How to lose your fear of tensor products

Timothy Gowers

July 2023

1 How to lose your fear of tensor products

If you are not in the slightest bit afraid of tensor products, then obviously you do not need to read this page. However, if you have just met the concept and are like most people, then you will have found them difficult to understand. The aim of this page is to answer three questions:

1. What is the point of tensor products?

2. Why are they defined as they are?

3. How should one answer questions involving them?

2 Why bother to introduce tensor products?

One of the best ways to appreciate the need for a definition is to think about a natural problem and find oneself more or less forced to make the definition in order to solve it. Here, then, is a very basic question that leads, more or less inevitably, to the notion of a tensor product. (If you *really* want to lose your fear of tensor products, then read the question and try to answer it for yourself.)

Let V, W and X be vector spaces over $\mathbb{R}$. (What I have to say works for any field $\mathbb{F}$, and in fact under more general circumstances as well.) A function $f : V \times W \to X$ is called *bilinear* if it is linear in each variable separately. That is, $f(av+bv', w) = af(v,w)+bf(v',w)$ and $f(v, cw+dw') = cf(v,w)+df(v,w')$ for all possible choices of a, b, c, d, v, v', w, w'.

I shall take it for granted that bilinear maps are worth knowing about—they crop up all over the place—and try to justify tensor products given that assumption.

Now, bilinear maps are clearly related to linear maps, and there are questions one can ask about linear maps that one can also ask about bilinear ones. For example, if $f : V \to W$ is a linear map between finite-dimensional vector spaces V and W, then one thing we like to do is *encode* it using a collection of numbers. The usual way to do this is to take bases of V and W and define a matrix A. To obtain the jth column of this matrix, one takes the jth basis vector e_i of V,

writes $f(e_j)$ as a linear combination of the vectors in the basis of W, and uses those coefficients.

The reason that the matrix encodes the linear map is that if you know $f(e_j)$ for every j then you know f: if v is a linear combination of the e_j then $f(v)$ is the corresponding linear combination of the $f(e_j)$.

This suggests two questions about bilinear maps:

1. Can bilinear maps be encoded in a natural way using just a few real numbers?

2. Let V, W and X be finite-dimensional vector spaces, let $f : V \times W \to X$ be a bilinear map, and let $\{(v_i, w_i) : i = 1, 2, \ldots, n\}$ be a collection of pairs of vectors in $V \times W$. When is f completely determined by the values of the $f(v_i, w_i)$?

The first question has an easy answer. Pick bases of V, W and X. If you know $f(v_i, w_j)$ whenever v_i and w_j are basis vectors then you know $f(v, w)$ for all pairs (v, w) in $V \times W$—by bilinearity. But each $f(v_i, w_j)$ is a vector in X and can therefore be written in terms of the basis of X. Thus, if the dimensions of V, W, and X are p, q and r respectively, it is enough to specify pqr numbers (in a sort of 3-dimensional matrix) in order to specify f. Furthermore, it is not hard to see that every p-by-q-by-r grid of numbers specifies a bilinear map: the number in position (i, j, k) tells us the kth coordinate of $f(v_i, w_j)$.

This observation provides a partial answer to the second question as well. If the pairs (v_i, w_i) run over all pairs (e_i, f_j), where (e_i) and (f_j) are bases of V and W, then the values of $f(v_i, w_i)$ determine f. However, this is not the only way to fix f. For example, let $V = W = \mathbb{R}^2$ and let $X = \mathbb{R}$. Let $e_1 = (1, 0)$ and $e_2 = (0, 1)$. If you know the values of $f(e_1, e_1)$, $f(e_2, e_2)$, $f(e_1 + e_2, e_1 + e_2)$ and $f(e_1 + e_2, e_1 + 2e_2)$, then you know f. Why? Well,

$$f(e_1 + e_2, e_1 + e_2) - f(e_1, e_1) - f(e_2, e_2) = f(e_1, e_2) + f(e_2, e_1) \qquad (1)$$

and

$$f(e_1 + e_2, e_1 + 2e_2) - f(e_1, e_1) - 2f(e_2, e_2) = 2f(e_1, e_2) + f(e_2, e_1) \qquad (2)$$

which allows us to work out $f(e_1, e_2)$ and $f(e_2, e_1)$, and hence determines f.

