Selected topics in Mathematical Physics: Quantum Information Theory

Talk 1: Overview and introduction

Markus Müller[1, *]
[1]Institut für Theoretische Physik, Universität Heidelberg
(Dated: October 15, 2013)

The goal of this seminar is to get an overview on current research which tries to understand the foundations of statistical mechanics from a quantum information point of view. To this end, we will start with some basics of quantum information theory (entanglement, its quantification and operational meaning), then turn to some basic group representation theory and Lévy's Lemma, which allows us to compute "Hilbert space averages" and prove concentration of measure for random quantum states. Finally, these technical tools will be applied to obtain strong statements about the thermalization of closed quantum systems.

If time permits, we will also discuss some closely related topics, such as Hayden and Preskill's work on the black hole information paradox, quantum pseudorandomness, or the evolving field of single-shot non-equilibrium thermodynamics.

This document summarizes the first talk, containing an overview on the seminar and some basics on the state space of quantum theory and the description of composite systems.

I. OVERVIEW

The seminar has a website: it can be found at `http://www.mpmueller.net/seminar.html`. There, you will find an overview on the topics of the seminar, as well as a long list of references. An important reference for the start is the book by Nielsen and Chuang [1].

II. CLASSICAL VERSUS QUANTUM STATE SPACES

In the following, we will be interested in the phenomenon of *entanglement*, and how it can be quantified and applied as a resource. One important point in understanding entanglement will be to see in what way quantum theory (QT) differs from classical probability theory (CPT): while CPT allows for correlations, too, entangled quantum states are still fundamentally different. We will soon see in what way this is the case [3].

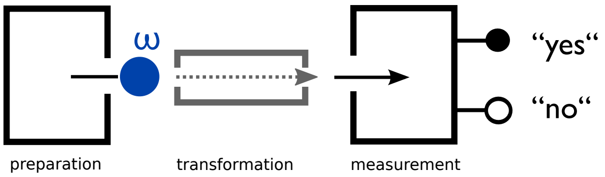

Figure 1: The basic operational/probabilistic setup which allows to understand the basic formalism of quantum theory and classical probability theory in a unified way.

In order to discuss classical and quantum state spaces in a unified way, we consider a simple schematic laboratory situation as in Figure 1. We have a *preparation device* (either an actual man-made device, or a physical system arising in nature, for example a distant star) that can emit physical systems, such as particles. Then the outgoing physical system can be subject to some *transformations*; in the simplest case, this would be time evolution according to the laws of physics; it could be a manufactured transformation device, such as a carefully designed magnet in a Stern-Gerlach setup, or, more abstractly, a circuit in a computation.

[*] Electronic address: m.mueller@thphys.uni-heidelberg.de

Finally, a *measurement* takes place. It has n different possible outcomes (in the picture, we would have $n = 2$). Only one of the outcomes will actually take place, but in general we will not know beforehand which one. This amounts to a *probabilistic* description.

In the end, by a *state* ω, we mean a description of the output of the preparation device which allows to predict the probabilities of all conceivable future measurements that can possibly be applied to the system. A *transformation* T will be described as a map, from states to states, and *measurements* will be modelled as linear functionals, yielding the probability of measurement outcomes if applied to a state.

This description unifies CPT and QT in a single picture:

- **Classical probability theory.** Think of the physical system as a die, and the preparation device as a player shaking the dice box. In the discrete case, we have $n \in \mathbb{N}$ levels ($n = 6$ for the die), and the state of the system will be described by a probability vector $p = (p_1, \ldots, p_n)$, $p_i \geq 0$, $\sum_i p_i = 1$.

