Tensors – minimal formalism for physicists

Jonas Daverio

December 21, 2023

This document is intended for physicists who are learning tensors for the first time in classical theories such as electrodynamics or continuum mechanics.

Tensors are often introduced with the following sentence: "tensors are objects that transform like tensors". This is obviously not very illuminating, and we are going to take a more mathematical view, maybe less common among physicists. The alternative obscure sentence will now be "tensors are multilinear objects". It is obscure because it is abstract, but it shouldn't bother you if you already accepted the definition of a vector you saw in linear algebra.

I'm going to assume you know what a vector space is, and in fact, we'll start our journey with $\mathbb{R}$-vector spaces $(V_1, \ldots, V_k, W)$ which will accompany us all along. The same exact document could be written with fields other than $\mathbb{R}$, but $\mathbb{R}$ will be of main interest in classical electrodynamics. Vector spaces on $\mathbb{C}$ are really important in quantum mechanics.

Section 1.1 can be read superficially if you don't care about the formal construction. It will be of no use in most of physics, but it may help some put their minds in order.

Contents

1 Formal definitions

1.1 Tensor product

1.1.1 Motivation

The whole idea of tensors comes from considering multilinear functions:

$$\begin{aligned} A: V_1 \times \cdots \times V_k &\to W \\ (\mathbf{v}_1, \ldots, \mathbf{v}_k) &\mapsto A(\mathbf{v}_1, \ldots, \mathbf{v}_k) \\ A(\mathbf{v}_1, \ldots, \mathbf{v}_i + \lambda \mathbf{v}_i', \ldots, \mathbf{v}_k) &= A(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k) \\ &\quad + \lambda A(\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_k). \end{aligned} \tag{1}$$

The general remark is that there are many different n-tuples in $V_1 \times \cdots \times V_k$ that map to the same $\mathbf{w} \in W$ *regardless* of the choice of A. For example, $(\mathbf{x}_1, \ldots \lambda \mathbf{x}_i, \ldots, \mathbf{x}_k)$ and $(\mathbf{x}_1, \ldots, \lambda \mathbf{x}_j, \ldots, \mathbf{x}_k)$ are mapped to the same element because of the multilinear property.

What we would like to achieve is to build a new space, which we'll call the tensor product $V_1 \otimes \cdots \otimes V_k$, in which there are exactly "enough" elements such that, for any two elements in it, you can always fine a function f that maps them to different $\mathbf{w}$'s in W. So we see that, in that sense, $V_1 \times \cdots \times V_k$ contains to many elements.

The other thing we would like to do is to transform this multilinear property, which is slightly annoying, into a simple linear property. The idea is that we may try to write

$$\begin{aligned} T &= [(\mathbf{v}_1, \ldots, \mathbf{v}_i + \lambda \mathbf{v}_i', \ldots, \mathbf{v}_k)] \\ &= [(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k)] + \lambda [(\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_k)] \\ &= T_1 + \lambda T_2 \end{aligned} \tag{2}$$

with T, T_1 and T_2 in $V_1 \otimes \cdots \otimes V_k$, and $[u]$ denoting the tensor associated with the tuple u, whatever $+$ will mean (we will formalize this in a moment). If we achieve this, the relation (1) simply becomes

$$\begin{aligned} \tilde{A}: V_1 \otimes \cdots \otimes V_k &\to W \\ T &\to A(T) \\ \tilde{A}(T_1 + \lambda T_2) &= \tilde{A}(T_1) + \lambda \tilde{A}(T_2) \end{aligned} \tag{3}$$

There is a last problem (I promise): for each space V_i, take any pair of elements $\mathbf{u}_i$ and $\mathbf{v}_i$. Consider the two following:

$$A(\mathbf{u}_1 + \mathbf{v}_1, \dots, \mathbf{u}_k + \mathbf{v}_k) \stackrel{?}{=} A(\mathbf{u}_1, \dots, \mathbf{u}_k) + A(\mathbf{v}_1, \dots, \mathbf{v}_k). \tag{4}$$

There is absolutely no reason to except them to be equal. We cannot apply (1) to go from one to the other. In fact, for a tuple $u = (\mathbf{u}_1, \dots, \mathbf{u}_k)$, take a multilinear function A such that $A(u) \neq 0$. Take $\mathbf{v}_i = \mathbf{u}_i$. You then have

$$A(2\mathbf{u}_1, \dots, 2\mathbf{u}_k) = 2^k A(u) \neq A(\mathbf{u}_1, \dots, \mathbf{u}_k) + A(\mathbf{u}_1, \dots, \mathbf{u}_k) = 2A(u). \tag{5}$$

Even worse, if the $\mathbf{v}_i$ are linearly independent of the $\mathbf{u}_i$, you can show (you can do that if you're into that kind of stuff) that you can always find some multlinear function f such that

$$A(\mathbf{u}_1, \dots, \mathbf{u}_k) + A(\mathbf{v}_1, \dots, \mathbf{v}_k) \neq A(u) \quad \forall u \in V_1 \times \cdots \times V_k. \tag{6}$$

However, the associated tensors

$$T_1 = [(\mathbf{u}_1, \dots, \mathbf{u}_k)] \text{ and } T_2 = [(\mathbf{v}_1, \dots, \mathbf{v}_k)]$$

must satisfy

$$\tilde{A}(T_1) + \tilde{A}(T_2) = \tilde{A}(T_1 + T_2) \tag{7}$$

with $T_1 + T_2$ being another tensor. The above argument means that the tensor $T_1 + T_2$ cannot be associated with any tuple from $V_1 \times \cdots \times V_k$.