On the other hand, f is not determined by $f(e_1, e_1)$, $f(e_2, e_2)$, $f(e_1 + e_2, e_1 + e_2)$, and $f(e_1 - e_2, e_1 - e_2)$. For example, if these values are all 0, then f could be identically 0 or it could be defined by $f((a, b), (c, d)) = ad - bc$.

How, then, are we to say which sets of pairs fix f and which do not? This is the point at which, if you do not know the answer already, I would suggest reading no further and trying to work it out for yourself.

3

There is no doubt that what we are looking for is something *like* a 'basis' of pairs (v, w) in $V \times W$. This should be 'independent' in the sense that the value

of f on any pair cannot be deduced from the value of f on the other pairs (or, equivalently, we can choose the values of f at the pairs however we like), and 'spanning' in the sense discussed above—that f is determined by its values on the given pairs.

Equally, there is no doubt that we are not looking for a basis of the vector space $V \times W$ itself. For example, if V and W are both $\mathbb{R}$, then $(1, 0)$ and $(0, 1)$ form a basis of $V \times W$, but to be told the values of a bilinear map $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ at $(1, 0)$ and $(0, 1)$ is to be told nothing at all, since they have to be 0.

To get a feel for what is happening, let us solve the problem in this special case (that is, when $V = W = \mathbb{R}$). Suppose we know the value of $f(a, b)$. We have just seen that this information is useless if either a or b is zero, but otherwise it completely determines f, since $f(x, y) = \frac{xy}{ab} f(a, b)$.

Perhaps that was too simple for any generalization to suggest itself, so let us try $V = W = \mathbb{R}^2$. Suppose that we know the value of $f(v, w)$. That tells us $f(av, bw)$ for any pair of scalars a and b, so if we want to introduce a second pair (v', w') which is independent of the first (not that we know quite what this means) then we had better make sure that either v' is not a multiple of v or w' is not a multiple of w. If we have done that, then what can we deduce from the values of $f(v, w)$ and $f(v', w')$? Well, we have the values of all $f(cv', dw')$, but unless v' is a multiple of v or w' is a multiple of w it seems to be difficult to deduce much else, because in order to use the fact that $f(x, y + z) = f(x, y) + f(x, z)$ or $f(x + y, z) = f(x, z) + f(y, z)$ we need to have one coordinate kept constant.

One thing we *can* say, which doesn't determine other values of f but at least places some restriction on them, is that

$$f(v + v', w + w') = f(v, w) + f(v, w') + f(v', w) + f(v', w'). \qquad (3)$$

Since we know $f(v, w)$ and $f(v', w')$, this means that $f(v, w')$, $f(v', w)$ and $f(v+v', w+w')$ cannot be freely and independently chosen—once you've chosen two of them it fixes the third. But this isn't terribly exciting, so let's look at more pairs.

I shall now change notation, and imagine that we know the values of $f(s, t)$, $f(u, v)$, $f(w, x)$ and $f(y, z)$. If s, u and w are all multiples of a single vector, then some of this information is redundant, since t, v and x are not linearly independent (they live in $\mathbb{R}^2$). So let us suppose that no three of the first vectors are multiples and no three of the second are. It follows easily that we can take two pairs, without loss of generality (s, t) and (u, v), and assume that s and u are linearly independent, and that so are t and v.

