The Mathematical Concept of Dirac Notation

Michael F. Herbst

22. July 2014

1 Motivation

Dirac notation or bra-c-ket notation is a common feature when describing states in quantum mechanics. Its great success originates from its structural simplicity whilst still being widely applicable. In some sense one could say it manages to hide the mathematical subtleties when dealing with infinite dimensional spaces, such that up to a large extent we can apply our intuitive logic from finite dimensions.
This talk tries to give some insight on the mathematical structure behind Dirac notation and tries to hint at why and how it works.

Before we go into the details of how to abstractly define the spaces on which Dirac notation lives, let us first get a starting point by considering the familiar *Euclidean space*.

1.1 The structure of Euclidean space

Euclidean space is the set of all Euclidean vectors, which in turn are the usual arrows, which are characterised by their length and direction. In the representation of $\mathbb{R}^3$, we can denote a vector $\underline{\boldsymbol{x}}$ as

$$\underline{\boldsymbol{x}} \equiv \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (x_1\ x_2\ x_3)^T \tag{1.1}$$

where the x_1, x_2 and x_3 are just the (real) coordinates of the point the arrow shifts the origin to. In other words, the Euclidean vectors are just position vectors of points in $\mathbb{R}^3$. For the rest of this chapter we will lift the distinction in the notation of points and position vectors and use $\underline{\boldsymbol{x}}$ and the coordinate triple $(x_1\ x_2\ x_3)$ to refer to both objects.

Basic operations If we accept the sum of two vectors $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ as simply pasting one arrow after the other (see fig. 1), we may write

$$\underline{\boldsymbol{x}} + \underline{\boldsymbol{y}} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}. \tag{1.2}$$

Furthermore we can also consider multiplication by a real scalar α, such that

$$\alpha\underline{\boldsymbol{x}} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{pmatrix}. \tag{1.3}$$

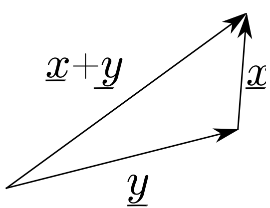

Figure 1: Illustration of vector addition by pasting $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ together.

Pictorially we can interpret this as pasting α versions of the vector $\underline{\boldsymbol{x}}$ together to get the resulting $\alpha\underline{\boldsymbol{x}}$.

Distance In order to formally describe the length of an Euclidean vector $\underline{\boldsymbol{x}}$, recall that it is spanned between two points: The origin and $(x_1\ x_2\ x_3)$. Applying Pythagoras' theorem in three space dimensions yields an expression for the length (or *norm*) of a vector:

$$\|\underline{\boldsymbol{x}}\| = \sqrt{x_1^2 + x_2^2 + x_3^2}. \tag{1.4}$$

It is easy to see that the vector $\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}$ is exactly the arrow that shifts $\underline{\boldsymbol{x}}$ to $\underline{\boldsymbol{y}}$. The distance between $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ is thus just $\|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}\|$.

Scalar product Using basic trigonometry one can show that the Euclidean scalar product

$$(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}) = \underline{\boldsymbol{x}} \cdot \underline{\boldsymbol{y}} = x_1 y_1 + x_2 y_2 + x_3 y_3 \tag{1.5}$$

may be used to compute the angle between two vectors $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ via the relation

$$\cos(\angle(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}})) = \frac{(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}})}{\|\underline{\boldsymbol{x}}\|\,\|\underline{\boldsymbol{y}}\|}. \tag{1.6}$$

This has two important consequences:

- Firstly it implies that the definition of a scalar product already gives some kind of an angle. Consider for example the familiar result that two non-zero vectors $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ are orthogonal iff[1] $(\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}}) = 0$.
- Secondly, note that in Euclidean space $\|\underline{\boldsymbol{x}}\| = \sqrt{(\underline{\boldsymbol{x}}, \underline{\boldsymbol{x}})}$. We say that the scalar product $(\cdot, \cdot)$ induces the norm $\|\cdot\|$.

For our ultimate goal of abstracting the structure of Euclidean space, we note that having the notion of some kind of scalar product appears to be enough to measure angles and lengths.

1.2 Calculus, limits and completeness

The final thing we will need to discuss about Euclidean space is why calculus works in $\mathbb{R}^3$. Calculus is loosely speaking the study of small changes — consider the derivative or the (Riemann) integral for example. Therefore the fundamental tool we need to demand from a space is the ability to take limits[2]. More specifically we need one property, that does indeed hold for Euclidean space, called *completeness*. Loosely speaking a space V is complete if every sequence $\{a_k \in V\}_{k\in\mathbb{N}}$ that converges[3], has a limit that is also a member of this space.
At first glance this seems to be a rather trivial statement, so let us look at a counterexample.

Example 1.1. Consider the sequence

$$a_k = \sum_{n=0}^{k} \frac{1}{n!}.$$

Each term of the sum is clearly a rational number. So for finite k, we surely have $a_k \in \mathbb{Q}$. In the limit of $k \to \infty$, the sum tends to Euler's number (think about the Taylor expansion of $\exp(x)$). In a mathematically formal way we write this as

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} \text{ such that } \left| \sum_{n=0}^{k} \frac{1}{n!} - e \right| < \varepsilon \qquad \forall k > N. \tag{1.7}$$

Note, however, that $e \notin \mathbb{Q}$, which implies that $\mathbb{Q}$ is not complete.

[1] if and only if
[2] i.e. that we can consider the $\lim_{x\to y}$ for some structure
[3] for those who know the definition: What is meant is that the sequence is Cauchy.

So say we really wanted to do calculus in $\mathbb{Q}$, how would we proceed? The obvious choice would be to form a new set, which not only contains $\mathbb{Q}$, but also all the limits of all possible sequences in $\mathbb{Q}$ as well. This process (called *completion* of $\mathbb{Q}$) is one way of constructing the set of real numbers $\mathbb{R}$, making them trivially complete. With this in mind it is relatively easy to show that Euclidean space $\mathbb{R}^3$ is complete as well, which we will not attempt to do here, however.

