: For  $E > 0$  and  $\epsilon > 0$ , the micro canonical ensemble is given by

$$\rho(E) = Z(E)^{-1} \sum_{E-\epsilon \leq E_j \leq E} |E_j\rangle\langle E_j|. \quad (4.84)$$

$Z(E)$  is called the micro canonical partition function.

### 4.2.2 Canonical ensemble

Similarly, we can define the canonical ensemble, belonging to situations at some given inverse temperature  $\beta > 0$ .

Canonical ensemble: For  $\beta > 0$ , the canonical ensemble is given by

$$\rho_\beta = Z^{-1} e^{-\beta H}. \quad (4.85)$$

This state is also called thermal state or Gibbs state. Here,

$$Z = \text{tr}(e^{-\beta H}) \quad (4.86)$$

is the canonical partition function.

We encounter here a matrix function. Generally, for a Hermitian matrix  $A$ , we can diagonalise it as

$$A = UDU^\dagger, \quad (4.87)$$



with  $U$  unitary and  $D$  being diagonal and real. For a real-valued function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , we can now define

$$f(A) = Uf(D)U^\dagger. \quad (4.88)$$

In terms of the above spectral decomposition of  $H$ , the canonical ensemble hence becomes

$$\rho = Z^{-1} \sum_j e^{-\beta E_j} |E_j\rangle\langle E_j|, \quad (4.89)$$

and

$$Z = \sum_j e^{-\beta E_j}. \quad (4.90)$$

This expression looks very similar to the classical one.

Both ensembles are invariant under time evolution, since any matrix function of  $H$  clearly commutes with  $H$ . We have for the canonical ensemble

$$e^{-iHt} \rho_\beta e^{iHt} = \rho_\beta, \quad (4.91)$$

and the same expression for the micro canonical ensemble. They are hence preserved in time.

We are also already in the position to formulate the main equation that allows to derive thermostatic properties from microscopic ensembles. We can identify the free energy with

$$F = -\frac{1}{\beta} \log \text{tr}(e^{-\beta H}), \quad (4.92)$$

from which can deduce all thermostatic properties.

### 4.2.3 Gibbs variational principle revisited

Let us introduce the functionals

$$F(H) = -\frac{1}{\beta} \log \text{tr}(e^{-\beta H}), \quad (4.93)$$

$$\tilde{S}(\rho) = -\text{tr}(\rho \log \rho). \quad (4.94)$$

Here, again, this is to be read as a matrix function. In analogy with the classical case, we expect an expression like

$$F(H) = \inf_{\rho} \left\{ \text{tr}(\rho H) - \frac{1}{\beta} \tilde{S}(\rho) \right\}. \quad (4.95)$$

And indeed, as before, the Gibbs or thermal state of the canonical ensemble assumes that minimum:

Quantum mechanical Gibbs variational principle: Among all states, the Gibbs state  $\rho_\beta$  of the canonical ensemble minimises  $\text{tr}(\rho H) - \tilde{S}(\rho)/\beta$ .

So as before, the canonical ensemble minimises the expression for the free energy. This was already true for classical densities, and we encountered a similar formula in the thermodynamic setting. In a similar before, we indeed find that the above pair is a pair of Legendre transforms, so we have

$$\tilde{S}(\rho) = \beta \inf_H \{ \text{tr}(\rho H) - F(H) \}. \quad (4.96)$$

#### 4.2.4 Von Neumann entropy

Let us spend some time with the entropy as such the entropy we just encountered, referred to as von Neumann entropy. Clearly, if we write the spectral decomposition of a state  $\rho$  as

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|, \quad (4.97)$$

we find that the entropy can be written as follows:

Von Neumann entropy:

$$\tilde{S}(\rho) = \text{tr}(\rho \log \rho) = - \sum_j p_j \log(p_j). \quad (4.98)$$

