12. The formalism of quantum mechanics

In this chapter, we discuss some mathematical issues of the theory of quantum mechanics. The starting point will be an axiomatic description of the formal structure of the theory. Before we look at this, let's discuss the structure of (Hamiltonian) classical mechanics in similar style, as a warm-up.

If the system consists of a single particle, then the possible *states* of this system are $(p, q) \in P = \mathbb{R}^6$, where $q \in \mathbb{R}^3$ is the position of the particle and $p \in \mathbb{R}^3$ is its momentum (= mass $\times$ velocity). The traditional name for P would be phase space, but I want to use the slightly more descriptive term *state space* instead. Note that the state refers to a fixed moment in time, not the evolving system. An *observable* is a continuous function $F : P \to \mathbb{R}$ on the state space. For example, the components of the *position* $F_j(p, q) = e_j \cdot q$ or the *angular momentum* $F_j(p, q) = e_j \cdot (p \times q)$ are observables. The time evolution of the state is governed by a system of ODEs, called *Hamilton's equations:* if $H(p, q)$ is the observable *energy,* then $\dot{p} = -\partial H/\partial q$, $\dot{q} = \partial H/\partial p$. Finally, a composite system, consisting of subsystems with state spaces P_1 and P_2, respectively, has state space $P = P_1 \times P_2$.

Now let's give a similarly structured description of quantum mechanics. We impose the following axioms, which we'll label S (states), O (observables), C (conditioning), and D (dynamics).

(S) The *state* space of a quantum mechanical system is $\mathbb{P}H$, the space of one-dimensional subspaces of a Hilbert space H. We usually do not take this distinction between states and vectors $\psi \in H$ very seriously and simply use normalized vectors ($\|\psi\| = 1$) to represent states.

(O) An *observable* is a self-adjoint operator on H. To actually observe the value of an observable, the observer must perform a *measurement.* The outcome of a measurement is random. If the system is in the state $\psi \in H$, $\|\psi\| = 1$, at the time of the measurement of the observable $A = A^*$, then the probability to observe a value in $M \subseteq \mathbb{R}$ is given by

$$P_A(M) = \|E_A(M)\psi\|^2,$$

where E_A denotes the spectral resolution of A: $A = \int_{\mathbb{R}} t \, dE_A(t)$

(C) When a measurement has been performed and a value in M of the observable A was observed, then the state must be updated according to

$$\psi_{\text{new}} = \frac{E_A(M)\psi}{\|E_A(M)\psi\|}.$$

(D) If no measurement is carried out, then the state evolves according to a group of unitary operators:

$$\psi(t) = U(t)\psi(0),$$
$$U(s+t) = U(s)U(t), \quad U(t)^*U(t) = U(t)U(t)^* = 1$$

We also demand that $\psi(t)$ is a continuous function of t for any choice of $\psi(0)$.

This is an exceedingly strange and philosophically challenging theory. Let's for now ignore (D), which describes the time evolution of our system, and focus on (S), (O), (C), which describe the system at a fixed point in time. The first surprise is the use of a Hilbert space to describe states. This can perhaps be given some plausibility, with the benefit of hindsight at least, if we also take into account that (O) introduces probabilities and randomness. If $x \in H$, $\|x\| = 1$, and $\{e_n\}$ is an ONB of H, then we can expand $x = \sum a_n e_n$, with coefficients $a_n = \langle e_n, x\rangle$. These will satisfy $\sum |a_n|^2 = \|x\|^2 = 1$, so could in principle serve as probabilities. This can be done for a given x for *any* ONB, so a Hilbert space could be viewed as a powerful book-keeping machine that can store many different probability measures in just one vector $x \in H$.

Exercise 12.1. In fact, any $x \in H$, $\|x\| = 1$, stores *all* such probability measures on $\mathbb{N}$ if H is infinite-dimensional. Prove the following precise version of this statement: let $x \in \ell^2$, $\|x\| = 1$, and let numbers $p_n \geq 0$ with $\sum p_n = 1$ be given. Then there is an ONB $\{e_n\}$ of ℓ^2 such that $|\langle e_n, x\rangle|^2 = p_n$.

(Recall also that $H \cong \ell^2$ for any separable H with $\dim H = \infty$, so choosing this space is no restriction.)

Now let's take a closer look at (O), (C). The easiest type of measurement involves observables $A = P$ that are projections. Note that in general, only values $t \in \sigma(A)$ are possible outcomes of a measurement of A (because E_A is supported by this set).

Exercise 12.2. Let $P \in B(H)$ be a projection, $P \neq 0, 1$. Show that $\sigma(P) = \{0, 1\}$, and the spectral resolution of P is given by $E(\{1\}) = P$, $E(\{0\}) = 1 - P$.

So a measurement of P has only two possible outcomes; we can think of it as a yes/no question. One could in fact argue that such measurements are sufficient in general since any question can in principle be broken up into a series of yes/no questions. If the system is in state

$x \in H$, $\|x\| = 1$, and P is measured, then the probabilities for 0 and 1 are $\|(1-P)x\|^2$ and $\|Px\|^2$, respectively.

All this is quite different from what a state (p, q) does for us in classical mechanics. There, it can be observed directly, as is clear to us without any theory, but is also confirmed by the formalism since (the components of) p, q are observables themselves. In quantum mechanics, the state can not be accessed directly; only observables can be measured, and the state provides information about measurement outcomes only in an indirect way.

We can now also get a better understanding of (C): if let's say 1 was obtained as the result of the measurement of P, then (C) instructs us to update the state x to $y = Px/\|Px\|$. This makes sure (since $(1-P)y = 0$ now) that if the measurement of P is repeated, immediately following the first measurement, then the second measurement will with certainty reproduce the outcome of the first measurement. This would of course not have happened if we had left x untouched; in that case, we would just have to use the original probabilities one more time.