Instead we take a brief look at what an equivalent version of (1.7) for the convergence of sequences of Euclidean vectors would look like. One says a sequence $\{\underline{\boldsymbol{x}}_n \in \mathbb{R}^3\}_{n\in\mathbb{N}}$ converges to a limit $\underline{\boldsymbol{l}}$ iff

$$\forall\varepsilon > 0 : \exists N \in \mathbb{N} \text{ such that } \|\underline{\boldsymbol{x}}_n - \underline{\boldsymbol{l}}\| < \varepsilon \qquad \forall k > N. \tag{1.8}$$

Note, that this statement does depend on the norm $\|\cdot\|$, i.e. the precise way we measure lengths and distances in the space.

It is easy to imagine that the definition of completeness therefore depends on the norm as well. In fact for many spaces there is more than one way to define a norm and it may well happen that the space is complete with respect to one norm, but not another. So to be precise we should say that Euclidean space $\mathbb{R}^3$ is complete with respect to the Euclidean norm (1.4).

2 Hilbert spaces

Next we turn our attention to Hilbert spaces, which will turn out to be the abstract counterparts of Euclidean space.

In order to know what we are talking about, it is necessary to properly define some mathematical structures. Usually one does this by listing the axioms which those structures should satisfy in a very abstract way. Whilst this surely can be quite boring, note that the key points here are only that you get the general idea why things are defined the way they are. The precise mathematical formulation is only of minor importance.

2.1 Formally building up a Hilbert space

With that word of warning, we will now formally build the Hilbert space by splitting up its definition into some (hopefully easier to digest) chunks. Let us start with a vector space:

Definition 2.1. A *vector space* over a field[4] $\mathbb{F}$ is a non-empty set V together with two operations

$$\begin{aligned} + : V \times V \to V : (x, y) \mapsto x + y & \qquad \text{"vector addition"} \\ \cdot : \mathbb{F} \times V \to V : (\alpha, x) \mapsto \alpha x, & \qquad \text{"scalar multiplication"} \end{aligned}$$

that satisfy some axioms listed below. The elements of V are referred to as vectors and the elements of the field $\mathbb{F}$ are called scalars.

The axioms to be satisfied are:

- $(V, +)$ is an abelian group, i.e. it holds

$$\forall x, y, z \in V : \quad x + (y + z) = (x + y) + z \qquad \text{"associativity of addition"} \tag{2.1}$$

$$\exists 0 \in V : \quad \forall x \in V : x + 0 = x \qquad \text{"existence of zero vector"} \tag{2.2}$$

$$\forall x \in V : \quad \exists -x \in V : x + (-x) = 0 \qquad \text{"existence of additive inverses"} \tag{2.3}$$

$$\forall x, y \in V : \quad x + y = y + x \qquad \text{"commutativity of addition"} \tag{2.4}$$

- The so-called compatibility of scalar multiplication and field multiplication:

$$\forall \alpha, \beta \in \mathbb{F}, x \in V : \quad (\alpha\beta)x = \alpha(\beta x) \tag{2.5}$$

- The scalar multiplication with the multiplicative identity of $\mathbb{F}$ (i.e. the scalar 1) should satisfy

$$\forall x \in V : \quad 1\, x = x \tag{2.6}$$

- The following distributivity relations:

$$\forall \alpha \in \mathbb{F}, x, y \in V : \quad \alpha(x + y) = \alpha x + \alpha y \tag{2.7}$$

$$\forall \alpha, \beta \in \mathbb{F}, x \in V : \quad (\alpha + \beta)x = \alpha x + \beta x \tag{2.8}$$

Remark 2.2. We are only going to look at real or complex vector spaces, so for our considerations we will from here on assume that $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$.

Note. As stated above, this is all rather technical, so let us sum briefly up the rationale and the key ideas:

- Recall the picture of (1.2) where we pasted arrows together in order to add them. It is obvious that such an operation should satisfy (2.1) and (2.4), just because the precise order in which we apply the vectors does not matter. Surely if one adds the zero vector to any arrow, that arrow is unchanged, rationalising (2.2). (2.3) is also easy to understand if one sees the inverse as the original arrow with head and tail swapped.

[4] A field is another basic mathematical structure. If you are unaware of the definition just replace $\mathbb{F}$ by the complex numbers $\mathbb{C}$ in definition 2.1. We will restrict ourselves in remark 2.2 to the special fields of complex or real numbers anyways.

- Condition (2.5) establishes a formal consistency when scaling vectors multiple times.
- (2.6) gives us a uniquely defined scalar multiplication, since it defines the outcome of the multiplication operator for one specific scalar. The result for other scalars can than be deduced by applying distributivity and (2.5). It is remotely comparable to choosing the potential offset when calculating the energy.
- The distributivity relations (2.7) and (2.8) just give us the intuitive result that the order of vector/scalar addition and scaling of vectors does not matter.

As a nice example we are now going to show (at least partly)

Proposition 2.3. *Euclidean space $\mathbb{R}^3$ is a vector space over $\mathbb{R}$.*

The essential idea behind the proof will be to use the well-known properties from $\mathbb{R}$ in order to deduce the corresponding axioms in $\mathbb{R}^3$. Since this is a pretty boring task we are only going to formally proof some of the axioms, illustrating this key point. The rest is left as an exercise for the reader.

Proof. First note, that $\mathbb{R}^3$ is non-empty, since $(0\ 0\ 0)^T \in \mathbb{R}^3$. Now choose arbitrary $x, y \in \mathbb{R}^3$ and arbitrary $\alpha, \beta \in \mathbb{R}$. This implies (recall (1.1)) that we can find some $x_1, x_2, x_3,\ y_1, y_2, y_3 \in \mathbb{R}$ with

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

- Now use the previously established rule (1.2) and commutativity in $\mathbb{R}$ to write

$$x + y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} = \begin{pmatrix} y_1 + x_1 \\ y_2 + x_2 \\ y_3 + x_3 \end{pmatrix} = y + x, \tag{2.9}$$

  which shows that we satisfy (2.4).