All of this means that $V_1 \times \cdots \times V_k$ is a really bad way of representing $V_1 \otimes \cdots \otimes V_k$, because it both has too many elements ($[(\mathbf{a}_1, \dots, \lambda\mathbf{u}_i, \dots \mathbf{u}_k)] = [(\mathbf{u}_1, \dots, \lambda\mathbf{u}_j, \dots, \mathbf{a}_k)]$) and misses some elements ($\forall u$, $[u] \neq T_1 + T_2$).

1.1.2 Explicit construction

To construct $V_1 \otimes \cdots \otimes V_k$, we will first fill in all the missing elements, and then remove all the duplicate. This is surprisingly easy.

The first step simply consist in considering the set $\mathcal{F}(V_1 \times \cdots \times V_k)$ of all possible *formal sums* of tuple in $V_1 \times \cdots \times V_k$. This means that we are looking at strings of tuples that we sum with each other

$$\lambda_1(\mathbf{u}_1, \dots, \mathbf{u}_k) + \lambda_2(\mathbf{v}_1, \dots, \mathbf{v}_k) + \cdots + \lambda_n(\mathbf{w}_1, \dots, \mathbf{w}_k) \tag{8}$$

but we never really think about what the + is meaning. We simply look at the string of character as the object we are manipulating. Formally $\mathcal{F}(S)$ is the set of all functions

$$g : S \to \mathbb{R} \tag{9}$$

that are non-zero for only finitely many values of S. We can then write any element like (8) in the form

$$\sum_i g(u_i)u_i, \quad u_i \in V_1 \times \cdots \times V_k. \tag{10}$$

Now that we have filled in the missing elements, we are going to subtract all the duplicates. Consider the following equivalence relation:

$$(\mathbf{v}_1, \ldots, \mathbf{v}_i + \lambda \mathbf{v}_i', \ldots, \mathbf{v}_k) \sim (\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k) + \lambda(\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_k) \tag{11}$$

The final tensor product is finally given by all the equivalence classes under this relation:

$$\begin{aligned} V_1 \otimes \cdots \otimes V_k &= \mathcal{F}(V_1 \times \cdots \times V_k)/\sim \\ &= \{[u]_\sim \mid u \in \mathcal{F}(V_1 \times \cdots \times V_k)\} \\ [u]_\sim &= \{v \in \mathcal{F}(V_1 \times \cdots \times V_k) \mid v \sim u\}. \end{aligned} \tag{12}$$

The equivalence class of a given tuple in $V_1 \times \cdots \times V_k$ is more commonly written by

$$[(\mathbf{v}_1, \ldots, \mathbf{v}_k)]_\sim = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k \in V_1 \otimes \cdots \otimes V_k. \tag{13}$$

Some elements of $V_1 \otimes \cdots \otimes V_k$ can be written in that form. They are called pure tensors. But some elements cannot, because of the previous argument, and they necessarily have to be written as a sum of pure tensors.

1.1.3 Basis of the tensor product

It should be clear that it follows directly from its construction that the tensor product is a vector space in its own right. It can be shown that, if we have a basis $(\mathbf{e}_j)_i$ for each V_j, we directly get a basis for $V_1 \otimes \cdots \otimes V_k$ for free, given by the set of pure tensors

$$\{(\mathbf{e}_1)_{i_1} \otimes \cdots \otimes (\mathbf{e}_k)_{i_k} \mid 1 \leq i_1 \leq \dim V_1, \ldots, 1 \leq i_k \leq \dim V_k\}. \tag{14}$$

Using Einstein's summation convention, this let us write any tensor $T \in V_1 \otimes \cdots \otimes V_k$ as

$$T = T_{i_1 \dots i_k} (\mathbf{e}_1)_{i_1} \otimes \cdots \otimes (\mathbf{e}_k)_{i_k}, \quad T_{i_1 \dots i_k} \in \mathbb{R}. \tag{15}$$

1.1.4 Multilinear property

Because of the equivalence relation construction, by writing

$$\begin{aligned} T &= \mathbf{v}_1 \otimes \cdots \otimes (\mathbf{v}_i + \lambda \mathbf{v}_i') \otimes \cdots \otimes \mathbf{v}_k \\ T_1 &= \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{v}_k \\ T_2 &= \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i' \otimes \cdots \otimes \mathbf{v}_k \end{aligned} \tag{16}$$

we automatically get the following property on tensors:

$$T = T_1 + \lambda T_2. \tag{17}$$

This finally shows us why tensor are useful as we'll see in the following section.

