Physics Notations

Dominik Tschimmel

Abstract

The language of (modern) physics is mathematics. Following the analogy of languages, physicists use some kind of dialect of mathematics. Often skipping rigor, even when there is a proper mathematical theory in the background. On the other hand, this "dialect" offers some remarkable tricks to recall and derive equations. This report focuses on some of the more prominent examples, intended for beginning undergraduate students.

Contents

1 Differentiation and Fractions

A function $f\colon \mathbf{R} \longrightarrow \mathbf{R}$ is called **differentiable** in $x_0 \in \mathbf{R}$, if the limit

$$\lim_{x \to x_0} \frac{f(x) - f_{x_0}}{x - x_0}$$

exists. In this case, one writes $f'(x_0)$ for the limit. Varying the point x_0, one obtains a function $f'\colon \mathbf{R} \longrightarrow \mathbf{R}$, $x \longmapsto f'(x)$. This function is called **derivative**. Among others, the following notations are common, not only in physics, to denote the derivative:

$$f'(x) \equiv \frac{d}{dx} f(x) \equiv \frac{df(x)}{dx} \equiv \frac{\partial f(x)}{\partial x} \equiv \partial_x f(x) \equiv D_x f \ .$$

Note, that some of these notations have different meanings in higher dimensions, but coincide in one dimension.

1.1 The differential and "fractions"

Historically, and still used that way in experimental physics lectures, one understands df as infinitesimal difference of f. In that sense, the notation $\frac{df}{dx}$ resembles the idea of the formal definition

$$\frac{df}{dx} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \ .$$

Remark 1.1.
In fact, there exist approaches to make infinitesimals rigorous, e.g. non standard analysis. Though calculus becomes reasonable simple in these approaches, having a rigorous mathematical theory requires constructions, that are far from simple, especially for first semester students. Furthermore, more advanced topics of mathematics rely on the standard epsilon-delta approach, with only sparse to no literature for the non standard analysis approach. For that reason, it is not advisable, to use infinitesimals for more than a vague intuition. So we continue by giving df a rigorous definition:.

Definition 1.2.
Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function. Then, the **differential** df is defined as

$$df(x) = f'(x)\ dx$$

Now, it is not a good habit, to define objects on yet to be defined objects. However, the underlying concepts would require more work. Formally df is a **differential one-form** and $(dx)_p$ is the basis for the cotangent space $T_p^*\mathbb{R} \cong \mathbb{R}$, defining a frame dx on the cotangent bundle $T^*\mathbb{R}$. Explaining the details would lead us far astray, so the reader can consider dx as the basis of a one dimensional vector space. So the vector $df(x)$ consists of the coefficient $f'(x)$ and the basis vector dx. Accordingly, it holds that

$$df(x) + dg(x) = f'(x)\ dx + g'(x)\ dx = (f'(x) + g'(x))\ dx \ .$$

The resemblance to integration $\int f(x)\ dx$ is no coincidence. In fact, differential forms are the natural integrands (in higher dimensions, on manifolds etc.).

Back to the "fractions". While it may be admissible as notation, there is no such thing as $\frac{1}{dx}$. The problems might not arise in one dimension, as there is a canonical embedding $\boldsymbol{e_1} \mapsto 1$ of the vector space $V = \mathbb{R}$ into the field $\mathbb{F} = \mathbb{R}$. However differential forms are defined for arbitrary dimensions:

$$df(\vec{x}) = \sum_{j=1}^{n} \partial_i f(\vec{x})\ dx_j\ .$$

Remark 1.3.
Vector spaces do not carry a product $V \times V \longrightarrow V$, nor is there a canonical construction. There do exist notions of products, which promote the vector space to an algebra or ring. However, even rings are only semi groups with respect to the multiplicative structure. That is, there is not division operation, which would be needed to give $\frac{1}{dx}$ a meaning.

1.2 Natural operations for fractions - the linearity of the derivative

Owing to the linearity of the derivative, the differential is also linear. This means, that

$$d(f(x) + g(x) = df(x) + dg(x) \qquad \text{and} \qquad d(c \cdot f(x)) = cdf(x)\ .$$

It comes as no surprise, especially since $\frac{df}{dx}$ is also a common notation for the derivative, that the linearity also holds for the these fractions.

$$\frac{df}{dx} + \frac{dg}{dx} = \frac{df + dg}{dx} = \frac{d(f+g)}{dx} \qquad \text{and} \qquad c \cdot \frac{df}{dx} = \frac{cdf}{dx} = \frac{d(c \cdot f)}{dx}\ .$$

However, that is as far, as the natural operations of fractions carry over to the differential fractions. The product rule leads to

$$d(f(x) \cdot g(x)) = g(x)\ df(x) + f(x)\ dg(x)\ .$$

But then, this does not lead to the product of the differential fractions:

$$\frac{df}{dx} \cdot \frac{dg}{dx} \neq \frac{df\,dg}{dx^2}\ .$$

1.3 Expanding/reducing fractions and the chain rule

Quite often, one finds expansions/reductions of fractions in derivations of the form

$$\frac{dg}{dx} = \frac{dg}{dx}\frac{df}{df} = \frac{dg}{df}\frac{df}{dx}\ ,$$

to obtain the chain rule. Although the left and right side are just the chain rule, one can not simply expand by df, nor by any function $f\colon \mathbb{R} \to \mathbb{R}$. Recall:

Theorem 1.4.
Let $f\colon I \subset \mathbb{R} \longrightarrow f(I) \subset \mathbb{R}$ and $g\colon J \subset \mathbb{R} \longrightarrow \mathbb{R}$ be functions, with $f(I) \subset J$, such that f is differentiable in $x_0 \subset I$ and g is differentiable in $f(x_0)$. Then the composition $g \circ f\colon I \longrightarrow \mathbb{R}$ is differentiable in x_0 and it holds that

$$(g \circ f)'(x_0) = g'(f(c_0)) \cdot f'(x_0) \ .$$

Here another habit of the physics literature appears. Most often, physicists do not care to write down the functional dependence explicitly. So the function f and the function values $f(x)$ are used interchangeably, although they are quite different objects. This "vague" notation allows to reduce equations considerably, and keeping the general structure of them for a wide range of cases. For that reason, not specifying dependencies does happen in different contexts.

On the other hand, the vague notation can be quite dangerous, as it hides the assumptions made to obtain the result, that may not hold in general. So it is often a good idea, to write the result with dependencies. For the expansion of fractions we obtain:

$$\frac{dg(x)}{dx} = \frac{dg}{df(x)}(f(x)) \cdot \frac{df(x)}{dx} \ .$$

A handy corollary for integration from theorem 1.4 is

$$dg(f(x)) = g'(f(x))df(x) = g'(f(x))f'(x)\ dx \ . \tag{1}$$

1.4 Inverting fractions and the rule for inverse functions

The last, though not as commonly encountered, fraction manipulation is the inversion of fractions. The mathematical precise version, is to find the derivative of an inverse function f^{-1} for a given function f. The rule for the inverse function states:

Theorem 1.5.
Let f be a bijective function, that is differentiable at x_0, such that $f'(x_0) \neq 0$. Then the inverse function f^{-1} is differentiable at $f(x_0)$ with

$$(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)} \ .$$

In terms of the differential notation, the theorem reads

$$\frac{df^{-1}}{df(x_0)}(f(x_0)) = \frac{1}{\frac{df}{dx}(x_0)} \ .$$

Next, identifying $x = f^{-1}$ as is usually done with $y(x)$ and $x(y)$, and removing the functional dependencies in the notationl, we obtain

$$\frac{dx}{df} = \frac{1}{\frac{df}{dx}} \ .$$

Guided by this notation, we observe that

$$f'(f(x_0)) \cdot f'(x_0) = \frac{dx}{df}\frac{df}{dx} = 1 \ .$$

Of course, this is a valid result, as can be seen from theorem 1.5.