  The set of all possible states, that is, the *state space* of a classical n-level system, is

  $$\mathcal{S}_n := \left\{ p = (p_1, \ldots, p_n) \;\middle|\; p_i \geq 0,\ \sum_i p_i = 1 \right\}.$$

  A very important property is the fact that this set is *convex*: if $p \in \mathcal{S}_n$ and $q \in \mathcal{S}_n$ are two probability distributions, and $\lambda \in [0,1]$ is a real number, then $\lambda p + (1-\lambda) q \in \mathcal{S}_n$ is a probability distribution, too. This has the following **interpretation**. Think of a preparation device D_p that prepares a physical system in state p, and another preparation device D_q that prepares a physical system in state q. Now build a new device D which internally tosses a (biased) coin, and prepares p with probability λ, or q with probability $(1-\lambda)$. The output of this preparation device D will be statistically indistinguishable from $\lambda p + (1-\lambda) q$; that is, it will be a physical system that is described by this state.

  So what are *transformations* T? Clearly, a transformation should map an ingoing state $p \in \mathcal{S}_m$ to an outgoing state $p' := T(p) \in \mathcal{S}_n$. The number of levels may change: for example, a device may just look at a die, and prepare a coin in its "heads" state of the die shows an even number, or prepare the coin in its "tails" state if it shows an odd number. This would be a transformation from $\mathcal{S}_6$ to $\mathcal{S}_2$. So $m \neq n$ is perfectly possible.

  However, there are some restrictions on what T may look like. Clearly, if p is a valid probability vector, then so should be p', i.e. p' should not contain negative numbers. This condition is already stated mathematically by demanding that T is a map from $\mathcal{S}_m$ to $\mathcal{S}_n$.

  There is however an additional requirement arising from the *statistical interpretation of convex mixtures*. Imagine the device D from above. Suppose we apply a transformation T after this preparation. There are now two possible ways to describe this situation:

  1. The device D prepares the state $\lambda p + (1-\lambda) q$, and then the transformation T is applied. In the end, this results in the state $T(\lambda p + (1-\lambda) q)$.
  2. The device D internally tosses a coin. With probability λ, it prepares state p, and the subsequent transformation produces $T(p)$. With probability $(1-\lambda)$, it prepares state q, and the subsequent transformation produces $T(q)$. That is, the output of the two-stage process is statistically indistinguishable from the state $\lambda T(p) + (1-\lambda) T(q)$.

  Since both are valid descriptions, they better give the same physical predictions. This is the case if and only if

  $$T(\lambda p + (1-\lambda) q) = \lambda T(p) + (1-\lambda) T(q),$$

  i.e. T is an *affine-linear* map. Basic linear algebra tells us that we can write T as a matrix (i.e. extend it to a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$); the conditions that T maps probability vectors to probability vectors implies that $T_{ij} \geq 0$, and $\sum_i T_{ij} = 1$. Matrices T with these properties are called *stochastic matrices*. They describe classical (noisy) information channels.

  Under some conditions, transformations T are *reversible*. This is the case if T^{-1} exists as a linear map, *and* is a valid transformation, too. Intuitively, reversible transformations are channels that do not destroy information. Classically, the reversible transformations are the *permutations*. They act on probability vectors as

  $$T(p_1, \ldots, p_n) = (p_{\pi(1)}, \ldots, p_{\pi(n)}),$$

  where π is a permutation, i.e. a bijective map from the discrete set $\{1, \ldots, n\}$ to itself. Formally, T is a representation of the permutation on the vector space that carries the probability vectors. This is a setting that we will encounter frequently in quantum theory, too.

It remains to describe measurements. A measurement with k possible outcomes is exhaustively described by specifying the probabilities of all its k outcomes on all possible input states. If p is any state, denote by $m_1(p), \ldots, m_k(p)$ the probabilities of the different outcomes. We have $m_i(p) \geq 0$, and $\sum_i m_i(p) = 1$. With the same argument as above (recall our device D?), all the m_i must be *affine-linear*. In other words, the m_i are *linear functionals*, and we can represent them as vectors such that $m_i(p) = m_i \cdot p$ under the standard Euclidean inner product.