1.1.5 Universal property

If we define the value of $\tilde{A}$ for pure tensors to be

$$\tilde{A}(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k) = A(\mathbf{v}_1, \dots, \mathbf{v}_k), \tag{18}$$

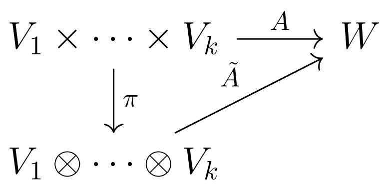
and for non-pure tensors to be

$$\tilde{A}(T_1 + T_2) = \tilde{A}(T_1) + \tilde{A}(T_2), \tag{19}$$

it is very easy to show that the multilinearity condition on A transfers to a simple linearity condition on $\tilde{A}$:

$$\tilde{A}(T_1 + \lambda T_2) = \tilde{A}(T_1) + \lambda \tilde{A}(T_2), \quad \forall T_1, T_2 \in V_1 \otimes \cdots \otimes V_k. \tag{20}$$

For those among you loving abstract nonsense, this is called a univeral property, and is associated with the following commutative diagram:



where $\pi(\mathbf{v}_1, \dots, \mathbf{v}_k) = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_k$.

1.2 Dual vector space – Alternative definition of the tensor product

1.2.1 Definition

You should all have seen the definition of the dual of vector space V. It is another vector space V^* containing all linear forms:

$$V^* = L(V, \mathbb{R}) = \{f : V \to \mathbb{R} \mid f(\mathbf{v}_1 + \lambda \mathbf{v}_2) = f(\mathbf{v}_1) + \lambda f(\mathbf{v}_2)\} \tag{21}$$

Elements of V^* are also called covectors for a reason explained in section 1.2.8.

You may have seem that the dual of the dual of a space is the original space. More precisely, $(V^*)^* \cong V$. The means that we can instead define V to be (up to isomorphism):

$$V \cong L(V^*, \mathbb{R}). \tag{22}$$

This seemingly random fact will appear to be very important to get some intuition of why tensors are used in physics.

1.2.2 Basis for the dual space

As you may have seen, V^* is isomorphic to V. However, there is no special way of choosing this isomorphism, and there is an infinite number of equivalent isomorphisms between V and V^*. However, if we fix a basis in V, it tells us a preferred way of choosing this isomorphism, called the canonical isomorphism.

Let's call $\mathbf{e}_i$ the basis of V, with $1 \leq i \leq \dim V$ (we'll omit the mention of the interval in which i lies from now on, and simply write the set of basis vectors as $\mathbf{e}_i$). We can define a basis $\boldsymbol{\varepsilon}^j$ (don't think too much about the upper and lower indices for now) in V^* by imposing

$$\boldsymbol{\varepsilon}^j(\mathbf{e}_i) = \delta_i^j. \tag{23}$$

Because any linear function on V is uniquely determined by the value it takes on all basis elements of V, all $\boldsymbol{\varepsilon}^j$ are uniquely determined. The canonical isomorphism between between V and V^* is now given by the unique (by the same argument) linear map

$$\begin{aligned} \pi : V &\to V^* \\ v^i \mathbf{e}_i &\mapsto v^i \boldsymbol{\varepsilon}^i \end{aligned} \tag{24}$$

1.2.3 Link with inner product and metric

An inner product, also known as a metric, is a positive definite and symmetric bilinear form:

$$\begin{aligned} g : V \times V &\to \mathbb{R} \\ (\mathbf{v}_1, \mathbf{v}_2) &\mapsto g(\mathbf{v}_1, \mathbf{v}_2). \end{aligned} \tag{25}$$

We may use any of the following notation:

$$g(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1 \cdot \mathbf{v}_2 = \langle \mathbf{v}_1 | \mathbf{v}_2 \rangle. \tag{26}$$

You should note that in physics, the positive definite assumption is often ignored, as, for example, the Minkowski metric can be negative.

The interesting thing about g is that it gives yet another isomorphism[1] between V and $V*$, given by

$$\begin{aligned} g : V &\to V^* \\ \mathbf{v} &\mapsto g(\mathbf{v}, \cdot), \end{aligned} \tag{27}$$

where $g(\mathbf{v}, \cdot)$ is the linear form

$$\begin{aligned} V &\to \mathbb{R} \\ \mathbf{u} &\mapsto g(\mathbf{v}, \mathbf{u}). \end{aligned} \tag{28}$$

Because it is an element of V^*, we can write it in the basis $\boldsymbol{\varepsilon}^j$, and decompose $\mathbf{v}$ in the basis $\mathbf{e}_i$:

$$g(\mathbf{v}, \cdot) = g(v^i \mathbf{e}_i, \cdot) = v^i g(\mathbf{e}_i, \cdot) \equiv v_j \boldsymbol{\varepsilon}^j \tag{29}$$

This is precisely the definition of the relation between v^i and v_i, and we will see how this generalizes below.

Moreover, we can see that we can either see g as a linear map $V \times V \to \mathbb{R}$ or $V \to V^*$. This idea is of great importance to understand intuitively what tensors are and why we use them in physics, and will be developped in the following section.

[1] It's an isomorphism if and only if the metric is positive definite. Since the Minkowski metric is not positive definite, it has a non-zero kernel and therefore maps some vectors to the trivial linear form 0.

1.2.4 Alternative definition of tensors (very important)

What is really interesting is that we can show without too much work that

$$V^* \otimes V^* \cong L(V \times V, \mathbb{R}). \tag{30}$$

The general idea is that you can show any bilinear map $V \times V \to \mathbb{R}$ can be represented by the sum of products of two linear maps $V \to \mathbb{R}$ in the following way:

$$f(\mathbf{u}, \mathbf{v}) = \sum_i^k g_i(\mathbf{u}) h_i(\mathbf{v}) \tag{31}$$

Next, because of the multilinear property, we will have many different ways to represent your bilinear map as a sum of product of two linear maps. In the end, the family of all the different way of representing the same bilinear map will look exactly the same as the equivalence relation defined in (11), and the isomorphism with $V^* \otimes V^*$ will be trivial.