We can now write $w = as + bu$, $x = ct + dv$, $y = es + gu$, $z = ht + kv$, and we know, by bilinearity, that

$$f(w, x) = acf(s, t) + adf(s, v) + bcf(u, t) + bdf(u, v) \qquad (4)$$

and

$$f(y, z) = ehf(s, t) + ekf(s, v) + ghf(u, t) + gkf(u, v). \qquad (5)$$

Since we know $f(s, t)$, $f(u, v)$, $f(w, x)$ and $f(y, z)$, this gives us two linear equations for $f(s, v)$ and $f(u, t)$. They will have a unique solution as long as

3

$adgh$ does not equal $bcek$. In this case we have determined f completely, since

$$f(a's + b'u, c't + d'v) = a'c'f(s,t) + a'd'f(s,v) + b'c'f(u,t) + b'd'f(u,v), \quad (6)$$

the pair on the left hand side can be anything and we know all about values of f on the right hand side.

Notice that if $adgh$ does equal $bcek$ then an appropriate linear combination of the above two equations (ek times the first minus ad times the second) gives a linear equation that is automatically satisfied by $f(s,t)$, $f(u,v)$, $f(w,x)$ and $f(y,z)$. In other words, the pairs (s,t), (u,v), (w,x) and (y,z) are not 'independent', in the sense that f of three of them determines f of the fourth.

So now we understand the case $V = W = \mathbb{R}^2$ reasonably well, but it is not quite obvious how to generalize the above argument to arbitrary spaces V and W. Before we try, let us think a little further about what we have already proved, and how we did it. In particular, let us see if we can be more specific about 'independence' of pairs in $V \times W$.

Why, for instance, did we say that (s,t), (u,v), (w,x) and (y,z) were not 'independent' when $adgh = bcek$? Was there some 'linear combination' that gave a 'dependence'? Well, I mentioned that there was a linear equation automatically satisfied by $f(s,t)$, $f(u,v)$, $f(w,x)$ and $f(y,z)$. To be specific, it is

$$ekf(w,x) - adf(y,z) = (ekac - adeh)f(s,t) + (ekbd - adgk)f(u,v). \quad (7)$$

This looks like a linear dependence between $f(w,x)$, $f(y,z)$, $f(s,t)$ and $f(u,v)$, but it isn't quite that because these are just four real numbers, and we are trying to say something more interesting than that the dimension of $\mathbb{R}$ is less than 4. What we are saying is more like this: if f is an *arbitrary* bilinear function, then the above linear equation will *always* be satisfied.

How can we express that statement in terms of straightforward linear algebra? Here are a few suggestions.

4 A first way of making sense of 'independence' of pairs.

We could think of f, in an expression like $f(u,v)$, as standing for all bilinear functions at once. So then we would make a statement like $f(u,v) = 2f(w,x)$ only if this equation was true for every f, rather than for some specific f. We could even make this formal in a rather naughty way as follows. Let B be the set of all bilinear maps defined on $V \times W$. (That's the naughtiness —B is too big to be a set, but actually we will see in a moment that it is enough to look just at bilinear maps into $\mathbb{R}$.) Now regard (u,v) as a function defined on B—if f is a map in B then $(u,v)(f)$ is just $f(u,v)$. Then the statement that $f(u,v)$ always equals $2f(w,x)$ is the statement that (u,v), considered as a function on B, is twice (w,x), considered as a function on B.

Just so that we don't have to keep writing the phrase 'considered as a function on B', let us invent some notation. When I want to think of (u, v) as a function on B I'll write $[u, v]$ instead. So now the dependence that I wrote earlier becomes

$$ek[w, x] - ad[y, z] = (ekac - adeh)[s, t] + (ekbd - adgk)[u, v]. \qquad (8)$$

This dependence really is genuine linear dependence, in the vector space of functions from B to... well, some rather complicated big sum of vector spaces or something. Instead of bothering to sort out that little difficulty, let's see why it is in fact enough to let B be the set of all bilinear functions to $\mathbb{R}$, otherwise known as bilinear forms.