This also explains the name *conditioning* for this postulate: you update your probabilities after new information has been obtained (in quantum mechanics, you actually update the state, but since this encodes probabilities, the net effect is the same). This kind of thing happens all the time in your daily life: for example, if tomorrow's forecast predicts 40% rain, 60% sunshine, then this is what you use for now. However, when you look out of the window the next morning and see sunshine, you instantaneously update to 0% rain, 100% sunshine.

I should also add that *conditioning* is in fact not the most common name for (C) though it is quite descriptive. Usually, (C) is referred to as the *reduction* or, rather more dramatically, the *collapse of the wave function,* and here *wave function* is just another name for the state $\psi \in H$. What collapses here are potentialities: for typical states ψ, any $t \in \sigma(A)$ is a possible outcome of the measurement of A, but after measuring and observing a specific value t, an immediately following re-measurement will confirm t, with the uncertainty gone or collapsed.

If discussed in this way, then (C) seems rather mundane, and the continuity property of the results of consecutive measurements that (C) expresses seems desirable or even indispensable. However, as a side effect, one is then more or less forced to view the state $\psi \in H$ correspondingly as knowledge or information about the system and nothing else. Such an interpretation is possible and is in fact close in spirit to what can probably be called the mainstream interpretation of quantum mechanics, the *Copenhagen interpretation.*

Other physicists and philosophers consider this approach to be unsatisfactory and too unkind towards the straightforward realistic world view that works reasonably well in classical physics, and they would prefer a more realistic interpretation of ψ also. Then (C) becomes potentially more troublesome since it introduces discontinuous instantaneous changes in ψ.

Another related issue that just refuses to go away is the *measurement problem.* This is not really a well defined question, and it means different things to different people. The common denominator is the felt need for a deeper understanding of what exactly is going on during a measurement. For instance one can imagine treating the system plus the measurement device as one big quantum mechanical system, which then presumably should evolve continuously according to (D). It is then not clear how (C) could ever come about. More precisely, it is clear that it can't; this version of the measurement problem is unsolvable.

Or try the following thought experiment, which in this version is due to Everett, but really just combines the two classics *Schrödinger's cat* and *Wigner's friend.* Physicist A performs a measurement of a quantum system S in a lab L. He also records the outcome by writing it down on a piece of paper. Physicist B is outside the lab L, which is completely isolated from its environment and forms a closed system. B describes the whole arrangement (A+S+L) quantum mechanically (using (D) throughout, obviously, as B does not conduct any measurements). Now A clearly performed a reduction of the state ψ_S of S, according to (C), at the time of the measurement, and he even has proof of this in the form of his note. From B's point of view, however, nothing special ever happened, and no reduction occurred. B uses a state ψ_{A+S+L} that contains a mixture of all the notes that A conceivably could have written. Finally, to make things even more confusing, B can ask A afterwards to confirm that for A, the reduction of ψ_S occurred when the measurement of S was performed and a unique note was written, not a surreal superposition of all possible notes. Despite all this convincing testimony, for B, no reduction can occur before the inspection of L and the interaction with A.

This is just some food for thought. I don't want to attempt a discussion of these issues here. Let me just say in closing that almost 100 years after the initial formulation of quantum mechanics, none of these very basic philosophical questions have received answers that have led to anything like a consensus on any aspect of them.

In classical mechanics, one can of course observe any number of observables simultaneously. It suffices to determine the state (p, q), and then one can in principle determine the values of *all* observables $F(p, q)$.

Quantum mechanics does not admit even the possibility of measuring more than one observable at the same time, if the axioms are taken at face value. However, one can get around this to some extent, as follows:

Definition 12.1. Let S, T be self-adjoint operators on H. We say that S, T are *simultaneously measurable* if there is a third observable $X = X^* \in B(H)$ such that $S = f(X)$, $T = g(X)$ for certain (Borel) functions f, g on $\mathbb{R}$.

In this situation, it indeed makes sense to interpret a measurement of X as a measurement of S, T also since one could in principle find these observables from X. This interpretation also works well when done carefully within the formalism.

Exercise 12.3. Let S, T, X be as in Definition 12.1. Suppose that a measurement of X gave a result in $M \subseteq \mathbb{R}$. Show that an immediately following measurement of S (in other words, update the state according to (C)) will then produce a value in $f(M)$, with probability 1, and similarly for T.

Even more ambitiously, could do *both,* measure S and then immediately afterwards T, following the measurement of X, and you would still be guaranteed results in $f(M)$ and $g(M)$, respectively.

Remark: This will probably require a substitution rule of sorts for $\int h\, dE$, to recover the spectral resolutions of S, T from that of X. Compare Exercise 11.20(a).

The question of what observables are simultaneously measurable is then clarified by the following result:

Theorem 12.2 (von Neumann)**.** *Let $S, T \in B(H)$ be self-adjoint operators. Then S, T are simultaneously measurable if and only if $[S, T] \equiv ST - TS = 0$.*

Sketch of proof. In more technical language, the result says that S, T are both functions of a third self-adjoint operator if and only if they commute. One direction is obvious since any two functions $f(X), g(X)$ of the same operator commute.

So assume now that $[S, T] = 0$. The version of the Spectral Theorem that we established in Theorem 10.4(a) is close to what we want. This result shows that if S, T commute, then $S = \int_\Delta \widehat{S}\, dE$, $T = \int_\Delta \widehat{T}\, dE$ for some common (to S and T) spectral resolution E. The only problem is that Δ is just an abstract space, which may be too big for our purposes. If we had $\Delta \subseteq \mathbb{R}$, then we could let $X = \int t\, dE$ and would obtain $S = f(X)$, $T = g(X)$, as desired, with the functions f, g being the Gelfand transforms of S and T, respectively.

