Mathematical QM - Lecture 6

Armin Scrinzi

February 4, 2020

Contents

1 States ≈ Density matrices

We had introduced general notion of a states on a C^* algebra of observables as the normalized positive linear functionals $\omega : \omega(B^*B) \geq 0$. In any Hilbert space representation states appear as *density matrices*: positive linear operators $\widehat{\rho}$ with $\mathrm{Tr}\widehat{\rho} = 1$ and

$$\omega(B) = \mathrm{Tr}\widehat{\rho}B. \tag{1}$$

We will make this statement more mathematically precise below. As $\widehat{\rho}$ is positive, we have a spectral representation. As it has finite trace, it is a fortiori compact and therefore has a purely discrete spectrum with a spectral representation as

$$\widehat{\rho} = \sum_i |i\rangle \rho_i \langle i| \tag{2}$$

The physical interpretation is that we have incomplete knowledge about many copies of the same system and that ρ_i is the relative frequency for finding a system picked from the ensemble in the state $|i\rangle$. It is important to remember the difference to a superposition state

$$|s\rangle = \sum_i e^{i\alpha}\sqrt{\rho_i}|i\rangle : \tag{3}$$

this describes a single system that is fully determined and whose density matrix is $\widehat{\rho}_s = |s\rangle\langle s|$.

For clarity, we will denote density *matrices* as $\widehat{\rho}$ to be distinguished from their eigenvalues

denoted as ρ_i.

1.1 Physics interpretation

Suppose we want to describe a system, about which we only know the probabilities ρ_i of finding it in one of the orthogonal states $|i\rangle$, but the system is not further determined. The density matrix $\widehat{\rho} = \sum_i |i\rangle\rho_i\langle i|$ exactly describes this situation

$$P[i] = \mathrm{Tr}\,|i\rangle\langle i|\widehat{\rho} = \rho_i. \tag{4}$$

If the introduction of the density matrix above appears somewhat *ad hoc*, it is useful to see that we can recover it from the more standard description in terms of vectors in a Hilbert space: the density matrix describes a subsystem of a larger system, if we restrict our observables to that sub-system. Restricting the observables to a subsystem means that we never measure anything outside the sub-system. We only require that the outside part exists. The total Hilbert space composed of the Hilbert space of the sub-system $\mathcal{H}_s$ and the outside world $\mathcal{H}_o$: $\mathcal{H} = \mathcal{H}_s \otimes \mathcal{H}_o$. Observables, where we do not observe anything $\mathcal{H}_o$ have the form

$$A = A_s \otimes \mathbf{1}. \tag{5}$$

Suppose we represent a given density matrix $\widehat{\rho}$ for $\mathcal{B}(\mathcal{H}_s)$ in diagonalized form

$$\widehat{\rho} = \sum_i |\Phi_i\rangle_s \rho_i \langle\Phi_i|_s, \tag{6}$$

emphasizing that the $|\Phi_i\rangle_s$ are vectors of the sub-system $\mathcal{H}_s$. This $\widehat{\rho}$ corresponds to a single state vector $|\Psi\rangle_\rho \in \mathcal{H}$

$$|\Psi\rangle_\rho = \sum_i \sqrt{\rho_i}|\Phi_i\rangle_s \otimes |x_i\rangle_o, \tag{7}$$

with an arbitrary orthonormal $|x_i\rangle_o \in \mathcal{H}_o$. Expectation values for any A are

$$\begin{aligned} \langle\Psi_\rho|A|\Psi_\rho\rangle &= \sum_{ij} \sqrt{\rho_i\rho_j}\langle\Phi_i| \otimes \langle x_i|(A\otimes 1)|\Phi_j\rangle_s \otimes |x_j\rangle_o & (8)\\ &= \sum_{ij} \sqrt{\rho_i\rho_j}\langle\Phi_i|A|\Phi_j\rangle \underbrace{\langle x_i|1|x_j\rangle_o}_{=\delta_{ij}} = \operatorname{Tr}\widehat{\rho}\widehat{A}. & (9)\end{aligned}$$

The density matrix can be viewed as a convenient way of describing observations that are restricted to a subsystem of a larger entity, where there are entanglements that we do not keep track of (as in the sum above). This accommodates also the situation where we have many repeated preparations of the same system, whose phase relations to each other we do not keep track of.