2 Integration in one dimension

Especially in later courses on quantum mechanics, there will appear a quit odd way of writing $\int_{\mathbb{R}} f(x)\ dx$ as $\int_{\mathbb{R}} dx\ f(x)$. One way of interpreting this notation, is to understand the part $\int_{\mathbb{R}} dx$ as map that acts on functions $f(x)$, assigning them the value $\int_{\mathbb{R}} f(x)\ dx$. More precisely, $\int_{\mathbb{R}} dx$ is the regular distribution T_1, defined by

$$T_1[\varphi] \coloneqq \int_{\mathbb{R}} 1 \cdot f(x)\ dx \equiv \int_{\mathbb{R}} dx\ f(x)\ .$$

So this notation has some connection to rigorous mathematics. In fact, the theory of distributions is the rigorous basis of objects like delta-functions. However, it is not always advisable to use this notation, as it hides some subtleties about integration.

2.1 Integration by substitution and the *dx*

It has already been mentioned, that differential forms are the natural integrands, not functions. Now, what is meant by this statement? In school, you might have already seen the example, why the dx is necessary. Here, we want to formalize this example further, embedding it into the context of differential forms.

Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be a differentiable map. This map can be regarded as coordinate transformation in the sense, that it rescales the x-axis. Then of course, the inverse, reverses the scaling, since $\phi \circ \phi^{-1} = \mathrm{Id}_{\mathbb{R}}$. If ϕ transforms coordinates, one has to account for this effect on the side of functions. This is done by $\phi^* f$, which just means

$$\phi^* f(x) = f(\phi(x))\ .$$

Understanding integration as measuring the oriented surface under the graph of a function, changing the scale of the x-axis, will change the result. To fix this, one applies the following rule:

$$\int_{\phi(\mathbb{R})} = \int_{\mathbb{R}} \phi^*\ .$$

Intuitively speaking, instead of changing the x-axis, one changes the functions.

Returning to the question about the dx. Let us now pretend, functions are the natural integrals. Then for $I = [a, b]$, we would get

$$\int_a^b f(x) = \int_I f(x) = \int_{\phi(\phi^{-1}(I))} f(x) = \int_{\phi^{-1}(I)} \phi^* f(x) = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(x))\ .$$

This however is not the substitution rule. You can check yourself by calculating some examples, that the above equation does not hold.

So at this point it becomes clear, that functions are not the proper integrand. On the other hand, one forms $f(x)\ dx$ are. To obtain the substitution rule, all we need to do, is to find the action fo ϕ on dx. One can show, that $\phi^* dx = d\phi(x) = \phi'(x)\ dx$. Then:

$$\boxed{\int_a^b f(x)dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} \phi^* f(x)\ dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(x))\phi'(x)\ dx\ ,}$$

which is the substitution rule.

Example 2.1.
So far, the dx can be dismissed as formality of mathematics. It turns out however, that the distinction between differential forms $f(x)\,dx$ and the corresponding (density) functions $f(x)$ is quite a physical one. Consider for example Planck's law. The intensity is usually written as

$$I(\nu) = \frac{2h\nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \quad \text{and} \quad I(\lambda) = \frac{2hc^3}{c^2}\frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \; .$$

When you use the frequency-wavelength relation of light, $\nu = \frac{c}{\lambda}$, you find, that the formulas seem to contradict. You are missing a factor of $\frac{c}{\lambda^2}$. With $d\nu = d(\frac{c}{\lambda}) = -\frac{c}{\lambda^2}\,d\lambda$, you obtain

$$I(\nu)\,d\nu = -I(\lambda)\,d\lambda \; .$$

Note, that the $-$ appears, because you change the orientation of the axis, i.e. increasing frequency corresponds to decreasing wave length.
This discrepancy on the function level hints at a very important physical interpretation. The intensity densities I depend on a choice of scaling. So they have no physical existence of their own. To fix this "gauge-variance", one has to add the scaling information by $d\nu$ or $d\lambda$. The physically relevant intensities $I(\nu)\,d\nu$ and $I(\lambda)\,d\lambda$ do not contradict (assuming oriented integration).

2.2 Examples of the integration operator

There are two prime examples for the integration operator and the fraction notation. The first is the derivation of the kinetic energy from the general definition. The second is the separation of variables technique to solve differential equations.

The definition of Energy is, to integrate the force (one-form) along the path. So in one dimension, $E = \int dx\; F(x)$. If the force is of the general form $F(x) = m\ddot{x}$, then

$$\begin{aligned} T &= \int dx\; F(x) = \int dt \frac{dx(t)}{dt}\; m \cdot \frac{d^2}{dt^2}x(t) = \int dt \frac{d}{dt}\left(\frac{1}{2}m\left(\frac{d}{dt}x(t)\right)^2\right) \\ &= \frac{1}{2}m\dot{x}(t)^2 + \text{const.} \; . \end{aligned}$$

Remark 2.2.
In the standard notation,using the substitution rule , the derivation looks as follows:

$$\begin{aligned} T &= \int F(x)\; dx = \int F(x(t))\; dx(t) = \int F(x(t))\; \dot{x}(t)\; dt = \int m \cdot \ddot{x}(t) \cdot x(t)\; dt \\ &= \int \frac{d}{dt}(\tfrac{1}{2}m\dot{x}(t)^2)\; dt = \int d(\tfrac{1}{2}m\dot{x}(t)^2) = \tfrac{1}{2}m\dot{x}(t)^2 + \text{const.} \; . \end{aligned}$$

For the second example, consider a differential equation $f'(x) = g(x) \cdot h(f(x))$. The

procedure of separating the variables works as follows:

$$f'(x) = g(x)h(f(x)) \quad \longrightarrow \quad \frac{df}{dx} = g(x)h(f) \quad \longrightarrow \quad \frac{1}{h(f)}\, df = g(x)\, dx$$
$$\longrightarrow \quad \int \frac{1}{h(f)} = \int g(x)\, dx$$

Remark 2.3.
Here, the general procedure is easily memorable. However, the proper mathematical procedure is not that difficult either, as it is only the substitution rule. Furthermore, it is more instructive about the dependencies. Assume that $h(f(x_0)) \neq 0$. Then there exists a neighborhood U around x_0, such that for all $x \in U$, the straightforward integration of $g(y) = \frac{f'(y)}{h(f(y))}$ is a solution. Then, all there is left to do, is to use the substitution rule:

$$\int_{x_0}^{x} g(y)\, dy = \int_{x_0}^{x} \frac{f'(y)}{h(f(y))}\, dy = \int_{f(x_0)}^{f(x)} \frac{1}{h(f)}\, df \ .$$

3 Tensors and tensor indices

Sooner or later, one will encounter terms like $\partial_\mu x^\mu$ or $R^\mu{}_\mu$, etc.. The concept behind this notation, or better called "set of rules to manipulate indices", is that of tensors and tensor fields. Asking a mathematician, tensors are introduced by a universal property. Asking a physicist, tensors are most likely described as components of "something", that transform in a certain way. The physicists point of view allows for a powerful set of rules for quick calculations. However, to get some insight, the mathematical approach is better suited.