We can then extend the argument to show that

$$V^* \otimes \cdots \otimes V^* \cong L(V \times \cdots \times V, \mathbb{R}), \tag{32}$$

and by using the fact that $V \cong (V^*)^* = L(V^*, \mathbb{R})$, we immediately deduce that

$$V \otimes \cdots \otimes V \cong L(V^* \times \cdots \times V^*, \mathbb{R}), \tag{33}$$

and finally that

Equivalent definition

$$\underbrace{V \otimes \cdots \otimes V}_{p \text{ copies}} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q \text{ copies}} \cong L(\underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}}, \mathbb{R}). \tag{34}$$

Elements of this tensor product space are called (p, q)-tensors of rank $(p + q)$, or p-times contravariant and q-times covariant tensors, for a reason that for now should be totally obscure, but that we'll explain in section 2.

1.2.5 A last important isomorphism

We will now give a last isomorphism which really is the highlight of the show (or one of them):

An important isomorphism

$$L(\underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}}, \mathbb{R}) \cong L(\mathbb{R}, \underbrace{V \times \cdots \times V}_{p \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{q \text{ copies}}) \tag{35}$$

which is a special case of

$$\begin{aligned} &L(\underbrace{V^* \times \cdots \times V^*}_{p_1 \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q_1 \text{ copies}}, \underbrace{V \times \cdots \times V}_{p_2 \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{q_2 \text{ copies}}) \\ &\cong L(\underbrace{V^* \times \cdots \times V^*}_{(p_1-k) \text{ copies}} \times \underbrace{V \times \cdots \times V}_{(q_1-j) \text{ copies}}, \underbrace{V \times \cdots \times V}_{(p_2+k) \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{(q_2+j) \text{ copies}}) \end{aligned} \tag{36}$$

This looks very ugly in its general form. Ugliness here is just the shear size of the expression, but what it says is actually pretty simple. It means that we can swap side for each V and V^* and take its dual (exchanging V and V^*). Let's have a look at those isomorphisms for rank 1 and 2 tensors:

$$V \cong L(V^*, \mathbb{R}) \cong L(\mathbb{R}, V) \tag{37}$$
$$V^* \equiv L(V, \mathbb{R}) \cong L(\mathbb{R}, V^*) \tag{38}$$
$$V \otimes V \cong L(V^* \times V^*, \mathbb{R}) \cong L(V^*, V) \tag{39}$$
$$V^* \otimes V^* \cong L(V \times V, \mathbb{R}) \cong L(V, V^*) \tag{40}$$
$$V \otimes V^* \cong L(V^* \times V, \mathbb{R}) \cong L(V, V) \tag{41}$$

The first two are not really important, but are almost trivial to show. Here is the explicitly constructed isomorphism:

$$\begin{aligned} V^{(*)} &\to L(\mathbb{R}, V^{(*)}) \\ \mathbf{v} &\mapsto \begin{cases} \mathbb{R} \to V^{(*)} \\ \lambda \mapsto a\mathbf{v} \end{cases} \end{aligned} \tag{42}$$

The last last two are very useful in that they appear very often in physics. We'll discuss them seperately and give an explicit construction of the isomorphism.

1. The first says that $(0, 2)$-tensors, bilinear forms (bilinear maps that map a pair of vectors to a scalar) can also be seen as maps between V and V^*. We already saw this in section 1.2.3. From a bilinear map

$$f : V \times V \to \mathbb{R}, \tag{43}$$

we can build a map

$$\begin{aligned} \tilde{f} : V &\to V^* \\ \mathbf{v} &\mapsto f(\mathbf{v}, \cdot). \end{aligned} \tag{44}$$

If f is positive definite then $\tilde{f}$ is an isomorphism between V and V^* and gives a way to transform a vector into a covector.

2. The second one says that $(1,1)$-tensors really are linear maps (also called endomorphisms). They are extremely important, and you'll have seen them all over the place in physics. To construct the isomorphism, we'll use the same idea, and transform our bilinear map

$$\begin{aligned} f : V^* \times V &\to \mathbb{R} \\ (\boldsymbol{\beta}, \mathbf{v}) &\mapsto f(\boldsymbol{\beta}, \mathbf{v}) \end{aligned} \tag{45}$$

into

$$\begin{aligned} \tilde{f} : V &\to L(V^*, \mathbb{R}) \\ \mathbf{v} &\mapsto (\cdot, \mathbf{v}). \end{aligned} \tag{46}$$

We'll use our isomorphism between $L(V^*, \mathbb{R})$ and V to conclude the proof.

1.2.6 Partial application argument

Both previous result can be interpreted in the following way: a $(0,2)$-tensor is something that wants to eat to vector. You can either feed it two vectors and it will spit you back some number, or you can feed it one vector and it will spit you something that want to eat a second vector. At the end, the number of vectors that something want to eat is the same: two.