As an example, suppose that we build a device that accepts a die (i.e. a classical 6-level system) as input, and has two possible outcomes (which are two lightbulbs that may flash). It works as follows: if the die shows a prime number (i.e. 2, 3 or 5), the first bulb flashed. Otherwise, the second bulb flashes. It is described by the functionals

$$m_1 = (0, 1, 1, 0, 1, 0), \qquad m_2 = (1, 0, 0, 1, 0, 1),$$

such that

$$m_1 \cdot p = p_2 + p_3 + p_5, \quad m_2 \cdot p = p_1 + p_4 + p_6.$$

All entries of every m_i must be in the interval $[0, 1]$, and we have $\sum_i m_i = \mathbf{1} := (1, 1, \ldots, 1)$, reflecting that the total probability is one.

- **Quantum theory.** States of a quantum n-level system are density matrices ρ; that is, the state space is

$$\mathcal{S}_n = \left\{ \rho \in \mathbb{C}^{n \times n} \mid \operatorname{tr}(\rho) = 1, \ \ \rho \geq 0 \right\},$$

where $\rho \geq 0$ means that ρ must be positive-semidefinite, that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all vectors $|\psi\rangle \in \mathbb{C}^n$. It can be shown that this implies that $\rho = \rho^\dagger$.

To see what the possible transformations are, note that the arguments that we have seen in the case of classical probability theory still apply here: transformations T must be linear in the density matrix, i.e. $T(\lambda\rho+(1-\lambda)\sigma) = \lambda T(\rho) + (1 - \lambda)T(\sigma)$. It turns out that the possible quantum transformations T are the *quantum operations*, that is, the completely positive, trace-preserving linear maps. We will hear more about them in a later talk.

The *reversible transformations* are the unitaries: that is, the maps

$$\rho \mapsto T(\rho) := U\rho U^\dagger, \qquad \text{where } U^\dagger U = \mathbf{1}.$$

Wigner's Theorem shows that these are the only possible reversible transformations, just by assuming that the state space is given by the set of density matrices. (The only other possible transformations, *antiunitary transformations*, act badly on entangled states and are thus not possible.)

Measurements with k outcomes are again described by k linear functionals. In CPT, the vector space which carried the states was $\mathbb{R}^n$, and so we were using the standard Euclidean inner product on $\mathbb{R}^n$ to write those functionals as vectors. Now, in QT, the vector space which carries the states (the density matrices) is the real vector space

$$V_n := \{A \in \mathbb{C}^{n \times n} \mid A = A^\dagger\}.$$

It has inner product $\langle A, B\rangle := \operatorname{tr}(A^\dagger B) = \operatorname{tr}(AB)$, the *Hilbert-Schmidt inner product*. Hence we can (and will) describe measurements with k outcomes by a set of Hermitian matrices $E_1, \ldots, E_k$, such that

$\langle E_i, \rho\rangle = \operatorname{tr}(E_i\rho)$ is the probability to obtain the i-th outcome if the measurement is applied to state ρ.

In order to obtain valid probabilities that sum to one, we must have $E_i \geq 0$ (i.e. the E_i must be positive-semidefinite) and $\sum_i E_i = \mathbf{1}$. A set of matrices E_i with these properties is called a *positive operator-valued measure*, or POVM. A special case of a POVM is a *projective measurement*, where every E_i is an orthogonal projector, i.e. $E_i = E_i^2$.

It is clear that the laboratory setting in Figure 1 allows for many other possible convex state spaces rather than CPT and QT. These are called *general-probabilistic theories* and are a fascinating research area on their own. Some of them allow for states that violate Bell inequalities even stronger than any quantum state, such as so-called *Popescu-Rohrlich boxes*. For a first overview, see, for example, [2].

An important difference between classical and quantum state spaces lies in the decomposition of mixed states into pure states. The following definition applies to state spaces $\mathcal{S}$ of both CPT and QT.