3.1 The dual space

To understand the difference between v_μ and v^μ, one needs to understand the concept of the dual space. Let V be an $\mathbb{F}$ vector space with finite dimension, for simplicity. Here, $\mathbb{F}$ can be either $\mathbb{R}$ or $\mathbb{C}$. One can consider linear maps $\ell \colon V \longrightarrow \mathbb{F}$. These form a vector space themselves, defining

$$(\alpha_1\ell_1 + \alpha_2\ell_2)(v) = \alpha_1\ell_1(v) + \alpha_2\ell_2(v) \qquad \forall\ v \in V \ .$$

Definition and Lemma 3.1.
The vector space $V^* \coloneqq \{\ell \mid \ell \text{ is a linear map } V \longrightarrow \mathbb{F}\}$ is called **dual (vector) space**. If $\{e_i\}_{i=1,\dots,n}$ is a basis of V, then $\{\vartheta_i\}_{i=1,\dots,n}$, where ϑ_i is defined by

$$\vartheta_i(e_j) \coloneqq \delta_{ij} \equiv \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j \end{cases}$$

is a basis of V^*, called **dual basis**. Dual vectors are also called **covectors**.

Proof 3.2.
Let $\ell \in V^*$. Then there are numbers $\alpha_i \in \mathbb{F}$ with $\ell(e_i) = \alpha_i$. It follows that $\ell = \sum_i \alpha_i \vartheta_i$. Indeed:

$$\ell(v) = \ell\left(\sum_j v_j e_j\right) = \sum_j v_j \ell(e_j) = \sum_j v_j \alpha_j = \sum_{i,j} v_j \alpha_i \delta_{ij} = \sum_{ij} v_j \alpha_i \vartheta_i(e_j)$$
$$= \sum_i \alpha_i \vartheta_i \left(\sum_j v_j e_j\right) = \sum_i \alpha_i \vartheta_i(v) \equiv \left(\sum_i \alpha_i \vartheta_i\right)(v) \ .$$

□

Remark 3.3.
In the usual $\mathbb{R}^n$ basis representation, a vector $v = \sum_i v_i e_i$ is written as column vector

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \ .$$

Then, linear maps are written as matrices. Dual vectors as linear maps from V in $\mathbb{F}$ are then just $1 \times n$ matrices $\varphi = (\varphi_1, \ldots, \varphi_n)$ and the action $\varphi(v)$ becomes the matrix product:

$$\varphi(v) = (\varphi_1, \ldots, \varphi_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \varphi_1 \cdot v_1 + \ldots + \varphi_n \cdot v_n \ .$$

This implies, that $\dim(V) = \dim(V^*)$. Furthermore, choosing a basis $\mathcal{B} = \{e_i\}$ yields an isomorphism $I_{\mathcal{B}} \colon V \longrightarrow V^* \ , \ e_i \longmapsto \vartheta_i$. However, this isomorphism does depend on the chosen basis, i.e. is not canonical. If the vector space carries a real inner product, i.e. is Euclidean, the isomorphism becomes independent of any such choice.

Lemma 3.4.
Let V be a vector space with Euclidean inner product $\langle \cdot, \cdot \rangle$, then $\ell \colon V \longrightarrow V^* \ , \ v \longmapsto I(v)$, defined by

$$I(v)(w) = \langle v, w \rangle$$

is an isomorphism.

Proof 3.5.
Linearity follows from the linearity of the inner product. It can be shown, that there is an orthonormal basis $\{e_i\}$. Then

$$I(e_i)(e_j) = \langle e_i, e_j \rangle = \delta_{ij} \quad \Rightarrow \quad I(e_i) = \vartheta_i \ .$$

So $I \colon e_i \longrightarrow \vartheta_i$, which is bijective. □

Remark 3.6.
A more abstract way to prove bijectivity that emphasizes the independence of the choice of basis, looks as follows: From $\langle v, 0\rangle = 0$, injectivity follows, and surjectivity is then a consequence of the rank nullity theorem.

Since the dual space is itself a vector space, one can consider the dual space of the dual space, called **bidual space** V^{**}. For finite dimensional vector spaces, $L\colon V \longmapsto V^{**}\ ,\ v \longmapsto L_v$, defined by $L_v(\varphi) = \varphi(v)$, is a (canonical) isomorphism. For that reason, we may define:

$$v(\varphi) = \varphi(v) \qquad \forall\ v \in V\ ,\ \varphi \in V^*\ .$$

3.2 Tensors and the tensor space

One can introduce tensors with a universal property. This approach allows to show, that the tensor product of two vector spaces always exists and that it is unique up to isomorphy. Yet, that approach is highly abstract and can deter from the use of tensors, if not used to universal properties. For that reason, we choose a more constructive approach here.

Let V and W be vector spaces (possibly dual spaces, too). For $v \in V$ and $w \in W$, the tensor product is just a pair $(v, w) = v \otimes w$, that is bilinear and distributive:

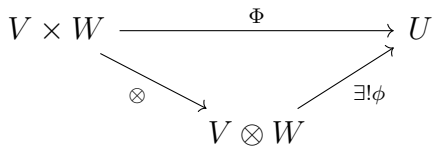
$$(\alpha v) \otimes w = v \otimes (\alpha w)\ , \qquad (v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$$

$$\text{and} \quad v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2\ .$$

One cannot reduce $v \otimes w$ any further, as it is just a special pair (v, w). Not, that the tensor product is not commutative in general, i.e. $v \otimes w \neq w \otimes v$.

Definition 3.7.
Let $\{e_i\}_{i=1,\dots,n}$ be a basis of V and $\{f_j\}_{j=1,\dots,m}$ be a basis of W. Then, the **tensor space** $V \otimes W$ is the linear span of the $e_i \otimes f_j$.

From the definition, it is not clear, if the tensor space depends on the choice of bases. However, this can be proven from the following universal property. Let $\Phi\colon V \times W \longrightarrow U$ be a bilinear map to another vector space. Then there exists a unique linear map $\phi\colon V \otimes W \longrightarrow U$, such that the following diagram commutes (i.e. $\phi \circ \otimes = \Phi$):



Since the tensor space is a vector space again, one can build $(U \otimes V) \otimes W$, etc. . It can also be shown, that the tensor product is associative, i.e.

$$(U \otimes V) \otimes W \cong U \otimes (V \otimes W) \equiv U \otimes V \otimes W\ .$$

Tensors as multilinear maps

With the definition of V^* as $\mathrm{Hom}(V, \mathbb{F})$ and the isomorphy $V^{**} \cong V$, given by $v(\varphi) = \varphi(v)$, tensors can be regarded as multilinear maps. Let for example $T = u \otimes \varphi \otimes v \in U \otimes V^* \otimes V$. Then one defines

$$T(\xi, x, \phi) = u(\xi) \cdot \varphi(x) \cdot v(\phi) = \xi(u) \cdot \varphi(x) \cdot \phi(v) \qquad \forall\ \xi \in U^*,\ x \in V,\ \phi \in V^* \ .$$

So all one does is to use the natural action $\varphi(v)$ and $v(\phi) = \phi(v)$ etc. component wise, and replace $\otimes$ with $\cdot$.

Tensors and linear maps

Having only the right part of tensors of the form $V_2 \otimes V_1^*$ act on a vector $v_1 \in V_1$, these tensors define linear maps $V_1 \longrightarrow V_2$. More precisely:

Lemma 3.8.
There is an embedding $V_2 \otimes V_1^*$ into $\mathrm{Hom}(V_1, V_2)$, defined by the following injective linear map

$$V_2 \otimes V_1^* \hookrightarrow \mathrm{Hom}(V_1, V_2) \ , \quad v_2 \otimes \vartheta_1 \mapsto \ell_{v_2, \vartheta_1} \ ,$$

where $\ell_{v_2,\vartheta_1}(v) = \vartheta_1(v) \cdot v_2$. This map can be extended linearly for $V_2 \otimes V_1^*$.