For $(1,1)$-tensor, it something that wants to eat one vector and one covector. You can feed it only one vector and it will spit you something that wants to eat a covector. But that's precisely what vector are, they are covector eaters. So, really, your tensor can also be seen as eating one vector and spitting back another vector: a linear application.

We can generalize both previous result to prove the general case (36). We transform our linear map

$$\begin{aligned} f : \underbrace{V^* \times \cdots \times V^*}_{p_1 \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q_1 \text{ copies}} \times V^{[*]} &\to \underbrace{V \times \cdots \times V}_{p_2 \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{q_2 \text{ copies}} \\ (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{p_1}, \mathbf{v}_1, \ldots, \mathbf{v}_{q_1}, \boldsymbol{\beta}) &\mapsto f(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{p_1}, \mathbf{v}_1, \ldots, \mathbf{v}_{q_1}, \boldsymbol{\beta}) \end{aligned} \tag{47}$$

into

$$f : \underbrace{V^* \times \cdots \times V^*}_{p_1 \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q_1 \text{ copies}} \to L(V^{[*]}, \underbrace{V \times \cdots \times V}_{p_2 \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{q_2 \text{ copies}})$$
$$(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{p_1}, \mathbf{v}_1, \dots, \mathbf{v}_{q_1}) \mapsto f(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{p_1}, \mathbf{v}_1, \dots, \mathbf{v}_{q_1}, \cdot) \tag{48}$$

Finally, we have to show that the codomain is equivalent to the one we need to obtain. It will be shown in the same way as everything else, but with a subtle technicality, which I don't think you are interested in.

This argument is called partial application, as we didn't feed our function with all of its argument, but we volontarily missed one. This in turn gives a new function that takes this missing argument and gives the same result as if we had not forgotten the missing argument in the first place.

1.2.7 Examples of tensors

$(1, 0)$-tensors are just vectors $(0, 1)$-tensors covectors. I don't think you need any example of why vectors are important, but covectors usually don't appear on their own. You will typically see them constructed from vectors by using a metric, as discussed in section 1.2.3. They also appear as differential form in differential geometry, which are the $\mathrm{d}x$ you are integrating, but a formal introduction to this would take some time.

$(1, 1)$-tensors are linear application, and an example of this is the inertia tensor in rigid body mechanics.

$(0, 2)$-tensors are bilinear forms, of which the greatest example is the metric, be it the Minkowski metric of flat spacetime, the euclidean metric of Newtonian mechanics or more exotic metrics in general relativity.

Finally, the electromagnetic tensor $F^{\mu\nu}$ is a $(2, 0)$ tensor, that eats a covector (typically a derivative ∂_μ) and spits a vector (with Maxwell's equation $\partial_\mu F^{\mu\nu} = J^\nu$).

Don't think to hard about the equations at this point.

1.2.8 Elements from category theory

The previous isomorphism theorem can be interpreted as exchanging the direction of arrows (swapping domain and codomain) in the category of tensors and swapping vector spaces with their dual space. Everytime we do such things in category theory, we like to add a prefix co- (domain, codomain), and it's precisely the reason of the name "covectors" for the elements of the dual space.

2 Transformation rules

We will now see why physicists say that "tensors are object that transform like tensors". Einstein's summation convention on repeated indices will be used throughout.

2.1 Vectors

2.1.1 Vector spaces in physics

We saw that a general tensor is an element of a tensor product space $V_1 \otimes \cdots \otimes V_k$. In physics, we will only consider V_i to be either V or V^*. V will typically be spacetime, a 4-dimensional real vector space.[2] The basis of V will correspond to the directions that a given observer would call time and the three directions of space. We won't think about curvilinear coordinates, so in our case, our vectors are just pointing in the direction of coordinate lines. For example, $\mathbf{e}_x$ is tangential to the line of constant y, z, and t (this last sentence is true even for curvilinear coordinates).

If you want your basis to be physically meaningfull, you don't have complete freedom of choice. If you select any three direction in spacetime, and you call them space, your are forced to select the fourth one if you want it to be purely oriented in time and towards the future for some physical observer, as to not violate special relativity. Moreover, you cannot select any arbitrary direction and call it space, you can only choose among space-like points of spacetime.

2.1.2 Change of basis

There are two ways of seeing a change of reference frame in physics:

- either you are transforming each vectors from V into a new vector of a new vector space V', and now, two observers would disagree on the

[2] In reality, spacetime in general relativity is a Riemannian manifold, and the vector space we are looking at is only the tangent space at a given point. Don't worry about it now.

different vectors, and, for example, the same spacetime event would be associated to two different vectors of their respective vector spaces,

- or each physical things (spacetime events, fields, forces) is associated to a unique vector of V, independent of the observer, and two different observers would simply use a different basis of V and thus get different component for each vector, but would agree on the vectors themselves.

The second point of view is the one more commonly seen in physics.

We take now two different basis $\mathbf{e}_i$ and $\tilde{\mathbf{e}}_i$ of V, and take an arbitrary vector $\mathbf{v} \in V$. You can show that you can always find an invertible linear map $A : V \to V$ such that

$$\tilde{\mathbf{e}}_i = A(\mathbf{e}_i). \tag{49}$$

Because of the property of a basis, we can always write

$$\mathbf{v} = v^i \mathbf{e}_i = \tilde{v}^j \tilde{\mathbf{e}}_j. \tag{50}$$

You still don't have to worry about upper and lower indices.