1.2 A mathematical definition and a few properties of $\widehat{\rho}$

It turns out that the definition of state as a normalized, positive linear functional on $\mathcal{B}(\mathcal{H})$ comprises more than just density matrices. In order to narrow the definition to density matrices we impose one more (legitimate) constraint on the linear functional, by supplementing the continuity requirement to make the state compatible with the formation of a supremum in certain subsets of the $\mathcal{B}(\mathcal{H})$ as follows:

Definition 1. *Ascending filter $F \subset \mathcal{B}(\mathcal{H})$ is a subset with bounded norm $||f|| < C\,\forall f \in F$ where for any two elements a, b there is a third one c with $a \leq c$ and $b \leq c$. The supremum $\sup F \in \mathcal{B}(\mathcal{H})$ is the smallest element such that $a \leq \sup F\ \forall a \in F$.*

Definition 2. *A state ω is* **normal** *if $\sup_{a\in F}\omega(a) = \omega(\sup F)$.*

To understand the purpose of such a definition, consider the following example: define a set of operators diagonal operators $v^{(n)}$ with the diagonals

$$(v_i^{(n)}) = (1, 1, \ldots, \underbrace{1}_{n'\text{th place}}, 0, 0, \ldots) \tag{10}$$

These obviously form an ascending filter. Define the state

$$\omega(v) = \lim_{i\to\infty} v_i \quad \text{for } v = (v_0, v_1, v_2, \ldots). \tag{11}$$

As a positive linear funtional, we know it is continuous, i.e. a state. But

$$\omega(v^{(n)}) = 0 \quad \forall n, \quad \sup(v^{(n)}) = (\text{ all 1's}), \quad \omega(\sup v) = 1 \neq \sup(\omega(v^{(n)})). \tag{12}$$

Theorem 1. *For a state ω on $\mathcal{B}(\mathcal{H})$ following two statements are equivalent*

1. ω is normal

2. ω is given by a **density matrix** $\widehat{\rho}$*:* $\omega(a) = Tr\widehat{\rho}a, \widehat{\rho} \geq 0, Tr\widehat{\rho} = 1$.

The proof $2 \to 1$ involves some reasoning with various shades of topologies, $1 \to 2$ is non-elementary.

1.3 Comparing density matrices

Our physical interpretation of $\widehat{\rho}$ is that we have only part of the complete possible knowledge of a system. In some cases we can compare two density matrices $\widehat{\rho}_1$ and $\widehat{\rho}_2$ and say whether one represents more knowledge than the other. Think of a system that can occur only in a finite set of N different quantum states $|i\rangle$ and consider the density matrix

$$\widehat{\rho}_x = \frac{1}{N}\sum_{i=1}^{N} |i\rangle\langle i|, \qquad \rho_i \equiv \frac{1}{N}, \tag{13}$$

which tells us that we know "nothing" about a given system other that it exists. The knowledge that is exists in only N states is *a priori*.

One should pause briefly, as it appears that we "know" that the system is equally likely in any of the states that it can assume at all. This "knowledge", however, is rather a concept connected to likelihood in general. The concept of likelihood is always based on instances or cases that are all equally likely: the faces of dice, the points in phase space, the states of a quantum system. Any deviation from this equal distribution is information that we have about our specific system.

In contrast the density matrix

$$\widehat{\rho}_{i_0} = \delta_{ii_0} \tag{14}$$

tells us all we can possibly know about the system, namely, that it is in a given state $|i_0\rangle$. It appears legitimate to try to sort the matrices by how little they tell us about the system, by the following symbol:

$$\widehat{\rho}_{i_0} \preceq \widehat{\rho}_x : \tag{15}$$

here $\widehat{\rho}_{i_0}$ is"less chaotic" than $\widehat{\rho}_x$, $\widehat{\rho}_x$ is more mixed than $\widehat{\rho}_{i_0}$. We generalize the above idea above by sorting the eigenvalues in decreasing order $\widehat{\rho}_1 \geq \widehat{\rho}_2 \geq \widehat{\rho}_3 \geq \ldots$ and considering the sums the first n eigenvalues for all n:

$$\rho(n) := \sum_{i=1}^{n} \rho_i. \tag{16}$$

Definition 3. *A density matrix $\tilde{\widehat{\rho}}$ is said to be* **more mixed** *(or "more chaotic") than $\widehat{\rho} : \widehat{\rho} \preceq \tilde{\rho}$, if*

$$\rho(n) \geq \tilde{\rho}(n) \quad \forall n. \tag{17}$$

One may picture this definition as that $\tilde{\rho}$ is more chaotic than $\widehat{\rho}$ if it consistently spread out over more states. As $\sum_i \rho_i = 1$, this implies that the eigenvalues $\tilde{\rho}_i$ of more chaotic state occupy a narrower interval $\subseteq [0, 1]$. Eigenvalues of the pure state $\widehat{\rho}_{i_0}$, $(1, 0, 0, \ldots)$ touch both ends of the interval, whereas for $\widehat{\rho}_x$: $(1/N, 1/N, 1/N, \ldots)$ all eigenvalues

cluster in a single point. You might say the more chaotic state has less "contrast" among its values ρ_i. One purpose of entropy it to connect a measure to this notion.

As always with vectorial objects, the ordering cannot be total, but only partial: the majority of density matrices cannot be sorted by their information content, but we *can* construct, e.g., sequences of density matrices with ever increasing chaos.

2 Entropy

Entropy is a measure about the fuzziness of our knowledge about the system, i.e. the distribution of the ρ_i between the extremes (0 entropy) = (0 fuzziness) = (perfect knowledge) = $\rho_i = \delta_{ii_0}$ = pure state and (maximal entropy) = (maximal fuzziness) = (no nontrivial knowledge) = $\rho_i = \rho_j \forall i, j$

By general desired properties as well as motivated through a probability interpretation one arrives a the von Neumann - entropy

$$S(\widehat{\rho}) = -\mathrm{Tr}\widehat{\rho}\log\widehat{\rho} = -\sum_i \rho_i \log \rho_i \tag{18}$$

2.1 Desired properties of entropy

2.1.1 Monotonicity w.r.t. $\preceq$

For our discussion of the time-evolution it was only important to observe that if the density matrix were to evolve from less pure to more pure, entropy decreases. Having defined an idea of "less pure" (or "more chaotic") by the partial ordering $\preceq$ we require

$$\widehat{\rho}_1 \preceq \widehat{\rho}_2 \Rightarrow S(\widehat{\rho}_1) \leq S(\widehat{\rho}_2). \tag{19}$$

This would be delivered by the trace of any concave function of the density matrix (or any monotonic function of that number). One can grasp this intuitively as follows: picture the order $\preceq$ as saying that, if a density matrix is more chaotic than some reference, its eigenvalues ρ_i are spread over a smaller interval of $[0, 1]$. With concavity

$$\alpha f(x_1) + (1-\alpha) f(x_2) \leq f(\alpha x_1 + (1-\alpha)x_2) \tag{20}$$

and

$$\mathrm{Tr}(f(\widehat{\rho})) = \sum_i f(\rho_i) \tag{21}$$

A proof of this fact involves a few, but elementary steps and can be found, e.g. in Thirring, vol 4.

2.1.2 Additivity for independent systems

If entropy is to be a measure of our knowledge (or our ignorance) about a system it must not depend on how we decide to organize that knowledge. Suppose we have the information about two subsystems by the density matrices $\widehat{\rho}_1$ and $\widehat{\rho}_2$. We can widen our horizon and regard the two subsystems as the two parts of a larger system, however, without

drawing any conclusions from relating measurements on both systems simultaneously. In that case, the density matrix on the wider horizon is just the tensor product $\widehat{\rho}_1 \otimes \widehat{\rho}_2$. This change of bookkeeping should not alter our measure of information, entropy should be additive:

$$S(\widehat{\rho}_1 \otimes \widehat{\rho}_2) = S(\widehat{\rho}_1) + S(\widehat{\rho}_2). \tag{22}$$

This suggests a logarithm for our measure, but it does not single out the von Neumann entropy. Actually, any