Definition 1. *A state $\omega \in \mathcal{S}$ is* mixed *if there exists $0 < \lambda < 1$ and states $\omega_1, \omega_2 \in \mathcal{S}$ with $\omega_1 \neq \omega_2$ such that*

$$\omega = \lambda\omega_1 + (1-\lambda)\omega_2.$$

Otherwise ω is pure.

In CPT, the pure states are $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0)$ etc., i.e. all states that have only zeroes and ones as entries. Every mixed state p has a *unique* decomposition into pure states:

$$(p_1,\ldots,p_n) = p_1(1,0,\ldots,0) + p_2(0,1,\ldots,0) + \ldots + p_n(0,\ldots,0,1).$$

In QT, a state ρ is pure if and only if there exists a normalized vector $|\psi\rangle \in \mathbb{C}^n$ such that $\rho = |\psi\rangle\langle\psi|$ (proof: exercise). As a main difference to the classical case, *decompositions into pure states are not unique*: for example, if $|0\rangle$ and $|1\rangle$ denote two orthonormal vectors in the Hilbert space $\mathbb{C}^2$, and $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, then the mixed state $\rho = \frac{1}{2}\mathbf{1}$ has the two different decompositions

$$\frac{1}{2}\mathbf{1} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|$$

and infinitely many others (many of them not into mutually orthogonal pure states, or more than two pure states).

III. COMPOSITE SYSTEMS

Now suppose we have *two* physical systems A and B (such as internal degrees of freedom of two particles, for example). Suppose that A is an m-level system, and B is an n-level system. How can we describe states on the joint system AB? Again, it turns out that many aspects of the quantum formalism can already be found in classical probability theory.

- **Classical probability theory.** The state of system A alone will be described by a probability vector $p_A \in \mathbb{R}^m$. It will simplify the following discussion if we use a particular convention for the entries of this vector. To this end, we imagine that the physical system A carries a random variable which we also denote A (abusing, but simplifying notation), and which can take on m different values $a_1,\ldots,a_m$ (these might be the first m natural numbers, or the $m = 6$ sides of a die). We can do the same for B. In other words,

  $$p_A = \big(P(A=a_1), P(A=a_2),\ldots,P(A=a_m)\big) \in \mathbb{R}^m, \qquad p_B = \big(P(B=b_1), P(B=b_2),\ldots,P(B=b_n)\big) \in \mathbb{R}^n.$$

  The state of the joint system AB will be a joint probability distribution:

  $$p_{AB} = \big(P(A=a_1,B=b_1), P(A=a_1,B=b_2),\ldots,P(A=a_m,B=b_n)\big) \in \mathbb{R}^{mn} \simeq \mathbb{R}^m \otimes \mathbb{R}^n.$$

  In the case where the systems A and B are *prepared independently*, we obtain that $P(A=a_i,B=b_j) = P(A = a_i)\cdot P(B=b_j)$ due to statistical independence of the random variables A and B. That is, the entries of p_{AB} are the products of the entries of p_A and p_B. In other words, if we choose the correct ordering of the vector entries, we obtain

  $$p_{AB} = p_A \otimes p_B.$$

  States of this form are called *uncorrelated*. Of course, we can also have correlated states in CPT. For example, imagine a pair of shoes (a left one, L, and a right one, R). Suppose that an experimenter puts them into two identically looking boxes, shuffles them randomly, and distributes them to two parties called Alice and Bob. Then we have $P(A=L,B=R) = P(A=R,B=L) = \frac{1}{2}$, and $P(A=L,B=L) = P(A=R,B=R) = 0$. That is, we have $p_{AB} = (0,\frac{1}{2},\frac{1}{2},0) \in \mathbb{R}^2 \otimes \mathbb{R}^2$, and this vector cannot be written in the form $p_A \otimes p_B$ – it is a correlated state. If Alice looks into her box and finds a left show, she will immediately know that Bob has a right shoe in his box.