- From (1.3) and associativity of the multiplication in $\mathbb{R}$ we show the compatibility condition (2.5).

$$(\alpha\beta)x = \begin{pmatrix} (\alpha\beta)x_1 \\ (\alpha\beta)x_2 \\ (\alpha\beta)x_3 \end{pmatrix} = \begin{pmatrix} \alpha(\beta x_1) \\ \alpha(\beta x_2) \\ \alpha(\beta x_3) \end{pmatrix} = \alpha(\beta x) \tag{2.10}$$

- Apply both rules and use distributivity in $\mathbb{R}$ to see that the axiom (2.7) holds.

$$\alpha(x + y) = \begin{pmatrix} \alpha(x_1 + y_1) \\ \alpha(x_2 + y_2) \\ \alpha(x_3 + y_3) \end{pmatrix} = \begin{pmatrix} \alpha x_1 + \alpha y_1 \\ \alpha x_2 + \alpha y_2 \\ \alpha x_3 + \alpha y_3 \end{pmatrix} = \alpha x + \alpha y \tag{2.11}$$

- and so on

□

The next step is to add a scalar product, also known as an inner product.

Definition 2.4. An *inner product space* over $\mathbb{F}$ is a vector space V (over the same field) that is further equipped with an inner product, i.e. a map

$$(\cdot, \cdot) : V \times V \to \mathbb{F}$$

that satisfies (for all vectors $x, y, z \in V$ and all $\alpha \in \mathbb{F}$)

$$(x, y)^* = (y, x) \qquad \text{"Conjugate symmetry"} \tag{2.12}$$

$$(\alpha x + y, z) = \alpha(x, z) + (y, z) \qquad \text{"Linearity in the first argument"} \tag{2.13}$$

$$(x, x) \geq 0 \quad \text{and} \quad (x, x) = 0 \ \Rightarrow \ x = 0 \qquad \text{"Positive-definiteness"} \tag{2.14}$$

here the asterisk * denotes complex conjugation.

Remark 2.5. Some other well-known properties of the inner product are in fact not axioms but can be easily deduced by combining two of the above:

- Conjugate symmetry (2.12) implies *antilinearity* (or *conjugate linearity*) in the second argument

$$(z, \alpha x + y) = \alpha^*(z, x) + (z, y) \tag{2.15}$$

  In other words the inner product is what is called a *sesquilinear form.*

- If $\mathbb{F} = \mathbb{R}$, then Conjugate symmetry (2.12) simplifies to proper symmetry, ie

$$(x, y) = (y, x).$$

  The antilinearity in the second argument then also reduces to proper linearity

$$(z, \alpha x + y) = \alpha(z, x) + (z, y), \tag{2.16}$$

  such that this time we get a so-called *bilinear form.*

Note. From your previous encounters with Dirac notation $\langle\cdot|\cdot\rangle$ you might think that it satisfies all these properties and should be regarded as an inner product as well. In a loose sense this is not actually a bad description and it probably suffices as working knowledge. From a mathematically precise perspective this is, however, not correct. Nevertheless both objects, the inner product $(\cdot, \cdot)$ and the Dirac bra-c-ket $\langle\cdot|\cdot\rangle$, are highly related. This will become clear from the formal definition of the latter, which we will present later.

As expected one can show:

Proposition 2.6. *Euclidean space* $\mathbb{R}^3$ *is an inner product space over* $\mathbb{R}$.

Proof. Left as an exercise for the reader. □

Next we proceed to discuss the norm:

Definition 2.7. Given a vector space over the field $\mathbb{F}$, a *norm* is a map

$$\|\cdot\| : V \to \mathbb{R}$$

such that the following axioms hold for all vectors $x, y \in V$ and all $\alpha \in \mathbb{F}$:

$$\|\alpha x\| = |\alpha| \, \|x\| \qquad \text{"Absolute scalability"} \tag{2.17}$$

$$\|x + y\| \leq \|x\| + \|y\| \qquad \text{"Triangle inequality"} \tag{2.18}$$

$$\text{If } \|x\| = 0 \quad \Rightarrow \quad x \text{ is the zero vector} \qquad \text{"Norm separates points"} \tag{2.19}$$

Remark 2.8. Whilst absolute scalability (2.17) seems like a reasonable thing to ask for when defining the norm (again think about scaling arrows), the other two might not be as obvious from the start.

- The abstract formulation of the triangle inequality (2.18) is — as the name suggests — a direct generalisation of the geometric constraints for a basic triangle: One side always has to be shorter than the sum of the two others. From the triangular structure that vector addition gives (see fig. 1 on page 1) the axiom can be easily illustrated in the case of Euclidean vectors.

- The norm separates points condition again is one of those axioms to give a uniquely defined structure, like (2.6) for the scaling of vectors. The choice is the most natural as it leads to a positive norm which can be seen as follows. From absolute scalability we have $\|-x\| = \|x\|$ for each vector $x \in V$ and furthermore for the zero vector $\|0\| = 0$. By the triangle inequality $0 = \|0\| = \|x - x\| \leq \|-x\| + \|x\| = 2\,\|x\|$. Dividing both sides by 2 gives the result $\|x\| \geq 0$.

Proposition 2.9. *For every inner product space there exists a norm given by*

$$\|x\| = \sqrt{(x, x)} \qquad \forall x \in V, \tag{2.20}$$

the so-called induced norm.

Proof. Consider arbitrary $x, y \in V$ and $\alpha \in \mathbb{F}$

- The point separating property (2.19) is a trivial consequence of (2.14).
- We get absolute scalability from linearity (2.13) and antilinearity (2.15) of the scalar product

$$\|\alpha x\| = \sqrt{(\alpha x, \alpha x)} = \sqrt{\alpha \alpha^* (x, x)} = |\alpha| \, \|x\|$$

- Using (2.19) and again linearity we find

$$\begin{aligned} 0 &\leq \|x \pm y\|^2 \\ &= (x \pm y, x \pm y) = (x, x) \pm (x, y) \pm (y, x) + (y, y) \\ &= (x \pm y, x \pm y) = (x, x) \pm (x, y) \pm (x, y)^* + (y, y) \\ &= \|x\|^2 + \|y\|^2 \pm 2 \operatorname{Re}(x, y) \end{aligned} \tag{2.21}$$