In other words, the von Neumann entropy is nothing but the information theoretic Shannon entropy for the eigenvalues of  $\rho$ . We will come back to that point later. The function

$$x \mapsto -x \log x \quad (4.99)$$

is concave, and indeed, even the von-Neumann entropy  $S$  is even concave on states. For pure states  $\rho$ , only one of the  $\{p_j\}$  equals to one, while all other values are zero, so without loss of generality  $p_1 = 1$ ,  $p_j = 0$  for  $j > 1$ . Hence, for pure states  $\rho$ , the entropy vanishes,

$$\tilde{S}(\rho) = 0. \quad (4.100)$$

For all other states, one has  $S(\rho) > 0$ . This is easy to see: From  $S(\rho) > 0$  it follows that all spectral values must be  $p_j \in \{0, 1\}$ . Because of

$$\sum_j p_j = 1, \quad (4.101)$$

however, only exactly one eigenvalue can be equal to 1. As mentioned before, the entropy is also concave, so for any states  $\rho_1, \rho_2$  and any  $\lambda \in [0, 1]$ , one has that

$$\tilde{S}(\lambda\rho_1 + (1 - \lambda)\rho_2) \geq \lambda\tilde{S}(\rho_1) + (1 - \lambda)\tilde{S}(\rho_2). \quad (4.102)$$

In other words, we have the following: Mixing, so convex combination of two states, can only enlarge the entropy, compared to the weighted sum of the entropies of the parts.

How does the von Neumann entropy behave under tensor products? Let us assume we have a state  $\rho^{(1)} \otimes \rho^{(2)}$ , reflecting totally independent preparations. Then it is easy to see that

$$\tilde{S}(\rho^{(1)} \otimes \rho^{(2)}) = \tilde{S}(\rho^{(1)}) + \tilde{S}(\rho^{(2)}). \quad (4.103)$$

That is, the von Neumann entropy is additive, or an extensive quantity. In fact, also the free energy is an additive quantity. If the parts do not interact, so if the Hamiltonian is of the form

$$H = H_1 \otimes \mathbb{1} + \mathbb{1} \otimes H_2, \quad (4.104)$$

then

$$e^{-\beta H} = e^{-\beta H_1} \otimes e^{-\beta H_2}, \quad (4.105)$$

and therefore,

$$\begin{aligned} F(H_1 \otimes \mathbb{1} + \mathbb{1} \otimes H_2) &= -\frac{1}{\beta} \log (\operatorname{tr} e^{-\beta H_1} \operatorname{tr} e^{-\beta H_2}) \\ &= F(H_1) + F(H_2). \end{aligned} \quad (4.106)$$

In a way, one can say the following. This is not meant to be a rigorous statement, but rather a general guideline.

Statistical physics of uncoupled systems: “The statistical physics of uncoupled systems is in many ways essentially that of single systems, up to simple sums and products.”

Finally, there is another thing one can show: If  $\rho_{\beta_1}$  and  $\rho_{\beta_2}$  are two Gibbs states of the canonical ensemble, with  $\beta_2 > \beta_1$ , then

$$\tilde{S}(\rho_1) > \tilde{S}(\rho_2). \quad (4.107)$$

In other words, the von Neumann entropy is a monotonous function in the temperature. This is again a manifestation of the von Neumann entropy quantifying mixedness: If a state is pure, it has vanishing entropy. The higher the temperature becomes, the “more mixed” is the state.

### 4.3 Third law of thermodynamics

Although this may be a bit late, we for a moment go back to thermodynamics. This, however, for a good reason. Before we entered the quantum description, there was hardly a way to motivate this last law of thermodynamics. Now there

is. We remember that the (thermodynamic) entropy was only determined after fixing the absolute temperature scale up to an additive constant. This was an undesired feature: In fact, it was in these additive constants that gave information about whether a system in a reaction would consume or deliver energy. In the context of phenomenological thermodynamics – so actually independent of statistical physics – the question emerged what happens at low temperatures with the entropy. Nernst postulated, that for

$$T \rightarrow 0 \tag{4.108}$$

the entropy converges to a constant value which is independent of all intensive state variables (such as the pressure  $p$ ). This postulate is usually called the third law of thermodynamics. As a convention, one can set this constant value to zero, and this is usually done. In this way, one has eliminated the last arbitrary aspect from the concept of the absolute entropy:

Third law of thermodynamics: For  $T \rightarrow 0$  the absolute entropy converges to 0.