The embedding is an isomorphy in the finite-dimensional case.

Corollary 3.9.
Let $\{e_i\}_i$ be a basis of a finite-dimensional vector space V and $\{\vartheta_j\}_j$ the dual basis. Then, every linear operator $L \in \mathrm{Hom}(V, W)$ can be written as tensor from $W \otimes V^*$:

$$L = \sum_i (Le_i) \otimes \vartheta_i \ .$$

Assuming L_{mn} are the coefficients of the matrix representation of L (in the basis $\{e_i\}$). Then it holds that $Le_i = \sum_m L_{mi} e_m$. One obtains

$$L = \sum_i (Le_i) \otimes \vartheta_i = \sum_{m,i} L_{mi} (e_m \otimes \vartheta_i) \ .$$

Remark 3.10.
On an abstract level, the terms $e_m \otimes \vartheta_i$ are convenient to work with:

$$\begin{aligned}(e_m \otimes \vartheta_i)(v) &= \vartheta_i(v) \cdot e_m = \vartheta_i \left(\sum_k v_k e_k \right) \cdot e_m = \sum_k v_k \vartheta_i(e_k) \cdot e_m = \sum_k v_k \delta_{ik} \cdot e_m \\ &= v_i \cdot e_m \ .\end{aligned}$$

This means, that $e_m \otimes \vartheta_i$ projects the i-th component of v on the m-th component axis. This calculation also reveals how to calculate the tensor product, when denoting vectors as columns $\begin{pmatrix} \vdots \end{pmatrix}$ and dual vectors as columns $(\cdots)$. For simplicity here in 2

dimensions:

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes (1,0) = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} \ , \qquad \begin{pmatrix}0\\1\end{pmatrix} \otimes (1,0) = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} \ ,$$

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes (0,1) = \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix} \ , \qquad \begin{pmatrix}0\\1\end{pmatrix} \otimes (0,1) = \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} \ .$$

For general (dual) vectors $\binom{v_1}{v_2}$ and (φ_1, φ_2), the tensor product works as follows:

$$\begin{pmatrix}v_1\\v_2\end{pmatrix} \otimes (\varphi_1, \varphi_2) = \left(\begin{pmatrix}v_1\\v_2\end{pmatrix}(\varphi_1) \quad \begin{pmatrix}v_1\\v_2\end{pmatrix}(\varphi_2)\right) = \begin{pmatrix}v_1\varphi_1 & v_1\varphi_2\\v_2\varphi_1 & v_2\varphi_2\end{pmatrix} \ .$$

3.3 Ricci calculus

In Ricci calculus, tensors are characterized by their coefficients (becoming a list of numbers). There are three important principles we need to understand:

1) Components define objects
2) Position of indices determines transformation behavior.
3) Summation convention.

3.3.1 Co-and contravarianz

Let V be a vector space, $\{e_i\}_{i=1,\dots,n}$ a basis and V^* be the dual space with dual basis $\{\vartheta^i\}_{i=1,\dots,n}$. A vector $v \in V$ is called **contravariant** and is described by coefficients with upper indices:

$$v = \sum_{i=1}^{n} v^i e_i \equiv v^i e_i \ .$$

A dual vector $\varphi \in V^*$ is called **covariant** and is described by coefficients with lower indices

$$\varphi = \varphi_i \vartheta^i \ .$$

In the last equation we have already used the summation convention. Over same indices, one upper and one lower, will always be summed (without having to write the summation symbol).

Changing the basis $e_i \to \tilde{e}_i$ does not change the element v, but its components:

$$v = v^i e_i = \tilde{v}^k \tilde{e}_k \ .$$

By definition there are coefficients, such that $e_i = A^j{}_i \tilde{e}_j$ can be written. Plugging in yields the connection between v^i and $\tilde{v}^k$:

$$v = \tilde{v}^k \tilde{e}_k = v^i e_i = v^i A^j{}_i \tilde{e}_j \qquad \Rightarrow \qquad \tilde{v}^k = A^k{}_i v^i \ .$$

For the coefficients of the matrix A, we have already used the Ricci convention. Still, the first index describes the row and the second index the columns, independent if it is an upper or lower index.

Remark 3.11 (Composition of linear maps)**.**
Let $A, B \in \mathrm{End}(V)$ for a finite dimensional vector space V. Due to the isomorphy $\mathrm{End}(V) \simeq V \otimes V^*$, these maps can be written as tensors:

$$A = Ae_i \otimes \vartheta^i = A^j{}_i e_j \otimes \vartheta^i \quad \text{and} \quad B = B^k{}_\ell e_k \otimes \vartheta^\ell \ .$$

Evaluating $(A \circ B)(v)$ for an arbitrary vector $v \in V$ yields

$$\begin{aligned}(A \circ B)(v) &= A^j{}_i e_j \cdot \vartheta^i \left(B^k{}_\ell e_k \cdot \vartheta^\ell(v)\right) = A^j{}_i B^k{}_\ell \vartheta^\ell(v)\vartheta^i(e_k) \cdot e_j \\ &= A^j{}_i B^k{}_\ell \vartheta^\ell(v)\delta^i_k \cdot e_j = A^j{}_i B^i{}_\ell \vartheta^\ell(v) \cdot e_j \\ &= \left(\left(A^j{}_i B^i{}_\ell\right) e_j \otimes \vartheta^\ell\right)(v) \ .\end{aligned}$$

From the last equality we can read off the coefficient behavior under composition:

$$\boxed{(A \circ B)^j{}_\ell = A^j{}_i B^i{}_\ell}$$

If A is a basis change matrix, there is an inverse A^{-1}. From the above remark we know, that this can be expressed by $A^i{}_j (A^{-1})^j{}_k = \delta^i_k$. Dual vectors are linear and defined by $\vartheta^i(e_j) = \delta^i_j$ and respectively $\tilde{\vartheta}^i(\tilde{e}_j) = \delta^i_j$, thus we find:

$$\delta^i_j = \vartheta_i(e_j) = \vartheta_i(A^k{}_j \tilde{e}_k) = A^k{}_j \vartheta_i(\tilde{e}_k) \ .$$

Since the dual vector space is also a vector space, there are coefficients such that $\vartheta^i = M^i{}_\ell \tilde{\vartheta}^\ell$. Plugging in results in the transformation behavior of covectors:

$$\delta^i_j = A^k{}_j \vartheta^i(\tilde{e}_k) = A^k{}_j M^i{}_\ell \tilde{\vartheta}^\ell(\tilde{e}_k) = A^k{}_j M^i{}_\ell \delta^k_\ell = A^k{}_j M^i{}_k = M^i{}_k A^k{}_j \quad \Rightarrow \quad M = A^{-1} \ .$$

$$\Rightarrow \qquad \varphi = \tilde{\varphi}_j \tilde{\vartheta}^j = \varphi_i \vartheta^i = \varphi_i (A^{-1})^i{}_k \tilde{\vartheta}^k \qquad \Rightarrow \qquad \tilde{\varphi}_j = (A^{-1})^i{}_j \varphi_i \ .$$

By definition, basis change matrices are orthogonal/unitary. That is $A^{-1} = A^\dagger \ / A^{-1} = A^T$. Summing up our findings:

	coefficients	**transformation basis vectors**
contravariant	$\tilde{v}^k = A^k{}_i v^i$	$\tilde{e}_k = (A^T)^i{}_k e_i$
covariant	$\tilde{\varphi}_k = (A^T)^i{}_k \varphi_i$	$\tilde{\vartheta}^k = A^k{}_i \vartheta^i$

3.3.2 Tensors in Ricci calculus

After we have seen the foundations of Ricci calculus we can use this formulation on tensors:

Definition 3.12.
A tensor, consisting of r vectors and s covectors

$$T \in \underbrace{V \otimes \ldots \otimes V}_{r \text{ times}} \otimes V^* \otimes \underset{s \text{ times}}{\ldots} \otimes V^*$$

is called **tensor of type (r, s)**. The number $r + s$ is called the **rank**, also for a

general order of vectors and covectors.