2.1.3 Transformation of components

Our goal will now be to relate v^i and $\tilde{v}^j$. This will teach us how to change our reference frame. We will first express $\tilde{\mathbf{e}}_j$ in term of the basis $\mathbf{e}_i$:

$$\tilde{\mathbf{e}}_j = A(\mathbf{e}_j) = A^i{}_j \mathbf{e}_i = (A^T)_j{}^i \mathbf{e}_i. \tag{51}$$

For now, think of $A^i{}_j$ to be equivalent to A_{ij}, the matrix component of the linear application A in the basis $\mathbf{e}_k$. If you are not convinced about the order that I put the indices in, and you think it is not consistent with how you know matrix multiplication is defined, we compute

$$\begin{aligned} \mathbf{w} = A(\mathbf{u}) = A(u^j \mathbf{e}_j) = u^j A(\mathbf{e}_j) &= u^j A^i{}_j \mathbf{e}_i && (52)\\ w^i \mathbf{e}_i &= A^i{}_j u^j \mathbf{e}_i && (53)\\ w^i &= A^i{}_j u^j, && (54) \end{aligned}$$

which really is consistent with the definition of matrix multiplication. Indeed, matrix multiplication acts on column-vectors that contains vector components in $\mathbb{R}^n$ and not vectors in V, and in fact, your matrix has to be transposed when you act on basis vectors instead of components.

Now, we can finally relate v^i and $\tilde{v}^j$:

$$v^i \mathbf{e}_i = \tilde{v}^j \tilde{\mathbf{e}}_j \tag{55}$$
$$= \tilde{v}^j A(\mathbf{e}_j) \tag{56}$$
$$= \tilde{v}^j A^i{}_j \mathbf{e}_i \tag{57}$$
$$v^i = A^i{}_j \tilde{v}^j \tag{58}$$
$$\tilde{v}^j = (A^{-1})^j{}_i v^i \tag{59}$$

We can finally understand the origin of the word “contravariant”: vector components transform with the inverse transormation that transform basis vectors, they “vary contrary” to basis vectors.

2.2 Covectors

We'll now do the same for covectors, the vectors of the dual space V^*. We recall that we can construct a basis $\boldsymbol{\varepsilon}^j$ in V^* from a basis of V by imposing

$$\boldsymbol{\varepsilon}^j(\mathbf{e}_i) = \delta^j_i. \tag{60}$$

2.2.1 Change of basis

If we change our basis of V to $\tilde{\mathbf{e}}_i$, we would like to change our basis of V^* to $\tilde{\boldsymbol{\varepsilon}}^j$ as to still satisfy

$$\tilde{\boldsymbol{\varepsilon}}^j(\tilde{\mathbf{e}}_i) = \delta^j_i \tag{61}$$

This yields

$$\tilde{\boldsymbol{\varepsilon}}^j(A(\mathbf{e}_i)) = \delta^j_i \tag{62}$$
$$\tilde{\boldsymbol{\varepsilon}}^j(A^k{}_i \mathbf{e}_k) = \delta^j_i \tag{63}$$
$$\tilde{\boldsymbol{\varepsilon}}^j(\mathbf{e}_k) A^k{}_i = \delta^j_i. \tag{64}$$

We express $\tilde{\boldsymbol{\varepsilon}}^j$ in the $\boldsymbol{\varepsilon}^l$ basis:

$$\tilde{\boldsymbol{\varepsilon}}^j = X_l{}^j \boldsymbol{\varepsilon}^l. \tag{65}$$

The $X_i{}^j$ are for now a set of unkown constant, and the choice of labeling them with upper and lower indices should seem arbitrary now, but will be clear later. Plugging this into (64) gives

$$X_l{}^j \boldsymbol{\varepsilon}^l(\mathbf{e}_k) A^k{}_i = \delta_i^j \tag{66}$$
$$X_l{}^j \delta_k^l A^k{}_i = \delta_i^j \tag{67}$$
$$X_k{}^j A^k{}_i = \delta_i^j \tag{68}$$

This, in matrix notation, is completly equivalent to (and you're encourage to verify that)

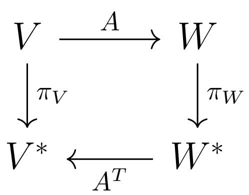
$$X^T A = \mathbb{1} \Leftrightarrow X = A^{-T} \tag{69}$$

where A^{-T} means that we take both the conjugate and the inverse of A.

Plugging this into (60), this means that the transformation rule for basis covectors is

$$\tilde{\boldsymbol{\varepsilon}}^j = A^{-T}(\boldsymbol{\varepsilon}^j) = (A^{-T})_i{}^j \boldsymbol{\varepsilon}^i = (A^{-1})^j{}_i \boldsymbol{\varepsilon}^i \tag{70}$$

For those greedy of some more abstract nonsense, this precisely reflects the definition of the transpose of linear map A, which is the unique linear map A^T that makes the following diagram commute:



where π are the isomorphisms described in (24).[3]

2.2.2 Transformation of components

Because V^* is a vector space in its own right, we can redo this exact same derivation as in section 2.1.3, by considering a covector expressed in both $\boldsymbol{\varepsilon}^i$ and $\tilde{\boldsymbol{\varepsilon}}^j = A^{-T}(\boldsymbol{\varepsilon}^j$:

$$\boldsymbol{\alpha} = \alpha_i \boldsymbol{\varepsilon}^i = \tilde{\alpha}_j \tilde{\boldsymbol{\varepsilon}}^j \tag{71}$$

The components will transform with the inverse of A^{-T}, which is just A^T:

$$\tilde{\alpha}_j = (A^T)_j{}^i \alpha_i \tag{72}$$

This tells us why covectors are called covariant, it's because their components "vary with" the basis vectors of V.