This can be remedied as follows. We apply Theorem 10.4(a) to specifically the C^*-subalgebra $A = C^*(S,T)$ of $B(H)$ that is generated by S, T. Then Δ will be the maximal ideal space of A. We can now see in the same way as in the proof of Theorem 9.12 that Δ is homeomorphic to a compact subset K of $\mathbb{R}^2$; a homeomorphism is given by mapping $\phi \mapsto (\phi(S), \phi(T)) = (\widehat{S}(\phi), \widehat{T}(\phi))$. Now we use the fact that there will be a Borel isomorphism $B_0 : K \to [0,1]$ onto a subset of $[0,1]$. By this I mean a bijective map with the property that both B_0 and B_0^{-1} are Borel functions. Results of this type hold in great generality; I don't want to discuss them further here.

We can then also compose with the homeomorphism between Δ and K to obtain such a Borel isomorphism $B : \Delta \to [0,1]$. Now use B to move everything over to $[0,1]$; then $S = \int_{[0,1]} f\, dF$, $T = \int_{[0,1]} g\, dF$, with $F(M) = E(B^{-1}(M))$, and from a substitution rule for these integrals we also know that $f = \widehat{S} \circ B^{-1}$, $g = \widehat{T} \circ B^{-1}$ (which we don't need here). We can take $X = \int_{[0,1]} t\, dF(t)$. $\square$

Exercise 12.4. Let $T \in B(H)$ be self-adjoint, and let $f, g : \mathbb{R} \to \mathbb{R}$ be bounded Borel functions. Show that $f(g(T)) = (f \circ g)(T)$.

Exercise 12.5. Give an elementary proof, from scratch, of Theorem 12.2 in the special case $H = \mathbb{C}^n$. *Suggestion:* Show that H has an ONB of simultaneous eigenvectors of S, T. Then define X by removing possible degeneracies (so if S, say, has simple spectrum, then $X = S$ would work).

Since typically operators don't commute, Theorem 12.2 puts strong restrictions on what simultaneous measurements are possible. A famous pair of non-commuting observables is given by the position operator $q = M_x$ (multiplication by x) and the momentum operator $p = -i\, d/dx$ on $H = L^2(\mathbb{R})$. (I am stretching our results somewhat here since q, p are *not* bounded, so Theorem 12.2 as stated doesn't apply.) If we ignore domains and proceed formally, then $[q, p] = i \neq 0$.

(Comp) The state space H of a composite quantum system, consisting of two subsystems with state spaces H_1 and H_2, is the tensor product $H = H_1 \otimes H_2$.

This looks similar typographically to the prescription $P = P_1 \times P_2$ for composite classical systems, but of course the tensor product is not the easiest way to construct a new Hilbert space out of H_1, H_2. That probably would have been $H_1 \oplus H_2$, but in fact there are immediate objections against this choice. For example $(\psi, 0) \in H_1 \oplus H_2$ does not associate a state to the second system; but if we rule out such vectors, then the state space of our system is no longer $\mathbb{P}H$, contradicting (S). A related problem is that there is no good way to make measurements on one subsystem exclusively depend only on the state this part of the system is in. The tensor product, on the other hand, handles these issues gracefully. As we will see, its use has striking consequences.

Let's first discuss the notion of the tensor product $V \otimes W$ of two vector spaces. This is easy to describe intuitively, but any precise definition will inevitably get a bit clumsy. We want $V \otimes W$ to contain vectors that we will write as $v \otimes w$, with $v \in V$, $w \in W$. Since $V \otimes W$ is supposed to be a vector space, we must also put (formal, at this point) linear combinations $\sum_{j=1}^n c_j v_j \otimes w_j$ into $V \otimes W$. Finally, as the notation suggests, we want $\otimes$ to behave like a product. More precisely, we want $(au + bv) \otimes w = a(u \otimes w) + b(v \otimes w)$, and similarly in the second factor.

This completes a slightly informal description of $V \otimes W$ that is in fact good enough for many purposes when combined with an important general fact that is formulated as Exercise 12.6 below. This will also show that $\dim V \otimes W = \dim V \dim W$.

A more rigorous definition can be given as follows. For any set S, we define the *free vector space* $F(S)$ over S as the set of finitely supported functions $f : S \to \mathbb{C}$, with its obvious vector space structure. You could also view the elements of $F(S)$ as formal (!) linear combinations $\sum_{j=1}^n c_j s_j$; the connection to the first version is obtained by viewing $f(s)$ as the coefficient of s. But here we must be careful to not misinterpret this as an actual linear combination, especially if S happens to have a natural vector space structure already, as in our intended application. In either version, $F(S)$ is simply a construction of a vector space with basis S (after identifying s with the function $f_s(s) = 1$, $f_s(t) = 0$ for $t \neq s$ in the first version, and with $1s$ in the second one).

Definition 12.3. The (algebraic) *tensor product* of the vector spaces V, W is $V \otimes W := F(V \times W)/R$, where R is the subspace of $F(V \times W)$ spanned by all vectors of the form

$$\begin{aligned} (u + v, w) - (u, w) - (v, w), &\quad (cv, w) - c(v, w), \\ (v, w + x) - (v, w) - (v, x), &\quad (v, cw) - c(v, w), \end{aligned}$$

with $u, v \in V, w, x \in W, c \in \mathbb{C}$.

We then return to the original notation and write the element of $V \otimes W$ that is represented by (v, w) as $v \otimes w$. The elements of $V \otimes W$ are also referred to as *tensors,* and such vectors $x = v \otimes w$ are called *pure* (or *simple*) *tensors.* The condition really is that x can be written in this form, so for example $u \otimes w + v \otimes w$ is a pure tensor.

Exercise 12.6. Let $\{e_\alpha : \alpha \in A\}$ and $\{f_\beta : \beta \in B\}$ be bases of V and W, respectively. Show that then $\{e_\alpha \otimes f_\beta : \alpha \in A, \beta \in B\}$ is a basis of $V \otimes W$. *Suggestion:* This is not as trivial as one might think it should be. To prove linear independence, construct linear functionals $F_{\gamma\delta} \in (V \otimes W)^*$ that satisfy $F_{\gamma\delta}(e_\alpha \otimes f_\beta) = \delta_{\alpha\gamma}\delta_{\beta\delta}$.