A tensor of type (r, s) can be expanded as follows:

$$T = T^{i_1 \dots i_r}{}_{j_1 \dots j_s}\, e_{i_1} \otimes \dots \otimes e_{i_r} \otimes \vartheta^{j_1} \otimes \dots \otimes \vartheta^{j_s} \ .$$

In Ricci calculus one agrees upon the following identification:

$$T = T^{i_1 \dots i_r}{}_{j_1 \dots j_s} \ .$$

A change of basis results in the following transformation behavior:

$$\tilde{T}^{n_1 \dots n_r}{}_{m_1 \dots m_s} = A^{n_1}{}_{i_1} \dots A^{n_r}{}_{i_r} (A^{-1})^{m_1}{}_{j_1} \dots (A^{-1})^{m_s}{}_{j_s}\ T^{i_1 \dots i_r}{}_{j_1 \dots j_s}$$

This behavior is used to define tensors in the physical literature.

3.3.3 Raising and lowering indices

In case of an Euclidean vector space, there is a fourth principle, induced by the inner product:

4) Raising and lowering indices

To understand the invariant meaning behind these manipulations, instead of just defining them, it is best to use the coordinate free formulation first. To keep the notation as close to the textbook notation, we write $\langle \cdot, \cdot \rangle = g(\cdot, \cdot)$ for the inner product.[1] We have already met the isomorphism

$$I : \text{vector} \longrightarrow \text{covector} \qquad v \longmapsto I(v) = g(v, \cdot) \ .$$

Remark 3.13.
In the literature, the isomorphism I and its inverse I^{-1} are called **flat-** and **sharp isomorphism** respectively. The usual notation is

$$I(v) = v^\flat \qquad \text{and} \qquad I^{-1}(\omega) = \omega^\# \ .$$

These isomorphisms can be applied to individual parts of the tensor, still defining an isomorphism between tensor spaces. For example, a $(1,1)$-tensor becomes a $(0,2)$-tensor if $I \otimes \mathbb{1}$ is applied, and a $(2,0)$-tensor, if $\mathbb{1} \otimes I^{-1}$ is applied.

Let e_μ be a basis and ϑ^ν be the dual basis . Defining the coefficients of the inner product by $g_{\mu\nu} = g(\partial_\mu, \partial_\nu)$, there is an inverse matrix (list of numbers) $g^{\mu\nu}$. By definition of inner products the matrices are symmetric: $g_{\mu\nu} = g_{\nu\mu}$ and hence $g^{\mu\nu} = g^{\nu\mu}$. For the isomorphisms I_1 and I_1^{-1} it follows that:

$$I(e_\mu) = g(e_\mu, \cdot) = g_{\mu\nu}\vartheta^\nu \qquad \text{and thus}[2] \qquad I^{-1}(\vartheta^\mu) = g^{\mu\nu} e_\nu \ .$$

The coefficients transform as follows:

$$I(v^\mu \partial_\mu) = g_{\mu\nu} v^\mu dx^\nu =: v_\nu dx^\nu \qquad \text{and} \qquad I^{-1}(u_\mu dx^\mu) = g^{\mu\nu} u_\mu \partial_\nu =: u^\nu \partial_\nu \ .$$

[1] This comes from the more general context of Riemannian manifolds, where g is the Riemannian metric.
[2] $e_\mu = I^{-1}(g_{\mu\nu}\vartheta^\mu) = g_{\mu\nu} I^{-1}(\vartheta^\mu) \quad \Rightarrow \quad g^{\mu\nu} g_{\mu\nu} I^{-1}(\vartheta^\mu) = I^{-1}(\vartheta^\mu) = g^{\mu\nu} e_\mu$.

Remark 3.14.
A contravariant vector v^μ becomes a covariant vector v_ν by lowering the index:

$$v_\nu = g_{\mu\nu} v^\mu \ .$$

Raising an index on the other hand, transforms a covector into a vector:

$$u^\nu = g^{\mu\nu} u_\mu \ .$$

The raising and lowering can be applied for indices of tensors separately:

$$g_{\mu\nu} A^{\cdots\mu\cdots}{}_{\cdots} = A^{\cdots}{}_{\mu}{}^{\cdots}{}_{\cdots} \ .$$

Remark 3.15.
The inner product of two vectors $u^\mu e_\mu$ and $v^\nu e_\nu$ can be written as composition:

$$g(u^\mu e_\mu, v^\nu e_\nu) = u^\mu v^\nu g_{\mu\nu} = u_\nu v^\nu = v_\mu u^\mu \ .$$

4 Dirac Notation and its pitfalls

The Dirac notation is usually introduced in introductory courses on quantum mechanics. Here, we are concerned about the notation scheme itself, rather than the context of its natural appearance. This is not only to focus on the notational aspects, but also since the generality in which the Dirac notation is used can be misleading. In quantum mechanics, one considers (infinite) dimensional vector spaces, that have a (hermitian) inner product, such that all Cauchy sequences converge w.r.t. the norm induced by it. These vector spaces are called Hilbert spaces.[3] In general, an arbitrary Hilbert space does not have an orthonormal basis. In fact, the concept of basis becomes more difficult in infinite dimensional spaces. For that reason, one defines Hilbert bases as an infinite orthonormal sequence of vectors, with which one can linear combine almost all vectors.[4] It is not given, that such a sequence exists. Indeed, one can construct examples, where this is not the case. Hilbert spaces, that admit such a Hilbert basis are then called separable. Assuming, the Hilbert spaces in quantum mechanics are separable, then the question is, if the results from linear algebra carry over. Some do, although for non-trivial reasons, while other ones do not.

To be not concerned with the mathematical subtleties in the background, we confine the considerations here to finite dimensional vector spaces.

4.1 Bras and Kets

In Dirac notation, vectors and covectors get some augmentation to visually represent their origin. This representation mimics the bracket $\langle \cdot, \cdot \rangle$ of an inner product.

[3] Note that all finite dimensional Euclidean/unitary vector spaces are Hilbert spaces.
[4] More precisely, the norm closure of the span of the Hilbert basis is the full Hilbert space

Definition 4.1.
Let V be a vector space and $v \in V$ as well as $\varphi \in V^*$. Vectors are denoted by $|v\rangle$ and are called **ket-vectors**. Covectors are denoted by $\langle\varphi|$ and are called **bra-vectors**.

The terminology originates from the bracket resemblance. In fact, writing $\bar{\varphi}$ is more than just a notation. In case of an Euclidean vector space, we have seen that $I\colon v \longmapsto I(v) \coloneqq \langle v, \cdot\rangle$ is an isomorphism. Instead of writing $\langle I(v)|$, one uses, that $\langle\cdot|$ is enough to show, that it is a covector, and hides the isomorphy.