[3] Because the π depends on the choice of basis of V and W, A^T will also depend on this choice.

2.3 Summary

This is important enough to be summarize in a nice table:

	Basis elements	Components
Vectors	$\tilde{\mathbf{e}}_i = (A^T)_i{}^j \mathbf{e}_j$	$\tilde{v}^i = (A^{-1})^i{}_j v^j$
Covectors	$\tilde{\boldsymbol{\varepsilon}}^i = (A^{-1})^i{}_j \boldsymbol{\varepsilon}^j$	$\tilde{\alpha}_i = (A^T)_i{}^j \alpha_j$

We see that components of vectors transform like basis covectors, and components of covectors like basis of vectors. This is pretty much the reason we write vector components and covectors with upper indices, and covector components and vectors with lower indices.

2.3.1 Physics convention

In physics, where our change of basis is mostly given by the lorentz transfromation Λ, we often define things slightly in reverse, and write $A^{-1} = \Lambda$, which gives us the slightly different table

	Basis elements	Components
Vectors	$\tilde{\mathbf{e}}_i = (\Lambda^{-T})_i{}^j \mathbf{e}_j$	$\tilde{v}^i = \Lambda^i{}_j v^j$
Covectors	$\tilde{\boldsymbol{\varepsilon}}^i = \Lambda^i{}_j \boldsymbol{\varepsilon}^j$	$\tilde{\alpha}_i = (\Lambda^{-T})_i{}^j \alpha_j$

2.4 More general tensors

2.4.1 Basis of physical tensor product space

In physics, we are mostly concerned about tensor product of V and V^* only,[4] such that we interpret tensors as multilinear transformation of product of V only. For example, $V \otimes V^*$ is the set of linear map $V \to V$.

We already saw in (14) how to get a basis for a general tensor product. If $\mathbf{e}_i$ is a basis of V and $\boldsymbol{\varepsilon}^j$ is a basis for V^*, we get a basis of $\underbrace{V \otimes \cdots \otimes V}_{p} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q}$ by the set $\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_p} \otimes \boldsymbol{\varepsilon}^{j_1} \otimes \cdots \otimes \boldsymbol{\varepsilon}^{j_q}$.

[4] This is not entirely true. In quantum mechanics, we often consider tensor product of arbitrary vector spaces.

Any tensor will be expressed as

$$T = T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \boldsymbol{\varepsilon}^{j_1} \otimes \dots \otimes \boldsymbol{\varepsilon}^{j_q} \tag{73}$$

By looking at the transformation rule, we'll finally understand why we place the indices in such a way.

2.4.2 Linear transformation

We'll first look at a $(1,1)$-tensor in $V \otimes V^*$:

$$T = T^i{}_j \mathbf{e}_i \otimes \boldsymbol{\varepsilon}^j. \tag{74}$$

This can be interpreted as a linear map by the partial application argument that we discussed in section 1.2.6:

$$\begin{aligned} &V \to V \\ &\mathbf{v} \mapsto T(\mathbf{v}) = T^i{}_j \mathbf{e}_i \boldsymbol{\varepsilon}^j(\mathbf{v}) \end{aligned} \tag{75}$$

By writing $\mathbf{v} = v^k \mathbf{e}_k$, we get

$$T(\mathbf{v}) = v^k T(\mathbf{e}_k) \tag{76}$$
$$= v^k T^i{}_j \mathbf{e}_i \boldsymbol{\varepsilon}^j(\mathbf{e}_k) \tag{77}$$
$$= v^k T^i{}_j \mathbf{e}_i \delta^j_k \tag{78}$$
$$= T^i{}_j v^j \mathbf{e}_i \tag{79}$$

We see at the end that we precisely get the matrix product between the components of T and $\mathbf{v}$. We would like to know how these components would have changed had we used a different basis $\tilde{\mathbf{e}}_i = A(\mathbf{e}_i)$. We'll first write $\mathbf{v} = \tilde{v}^k \tilde{\mathbf{e}}_k$ with $\tilde{v}^k = (A^{-1})^k{}_l v^l$, and then express $\tilde{\mathbf{e}}_k$ in term of $\mathbf{e}_i$, with $\tilde{\mathbf{e}}_k = A^j{}_k \mathbf{e}_j$:

$$T(\mathbf{v}) = \tilde{v}^k T(\tilde{\mathbf{e}}_k) \tag{80}$$
$$= \tilde{v}^k T(A^j{}_k \mathbf{e}_j) \tag{81}$$
$$= \tilde{v}^k A^j{}_k T(\mathbf{e}_j) \tag{82}$$
$$= \tilde{v}^k A^j{}_k T^i{}_j \mathbf{e}_i \tag{83}$$