Exercise 12.7. (a) Show that every $x \in V \otimes W$ can be written as a sum of pure tensors $x = \sum_{j=1}^n v_j \otimes w_j$.
(b) Show that $0 \otimes w = v \otimes 0 = 0$ for any v, w, with the last zero denoting the zero vector of $V \otimes W$, of course.
(c) If none of the individual vectors is the zero vector, show that $v \otimes w = x \otimes y$ if and only if $v = cx$, $y = cw$ for some $c \in \mathbb{C}$.

Exercise 12.8. Let $V = \mathbb{C}^2$. For two linearly independent vectors $e, f \in V$, define

$$v(e, f) = e \otimes f - f \otimes e \in V \otimes V.$$

Now apply a matrix $S \in \mathbb{C}^{2\times 2}$ to e, f. Show that then

$$v(Se, Sf) = (\det S)v(e, f).$$

(This remarkable invariance of $v(e, f)$ will play a key role in our discussion of the Einstein-Podolsky-Rosen paradox.) Also, show that v is not a pure tensor.

We have a natural map $\varphi : V \times W \to V \otimes W$, $\varphi(v, w) = v \otimes w$. This map is bilinear (= linear in each argument separately). One can base an alternative definition of $V \otimes W$ on this and the following universal property: any bilinear map $B : (V, W) \to Z$ to a third vector space Z factors through $V \otimes W$ in a unique fashion: there is a unique linear map $L : V \otimes W \to Z$ such that $B = L \circ \varphi$. This is a bit too abstract for our purposes, but it does draw attention to the fact that this bilinear map $(v, w) \mapsto v \otimes w$ is a crucial part of the whole construction. Since a vector space doesn't have any structure other than its dimension, it would otherwise have been perfectly pointless to go to these lengths to construct $\mathbb{C}^m \otimes \mathbb{C}^n$, say, when we have the much more straightforward isomorphic space $\mathbb{C}^{mn}$ available. Or, to say essentially the same thing in more down-to-earth fashion, on a tensor product, I can not only point to Exercise 12.6 and say that the general vector is a linear combination of the basis vectors (this I can do in any vector space), I also gave meaning to $v \otimes w$ for arbitrary v, w.

If we now have two Hilbert spaces H_1, H_2, then $H_1 \otimes H_2$ can also be given a scalar product, as follows: for pure tensors, we define

$$\langle x_1 \otimes x_2, y_1 \otimes y_2 \rangle = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle , \tag{12.1}$$

and then we extend to general tensors by making $\langle \cdot, \cdot \rangle$ sesquilinear, as we must. Since different representations of tensors such as $(x+y)\otimes z = x \otimes z + y \otimes z$ lead to different options of evaluating scalar products, it must be checked that this is well defined. (In fact, a first small issue concerns (12.1) itself because u, v are not uniquely determined by $u \otimes v$. Rather, non-zero factors can be moved around at will; compare Exercise 12.7(c). It is clear, of course, that this at least is harmless.)

We can address these issues by taking seriously the set-up from Definition 12.3: (12.1) does define a map on $B \times B$, with $B = H_1 \times H_2$, and then also a unique sesquilinear map $S : F(B) \times F(B) \to \mathbb{C}$. More explicitly,

$$S\left(\sum a_j (x_j^{(1)}, x_j^{(2)}), \sum b_k (y_k^{(1)}, y_k^{(2)}) \right) = \sum \overline{a_j} b_k \left\langle x_j^{(1)}, y_k^{(1)} \right\rangle \left\langle x_j^{(2)}, y_k^{(2)} \right\rangle .$$

We need a sesquilinear map on $H \times H$, with $H = H_1 \otimes H_2$, and since $H = F(B)/R$, what we must check is that $S(v, w) = 0$ if $v \in R$ or $w \in R$. By (anti-)linearity, it suffices to check this on the spanning set of R from Definition 12.3, and this is straightforward.

Let's also verify that $\langle t, t \rangle \geq 0$ for any tensor $t \in H_1 \otimes H_2$, with equality precisely when $t = 0$. Any tensor is a sum of pure tensors: $t = \sum x_j \otimes y_j$. Fix ONBs $\{e_j\}$, $\{f_k\}$ of the (finite-dimensional) subspaces that are spanned by the x_j and the y_j, respectively, and expand each x_j and y_j in terms of these. After multiplying out, we then obtain a representation of the form

$$t = \sum_{j,k} c_{jk} e_j \otimes f_k,$$

with coefficients $c_{jk} \in \mathbb{C}$. Now plug this into $\langle t, t \rangle$ and again multiply out. This will leave us with a sum of many terms, but now all scalar products are among the vectors $e_j \otimes f_k$. We clearly have

$$\langle e_j \otimes f_k, e_m \otimes f_n \rangle = \begin{cases} 1 & j = m, k = n \\ 0 & \text{otherwise} \end{cases} .$$

This finally leads to $\langle t, t \rangle = \sum_{j,k} |c_{jk}|^2$. We conclude that indeed $\langle t, t \rangle \geq 0$ and $= 0$ if and only if $c_{jk} = 0$ for all j, k, which is equivalent to $t = 0$ since the $e_j \otimes f_k$ are linearly independent.

Exercise 12.9. Give a more explicit version of this argument. Also, if desired, provide a careful complete proof that $\langle\cdot,\cdot\rangle$ is a scalar product on $H_1\otimes H_2$.