Now define normalised vectors $\tilde{x} = \frac{x}{\|x\|}$ and $\tilde{y} = \frac{y}{\|y\|}$. In an analogous way to the negative branch in (2.21) above, we have

$$0 \leq \|\tilde{x}\|^2 + \|\tilde{y}\|^2 - 2 \operatorname{Re}(\tilde{x}, \tilde{y}) = 2 - 2 \operatorname{Re}(\tilde{x}, \tilde{y}),$$

such that $\operatorname{Re}(\tilde{x}, \tilde{y}) \leq 1$. Now consider

$$\begin{aligned} (\|x\| + \|y\|)^2 - \|x + y\|^2 &\overset{(2.21)}{=} 2 \|x\| \|y\| - 2 \operatorname{Re}(x, y) \\ &= 2 \|x\| \|y\| - 2 \operatorname{Re}(\|x\| \tilde{x}, \|y\| \tilde{y}) \\ &= 2 \|x\| \|y\| [1 - \operatorname{Re}(\tilde{x}, \tilde{y})] \\ &\geq 0 \end{aligned} \tag{2.22}$$

where we used linearity in the penultimate step, to get the (real) factors $\|x\|$ and $\|y\|$ out front. Taking the square root proves the triangle inequality (2.18).

□

Corollary 2.10 (Cauchy-Schwarz Inequality)**.** *Let V be an inner product space with inner product $(\cdot, \cdot)$ and induced norm $\|\cdot\|$. Then for all $x, y \in V$*

$$\|x\| \|y\| \geq \operatorname{Re}(x, y)$$

and furthermore

$$\sqrt{2} \|x\| \|y\| \geq |(x, y)| \, .$$

Proof. The first part has been implicitly achieved in (2.22). For the second part, note

$$|(x, y)|^2 = \Big(\operatorname{Re}(x, y) \Big)^2 + \Big(\operatorname{Im}(x, y) \Big)^2 = \Big(\operatorname{Re}(x, y) \Big)^2 + \Big(\operatorname{Re}(y, x) \Big)^2 \leq 2 \|x\|^2 \|y\|^2 \, .$$

Taking the square-root shows the claim. □

Finally, we arrive at the definition of a *Hilbert space*:

Definition 2.11. A Hilbert space V is an inner product space over the field $\mathbb{F}$, that is complete with respect to the induced norm.

Note. A formal discussion of completeness is omitted here. The picture we developed in section 1.2 on page 2 will be sufficient for our purposes and shall serve as an intuitive guideline whenever the term *complete* arises.

As we said in the last chapter, Euclidean space is indeed complete under its induced norm and can hence be regarded as a Hilbert space. We therefore succeeded in finding an abstract formalism that represents large parts of the internal mathematical structure of Euclidean space.

2.2 Application: L^2 — the Lebesgue space of square-integrable functions

In this subsection we will try to get a little insight into the space of square-integrable functions L^2. As it will turn out later, this space is the mathematical object on which Dirac Notation lives in most practical scenarios. In a similar fashion to completeness, the precise formulation of *integrability* is well out of scope here[5]. For us this section is merely meant to scratch the surface and introduce the formalism we will need later on.

The first problem one runs into very quickly when considering square-integrable functions is that our familiar Riemann integral definition is too tight. This means that there are too few functions that can be properly Riemann-integrated in order to prove the theorems in this subsection. The way to make progress is to broaden the definition of the integral and integrability. Hence we consider

Remark 2.12 (The Lebesgue integral)**.** Instead of a formal definition of the Lebesgue integral, we will just state a few of its properties in order to get a feeling for how it works.

- Whenever the Riemann integral exists, the values of the Riemann and the Lebesgue integrals agree. The only difference between the two integral definitions is that there exist some "problematic cases" where the Lebesgue integral still allows integration, but the Riemann integral does not. So if the Riemann integral exists (i.e. if we can do the usual integration and the result is a finite value), we can evaluate it and we are done for the calculation of the integral in both definitions.

- For a Lebesgue integral over a domain $\Omega \subset \mathbb{R}^1$, changing the value of the integrand at up to countably infinite number of points (in an arbitrary way) does not change the value of the integral. Note, that this means for example

$$\begin{aligned} &\int_{\mathbb{R}} \left\{ \begin{array}{ll} 1 & \text{for } x \in \mathbb{Q} \\ 0 & \text{for } x \in \mathbb{R}\backslash\mathbb{Q} \end{array} \right\} \mathrm{d}x && \text{only Lebesgue-integrable} \\ = &\int_{\mathbb{R}} 0 \, \mathrm{d}x && \text{Lebesgue- and Riemann-integrable} \\ = &\int_{\mathbb{R}} \left\{ \begin{array}{ll} 1 & \text{for } x \in \{\pi, e, 3, 15, 42\} \\ 0 & \text{else} \end{array} \right\} \mathrm{d}x && \text{Lebesgue- and Riemann-integrable,} \end{aligned} \tag{2.23}$$

  which trivially evaluates to zero (Consider the Riemann-integrable second term). For higher dimensions analogous statements are true: e.g. for integrals over subsets of $\mathbb{R}^2$, we can change the value of the integrand at up to countably infinite number of lines or points; similarly for 3D integrals at planes, lines or points, ... Often this allows us to alter an integrand in a way that integration in a Riemann sense is possible. For example in the first integral of (2.23), we changed the points where the function is 1 to 0 in order to get a Riemann-integrable version. This is possible since $\mathbb{Q}$ is a set with only up to countably infinite number of elements[6].

- A direct consequence of the above point is that for each $\Omega \subset \mathbb{R}^n$ there exist an infinitely large number of functions $f : \Omega \to \mathbb{C}$, such that $\int_\Omega f \, \mathrm{d}x = 0$. This will be important very soon.

- From now on all integrals in this script are to be understood as Lebesgue integrals.

That being said, let us consider complex integrability.

Definition 2.13. We call a *complex-valued* function $g : \Omega \to \mathbb{C}$ *integrable* on the domain $\Omega \subset \mathbb{R}^n$ if

$$\left| \int_\Omega g(x) \, \mathrm{d}x \right| < \infty, \tag{2.24}$$

i.e. if the modulus of the integral over the domain remains finite.