α-entropy

$$S_\alpha(\widehat{\rho}) = \frac{1}{1-\alpha} \log \mathrm{Tr} \widehat{\rho}^\alpha \tag{23}$$

has the

General properties of an entropy

- $0 \leq S_\alpha(\widehat{\rho}) \leq \log \dim \mathcal{H}$
- $\widehat{\rho}_1 \preceq \widehat{\rho}_2 \Rightarrow S(\widehat{\rho}_1) \leq S(\widehat{\rho}_2)$.
- $S(\widehat{\rho}_1 \otimes \widehat{\rho}_2) = S(\widehat{\rho}_1) + S(\widehat{\rho}_2)$
- For $\widehat{\rho}^{(n)}$ with the sorted eigenvalues $(\underbrace{1/n, \ldots, 1/n}_{n \text{ times}}, 0, \ldots)$, $S(\widehat{\rho}^{(n)}) = \log(n)$.

2.1.3 Compatibility with sub-space partitioning

There is one more property that one considers desirable for a measure of information. Suppose we (arbitrarily) group the vectors from an orthogonal basis $\{|i\rangle\}$ of $\mathcal{H}$ into subsets $\{|i\rangle^{(\alpha)}\} \subset \mathcal{H}_\alpha$, i.e. $\mathcal{H} = \bigoplus_\alpha \mathcal{H}_\alpha$ and suppose that we know the probabilities p_α for the system being in the respective $\mathcal{H}_\alpha$. In addition, we know that if the system is in $\mathcal{H}_\alpha$ it has a density matrix $\widehat{\rho}_\alpha$. The density matrix for this organization of our knowledge is

$$\widehat{\rho} = \bigoplus_\alpha p_\alpha \widehat{\rho}_\alpha. \tag{24}$$

If we want to measure our ignorance in this situation, it should measure our ignorance about which $\mathcal{H}_\alpha$ the system is in. In addition, it should measure the ignorance $S(\widehat{\rho}_\alpha)$ within each $\mathcal{H}_\alpha$, weighted with the probability p_α of being in $\mathcal{H}_\alpha$ to begin with. For a reasonable measure of information, we require a stronger form of additivity by

$$S(\widehat{\rho}) = \sum_\alpha p_\alpha S(\widehat{\rho}_\alpha) + S(\widehat{p}), \tag{25}$$

where $\widehat{p}_{\alpha\alpha'} = \delta_{\alpha\alpha'} p_\alpha$ is the diagonal matrix of probabilities. That specific form of the requirement may not be convincing until we realize that additivity for the tensor product $\widehat{\rho}_1 \otimes \widehat{p}$ maps onto the special case $\mathcal{H}_\alpha = \mathcal{H}_1 \, \forall \alpha$: a tensor product of two spaces can be recast as a large direct sum of as many of copies of the first factor as the second factor has dimensions. A tensor product of density matrices is exactly the weighted direct sum as above and, of course, $\sum_\alpha p_\alpha = 1$. Normalization and continuity finally leave the von Neumann entropy as the single candidate:

Von Neumann entropy The function $S(\widehat{\rho}) = -\mathrm{Tr}\widehat{\rho}\log\widehat{\rho}$ is the only entropy with the properties

1. $S(\widehat{\rho})$ is a continuous function of the eigenvalues ρ_i
2. $S\left[\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\right] = \log 2$
3. The strong form of additivity as expressed in (25).

2.2 A probabilistic argument for von Neuman entropy

In the spirit of Boltzmann, we try to assign a likelihood to a given statistical state $\widehat{\rho}$. We construct the likelihood as proportional to the number of ways how $\widehat{\rho}$ can be realized in a long series of measurements (or in many measurements of randomly picked systems). As above, rather than assigning a density matrix for a single system describing the ensemble of all N preparations, we assign a *global* state vector to a series of N measurements. Any observation, however, remains restricted to a single sub-system, i.e. a single preparation in the whole series of preparations. A given observable does not connect different measurements: we do not, say, try to interfere phenomena in one measurement with phenomena in the next. The corresponding operator for the whole series of measurements is

$$A := \frac{1}{N}\sum_{n=1}^{N} \underbrace{\mathbf{1}\otimes\ldots\otimes\mathbf{1}}_{n-1\,\text{times}} \otimes A_s \otimes \mathbf{1}\otimes\ldots\otimes\mathbf{1} \tag{26}$$

i.e. we measure in each sub-system and take the average.