Example 12.1. Let's now try out all of this for $H_1=H_2=H=\ell^2(\mathbb{Z})$. If we just run through the above constructions for these data, we won't get any additional insight, so let's right away bring into play a natural map $T:H\otimes H\to\ell^2(\mathbb{Z}^2)$ that is available here. On pure tensors, we define it as $T(x\otimes y)(m,n)=x(m)y(n)$, and then I want to extend linearly. This runs into an issue similar to the one above, when we introduced a scalar product on $H_1\otimes H_2$: there are linear relations among pure tensors, so the map is already overdetermined. For example, we have several competing options how to evaluate $T((x+y)\otimes z)$, and we should check that these are consistent. This can be done in pretty much the same way as above, and I don't want to say more on it here.

Notice then that T preserves scalar products among pure tensors:

$$\begin{aligned}\langle T(u\otimes v),T(x\otimes y)\rangle &= \sum_{m,n\in\mathbb{Z}}\overline{u(m)v(n)}x(m)y(n)\\ &= \sum_{m\in\mathbb{Z}}\overline{u(m)}x(m)\sum_{n\in\mathbb{Z}}\overline{v(n)}y(n)\\ &= \langle u,x\rangle\,\langle v,y\rangle=\langle u\otimes v,x\otimes y\rangle\end{aligned}$$

To carefully justify passing to the second line, note that the double series converges absolutely, by the Cauchy-Schwarz inequality. Since pure tensors span the tensor product, it follows that T is an isometry on $H\otimes H$. Moreover, if $x=\delta_m$, $y=\delta_n$, then $T(x\otimes y)=\delta_{(m,n)}$, and these latter vectors form an ONB of $\ell^2(\mathbb{Z}^2)$. So $T(H\otimes H)$ is dense in $\ell^2(\mathbb{Z}^2)$. However, T is not onto:

Exercise 12.10. Prove that (for example)

$$u(m,n)=\begin{cases}2^{-|n|} & m=n\\ 0 & \text{otherwise}\end{cases}$$

is not in $T(H\otimes H)$.

As a consequence, $H\otimes H$ is not complete. This is a slight setback, and it forces us to modify the definition in the Hilbert space setting.

Definition 12.4. Let H_1,H_2 be Hilbert spaces. The (Hilbert space) *tensor product* $H_1\otimes H_2$ is the completion of the (algebraic) tensor product of H_1,H_2, endowed with the scalar product (12.1).

The notation has now become somewhat overloaded since the symbol $\otimes$ could stand for the algebraic or the Hilbert space tensor product.

This is harmless, though, as we will always use the Hilbert space tensor product for Hilbert spaces. On top of this, our main application will involve finite-dimensional spaces only, and then the distinction disappears.

We can now also finish the discussion of Example 12.1. We have $\ell^2(\mathbb{Z}) \otimes \ell^2(\mathbb{Z}) \cong \ell^2(\mathbb{Z}^2)$ (Hilbert space tensor product, obviously), with a natural unitary map between these spaces, which maps pure tensors $x \otimes y \mapsto x(m)y(n)$.

Similarly, one can show that $L^2(X,\mu) \otimes L^2(Y,\nu) \cong L^2(X \times Y, \mu \otimes \nu)$, by sending $f \otimes g \mapsto f(x)g(y)$ (and $\mu \otimes \nu$ is the product measure, of course, not a tensor product).

Exercise 12.11. Define

$$\ell^2(\mathbb{N};\mathbb{C}^d) = \left\{ x : \mathbb{N} \to \mathbb{C}^d : \sum_{n=1}^{\infty} x(n)^* x(n) < \infty \right\}$$

("square summable sequences taking values in $\mathbb{C}^d$"). This becomes a Hilbert space with the scalar product $\langle x, y \rangle = \sum x(n)^* y(n)$. (You can just assume this plausible fact, or try to prove it if you want to.) Show that $\ell^2(\mathbb{N}) \otimes \mathbb{C}^d \cong \ell^2(\mathbb{N};\mathbb{C}^d)$, with a natural unitary map between these spaces (find this map!).

Exercise 12.12. Let H be an arbitrary Hilbert space. Show that

$$U(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} x_j \otimes e_j$$

defines a unitary map $U : H \oplus H \oplus \ldots \oplus H \to H \otimes \mathbb{C}^n$. Conclude that the algebraic tensor product $H \otimes \mathbb{C}^n$ agrees with the Hilbert space tensor product in this case.

Exercise 12.13. Let $\{e_j\}$ and $\{f_n\}$ be ONBs of H_1 and H_2, respectively. Show that $\{e_j \otimes f_k\}$ is an ONB of $H_1 \otimes H_2$.

As our final general topic related to tensor products, I want to introduce the tensor product of two operators. Suppose that $A_j \in B(H_j)$. For pure tensors, we then define $T(x_1 \otimes x_2) = A_1 x_1 \otimes A_2 x_2$. We next want to extend this linearly to the *algebraic* tensor product $H_1 \otimes H_2$ in a first step. Here we encounter for the third time the technical problem that a given $v \in H_1 \otimes H_2$ can be written in many ways as a linear combination of pure tensors, and we must check the consistency of our intended definition. We discussed the analogous issue for the scalar product more explicitly, and I want to leave it at that.

Proposition 12.5. *This operator T on the algebraic tensor product is bounded, with operator norm $\|T\| = \|A_1\| \, \|A_2\|$.*

This implies that T has a unique continuous extension to all of $H_1 \otimes H_2$ (Hilbert space tensor product now), and this operator we will denote by $A_1 \otimes A_2$. It still satisfies $\|A_1 \otimes A_2\| = \|A_1\| \, \|A_2\|$.