  This shows us two things. First, it shows us that the mere fact that there is correlation does not in itself imply "magic action at a distance". Entanglement in QT is often described in this way: if one party measures and obtains an outcome, she knows immediately the outcome of the other party. But this is a rather mundane effect which – as we have seen – is ubiquitous in CPT: it can be simulated by two shoes in a box. There must be *something else* which is special about entanglement, and we will come back to this point very soon.

  Second, it turns out that the correlated probability vector p_{AB} is a *mixed* state, not a pure state. This is because it has entries that are not zero or one, but $1/2$ (recall our characterization of pure classical states above). It turns out that this is unavoidable:

Lemma 2. *Classically, all pure states on AB are uncorrelated.*

Proof. If p_{AB} is pure, then all its entries are either zero or one: $P(A=a_i, B=b_j) \in \{0,1\}$ for all i,j. Since probabilities sum up to one, there must be exactly one pair of indices i_0, j_0 such that $P(A=a_{i_0}, B=b_{j_0})=1$, and all others are zero. Define p_A and p_B by

$$p_A = \left(\underbrace{P(A=a_1)}_{0}, \ldots, \underbrace{P(A=a_{i_0})}_{1}, \ldots, \underbrace{P(A=a_m)}_{0} \right), \qquad p_B = \left(\underbrace{P(B=b_1)}_{0}, \ldots, \underbrace{P(B=b_{j_0})}_{1}, \ldots, \underbrace{P(B=b_n)}_{0} \right),$$

then $p_{AB} = p_A \otimes p_B$, and thus p_{AB} is uncorrelated. □

This will be quite different in quantum theory.

Finally, if we are given a state p_{AB}, then we can compute the "local state" p_A by marginalization:

$$P(A=a_i) = \sum_{b_j} P(A=a_i, B=b_j).$$

If we interpret p_{AB} by saying that it is a state that is held by two parties, Alice and Bob, then p_A describes the state that Alice sees "locally" in *her* laboratory. From p_A, we can compute all probabilities of all possible measurements that Alice can perform locally in her lab.

- **Quantum theory.** The local Hilbert spaces are $\mathcal{H}_A = \mathbb{C}^m$ and $\mathcal{H}_B = \mathbb{C}^n$; the Hilbert space for AB is then $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^{mn}$. Exactly as in the classical case, if states ρ_A and ρ_B are independently prepared on A resp. B, we obtain

$$\rho_{AB} = \rho_A \otimes \rho_B,$$

an uncorrelated state. Here is a difference to the classical case:

Lemma 3. *In quantum theory, there are correlated pure states on AB, for example*

$$\rho_{AB} = |\psi_-\rangle\langle\psi_-|, \qquad \textit{where } |\psi_-\rangle := \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right). \tag{1}$$

Pure correlated states (i.e. ones which are pure and cannot be written in the form $\rho_A \otimes \rho_B$) are called *entangled*. There are also mixed entangled states. We will hear more about this next time.

The last point to discuss is the quantum version of *marginalization*. Suppose we are given a bipartite state ρ_{AB}. How do we know what it "looks like" on system A? I.e. how do we compute ρ_A, the state that predicts outcomes of measurements on one of the two physical systems?

Clearly, we can think of "throwing away system B" as an actual physical process – therefore, it should satisfy the postulates of a physical transformation. In particular, as discussed above, it should be *linear* in the density matrix. Furthermore, it should map uncorrelated states $\rho_A \otimes \rho_B$ to ρ_A. It turns out that these two properties already determine the map uniquely.