The state vector for a measurement series is an N-fold tensor product

$$|M\rangle_N := |\Phi_{j_1}\rangle\otimes\ldots\otimes|\Phi_{j_N}\rangle \tag{27}$$

and the expectation value is

$$\langle M|A|M\rangle_N = \frac{1}{N}\sum_n \langle\Phi_{j_n}|A_s|\Phi_{j_n}\rangle \tag{28}$$

For $j = 1,\ldots,n$, the total dimension of that space is n^N, a huge number even if the single system space has some small finite dimension n. We know that in the measurement series (or an ensemble) corresponding to the density matrix $\widehat{\rho}$, each state $|\Phi_i\rangle$ should occur $N_i = \rho_i N$ times (this is what the density matrix says). One measurement series complying with

$$\widehat{\rho} = \sum_{i=1}^{r} |\Phi_i\rangle\rho_i\langle\Phi_i| \tag{29}$$

could be

$$|M\rangle_{\rho,N} := \underbrace{|\Phi_1\rangle\otimes|\Phi_1\rangle\otimes\ldots\otimes|\Phi_1\rangle}_{N_1\,\text{times}} \otimes \underbrace{|\Phi_2\rangle\otimes\ldots\otimes|\Phi_2\rangle}_{N_2\,\text{times}} \otimes|\Phi_3\rangle\otimes\ldots \tag{30}$$

with the expectation value of a given state $A_s = |\Phi_{i_0}\rangle\langle\Phi_{i_0}|$ of

$$\langle M|A|M\rangle_{\rho,N} = \frac{N_i}{N} \sim \rho_i. \tag{31}$$

Of course, any permutation with the same frequency of $|\Phi_i\rangle$ also corresponds to the same $\widehat{\rho}$. For any set of N_i there are

$$\binom{N}{N_1, N_2, \ldots, N_r} = \frac{N!}{\prod_i N_i!} \tag{32}$$

such permutations. We assign to each state in the Hilbert space of the measurement series the same likelihood. This implies that we have no further knowledge about the Hilbert space as whole, say, we did not exclude certain states by some prior manipulation. Then the chance for finding a $|s\rangle_{\rho,N}$ with partioning $\{N_1, N_2, \ldots, N_N\}$ is proportional to the number of permutations that leave the partioning invariant. In the limit of large numbers we apply the Stirling formula $\log N! \approx N \log N$ to find

$$\ln\left[\frac{N!}{\prod_i N_i!}\right] = N \ln N - \sum_i N_i \ln N_i = -N \sum_i \rho_i \ln \rho_i. \tag{33}$$

2.3 Sub-additivity of entropy

If one has some information about the full system in the form of the density matrix $\widehat{\rho}$, one obviously will lose information if one discards the information referring to part of the system. E.g., we have measured all the components in a multi-particle break-up of some elementary particle, but when analyzing one considers only the properties of one fragment while integrating over (aka disregarding) all others.

In a density matrix this amounts to taking the

Definition 4. *Partial trace: Consider $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and $\widehat{\rho} \in \mathcal{B}(\mathcal{H})$ with finite trace. Let $\{|j\rangle\}$ be an orthonormal basis in $\mathcal{H}_1$. Then the* partial trace *over the factor $\mathcal{H}_1$ is defined as*

$$Tr_1\widehat{\rho} := \sum_j \langle j|\widehat{\rho}|j\rangle \in \mathcal{B}(\mathcal{H}_2) \tag{34}$$

and analogously for $1 \leftrightarrow 2$.

In this way, information about the complete system $\widehat{\rho}$ can be reduced to information about the respective parts

$$\widehat{\rho}_i = \mathrm{Tr}_i\widehat{\rho}. \tag{35}$$

The trivial fact that the information about all parts is not the same as the information about the whole is reflected in the equally trivial mathematical statement that in general

$$\widehat{\rho} \neq \widehat{\rho}_1 \otimes \widehat{\rho}_2. \tag{36}$$

In fact, we expect that by cutting a system into independent parts we loose information, even if we keep all partial information. This is dutifully reflected in the

Sub-additivity of entropy Let $\widehat{\rho}_{1,2} = \mathrm{Tr}_{1,2}(\widehat{\rho})$, then $S(\widehat{\rho}) \leq S(\widehat{\rho}_1) + S(\widehat{\rho}_2)$.