5   Reducing to the case of bilinear maps into $\mathbb{R}$.

Here again there is an analogy with linear maps. Suppose that V and W are finite-dimensional vector spaces and $f : V \to W$ is a linear map. Let $w_1, \ldots, w_m$ be a basis for W. If we write vectors in W in coordinate form using this basis, then we will write $f(v)$ as $(f_1 v, \ldots, f_m v)$, and in that way we see that a linear map to an m-dimensional vector space can be thought of as a sequence of m linear maps to $\mathbb{R}$.

Exactly the same is true of a bilinear map $f : V \times W \to X$ if X is m-dimensional—we can think of it as a sequence $(f_1, \ldots, f_m)$ of bilinear maps from $V \times W$ to $\mathbb{R}$. From this observation we can make the following simple deduction. If

$$a_1 f(v_1, w_1) + \ldots + a_n f(v_n, w_n) = 0 \qquad (9)$$

for every bilinear map $f : V \times W \to \mathbb{R}$, then it is zero for every bilinear map from $V \times W$ to *any* finite-dimensional vector space X.

In fact, one can even do away with the condition that X should be finite-dimensional, as follows. If $f : V \times W \to X$ is a bilinear map such that

$$a_1 f(v_1, w_1) + \ldots + a_n f(v_n, w_n) = x \qquad (10)$$

for some non-zero vector x, then let g be a linear map from X to $\mathbb{R}$ such that $g(x)$ is not zero. The existence of this map can be proved as follows. Using the axiom of choice, one can show that the vector x can be extended to a basis of X. Let $g(x) = 1$, let $g(y) = 0$ and extend linearly. Once we have g, we have a bilinear map $gf : V \times W \to \mathbb{R}$ such that

$$a_1 gf(v_1, w_1) + \ldots + a_n gf(v_n, w_n) \qquad (11)$$

is non-zero.

This use of the axiom of choice was not very pleasant, but we shall soon see that it can be avoided.

6 Back to the main discussion.

Let us therefore redefine B to be the set of all bilinear maps from $V \times W$ to $\mathbb{R}$, and regard $[v, w]$ as notation for the function from B to $\mathbb{R}$ defined by $[v, w](f) = f(v, w)$. (This is the same definition as before, apart from the restriction of B to real-valued bilinear maps.) Now that $[v, w]$ is a function from B to $\mathbb{R}$, it is completely clear in what sense it lives in a vector space, what is meant by linear dependence and so on. Everything takes place inside the vector space of all real-valued functions on B.

The idea of defining functions such as $[v, w]$ was that it gave us a way of reformulating our original problem—when does a set of pairs (v_i, w_i) fix all bilinear maps?—in terms of concepts we know from linear algebra—it fixes all bilinear maps if and only if the vector space spanned by the functions $[v_i, w_i]$ contains all functions of the form $[v, w]$. Moreover, the set of pairs contains no redundancies if and only if the functions $[v_i, w_i]$ are linearly independent.

7 A second way of converting the problem into linear algebra.

Have we really said anything at all? It might seem not. After all, if we are given a set of pairs (v_i, w_i) and asked whether the corresponding functions $[v_i, w_i]$ span all functions $[v, w]$, we will find ourselves making exactly the same calculations that we would make if instead we asked whether the values of $f(v_i, w_i)$, for some unknown bilinear function f, determined the value of $f(v, w)$.

However, turning the problem into linear algebra does clarify it somewhat, especially if we can say something about the vector space generated by the functions $[v, w]$ (which contains all the functions on B we will ever need to worry about). It also gives us a more efficient notation.

Let me write down a few facts about this vector space.

1. $[v, w + w'] = [v, w] + [v, w']$

2. $[v + v', w] = [v, w] + [v', w]$

3. $[av, w] = a[v, w]$

4. $[v, aw] = a[v, w]$

The above relations hold for all vectors v, v' in V and w, w' in W, and all real numbers a. Now if we regard an equation like $[v, w + w'] = [v, w] + [v, w']$ as nothing but a shorthand for the statement that $f(v, w + w') = f(v, w) + f(v, w')$ for all bilinear maps f, then it would seem that everything about the space spanned by the functions $[v, w]$ ought to follow from these four facts. After all, the linear relationships between the functions $[v, w]$ are supposed to be the ones that we can deduce about the corresponding vectors $f(v, w)$ when *all* we know about f is that it is bilinear. So they ought to follow from the axioms for bilinearity—which translate into facts 1-4 above.