[5]In fact the whole topic of measure theory deals with just this definition.

[6]It is not necessarily intuitive to understand, what this means. Just imagine there are two kinds of infinities: For one kind one can abstractly speaking find a way to enumerate each element (the "countable infinity"), for the other one this is not possible.

Definition 2.14. A complex-valued function $f : \Omega \to \mathbb{C}$ is called *square-integrable* on the domain $\Omega \subset \mathbb{R}^n$ if the function $|f|^2 : \Omega \to \mathbb{R} : x \mapsto |f(x)|^2$ is integrable on that very same domain. This means nothing else but

$$\int_\Omega |f|^2 \, \mathrm{d}x < \infty.$$

Note. Usually one denotes the set of square-integrable functions as

$$L^2(\Omega) := \big\{ f : \Omega \to \mathbb{C} \;\big|\; f \text{ is square-integrable} \big\}. \tag{2.25}$$

Before we will eventually proceed to show that $L^2(\Omega)$ is actually a Hilbert space we need a few more intermediate results.

Lemma 2.15 (Minkowski's inequality in L^2)**.** *Let $f, g \in L^2(\Omega)$, i.e. square-integrable, then*

- *$f + g$ is square-integrable on Ω as well*
- $\sqrt{\int_\Omega |f+g|^2 \, \mathrm{d}x} \le \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x} + \sqrt{\int_\Omega |g|^2 \, \mathrm{d}x}$

Proof. See for example [1]. □

Note. When we say the resulting function $f \circ g$ of an operation $\circ$ is to be *defined pointwise*, we mean that $f \circ g$ can be defined by considering the action of $\circ$ in each point $x \in \Omega$ on the function values $f(x)$ and $g(x)$. For example the function $f + g$ from lemma 2.15 could be defined as the function for which

$$\Big(f + g \Big)(x) = f(x) + g(x) \qquad \forall x \in \Omega.$$

Similarly $\operatorname{Re} f$ is the function with

$$\Big(\operatorname{Re} f \Big)(x) = \operatorname{Re}\Big(f(x) \Big) \qquad \forall x \in \Omega.$$

Lemma 2.16. *For all complex-valued functions f, g, which are defined and integrable on $\Omega \subset \mathbb{R}^n$*

$$\int_\Omega |fg| \, \mathrm{d}x \le \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x} \cdot \sqrt{\int_\Omega |g|^2 \, \mathrm{d}x}.$$

Proof. This is a special case of the so-called Hölder's inequality, a proof of which can be found in [1]. □

From this we deduce

Lemma 2.17. *For all complex-valued functions f, g integrable on $\Omega \subset \mathbb{R}^n$, it holds*

$$\left| \int_\Omega f g^* \, \mathrm{d}x \right| \le \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x} \cdot \sqrt{\int_\Omega |g|^2 \, \mathrm{d}x}$$

and

$$f, g \in L^2(\Omega) \quad \Rightarrow \quad fg^* \textit{ is integrable},$$

where the function fg^* is to be constructed from f and g in a pointwise sense.

Proof. Consider first an integrable function h which is defined on $\Omega \subset \mathbb{R}^n$, we have

$$\int_\Omega |h| \, \mathrm{d}x = \int_\Omega \sqrt{(\operatorname{Re} h)^2 + (\operatorname{Im} h)^2} \, \mathrm{d}x \ge \int_\Omega |\operatorname{Re} h| \, \mathrm{d}x.$$

Noting that $|\operatorname{Re} h| \ge \pm \operatorname{Re} h$ and thus

$$\int_\Omega |\operatorname{Re} h| \, \mathrm{d}x \ge \pm \int_\Omega \operatorname{Re} h \, \mathrm{d}x$$

we get

$$\int_\Omega |h| \, \mathrm{d}x \geq \left| \int_\Omega \operatorname{Re} h \, \mathrm{d}x \right|. \tag{2.26}$$

Now note that for each complex number z we may find a $\phi \in \mathbb{R}$ such that $z \exp(\mathbf{i}\,\phi)$ is a real number. So specifically we can choose ϕ such that

$$\begin{aligned}
\exp(\mathbf{i}\,\phi) \int_\Omega h \, \mathrm{d}x &= \operatorname{Re}\left(\exp(\mathbf{i}\,\phi) \int_\Omega h \, \mathrm{d}x \right) \\
&= \operatorname{Re}\left(\int_\Omega \exp(\mathbf{i}\,\phi) h \, \mathrm{d}x \right) \\
&= \int_\Omega \operatorname{Re}\Big(\exp(\mathbf{i}\,\phi) h \Big) \, \mathrm{d}x
\end{aligned}$$

and therefore

$$\left| \int_\Omega h \, \mathrm{d}x \right| = \left| \exp(\mathbf{i}\,\phi) \int_\Omega h \, \mathrm{d}x \right| = \left| \int_\Omega \operatorname{Re}\Big(\exp(\mathbf{i}\,\phi) h \Big) \, \mathrm{d}x \right| \overset{(2.26)}{\leq} \int_\Omega |\exp(\mathbf{i}\,\phi) h| \, \mathrm{d}x = \int_\Omega |h| \, \mathrm{d}x$$

Now set $h = fg^*$ and use lemma 2.16 to obtain the first result

$$\left| \int_\Omega fg^* \, \mathrm{d}x \right| \leq \int_\Omega |fg^*| \, \mathrm{d}x = \int_\Omega |fg| \, \mathrm{d}x \leq \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x} \cdot \sqrt{\int_\Omega |g|^2 \, \mathrm{d}x}. \tag{2.27}$$