Proof. We have $T(A_1, A_2) = T(A_1, 1)T(1, A_2)$, so to prove that the original T is bounded, it suffices to show that these factors are bounded. We consider $T = T(A, 1)$. Let $u = \sum_{j=1}^n x_j \otimes y_j$ be a tensor. As we saw earlier, in our discussion of the scalar product, we can expand each x_j and y_j in terms of an ONB. Put differently, we may assume that both $\{x_j\}$ and $\{y_j\}$ are orthogonal systems. Then

$$\begin{aligned} \|Tu\|^2 &= \left\langle \sum_{j=1}^n Ax_j \otimes y_j, \sum_{k=1}^n Ax_k \otimes y_k \right\rangle = \sum_{j=1}^n \|Ax_j\|^2 \|y_j\|^2 \\ &\leq \|A\|^2 \sum_{j=1}^n \|x_j\|^2 \|y_j\|^2 = \|A\|^2 \|u\|^2; \end{aligned}$$

the last equality follows by evaluating $\|u\|^2 = \langle u, u\rangle$ by a similar calculation.

Using this and the analogous bound $\|T(1, B)\| \leq \|B\|$, we see that $T = T(A_1, A_2)$ is indeed bounded, and $\|T\| \leq \|A_1\| \, \|A_2\|$. To prove the opposite inequality, let $\epsilon > 0$ be given. Pick $x_j \in H_j$, $\|x_j\| = 1$, such that $\|A_j x_j\| > \|A_j\| - \epsilon$. Then also $\|x_1 \otimes x_2\| = 1$, and $\|T(x_1 \otimes x_2)\| \geq (\|A_1 - \epsilon)(\|A_2\| - \epsilon)$. ☐

We can now also add some precision to (Comp). Of course, from a very abstract point of view, any Hilbert space is as good as any other of the same dimension since they are all isomorphic. What we really mean when we say that the tensor product should be used to describe composite systems is that the individual factors still refer to the corresponding subsystems. In particular, if A_1 is an observable for the first system, then this measurement can still be carried out on the composite system, and the corresponding observable is $A_1 \otimes 1$.

Exercise 12.14. Suppose that $A_j \in B(H_j)$, $A_j \neq 0$. Prove that $A_1 \otimes A_2$ is self-adjoint if and only if both A_1 and A_2 are.

Exercise 12.15. Let $A_1 \in B(H_1)$ be self-adjoint, with spectral resolution E_1. Show that the spectral resolution of $A = A_1 \otimes 1$ is $E_1 \otimes 1$. Conclude that $\sigma(A) = \sigma(A_1)$.

After these somewhat lengthy preparations, we are now finally ready for a discussion of Bohm's version of the famous EPR paradox (after A. Einstein, B. Podolsky, N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, *Phys. Review* **47** (1935), 777–780). This is a thought experiment. We imagine a physical system consisting of two subsystems, each described by the Hilbert space $H = \mathbb{C}^2$. By (Comp), the Hilbert space of the full system is the four-dimensional space $H \otimes H$. Let $e = (1,0)^t$, $f = (0,1)^t$ be the standard ONB of H, and form the *EPR state*

$$v = \frac{1}{\sqrt{2}} (e \otimes f - f \otimes e) . \tag{12.2}$$

Compare Exercise 12.8; the additional factor $1/\sqrt{2}$ makes sure that $\|v\| = 1$.

We now perform the measurement corresponding to the projection P onto $L(e)$ on the first space. More precisely, we measure the observable $P \otimes 1$. This is a projection itself, projecting onto the subspace $\{e \otimes v : v \in H\}$. Thus the probability for the outcome 1 is given by

$$\|(P \otimes 1)v\|^2 = \frac{1}{2}\|e \otimes f - 0 \otimes e\|^2 = \frac{1}{2}.$$

If that indeed is the outcome of the measurement, then, according to (C), the updated state will be

$$w(1) = \frac{w_0(1)}{\|w_0(1)\|}, \quad w_0(1) = (P \otimes 1)v = \frac{1}{\sqrt{2}}\, e \otimes f,$$

so $w(1) = e \otimes f$.

Now suppose we immediately follow up with the same measurement, but on the second subsystem. The corresponding observable is $1 \otimes P$. Since $Pf = 0$, so $(1 \otimes P)w(1) = 0$, the outcome of this measurement is 0, with certainty. Similarly, if the first measurement gives the answer 0, then the updated state is the normalized version of $(1 - P \otimes 1)v$, which is $w(0) = -f \otimes e$. In this case, the second measurement (of $1 \otimes P$) will now produce the answer 1, again with certainty.

This is already quite remarkable since the subsystems could be well separated spatially. Imagine two electrons, call them a and b, moving in opposite directions, and wait until the distance between a and b is 10 light years, say. The two-dimensional Hilbert space $\mathbb{C}^2$ is appropriate to model the spin of an electron. In more physical terms, our calculations can thus be rephrased as follows: physicist A measures the spin of electron a. The outcome of this measurement is genuinely random, and the two possible outcomes, which in physical units would actually be $\pm\hbar/2$ (but, being mathematicians, let us call this ± 1, for convenience), have probability $1/2$ each. Let's say A measures spin $+1$. Now B performs the same measurement at electron b, one nanosecond after A's measurement (as A, B have agreed on in advance). As we showed, the outcome of this measurement is no longer random; rather, B will obtain the opposite of A's result with certainty, which in our example is spin -1.

Moreover, the two factor Hilbert spaces play symmetric roles. Thus we can also let B start and then A follow up immediately with the same measurement. This time, B's measurement result is random, with probability $1/2$ assigned to each outcome, while A's result can then be predicted with certainty: it will be the opposite of what B obtained.

A good summary of the whole situation is to say that there seems to be a conservation law at work: the total spin of both electrons together is always 0.

All this has already some shock value, especially if viewed against the background of Einstein's own theory of relativity, which predicts that no signal can be transmitted faster than the speed of light. There seems to be some instantaneous non-local interaction between electrons a, b, or at least quantum mechanics gives the appearance of some such effect. Even spatially widely separated subsystems must not be viewed in isolation. They can only be understood as parts of the full system.