Let us try to make this hunch precise and prove it. To do this we must ask ourselves which linear dependences can be deduced from 1-4, and then whether those are all the linear relationships that hold for every bilinear function. In a moment I will express this less wordily, but for now let us note that the first question is easy. Facts 1-4 give us four linear equations, which we can make a bit more symmetrical by rewriting them as

1. $[v, w + w'] - [v, w] - [v, w'] = 0$

2. $[v + v', w] - [v, w] - [v', w] = 0$

3. $[av, w] - a[v, w] = 0$

4. $[v, aw] - a[v, w] = 0$

Which other linear combinations of pairs must be zero if these ones are? Answer: all linear combinations of these combinations, and nothing else. This is because the only method we have of deducing linear equations from other linear equations is forming linear combinations of those equations.

There was something not quite satisfactory about what I have just said. It does seem to be true, but let us try to state and prove it more mathematically, without appealing to phrases like 'method of deducing'. It is certainly clear that if

$$a_1[v_1, w_1] + a_2[v_2, w_2] + \ldots + a_n[v_n, w_n] \tag{12}$$

is a linear combination of functions of the forms on the left-hand sides of 1-4 (in the second version), then

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) = 0 \tag{13}$$

for every bilinear function f. What is not quite so obvious is the converse: that for every *other* function of the form

$$a_1[v_1, w_1] + a_2[v_2, w_2] + \ldots + a_n[v_n, w_n] \tag{14}$$

we can find some bilinear map f such that

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) \tag{15}$$

does not equal 0. However, it turns out that there is an almost trivial way to show this. For every pair (v, w) in $V \times W$, regard $[[v, w]]$ as a meaningless symbol. We can define a rather large vector space Z by taking formal linear combinations of these symbols. By that I mean that Z consists of all expressions of the form

$$a_1[[v_1, w_1]] + a_2[[v_2, w_2]] + \ldots + a_n[[v_n, w_n]] \tag{16}$$

with obvious definitions for addition and scalar multiplication.

Next, we let E be the subspace of Z generated by all vectors of one of the following four forms (which you should be able to guess):

1. $[[v, w + w']] - [[v, w]] - [[v, w']]$

2. $[[v + v', w]] - [[v, w]] + [[v', w]]$

3. $[[av, w]] - a[[v, w]]$

4. $[[v, aw]] - a[[v, w]]$

We want everything in E to 'be zero', in some appropriate sense. The standard way to make that happen is to take a quotient space Z/E. (If you are hazy about the definition, here is a reminder. Two vectors z and z' in Z are regarded as equivalent if $z - z'$ belongs to E. The vectors in Z/E are equivalence classes, which can be written in the form $z + E$. That is, if K is an equivalence class and z belongs to K, then it is easy to see that $K = \{z + e : e \in E\} = z + E$. Addition and scalar multiplication are defined by $(z + E) + (z' + E) = z + z' + E$ and $a(z + E) = az + E$. It is not hard to check that these are well defined—that is, independent of the particular choices of z and z'.)