If the vector space V is unitary, the map $I\colon |v\rangle \longmapsto \langle v|$ is no longer an isomorphism. It is still a bijection, but not linear:

$$I(\alpha v)(w) = \langle \alpha v, w\rangle = \overline{\alpha}\langle v, w\rangle = \overline{\alpha} I(v)(w) \neq \alpha I(v)(w) \qquad \forall\ \alpha \in \mathbb{C}\ .$$

Yet, choosing a basis (since V is finite dimensional), we can construct an isomorphism, so $V \cong V^*$. In infinite dimensions, the map $I\colon |v\rangle \longmapsto \langle(|v)$ retains bijectivity, but is also no isomorphism. This result is known as **Fréchet-Riesz representation**.[5]

For the rest of this section, we assume V to be unitary. The Euclidean case is similar, if not to say easier, since I becomes an isomorphism in that case. To summarize the notation, we consider the bra and ket-vectors of $\alpha v + \beta w \in V$:

$$|\alpha v + \beta w\rangle = \alpha|v\rangle + \beta|w\rangle \qquad \text{and} \qquad \langle \alpha v + \beta w| = \overline{\alpha}\langle v| + \overline{\beta}\langle w|\ .$$

4.2 Scalar product and operators

As much as the Dirac notation suggests the inner product, it makes some intermediate steps, a beginner might need, more obscure. To fix this problem for the moment, we denote the hermitian inner product by $\langle\cdot,\cdot\rangle \equiv (\cdot,\cdot)\colon V \times V \longrightarrow \mathbb{C}$, when inserting ket-vectors. Then, as an inner product takes two vectors of the same vector space as arguments, we find:

$$(|v\rangle, |w\rangle) = (v, w) \equiv \langle v, w\rangle\ .$$

In the literature, one often finds the statement, that a bra- and a ket-vector, join together to the inner product:

$$\langle v||w\rangle = \langle v|w\rangle\ .$$

Yet, this is defining a map $V^* \times V \longrightarrow \mathbb{C}$, not $V \times V \longrightarrow \mathbb{C}$. In terms of the standard notation, we find:

$$\langle v||w\rangle = I(v)(w) = \langle v, w\rangle = \langle v \mid w\rangle\ .$$

So joining bra- and ket-vectors is not a scalar product, but the Fréchet-Riesz representation, which uses the inner product. With the notation set up here, we can give the intermediate steps, to calculate the scalar product between $\alpha_1|v_1\rangle + \alpha_2|v_2\rangle$ and $|w\rangle$:

$$(\alpha_1|v_1\rangle + \alpha_2|v_2\rangle, |w\rangle) = \overline{\alpha_1}(|v_1\rangle, |w\rangle) + \overline{\alpha_2}(|v_2\rangle, |w\rangle) = \overline{\alpha_1}\langle v \mid w\rangle + \overline{\alpha_2}\langle v_2 \mid w\rangle\ .$$

Using only Dirac notation, the calculation looks as follows:

$$\alpha_1|v_1\rangle + \alpha_2|v_2\rangle = |\alpha_1 v_1 + \alpha_2 v_2\rangle$$

[5] For separable Hilbert spaces $\mathcal{H}$ it furthermore also holds that $\mathcal{H} \cong \mathcal{H}^*$

$$\rightsquigarrow \quad \langle \alpha_1 v_1 + \alpha_2 v_2 \mid w\rangle = \langle \alpha_1 v_1 + \alpha_2 v_2 \,||\, w\rangle = (\overline{\alpha_1}\langle v_1| + \overline{\alpha_2}\langle v_2|)|w\rangle = \overline{\alpha_1}\langle v \mid w\rangle + \overline{\alpha_2}\langle v_2 \mid w\rangle \ .$$

A further notation rule enters, when dealing with operators. In finite dimensions, there is no distinction between operator and linear map.[6] So let $A\colon V \longrightarrow V$ be a linear map. Then one writes $|Aw\rangle = A|w\rangle$. It is also common, to emphasize the abstract meaning of $A|w\rangle$, which requires to know how A acts on w, by writing $\widehat{A}|w\rangle = |Aw\rangle$. The linear map A is then regarded as a representation of the abstract operator $\widehat{A}$, similarly to matrices being basis representations of abstract linear maps. Then, the notation is

$$\langle w \mid \widehat{A} \mid v\rangle = \langle w \mid Av\rangle \ .$$

In quantum mechanics, the **hermitian adjoint** is defined by

$$\langle Aw \mid v\rangle = \langle w \mid A^\dagger v\rangle = \langle w \mid \widehat{A}^\dagger \mid v\rangle \ .$$

Notice, that the definition of $A^\dagger$ does not depend on the position in the inner product:

$$\langle w \mid Av\rangle = \overline{\langle Av \mid w\rangle} = \overline{\langle v \mid A^\dagger w\rangle} = \langle A^\dagger w \mid v\rangle \ .$$

Furthermore, the hermitian adjoint is an involution, i.e. $(A^\dagger)^\dagger = A$:

$$\langle Aw \mid v\rangle = \langle w \mid A^\dagger v\rangle = \overline{\langle A^\dagger v \mid w\rangle} = \overline{\langle v \mid (A^\dagger)^\dagger w\rangle} = \langle (A^\dagger)^\dagger w \mid v\rangle \ .$$

In fact, to be more precise, it is called anti-involution for it exchanges the order of operators, $(AB)^\dagger = B^\dagger A^\dagger$:

$$\langle w \mid (AB)^\dagger v\rangle = \langle ABw \mid v\rangle = \langle Bw \mid A^\dagger v\rangle = \langle w \mid B^\dagger A^\dagger v\rangle \ .$$

Remark 4.2.
A common interpretation is then, that $A^\dagger$ acts on $\langle w|$ by $\langle w|A^\dagger = \langle Aw|$. In case of vector spaces with inner product, this interpretation may make sense. On a fundamental level however, $A^\dagger$ is a map $V \longrightarrow V$, while $\langle w| \in V^*$. There is another derived operator from A, acting on V^*, needing no inner product. The dual operator A^* is defined by

$$(A^*\varphi)(v) \coloneqq \varphi(Av) \qquad \forall\ v \in V,\ \varphi \in V^* \ .$$

In this case, we have a hermitian inner product and thus the Fréchet-Riesz representation $I\colon V \longrightarrow V^*$. Let know $\varphi = I(w)$, then we find

$$(A^*\varphi)(v) = \varphi(Av) = I(w)(Av) = \langle w, Av\rangle = \langle A^\dagger w, v\rangle = I(A^\dagger w)(v) \ .$$

$$\Rightarrow \qquad A^*I(w) = A^*\varphi = I(A^\dagger w) = I(A^\dagger w) \qquad \Rightarrow \qquad I^{-1}(A^*(I(w))) = (I^{-1}\circ A^*\circ I)w = A^\dagger w$$

$$\Rightarrow \qquad A^\dagger = I^{-1}\circ A^*\circ I \qquad \Leftrightarrow \qquad I\circ A^\dagger = A^*\circ I \ ,$$

which is the usual way, transformations of vectors act on linear maps. Recalling the Dirac notation $I\colon |v\rangle \longmapsto \langle v|$, we find:

$$\langle A^\dagger w| = I(A^\dagger|w\rangle) = I\circ A^\dagger(|w\rangle) = A^*\circ I(|w\rangle) = A^*\langle w| \ .$$

Another notation used in physics is $|v\rangle^\dagger = \langle v|$ and $\langle v|^\dagger = |v\rangle$.

[6] Most of the complicated concepts become trivial in finite dimensions. Every linear map between finite dimensional vector spaces is automatically continuous, bounded, compact etc. .