Nernst would not have suggested that based on classical mechanics. In fact, for the ideal gas,  $S$  converges for  $T \rightarrow 0$  to  $-\infty$  rather than to 0. For  $\beta \rightarrow \infty$  the function  $\gamma \mapsto e^{-\beta H(\gamma)}$  is more and more concentrated to the points in which  $H$  takes its absolute minimum  $\lambda_{\min}$ . Since the canonical ensemble is normalized, it actually does not matter at which point this minimum is assumed. All  $\gamma$ , for which

$$H(\gamma) > \lambda_{\min} + \epsilon \tag{4.109}$$

for some  $\epsilon > 0$  are suppressed at least by a factor of  $e^{-\beta\epsilon}$ . In the region of small  $H(\gamma)$ , where the canonical ensemble  $\rho_\beta$  is focused, all of the function values  $\log \rho_\beta(\gamma)$  are large as well. Hence, the integral diverges to  $-\infty$ . Hence, within the validity of classical mechanics, the third law of thermodynamics is not valid.

How is the situation in a quantum mechanical description? There is one thing that makes us certain that the situation is different here: We have

$$\tilde{S}(\rho) \geq 0 \tag{4.110}$$

for all states  $\rho$ . Hence, the entropy cannot diverge to  $-\infty$ . Since we already know that the von Neumann entropy is monotonous, we find that it converges to a finite value. But let us check whether really

$$\lim_{\beta \rightarrow \infty} \tilde{S}(\rho_\beta) = 0 \tag{4.111}$$

holds true. For that purpose, we have to again look at  $\rho_\beta$  for large  $\beta$ .

Again, of course, only the lower energy values matter. Let us write

$$H = \sum_j E_j P_j, \tag{4.112}$$

for  $E_0 < E_1 < E_2 < \dots$ , where  $P_j$  are the projections onto the eigenspaces with eigenvalue  $E_j$ . If the space is degenerate, then this is captured by this projector, and the respective projector is no longer one-dimensional. In these terms, we can write the canonical ensemble

$$\begin{aligned} \rho_\beta &= \frac{\sum_j e^{-\beta(E_j - E_0)} P_j}{\sum_j e^{-\beta(E_j - E_0)} \text{tr}(P_j)} \\ &= \mathcal{N} \left( P_0 + e^{-\beta(E_1 - E_0)} \sum_{j \geq 1} e^{-\beta(E_j - E_1)} P_j \right), \end{aligned} \quad (4.113)$$

with  $\mathcal{N}$  denoting normalization. The sum in the last expression converges for all  $\beta$  and is a monotone falling function of  $\beta$ . Together with  $e^{-\beta(E_1 - E_0)}$  the second term goes to zero, exponentially quickly. The same happens for the normalization  $\mathcal{N}$ , so that

$$\lim_{\beta \rightarrow \infty} \rho_\beta = \frac{P_0}{\text{tr}(P_0)} = \rho_\infty. \quad (4.114)$$

Therefore,

$$\lim_{\beta \rightarrow \infty} \tilde{S}(\rho_\beta) = \tilde{S}(\rho_\infty) = \log \text{tr}(P_0). \quad (4.115)$$

Now  $\text{tr}(P_0)$  is just the degeneracy of the ground state of  $H$ . Usually, this ground state is non-degenerate, so that indeed,

$$\lim_{\beta \rightarrow \infty} \tilde{S}(\rho_\beta) = 0. \quad (4.116)$$

But even if the ground state is degenerate, the degeneracy is usually small, and does not grow linearly with the system size. In the thermodynamic limit, the entropy density hence still goes to zero.