If furthermore $f, g \in L^2(\Omega)$ then the integrals in the RHS of (2.27) are finite (since the functions are square-integrable) and thus

$$\left| \int_\Omega fg^* \, \mathrm{d}x \right| \leq \int_\Omega |fg| \, \mathrm{d}x \leq \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x} \cdot \sqrt{\int_\Omega |g|^2 \, \mathrm{d}x} < \infty,$$

which implies the second result, that fg^* is integrable. □

Remark 2.18. For this lemma it is easy to see that an equivalent version cannot be constructed if the Riemann integral definition is to be used. To see this first define the two complex-valued functions $f, g : \mathbb{R} \to \mathbb{C}$ with

$$f(x) = \begin{cases} -1 & \text{for } x \in \mathbb{Q} \\ 1 & \text{for } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases} \qquad\qquad g(x) = \begin{cases} \mathbf{i} & \text{for } x \in \mathbb{Q} \\ 1 & \text{for } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases}.$$

Clearly

$$\forall x \in \mathbb{R} : \qquad |f(x)| = |g(x)| = 1,$$

such that in both the Riemann and the Lebesgue sense

$$\int_{[0,1]} |f|^2 \, \mathrm{d}x = \int_{[0,1]} 1 \, \mathrm{d}x = 1 = \int_{[0,1]} 1 \, \mathrm{d}x = \int_{[0,1]} |g|^2 \, \mathrm{d}x.$$

This makes f and g square-integrable functions on the domain $\Omega = [0, 1]$. On the other hand

$$(fg^*)(x) = \begin{cases} \mathbf{i} & \text{for } x \in \mathbb{Q} \\ 1 & \text{for } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases},$$

which makes (fg^*) Lebesgue, but not Riemann integrable. For lemma 2.17 this provides a counterexample if the Riemann integral definition is used.

Lemma 2.19. *For $\Omega \subset \mathbb{R}^n$, $L^2(\Omega)$ is an inner product space over $\mathbb{C}$ with the inner product given by the form*

$$(u, v)_\Omega = \int_\Omega uv^* \, \mathrm{d}x. \tag{2.28}$$

and a pointwise definition of vector addition and scalar multiplication, i.e. for functions f and g

$$\Big(f+g\Big)(x) = f(x) + g(x) \qquad \forall x \in \Omega$$

and a scalar multiple of f with $\alpha \in \mathbb{C}$

$$\Big(\alpha f\Big)(x) = \alpha f(x) \qquad \forall x \in \Omega.$$

Proof. First of all we need to verify that the pointwise vector operations are well-defined. This is the case if for all $f, g \in L^2(\Omega)$ and all $\alpha \in \mathbb{C}$ uniquely defined functions αf and $f + g$ exist that are also a member of $L^2(\Omega)$. Uniqueness follows from the properties of $\mathbb{C}$. For the sum $f + g$ the first part of Minkowski's inequality (2.15) guaranties that $f + g \in L^2(\Omega)$ if $f, g \in L^2(\Omega)$. For the scalar product this is a consequence of the linearity of the integral, i.e.

$$\begin{aligned} & f \text{ is square-integrable on } \Omega \\ \Leftrightarrow\quad & \int_\Omega |f(x)|^2 \,\mathrm{d}x < \infty \\ \Leftrightarrow\quad & \int_\Omega |\alpha f(x)|^2 \,\mathrm{d}x = |\alpha|^2 \underbrace{\int_\Omega |f(x)|^2 \,\mathrm{d}x}_{<\infty} < \infty \qquad \forall \alpha \in \mathbb{C} \\ \Leftrightarrow\quad & \alpha f \text{ is square-integrable on } \Omega \qquad \forall \alpha \in \mathbb{C} \end{aligned}$$

Showing now that $L^2(\Omega)$ is a vector space is pretty simple. The zero vector is just the zero function and the additive inverse to f is $-f$, i.e. the function with the signs of all function values reversed. To illustrate the argument we will show that the distributivity relation (2.7) holds. The remaining steps are left as an exercise for the reader.
So consider arbitrary $\alpha \in \mathbb{C}$ and arbitrary $f, g \in L^2(\Omega)$. We have

$$\forall x \in \Omega: \qquad \Big(\alpha(f+g)\Big)(x) = \alpha(f(x)+g(x)) = \alpha f(x) + \alpha g(x) = \Big(\alpha f + \alpha g\Big)(x).$$

Therefore the functions given by $\alpha(f + g)$ and $\alpha f + \alpha g$ have to be identical (since they are identical at all points x), which indeed proves (2.7).

Now we try to show that $L^2(\Omega)$ is an inner product space. The integral in the definition of the inner product (2.28) exists due to the second part of lemma 2.17, which gives that uv^* is integrable. Conjugate symmetry (2.12) and linearity in the first argument (2.13) are trivially satisfied by (2.28), due to the properties of the integral itself. Furthermore for all $f \in L^2(\Omega)$

$$(f, f)_\Omega = \int_\Omega f f^* \,\mathrm{d}x = \int_\Omega |f|^2 \geq 0.$$

The remaining point to show in order to verify that (2.28) gives indeed a positive-definite inner product is

$$(f, f)_\Omega = 0 \Rightarrow f = 0. \tag{2.29}$$

Recall the properties of the Lebesgue integral from remark 2.12. If the integral in $(f, f)_\Omega$ is zero for one specific function f then we can construct arbitrarily many functions for which this is the case as well. In a strict sense statement (2.29) can hence never be satisfied. The mathematical trick we employ here to overcome this limitation is to redefine the meaning of the equality = when comparing functions from $L^2(\Omega)$. One calls this new meaning of = "equality almost everywhere" and it describes something along the lines[7] of

$$f = g \quad \text{almost everywhere} \qquad \Leftrightarrow \qquad \int_\Omega f \,\mathrm{d}x = \int_\Omega g \,\mathrm{d}x.$$

In this weaker sense of equality between the $L^2(\Omega)$ functions f and 0, the relation (2.29) can indeed be proved, which makes $L^2(\Omega)$ a proper inner product space. □

[7] The mathematical precise formalism uses the notion of an equivalence relation and factor groups for the proper definition of $L^2(\Omega)$.