Remark 4.3.
What is meant with this is of course the Fréchet-Riesz representation. The hermitian adjoint makes no sense here. One also encounters the notations $(\widehat{A}|v\rangle)^\dagger = \langle v|A^\dagger$ and $(\langle v|A)^\dagger = A^\dagger|v\rangle$. Again, all that is happening here is the Fréchet-Riesz representation. For example:

$$I(A|v\rangle) = (I \circ A)|v\rangle = (I \circ (A^\dagger)^\dagger)|v\rangle = ((A^\dagger)^* \circ I)|v\rangle = (A^\dagger)^*\langle v| =: \langle v|A^\dagger \ .$$

4.3 Tensors in Dirac notation

We have already seen, that tensors of the form $e_i \otimes \vartheta_j$ are linear maps $V \longrightarrow V$. In fact, they are a basis of End(V). Applying Dirac notation, one would assume to write $|e_i\rangle \otimes \langle e_j|$, where we used that $I_(e_j) = \vartheta_j$, and the notation convention $I\colon |v\rangle \to \langle v| = I(v)$. Now, in Dirac notation, it is common practice, to understand the connection of two angled brackets always as tensor product, and to suppress the symbol $\otimes$:

$$|v\rangle \otimes \langle w| = |v\rangle\langle w| \ .$$

The application of $|v\rangle\langle w|$ as linear map becomes quite natural in this notation (at least as long as only one particle is considered). As a reminder, in the usual notation, we would need to write :

$$(v \otimes I(w))(u) = I(w)(u) \cdot v = \langle w, u\rangle v \ .$$

In Dirac notation, this becomes

$$|v\rangle\langle w||u\rangle = |v\rangle\langle w \mid u\rangle = \langle w \mid u\rangle|v\rangle \ .$$

Observe, that $\langle \cdot \mid \cdot \rangle$ is only a scalar, and the tensor product is just the usual product.[7]

Lemma 4.4.
Let $v \otimes \varphi \in V \otimes V^* \cong End(V)$, then it holds that $(v \otimes \varphi)^* = \varphi \otimes v$ and $(v \otimes \varphi)^\dagger = I^{-1}(\varphi) \otimes I(v)$.

Proof 4.5.
For the first equation, it has to be shown by definition, that $((\varphi \otimes v)\omega)(u) = \omega((v \otimes \varphi)u)$:

$$\begin{aligned}((\varphi \otimes v)\omega)(u) &= (v(\omega) \cdot \varphi)(u) \equiv (\omega(v) \cdot \varphi)(u) = \omega(v) \cdot \varphi(u) = \omega(\varphi(u) \cdot v) \\ &= \omega((v \otimes \varphi)u) \ .\end{aligned}$$

The second equation can be shown with the definition of the Fréchet-Riesz representation $I(v)(x) = I(v)(x) = \langle v \mid x\rangle$. Then:

$$\begin{aligned}\langle y \mid (v \otimes \varphi)x\rangle &= \langle y \mid \varphi(x)v\rangle = \varphi(x)\langle y \mid v\rangle = \varphi(x)\overline{\langle v \mid y\rangle} = \langle I^{-1}(\varphi) \mid x\rangle\overline{I(v)(y)} \\ &= \overline{I(v)(y)}\langle I^{-1}(\varphi) \mid x\rangle = \langle I(v)(y) \cdot I^{-1}(\varphi) \mid x\rangle \\ &= \langle (I^{-1}(\varphi) \otimes I(v))y \mid x\rangle \ .\end{aligned}$$

□

[7] This comes form the isomorphism $\mathbb{F} \otimes_{\mathbb{F}} V \cong V$, defined by $\alpha \otimes_{\mathbb{F}} v \longmapsto \alpha \cdot v$.

In Dirac notation, the main result reads $(|v\rangle\langle w|)^\dagger = |w\rangle\langle v|$. In fact, this result fits in nicely with the physics notation $|v\rangle^\dagger = \langle v|$ and $\langle w|^\dagger = |w\rangle$, as the hermitian adjoint is an anti-involution:

$$(|v\rangle\langle w|)^\dagger = \langle w|^\dagger |v\rangle^\dagger = |w\rangle\langle v| \ .$$

Remark 4.6.
Again, $|v\rangle^\dagger$ etc. are not well defined. However $(|v\rangle\langle w|)^\dagger$ is indeed well defined, understanding $|v\rangle\langle w| \in V \otimes V^* \cong \mathrm{End}(V)$.

Corollary 3.9 has the important consequence, that abstract operators can be expressed as (choosing a orthonormal basis $|e_n\rangle$):

$$L = \sum_n L|e_n\rangle\langle e_n| \ .$$

A common way to denote a Hilbert basis/orthonormal basis, is to just use the index $|n\rangle \equiv |e_n\rangle$. Operators can then be written as

$$L = \sum_n L|n\rangle\langle n| \ .$$

Since we are in finite dimensions, we can fully utilize linearity.[8] Consider the unit operator $\mathbb{1} \in \mathrm{End}(V)$, defined by $\mathbb{1}|v\rangle = |v\rangle$. It follows that

$$\mathbb{1} = \sum_n |n\rangle\langle n| \ .$$

For Hilbert spaces, this is called **completeness relation**. For the moment, we are in a finite dimension vector space, such that sums and liner maps commute. This holds especially for bra-vectors as linear maps $V \longrightarrow \mathbb{C}$ and ket-vectors as linear maps $V^* \longrightarrow \mathbb{C}$. A general procedure is now, to insert unit operators, to obtain easier to calculate expressions:

$$|v\rangle = \mathbb{1}|v\rangle = \sum_n |n\rangle\langle n \mid v\rangle \ , \qquad \langle v| = \langle v|\mathbb{1} = \langle v| \sum_n |n\rangle\langle n| = \sum_n \langle v \mid n\rangle\langle n| \ .$$

These are just the basis expressions for $|v\rangle$ and $\langle v|$. With this, also inner product can be manipulated:

$$\langle v \mid w\rangle = \langle v|\mathbb{1}|w\rangle = \langle v| \sum_n |n\rangle\langle n \mid w\rangle = \sum_n \langle v \mid n\rangle\langle v \mid w\rangle \ .$$

Remark 4.7.
What we have used here, is linearity. For infinite dimensional Hilbert spaces, this does not work, since one is dealing with limits. However, if the Hilbert space is

[8] Let A be a linear operator, and v_n be vectors. Then $\sum_{n=1}^N Av_n = A\sum_{n=1}^N v_n$. However, linear operators and limits do not commute in general (unless it is also continuous). Hence:

$$\sum_{n=1}^\infty Av_n = \lim_{N\to\infty} \sum_{n=1}^N Av_n = \lim_{N\to\infty} A\sum_{n=1}^N v_n \neq \lim_{N\to\infty} A\sum_{n=1}^N v_n = A\sum_{n=1}^\infty v_n \ .$$

separating, i.e. has a Hilbert basis, it is an important theorem, that the completeness relations hold, and vice versa. However, careful application of this theorem yields in most cases the same result, as the physics completeness relation.

4.4 Position representation and abuse of notation

In the end, we do not get around infinite dimensional spaces to give full account of the Dirac notation, as used in physics. Most of the time, when working with quantum mechanics, one is considering one of the following two cases. The Hilbert space is not specified further, but treated as abstract vector space with the Hilbert basis $\{|n\rangle\}$. The second one is, to consider the position representation with wave functions $\psi(x)$.