2.3.1 Remarks

The concept of entropy is not quite aligned with everyday thinking habits.

The whole and its parts Naïvely one might expect that our ignorance about the whole is necessarily greater that our ignorance if we content ourselves with the information we have on a small subsystem. But that is obviously not the case, most obviously if we we have a pure state for the full system, which has entropy =0. In fact, if one splits a pure state $\widehat{\rho} = |i\rangle\langle i|$ by forming two partial traces, the entropy of the subsystems is equal

$$S(\mathrm{Tr}_1\widehat{\rho}) = S(\mathrm{Tr}_2\widehat{\rho}). \tag{37}$$

Actually, the two partial density matrices agree in their non-zero eigenvalues: each is lacking the information about the relation to the other, and that deficiency is symmetric between the two parts.

Classical entropy The impossibility of mapping quantum mechanics into classical concepts fails also for entropy, notwithstanding the probabilistic character of the concept. Entropy in terms of classical phase-space densities is formally exactly analogous to the von Neumann entropy:

$$S_{cl} = -\int dq\, dp \rho(q,p) \log(q,p) \tag{38}$$

Taking very narrow distribution $\rho(q,p)$, one readily sees that classical entropy has no lower bound. (Classically, our ignorance is always infinite...).

2.3.2 Continuity properties

If no further constraints are imposed, almost all density matrices have infinite entropy. In particular, for any $\widehat{\rho}$ with finite $S(\widehat{\rho}) = s$, one can construct a $||\cdot||_1$-convergent sequence $\widehat{\rho}_N \to \widehat{\rho}$ whose entropy diverges. Such a sequence can be readily constructed by observing that on an infinite-dimensional Hilbert space there are density matrices with arbitrarily large entropy, e.g. the maximally mixed state on a C-dimensional subspace with its entropy

$$-C\frac{1}{C}\log\frac{1}{C} = \log C. \tag{39}$$

We now pick a sequence of density matrices such that $S(\sigma_N) = N^2$ and consider

$$\widehat{\rho}_N = \frac{1}{N}\sigma_N + (1-\frac{1}{N})\widehat{\rho} \overset{||\cdot||_1}{\to} \widehat{\rho} \tag{40}$$

As entropy is a concave function of the trace (and the trace is linear), $S(\widehat{\rho}_N)$ grows beyond all bounds

$$S(\widehat{\rho}_N) = S(\frac{1}{N}\sigma_N + (1-\frac{1}{N})\widehat{\rho}) \geq \frac{1}{N}S(\sigma_N) + (1-\frac{1}{N})S(\widehat{\rho}) = N + (1-\frac{1}{N})s. \tag{41}$$

The feeling that states with very large entropy may be pathologically spread out is correct: they necessarily cover such a large part of the Hilbert space that any reasonable energy diverges. "Reasonable" for the energy means that it is bounded from below and keeps growing sufficiently strongly such that $\mathrm{Tr}\exp(-\beta H)$ exists for some $\beta > 0$. (The physicists will recognize a Boltzmann-distribution or Gibbs-ensemble for temperature $T = (k\beta)^{-1}$, i.e. such an H allows a definition of temperature.)

Under the additional constraint that the expectation value of energy is finite, entropy *is* a continuous function of $\widehat{\rho}$.

Continuity of entropy at finite energies Assume a semibounded Hamiltonian $H \geq 0$ with $\mathrm{Tr}\exp(-\beta H) < \infty$ for some $\beta > 0$. Define the norm $||\widehat{\rho}||_H : \mathrm{Tr}\widehat{\rho}(H+1)$ and designate the space of density matrices with finite $||\widehat{\rho}||_H$ as C_H. Then the $S(\widehat{\rho})$ is a continuous map $C_H \to R_+$ w.r.t. to the $||\cdot||_H$-topology.