The last step — showing that $L^2(\Omega)$ is a Hilbert space — is pretty involved and we will just state the result here:

Theorem 2.20. *For (most)* $\Omega \subset \mathbb{R}^n$*,* $L^2(\Omega)$ *is a Hilbert space under the induced norm*

$$\|f\|_{L^2(\Omega)} = \sqrt{(f,f)_\Omega} = \sqrt{\int_\Omega |f|^2 \, \mathrm{d}x}. \tag{2.30}$$

Proof. See [1] □

Remark 2.21. Now what have these 4 pages of (more or less) rigorous mathematical labour given us? The important point is that we were able to show how the infinite dimensional function space $L^2(\Omega)$ behaves abstractly very similar to the finite dimensional Euclidean space. In other words one could well do analysis or linear algebra with functions instead of numbers or vectors. This might seem a little odd in the first place, but the fact that $L^2(\Omega)$ is a Hilbert space lets us deal with this in a fairly intuitive way. For Quantum Mechanics, where the wave functions are pretty much all members of (subspaces of) $L^2(\Omega)$, getting an idea of the structure of $L^2(\Omega)$ and the operations we can perform in this space can become quite important.

3 Riesz representation theorem

In this final section we turn our attention towards understanding the Riesz representation theorem, which will end up being the precise mathematical justification for Dirac notation itself. Abstractly speaking this theorem shows that there exists a so-called anti-isomorphic mapping between the actual Hilbert space V and something called its *dual space*. The latter is the set of some specific maps from V back into the underlying field $\mathbb{F}$ — the so-called *continuous linear one-forms*.

Definition 3.1. Let V, W be vector spaces over the field $\mathbb{F}$, equipped with respective norms $\|\cdot\|_V$ and $\|\cdot\|_W$. A mapping $\varphi : V \to W$ is called

- *linear* if
$$\forall \alpha \in \mathbb{F}, x, y \in V: \qquad \varphi(\alpha x + y) = \alpha\varphi(x) + \phi(y). \tag{3.1}$$

- *bounded* if
$$\forall x \in V: \qquad \sup_{x\in V, x\neq 0} \frac{\|\varphi(x)\|_W}{\|x\|_V} < \infty, \tag{3.2}$$
where sup refers to the *supremum*, ie the least upper bound of all possible numbers the fraction can take[8]. For many cases (and all cases we will consider) this condition is equivalent to the mapping being *continuous*.[9] Therefore the terms continuity and boundedness are often used interchangeably in the context of linear maps.

- a *continuous linear operator* if it is both bounded and linear.

- a *continuous linear one-form* if it is a continuous linear operator $\varphi : V \to \mathbb{F}$.
In the literature the alternative term *continuous linear functional* is used fairly frequently for these types of operators as well.

In the following proposition we define the dual space and state some of its fundamental properties.

Proposition 3.2 (Continuous dual space)**.** *Let V be a vector space over the field $\mathbb{F}$ equipped with a norm $\|\cdot\|_V$. We denote with*
$$V' = \{\varphi | \varphi : V \to \mathbb{F} \text{ is bounded and linear}\}$$
the set of continuous linear one-forms. Under pointwise defined addition and scalar multiplication, i.e.
$$\begin{aligned} \forall \varphi, \psi \in V': &\quad (\varphi + \psi)(x) = \varphi(x) + \psi(x) \\ \forall \varphi \in V', \alpha \in \mathbb{F}: &\quad (\alpha\varphi)(x) = \alpha\left(\varphi(x)\right), \end{aligned}$$
V' is a vector space over the field $\mathbb{F}$ as well. A norm on V' is given by
$$\|\varphi\|_{V'} = \sup_{x\in V, x\neq 0} \frac{|\varphi(x)|}{\|x\|_V}. \tag{3.3}$$
Thus V' is usually called the topological dual space *or* continuous dual space *(or sometimes just* dual space*) of V.*

Proof. Proving that V' is a vector space is pretty straightforward and left as an exercise for the reader. We will just show that (3.3) gives indeed a norm.
Absolute scalability (2.17) of $\|\cdot\|_{V'}$ is a trivial consequence of the pointwise defined scalar multiplication. Now consider arbitrary $\varphi, \psi \in V'$. The triangle equality (2.18) follows from the triangle inequality in $\mathbb{C}$:
$$\|\varphi + \psi\|_{V'} = \sup_{x\in V, x\neq 0} \frac{|\varphi(x) + \psi(x)|}{\|x\|_V} \leq \sup_{x\in V, x\neq 0} \frac{|\varphi(x)| + |\psi(x)|}{\|x\|_V} = \|\varphi\|_{V'} + \|\psi\|_{V'}$$

[8] In other words $\sup_i m_i = a$ iff $\forall i : m_i \leq a$ and there is no smaller a to achieve this.
[9] Here continuity is meant in the slightly more abstract topological sense. For those who are interested: The topological definition states that a function $f : V \to W$ is continuous if for every open subset $Y \subset W$, $f^{-1}(Y) \subset V$ is open as well.

Furthermore

$$\begin{aligned} 0 = \|\varphi\|_{V'} \quad &\Rightarrow \quad 0 = \sup_{x\in V, x\neq 0} |\varphi(x)| \\ &\Rightarrow \quad \forall 0 \neq x \in V : |\varphi(x)| \leq 0 \\ &\Rightarrow \quad \forall 0 \neq x \in V : \varphi(x) = 0 \\ &\Rightarrow \quad \varphi = 0 \end{aligned}$$

where we used the continuity of φ in the last step to fix the value at $x = 0$. This last relation is exactly condition (2.19), which makes (3.3) indeed a norm. □

Theorem 3.3 (Riesz representation theorem)**.** *Let V be a Hilbert space over $\mathbb{F}$ with the inner product denoted by $(\cdot, \cdot)$.*

i) For every $\varphi \in V'$ there exists a unique $x_\varphi \in V$ such that[10]

$$\forall y \in V : \qquad \varphi(y) = (y, x_\varphi) \tag{3.4}$$

ii) The corresponding map $\tau : V' \to V : \varphi \mapsto x_\varphi$ is a so-called isometric anti-isomorphism*, which means:*

- *It is a bijection*[11]*.*
- *It is an isometry, $\|x_\varphi\|_V = \|\varphi\|_{V'}$*
- *It is antilinear, $\tau(\alpha\varphi + \psi) = \alpha^*\tau(\varphi) + \tau(\psi)$*

where $\varphi, \psi \in V, \alpha \in \mathbb{F}$.