Example 4.8.
In finite dimensions, this is similar to the treatment of vector spaces. Let V be an n-dimensional abstract vector space. As such, one can choose a basis $\{v_j\}$. This defines an isomorphism to $\mathbb{R}^n$ by $v_j \longmapsto \vec{e}_j$. Then arbitrary vectors v can be represented as tuples $(v_1, \dots, v_n)^T$. Another example, more reminiscent of the position representation is $K_{n-1}[\mathbb{R}]$. This is the rather concrete vector space of polynomials $p = \sum_{j=0}^{n-1} \alpha_j x^j$. A basis of this space is $v_n = x^n$ for $n = 0, \dots, n-1$.

The Hilbert space used for the position representation is $L^2(\mathbb{R}^n)$. This is the space of square integrable (complex valued) functions over $\mathbb{R}^n$. So

$$f \in L^2(\mathbb{R}^n) \qquad \Leftrightarrow \qquad f \text{ is integrabel and } \int_{\mathbb{R}^n} |f(x)|^2 \, dx^n < \infty \ .$$

Remark 4.9.
To be more precise, $\mathcal{L}^2(\mathbb{R}^n)$ is actually the set of square integrable functions. The integral used in this context is the Lebesgue integral, a generalization of the Riemann integral. If the Riemann integral exists however, both integrals coincide. The Lebesgue integral is not sensitive of single points (more generally sets with measure zero). To create a Hilbert space, one has to define equivalent classes $[f]_\sim$ of functions. Another function is in the same class $g \in [f]_\sim$, if $g \sim f$. The equivalence relation $\sim$ used here is, that g is equal to f, with possible exceptions on zero sets, which the Lebesgue integral does not see anyway. In that sense, in actually makes not sense to write something like $f(x)$ because for $g \in [f]_\sim$ it can happen that $g(x) \neq f(x)$. Hence the values on single positions are not uniquely defined.

The inner product on $L^2(\mathbb{R}^n)$ is defined as follows:

$$\langle f \mid g \rangle_{L^2} = \int_{\mathbb{R}^n} \overline{f(x)} \cdot g(x) \, dx^n \ .$$

With $\overline{z} \cdot z = |z|^2$ it follows that square integrability means finite norm:

$$\| |f\rangle \|_{L^2}^2 = \langle f \mid f \rangle_{L^2} = \int_{\mathbb{R}^n} \overline{f(x)} f(x) \, dx^n = \int_{\mathbb{R}^n} |f(x)|^2 \, dx^n < \infty \ .$$

For the rest of the section, we consider the special case $n = 1$, i.e. $L^2(\mathbb{R})$, which introductory quantum mechanics restrict to. A Hilbert basis of $L^2(\mathbb{R})$ is given by the **Hermite polynomials** $|H_n\rangle$, where

$$H_n(x) = e^{\frac{x^2}{2}} \left(-\frac{d}{dx} + x\right)^n e^{\frac{x^2}{2}} .$$

Yet, these polynomials are rather cumbersome. Although there exist proper Hilbert bases for $L^2(\mathbb{R})$, physicists often use another "basis" instead. Physicists define the so called **position states** $|x\rangle$ and **momenta states** $|p\rangle$

$$|x\rangle = \delta(x - \cdot) \qquad \text{and} \qquad |p\rangle = e^{\frac{ipx}{\hbar}} .$$

These states are problematic as they do not belong to the Hilbert space, are neither a basis, nor a Hilbert basis. Indeed,

$$\int_{\mathbb{R}} |e^{\frac{ipx}{\hbar}}|^2 \, dx = \int_{\mathbb{R}} 1 \, dx = \infty .$$

Though the delta function might result a finite integral, it is not a function at all. It can even be proven, that no function exists, that exhibits all the demanded properties. There does exist a framework, where delta functions make perfect sense. Delta functions are singular, i.e. not as function representable, distribution. Using distribution theory, the states $|x\rangle$ and $|p\rangle$ can be used rigorously.

Remark 4.10.
As is often the case, the rigorous treatment is more complicated, and generally not allowing the same rules for calculations as linear algebra does. Most of the notational rules, that will follow, can be derived in a proper way. However, from this point on, one will reach the point, where there is a mathematical theory yet to be found. One example is quantum field theory, where manipulations with these states and delta functions for the fields would need product of distributions. An operation, that is not well defined in general, and does not cover all the operations needed for qft. The other major formulation, path integrals also have no complete mathematical description yet. This is more puzzling, as some of the most well measured constant are in accordance to the by qft calculated values.

The position and momenta states are sometimes called **pseudo eigen states**. Pseudo in the sense, that they are not states in the Hilbert space $L^2(\mathbb{R})$. As we have seen in subsection 4.2, one abstractly writes $\widehat{\boldsymbol{x}}|f\rangle$ and $\widehat{\boldsymbol{p}}|f\rangle$ for the position and momentum operators. Since we have a concrete Hilbert space, one needs to specify, what the concrete representation $\boldsymbol{x}$ and $\boldsymbol{p}$ are:

$$\boldsymbol{x}f(x) = x \cdot f(x) \qquad \text{and} \qquad \boldsymbol{p}f = -i\hbar\partial_x f(x) .$$

Then, we find, that

$$\widehat{\boldsymbol{p}}|q\rangle = -i\hbar\partial_x e^{\frac{iqx}{\hbar}} = -i\hbar(i\hbar)qe^{\frac{iqx}{\hbar}} = qe^{\frac{iqx}{\hbar}} = q|q\rangle .$$

For the position state, we note that a property of the delta function is, that $f(x)\delta(x-y) = f(y)\delta(y-x)$. So

$$\widehat{\boldsymbol{x}}|y\rangle = x \cdot \delta(y - x) = y \cdot \delta(y - x) = y|y\rangle \quad .$$

The pseudo eigen states are orthonormal w.r.t. the delta function orthonormality:[9]

$$\langle x \mid y \rangle = \int_{\mathbb{R}} \delta(x-q)\delta(y-q)\ dq = \delta(x-y)\ ,$$

$$\begin{aligned}\langle p \mid q \rangle &= \int_{\mathbb{R}} \overline{e^{\frac{ipx}{\hbar}}} e^{\frac{iqx}{\hbar}}\ dx = \int_{\mathbb{R}} e^{\frac{-ipx}{\hbar}} e^{\frac{iqx}{\hbar}}\ dx = \int_{\mathbb{R}} e^{\frac{i(q-p)x}{\hbar}}\ dx = \int_{\mathbb{R}} e^{i(q-p)y}\ d(\hbar y) \\ &= h\overline{\int_{\mathbb{R}}} e^{i(q-p)y}\ dy = 2\pi\hbar\delta(q-p)\ .\end{aligned}$$

Another notation, inspired by the action of the delta function is:

$$\langle x \mid f \rangle = \int_{\mathbb{R}} \delta(x-y)f(y)\ dy = f(y) \qquad \text{and} \qquad \langle f \mid x \rangle = \overline{f}(x)\ .$$

Finally, the completeness relation is used as

$$\mathbb{1} = \int_{\mathbb{R}} dx\ |x\rangle\langle x|\ .$$

Remark 4.11.
In quantum mechanics courses, one discusses, that the Fourier transformation $\mathcal{F}$ leads to a momentum representation. I.e. the function argument is the momentum $\langle p \mid f \rangle = f(p)$. In the momentum space, the completeness relation reads:

$$\mathbb{1} = \int_{\mathbb{R}} \frac{dp}{2\pi\hbar} |p\rangle\langle p|\ .$$

[9] It can be shown, that $\int_{\mathbb{R}} e^{i(q-p)x}\ dx = 2\pi\delta(q-p)$.