Finally, we'll express $\mathbf{e}_i = A^{-1}(\tilde{\mathbf{e}}_i)$ back in the $\tilde{\mathbf{e}}_l$ basis, with $\mathbf{e}_i = (A^{-1})^l{}_i \tilde{\mathbf{e}}_l$:

$$T(\mathbf{v}) = A^j{}_k T^i{}_j (A^{-1})^l{}_i \tilde{v}^k \tilde{\mathbf{e}}_l \tag{84}$$

But the matrix expression of T in the $\tilde{\mathbf{e}}_i$ basis is precisely the matrix such that

$$T(\mathbf{v}) = T(\tilde{v}^k \tilde{\mathbf{e}}_k) = \tilde{T}^l{}_k v^k \tilde{\mathbf{e}}_l \tag{85}$$

thus we must have

$$\tilde{T}^l{}_k = A^j{}_k T^i{}_j (A^{-1})^l{}_i = (A^T)_k{}^j (A^{-1})^l{}_i T^i{}_j \tag{86}$$

We see that we have to multiply it once by A^T and once by A^{-1}, which justify the name one time contravariant and one time covariant.

I would urge you not to absolutely want to write it in matrix form. This will force you to write some term on the left or on the right of T, and to put some random transposes at the right place. In the end, this is only a matter of convention, but sticking to index notation is much more common in physics (and I would say, less prone to error).

2.4.3 General case

We will know look at a general (p, q)-tensor

$$T = T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \boldsymbol{\varepsilon}^{j_1} \otimes \dots \otimes \boldsymbol{\varepsilon}^{j_q}. \tag{87}$$

Because we know the transformation of each of the basis vectors and covectors, the result is almost instantaneous. We use the inverse relations $\mathbf{e}_i = (A^{-T})_i{}^j \tilde{\mathbf{e}}_j = (A^{-1})^j{}_i \tilde{\mathbf{e}}_j$ and $\boldsymbol{\varepsilon}^i = A^i{}_j \tilde{\boldsymbol{\varepsilon}}^j = (A^T)_j{}^i \tilde{\boldsymbol{\varepsilon}}^j$:

$$T = T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \left(\mathbf{e}_{i_1} \otimes \dots \otimes \mathbf{e}_{i_p} \otimes \boldsymbol{\varepsilon}^{j_1} \otimes \dots \otimes \boldsymbol{\varepsilon}^{j_q} \right) \tag{88}$$

$$= T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \Big((A^{-1})^{k_1}_{\ i_1} \tilde{\mathbf{e}}_{k_1} \otimes \dots \otimes (A^{-1})^{k_p}_{\ i_p} \tilde{\mathbf{e}}_{k_p} \otimes (A^T)_{l_1}^{\ j_1} \tilde{\boldsymbol{\varepsilon}}^{l_1} \otimes \dots \otimes (A^T)_{l_q}^{\ j_q} \tilde{\boldsymbol{\varepsilon}}^{l_q} \Big) \tag{89}$$

$$= T^{i_1 \dots i_p}{}_{j_1 \dots j_q} (A^{-1})^{k_1}_{\ i_1} \dots (A^{-1})^{k_p}_{\ i_p} (A^T)_{l_1}^{\ j_1} \dots (A^T)_{l_q}^{\ j_q} \left(\tilde{\mathbf{e}}_{i_1} \otimes \dots \otimes \tilde{\mathbf{e}}_{i_p} \otimes \tilde{\boldsymbol{\varepsilon}}^{j_1} \otimes \dots \otimes \tilde{\boldsymbol{\varepsilon}}^{j_q} \right) \tag{90}$$

$$\equiv \tilde{T}^{k_1 \dots k_p}{}_{l_1 \dots l_q} \left(\tilde{\mathbf{e}}_{i_1} \otimes \dots \otimes \tilde{\mathbf{e}}_{i_p} \otimes \tilde{\boldsymbol{\varepsilon}}^{j_1} \otimes \dots \otimes \tilde{\boldsymbol{\varepsilon}}^{j_q} \right) \tag{91}$$

where we used the fundamental multilinear property of tensors to go from (89) to (90). So, we have the very general transformation rule for its component:

Transformation rule

$$\tilde{T}^{k_1 \dots k_p}{}_{l_1 \dots l_q} = (A^{-1})^{k_1}{}_{i_1} \dots (A^{-1})^{k_p}{}_{i_p} (A^T)_{l_1}{}^{j_1} \dots (A^T)_{l_q}{}^{j_q} T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \tag{92}$$

$$\tilde{T}^{k_1 \dots k_p}{}_{l_1 \dots l_q} = \Lambda^{k_1}{}_{i_1} \dots \Lambda^{k_p}{}_{i_p} (\Lambda^{-T})_{l_1}{}^{j_1} \dots (\Lambda^{-T})_{l_q}{}^{j_q} T^{i_1 \dots i_p}{}_{j_1 \dots j_q} \tag{93}$$

So, we see that for each p upper indices, we have p repeated A^{-1} factors (or Λ in physics convention), and for each q lower indices q repeated A^T factors (or Λ^{-T} in physics). It is thus p times contravariant, and q times covariant.

We could have use the proof we did in this section to prove the special case of linear transformations of the previous section much more easily.