But it gets even better (or worse, depending on your preferences). We can replace e, f by the rotated vectors $e' = Ue$, $f' = Uf$ for a $U \in \mathrm{SU}(2)$ (a unitary matrix with $\det U = 1$). By Exercise 12.8, the corresponding state

$$v = \frac{1}{\sqrt{2}} \left(e' \otimes f' - f' \otimes e' \right)$$

is still the original EPR state from (12.2), for any $U \in \mathrm{SU}(2)$. So the whole story can be retold with measurements corresponding to the projection onto e' rather than e. Physically, these still correspond to the electron spin, but in a different direction in physical space $\mathbb{R}^3$. (The full spin, being an internal angular momentum of sorts, is really a vector and thus has three components; what we measured above was one fixed component of it.) Now these distinct spin components, as observables, will not commute. This is clear because commuting self-adjoint matrices have the same eigenvectors, but the eigenvectors of the projection $P' = UPU^*$ onto $L(e')$ are e', f'. In particular, it is not possible to measure the full spin vector in quantum mechanics since the components don't commute. See again Theorem 12.2.

So now the situation is as follows: A can decide, spur of the moment, which of infinitely many non-commuting observables to measure. Immediately following this measurement, B measures a spin component of electron b. There is now one component (the one that A just measured for electron a) that will with certainty give a result that A can predict; the other components will give random outcomes. So A's measurement decides which of the possible measurements that B can perform will have a sharp value and this is the case even though there is no quantum mechanical state that will give such a sharp value to more than one spin component.

It appears that one can no longer pretend that electron spins actually have values that are just read off by the measurement process; rather, the measurement itself produces these values. Moreover, it substantially disturbs the whole system, as predicted by (C), including (quite shockingly) parts that are very far away from the location of the measurement.

If one wants to avoid such conclusions but still admits that quantum mechanically computed probabilities make correct predictions, then one can hope that quantum mechanics is an incomplete theory in about the same way as statistical mechanics. That is, one postulates that there are *hidden variables* (this is the technical term) that we haven't discovered yet. In this picture, the electrons really have definite spins, in all directions and at all times, and I could tell what they are if I knew the values of those hidden variables. Or, to make things simpler, we can just say that those spin components *are* the hidden variables of this system. The quantum mechanical probabilities work because they correctly model my ignorance of these variables, in much the same as assigning probability $1/2$ to the event *heads* for the toss of a fair coin correctly models my (only practical) inability to give a complete physical description of the in principle deterministic process of the throw of the coin.

These assumptions can be tested mathematically. Let's do this for the EPR system. I want to focus on three spin components, in three distinct but not necessarily orthogonal directions in physical space $\mathbb{R}^3$. I'll describe these by points on the unit sphere, and I use the angles $0 \leq \phi \leq \pi$, $0 \leq \theta < 2\pi$ from the spherical coordinates. I'll write $\omega = (\phi, \theta)$, and $S_e(\omega)$, $e = a, b$, will denote the spin of electron e in direction ω. My hidden variables are then the six quantities $S_e(\omega)$, for three directions $\omega = \alpha, \beta, \gamma$ and the two electrons $e = a, b$. Since the possible values of each $S_e(\omega)$ are ± 1, the system has 2^6 states. If the hidden variable theory plus an ignorance assumption about the true values of the hidden variables is supposed to produce the quantum mechanical probabilities, then we must rule out most of these states (or rather assign them probability zero): we saw above that two consecutive spin measurements along the same direction for both electrons always give opposite results. In other words, $S_b(\omega) = -S_a(\omega)$ for any direction ω. We are thus left with only 2^3 states.

Theorem 12.6 (Bell inequality)**.** *Define*

$$p(\alpha, \beta) = P(S_a(\alpha) = S_b(\beta) = 1).$$

Then for any three directions α, β, γ and any probability distribution of the hidden variables, we have

$$p(\alpha, \gamma) \leq p(\alpha, \beta) + p(\beta, \gamma). \tag{12.3}$$

More precisely, this is one of many versions of the famous Bell inequalities. As we'll see, it is trivial mathematically; it is of great interest nevertheless because of its relevance in our current discussion.

Proof. Let (s, t, u) denote the state with $S_a(\alpha) = s$, $S_a(\beta) = t$, $S_a(\gamma) = u$. Then the event from the left-hand side of Bell's inequality is $A = \{(1, 1, -1), (1, -1, -1)\}$ while those on the right-hand side are $B = \{(1, -1, 1), (1, -1, -1)\}$, $C = \{(1, 1, -1), (-1, 1, -1)\}$. Obviously, $A \subseteq B \cup C$, so $P(A) \leq P(B \cup C) \leq P(B) + P(C)$. $\square$

Now let's compare this with the situation in quantum mechanics. We will now need more precise information on what observables exactly correspond to a given spin direction ω in physical space. As we discussed above, we can use a projection P on $\mathbb{C}^2$, and the precise connection is as follows: The direction ω corresponds to the projection $P(\omega)$ onto $L(e(\omega))$, with

$$e(\omega) = \begin{pmatrix} \cos(\phi/2) \\ e^{i\theta} \sin(\phi/2) \end{pmatrix}. \tag{12.4}$$

This is not nearly as arbitrary as it might look at first. One intriguing observation is that this is the stereographic projection $S^2 \to \mathbb{C}_\infty$ in mild disguise: more precisely, (12.4) becomes this map if we view $e(\omega)$ as a point in projective space and then identify it with $e^{-i\theta} \cot(\phi/2) \in \mathbb{C}_\infty$. More importantly, (12.4) implements the SO(3) rotational symmetry of the physical space $\mathbb{R}^3$ in a natural way. One would expect that a change of coordinate system in $\mathbb{R}^3$ by a rotation matrix $R \in \mathrm{SO}(3)$ should leave the physics of the system unchanged. The structure preserving maps on the state space $H = \mathbb{C}^2$ are the unitary maps. There is a natural group isomorphism $\varphi : \mathrm{SU}(2)/\{-1, 1\} \to \mathrm{SO}(3)$, and a transformation of $e = (1, 0)^t$ by these unitary maps leads to (12.4).