Proof. For the proof of part i) and the fact that τ is an isometry see [2]. For the rest note:

- τ is antilinear since the inner product is antilinear in the second argument (2.15).
- τ is injective since it is an isometry: For $\varphi, \psi \in V'$ with $\phi \neq \psi$:

$$\|\tau(\varphi) - \tau(\psi)\|_V = \|\tau(\varphi - \psi)\|_V = \|\varphi - \psi\|_{V'} \neq 0$$

hence $\tau(\varphi) \neq \tau(\psi)$.

- Now for each $u \in V$, we can define a one-form

$$\theta_u(x) = (x, u) \qquad x \in V.$$

This form is linear due to the linearity of the inner product in the first argument (2.13). Furthermore θ_u is bounded, since

$$\begin{aligned} \|\theta_u\|_{V'} &= \sup_{x\in V, x\neq 0} \frac{|\theta_u(x)|}{\|x\|_V} = \sup_{x\in V, x\neq 0} \frac{|(x,u)|}{\|x\|_V} \\ &\leq \sup_{x\in V, x\neq 0} \frac{\sqrt{2}}{\|x\|_V} \|u\|_V \|x\|_V = \sup_{x\in V, x\neq 0} \sqrt{2}\, \|u\|_V \\ &= \sqrt{2}\, \|u\|_V \end{aligned}$$

where we used Cauchy-Schwarz (corr. 2.10) in the second line. As a result $\theta_u \in V'$ and trivially $\tau(\theta_u) = u$. By this family of one-forms θ_u, we have $\tau(V') = V$, which makes τ surjective.

□

[10]Note that this is the "physicists" version of the Riesz Representation Theorem. Other versions — usually preferred from Mathematicians — write (3.4) as $\varphi(y) = (x_\varphi, y)$, ie the vectors in the inner product are reversed. This changes the properties of the map τ slightly, but not the overall meaning of the theorem.

[11]A bijective map establishes a one-to-one correspondence between the elements of two sets, such that every element of one set is paired exactly with one element from the other set.

Finally we can state

Definition 3.4 (Dirac notation)**.** Let V be a Hilbert space with corresponding (continuous) dual space V'. By the Riesz representation theorem for each linear one-form $\varphi \in V'$ we can find a unique $X \in V$, such that $\varphi(Y) = (Y, X)$ for all $Y \in V$. This means that the vector X uniquely represents the linear one-form φ and we may uniquely label the latter by the former. In the notion of Dirac notation this is done by writing $\langle X| := \varphi$. One refers to $\langle X|$ as a *bra vector* or just *bra* and to V' as the *bra space*. Furthermore for convenience the vectors $Y \in V$ are allowed to be written as so-called *kets* $|Y\rangle$ and V is sometimes given the name *ket space* as well.

Finally we define the result of applying the one-form $\langle X|$ to the ket $|Y\rangle$ as the *Dirac bra-c-ket*

$$\langle X|Y\rangle := \langle X|\Big(|Y\rangle\Big) = \phi\Big(|Y\rangle\Big) = \phi(Y) = (Y, X). \tag{3.5}$$

Remark 3.5 (Properties of Dirac Notation)**.** Let V be a Hilbert space with inner product $(\cdot, \cdot)$ and Dirac bracket $\langle\cdot|\cdot\rangle$.

- The most important result we get from the Rnesz theorem 3.3 is that the mathematical structure of the ket space V and the bra space V' are very much related (they are anti-isomorph). This is a direct consequence of the fact that the map τ is bijective and antilinear. For example consider a ket $|X\rangle$, that is shifted by a small shift $\varepsilon \in \mathbb{F}$ in the direction $|Y\rangle$. Since for each ket, we can find a corresponding bra (τ is bijective, ie a one-to-one mapping), we have

$$\tau\Big(|X\rangle + \varepsilon|Y\rangle\Big) = \tau\Big(|X\rangle\Big) + \varepsilon^*\tau\Big(|Y\rangle\Big) = \langle X| + \varepsilon^*\langle Y|$$

  So the motion of passing through the vector $|X\rangle$ along a direction $|Y\rangle$ translates to a similar motion with the corresponding kets — just mirrored along the real line (due to the complex conjugation). This means that all kinds of curves are identically translated (up to the complex conjugation) between bra and ket space. All physical "processes" can hence be equivalently described in both spaces — whatever turns out to be more convenient.

- From the properties of the inner product we can show for all $|X\rangle, |Y\rangle, |Z\rangle \in V$ and all $\alpha \in \mathbb{F}$

$$\begin{aligned}
&\langle X|Y\rangle = \langle Y|X\rangle^* \\
&\langle \alpha X + Y|Z\rangle = \alpha^*\langle X|Z\rangle + \langle Y|Z\rangle \\
&\langle X|\alpha Y + Z\rangle = \alpha\langle X|Y\rangle + \langle X|Z\rangle \\
&\langle X|X\rangle \geq 0 \quad \text{and} \quad \langle X|X\rangle = 0 \;\Rightarrow\; X = 0
\end{aligned}$$

  In other words the Dirac bracket has very similar properties to the inner product itself.

Remark 3.6. Some concluding remarks:

- In principle any Hilbert space is sufficient for Dirac notation. In quantum mechanics most differential equations produce solutions (ie states) that are members of the space of square-integrable functions $L^2(\Omega)$. We could, however, also choose our familiar Euclidean space $\mathbb{R}^3$. Phenomenologically this means that again we could get an intuitive understanding of Dirac notation by considering it in the context of $\mathbb{R}^3$.
- Not all important differential equations give square-integrable functions. Notably the plane waves $\exp\Big(-i(k,x)\Big)$ are not square-integrable and not even members of any Hilbert space. But as it turns out more general versions of the Riesz representation theorem 3.3 exist, which justify the use of Dirac notation in those cases as well.
- We did not consider the quantum mechanical notion of operators and how they might be applied to kets from the left and to bras "from the right".

References

[1] R. A. Adams. *Sobolev Spaces.* Academic Press, 1978.

[2] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods.* Springer, 1994.