Lemma 4. *There is a unique linear map, called the "partial trace" Tr_B, from the set of Hermitian matrices on AB to the set of Hermitian matrices on A, which satisfies the condition $\mathrm{Tr}_B(\rho_A \otimes \rho_B) = \rho_A$ for all density matrices ρ_A, ρ_B. It has a unique linear extension to the set of* all *matrices on AB, where it satisfies $\mathrm{Tr}_B(M_A \otimes M_B) = M_A \mathrm{tr}(M_B)$.*

Proof. Denote by V_{AB} the real vector space of Hermitian matrices on AB (and similarly V_A resp. V_B). The density matrices $\rho \in V_A$ linearly span V_A (i.e. they do not all live in a proper linear subspace), and similarly for B. Furthermore, it turns out that $V_{AB} = V_A \otimes V_B$, and so the matrices of the form $\rho_A \otimes \rho_B$, with $\rho_A \in V_A$, $\rho_B \in V_B$, ρ_A, ρ_B density matrices, linearly span V_{AB}. Consider any basis of density matrices of this form, define $\mathrm{Tr}_B(\rho_A \otimes \rho_B) := \rho_A$ for all of them, and extend this map linearly to all of V_{AB}. Clearly, this map has a unique extension to the set of all matrices on AB, because this set of matrices is $V_{AB} + iV_{AB}$.

It remains to show that $\mathrm{Tr}_B(M_A \otimes M_B) = M_A \mathrm{tr}(M_B)$ for all matrices M_A, M_B, for example for density matrices that are not in the chosen basis. To this end, develop M_A into basis elements, $M_A = \sum_i \alpha_i \rho_A^i$ with $\alpha_i \in \mathbb{C}$, and $M_B = \sum_b \beta_j \rho_B^j$. Then

$$\mathrm{Tr}_B(M_A \otimes M_B) = \sum_{i,j} \alpha_i \beta_j \mathrm{Tr}_B(\rho_A^i \otimes \rho_B^j) = \sum_{i,j} \alpha_i \beta_j \rho_A^i = M_A \sum_j \beta_j = M_A \mathrm{tr}(M_B),$$

and we are done. □

Similarly as for classical marginalization, computing Tr_B amounts to "summing over all degrees of freedom of the system B". It is easy to see that if ρ_{AB} is a density matrix, then so is $\rho_A := \mathrm{Tr}_B \rho_{AB}$. Furthermore, for any observable W_A, we have the local expectation value

$$\mathrm{tr}(W_A \rho_A) = \mathrm{tr}(W_A \mathrm{Tr}_B \rho_{AB}) = \mathrm{tr}(W_A \otimes \mathbf{1}_B \rho_{AB}).$$

We only need the properties of Tr_B mentioned in Lemma 4 to compute it in concrete examples. Consider the state $\rho_{AB} = |\psi_-\rangle\langle\psi_-|$, with $|\psi_-\rangle$ defined in (1). To compute ρ_A, write out ρ_{AB} from its definition,

$$\rho_{AB} = |\psi_-\rangle\langle\psi_-| = \frac{1}{2}\left(|01\rangle\langle 01| - |01\rangle\langle 10| - |10\rangle\langle 01| + |10\rangle\langle 10|\right),$$

and apply Tr_B term by term, which we may do due to linearity. Since $\mathrm{tr}|\varphi\rangle\langle\psi| = \langle\psi|\varphi\rangle$, we can use manipulations like the following

$$\mathrm{Tr}_B |01\rangle\langle 10| = \mathrm{Tr}_B \left(|0\rangle\langle 1| \otimes |1\rangle\langle 0|\right) = |0\rangle\langle 1| \mathrm{tr}|1\rangle\langle 0| = |0\rangle\langle 1| \cdot \langle 0|1\rangle = 0$$

to arrive at the result $\rho_A = \frac{1}{2}\mathbf{1}$. That is, even though the *global* state is pure, the *local* state turns out to be mixed. This is impossible in classical probability theory.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.

[2] J. Barrett, *Information processing in generalized probabilistic theories*, Phys. Rev. A **75**, 032304 (2007); arXiv:quant-ph/0508211.

[3] Unfortunately, we won't have time here to discuss the violation of Bell inequalities, which is arguably the most fundamental difference between quantum and classical probability theory.