Exercise 12.16. Notice that $S(\omega) = 2P(\omega) - 1$ has eigenvectors $e(\omega), f(\omega)$ with corresponding eigenvalues $1, -1$. So this is the observable that describes a measurement according to (O) of what we already called $S(\omega)$ above. Use (12.4) to show that

$$S(\omega) = \begin{pmatrix} \cos\phi & e^{-i\theta}\sin\phi \\ e^{i\theta}\sin\phi & -\cos\phi \end{pmatrix}. \tag{12.5}$$

We can now be slightly more precise about the isomorphism mentioned above: transforming such a spin matrix by a $U \in \mathrm{SU}(2)$, as follows, $S' = US(\omega)U^*$, is the same as rotating $x(\omega) \in \mathbb{R}^3$ by the rotation matrix $R = \varphi(U) \in \mathrm{SO}(3)$ that corresponds to U: we have $S' = S(\omega')$, $x(\omega') = Rx(\omega)$.

The point $x(\omega) \in S^2 \subseteq \mathbb{R}^3$ described by ω is explicitly given by $x(\omega) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$. Introduce the *Pauli spin matrices*

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then we can rephrase (12.5) as follows: $S(\omega) = x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3$, and here we wrote $x(\omega) = (x_1, x_2, x_3)$. Or, in more compact notation, $S(\omega) = x(\omega) \cdot \sigma$.

Exercise 12.17. Compute the commutators $[\sigma_j, \sigma_k]$. Conclude again that no two orthogonal spin components can be measured simultaneously.

Exercise 12.18. Show that $P(S_a(\alpha) = 1 | S_a(\beta) = 1) = \cos^2(\vartheta/2)$, where ϑ denotes the angle between $x(\alpha), x(\beta)$: $\cos\vartheta = x(\alpha) \cdot x(\beta)$.

What I wrote as a conditional probability really refers to the probability of measuring $S_a(\alpha) = 1$, immediately following a measurement of $S_a(\beta)$ that gave the result 1. So you want to work out the quantum mechanical probability $P(S(\alpha) = 1)$ for a system in state $e(\beta) \in \mathbb{C}^2$.

I now want to discuss Bell's inequality (12.3) quantum mechanically. First of all, observe that these probabilities also make sense in this setting because $S_a(\alpha), S_b(\beta)$ are simultaneously measurable. As we discussed above, we can use the operator $P(\alpha) \otimes 1$ (or $S(\alpha) \otimes 1$) to measure $S_a(\alpha)$ and then similarly $1 \otimes P(\beta)$ to measure $S_b(\beta)$. Clearly these two operators commute. In fact, we don't really need the general theory (Theorem 12.2) here. Since we are only interested in the outcome $S_a(\alpha) = S_b(\beta) = 1$, we can also just observe that the ONB $v \otimes w$, $v = e(\alpha), f(\alpha)$, $w = e(\beta), f(\beta)$ of $H \otimes H$ consists of simultaneous eigenvectors of $P(\alpha) \otimes 1$ and $1 \otimes P(\beta)$, and what we are interested in, namely that both operators have eigenvalue 1, happens precisely on the eigenvector $e(\alpha)\otimes e(\beta)$. This means that we can just measure the observable $P(\alpha)\otimes P(\beta)$, which is the projection onto this vector.

Thus, if our system is in state $v\in H\otimes H$, then

$$p(\alpha,\beta)=\|(P(\alpha)\otimes P(\beta))v\|^2.$$

I mentioned (but did not carefully establish) above that the spin observables $S(\omega)$ reflect the rotational symmetry of $\mathbb{R}^3$ that is expressed by the action of SO(3). We can therefore hope that it is permitted to simplify the calculation by assuming that $\alpha=(0,0)$, the z direction in $\mathbb{R}^3$. So $e(\alpha)=e=(1,0)^t$. If now v is the EPR state (12.2), then $(P(\alpha)\otimes P(\beta))v=(1/\sqrt{2})e\otimes P(\beta)f$, so if $\beta=(\phi,\theta)$, then $p(\alpha,\beta)=(1/2)\sin^2(\phi/2)$. Since ϕ is the angle between α and β, it seems reasonable to expect the following result in the general case.

Proposition 12.7. *If the system is in the EPR state* (12.2)*, then*

$$p(\alpha,\beta)=\frac{1}{2}\sin^2\frac{\vartheta}{2}. \qquad (12.6)$$

Here ϑ again denotes the angle in $\mathbb{R}^3$ between the directions α,β.

Exercise 12.19. Establish (12.6) by an honest calculation, without making the simplifying assumption that $\alpha=(0,0)$.

For the comparison of this result with Bell's inequality, let's choose all three directions α,β,γ in the xy plane. Let's take α as the direction of the positive x axis, and let's then go counterclockwise to successively obtain β,γ. More formally, we can say that $\alpha=(\pi/2,0)$, $\beta=(\pi/2,2\theta)$, $\gamma=(\pi/2,2(\theta+\delta))$. Using (12.6), we can then rewrite Bell's inequality as

$$\sin^2(\theta+\delta)\le\sin^2\theta+\sin^2\delta.$$

This is false! It fails for all small $\theta,\delta>0$, or we could take $\delta=\theta$, and then the inequality says that $\sin^2 2\theta=4\sin^2\theta\cos^2\theta\le 2\sin^2\theta$ or $\cos^2\theta\le 1/2$, which is obviously not always true. A hidden variable theory of the type envisaged makes predictions that are in conflict with quantum mechanics. This leads to the unexpected possibility of subjecting such assumptions to experimental tests.

Exercise 12.20. Show that the EPR state is the only state that obeys the *total spin in any direction equals zero* conservation law that was mentioned above. In more formal language, show that there is a unique state in H such that a measurement of $S(\omega)\otimes 1+1\otimes S(\omega)$ will give the result 0 with certainty, for any direction ω, and that this state is the EPR state.