Why does this quotient space Z/E help us? Because it gives us a trivial proof of the assertion we wanted before. If

$$a_1[v_1, w_1] + a_2[v_2, w_2] + \ldots + a_n[v_n, w_n] \tag{17}$$

is not a linear combination of expressions of the form

1. $[v, w + w'] - [v, w] - [v, w']$

2. $[v + v', w] - [v, w] + [v', w]$

3. $[av, w] - a[v, w]$

4. $[v, aw] - a[v, w]$

then trivially

$$z = a_1[[v_1, w_1]] + a_2[[v_2, w_2]] + \ldots + a_n[[v_n, w_n]] \tag{18}$$

is not a linear combination of vectors of the form

1. $[[v, w + w']] - [[v, w]] - [[v, w']]$

2. $[[v + v', w]] - [[v, w]] + [[v', w]]$

3. $[[av, w]] - a[[v, w]]$

4. $[[v, aw]] - a[[v, w]]$

In other words, z does not belong to the subspace E. In other words again, $z + E$ is not zero in the quotient space Z/E. To complete the proof, it is enough to find a bilinear map f from $V \times W$ to Z/E such that

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) = z + E \tag{19}$$

and in particular is non zero. What is the most obvious map one can possibly think of? Well, $f(v, w) = [[v, w]] + E$ seems a good bet. Is it bilinear? Yes, by

8

the way we designed E. For example, $f(v, w + w') = [[v, w + w']] + E$, and since $[[v, w + w']] - [[v, w]] - [[v, w']]$ belongs to E, we may deduce that

$$f(v, w + w') = [[v, w]] + [[v, w']] + E = ([[v, w]] + E) + ([[v, w']] + E) \\ = f(v, w) + f(v, w') \tag{20}$$

To sum up, we have just proved the following (not very surprising) proposition.

Proposition. *A linear combination of functions of the form $[v, w]$ is zero if and only if it is generated by functions of the form $[av, w] - a[v, w]$, $[v, aw] - a[v, w]$, $[v, w + w'] - [v, w] - [v, w']$ and $[v + v', w] - [v, w] - [v', w]$.*

8   How to think about tensor products.

What has all this to do with tensor products? Now is the time to admit that I have already defined tensor products—in two different ways. They are a good example of the phenomenon discussed in my page about definitions: exactly how they are defined is not important: what matters is the properties they have.

The usual notation for the tensor product of two vector spaces V and W is $V \otimes W$. A typical element of $V \otimes W$ will be a linear combination of elements written $v \otimes w$. You can regard $v \otimes w$ as an alternative notation for $[v, w]$, or for $[[v, w]] + E$—it doesn't matter which as the above discussion shows that the space spanned by $[v, w]$ is isomorphic to the space spanned by $[[v, w]] + E$, via the (well-defined) linear map that takes $[v, w]$ to $[[v, w]] + E$ and extends linearly.

A tempting mistake for beginners is to think that every element of $V \otimes W$ is of the form $v \otimes w$, but this is just plain false. For example, if v, v', w and w' are vectors with no linear dependences, then $v \otimes w + v' \otimes w'$ cannot be written in that form. (If it could then there would be some pair (v'', w'') such that a bilinear map f satisfied $f(v'', w'') = 0$ if and only if $f(v, w) + f(v', w') = 0$. It is not hard to convince yourself that there is no such pair—indeed I more or less proved it in the discussion above about two-dimensional spaces.)

Another tempting mistake is to pay undue attention to how tensor products are constructed. I should say that the standard construction is the second one I gave, that is, the quotient space. Suppose that we try to solve problems by directly using this definition, or rather construction. They suddenly seem rather hard. For example, let v' and w' be non-zero vectors in V and W. How can we show that $v' \otimes w'$ is not zero? Well, to do so directly from the quotient space definition, we need to show that the pair $[[v', w']]$ does not belong to the space E defined earlier. In order to prove that, we somehow need to find a property of $[[v', w']]$ that is not shared by any linear combination of vectors of the four forms listed above.

Let us ask ourselves a very general question: how does one ever show that a certain point v does not lie in a certain subspace W of a vector space? If the space is $\mathbb{R}^n$ and we are given a basis of the subspace, then our task is to show

that a system of linear equations has no solution. In a more abstract set-up, the natural method—in fact, more or less the only method—is to find a linear map T from V to some other vector space ($\mathbb{R}$ will always be possible) such that Tv is not zero but Tw is zero for every w in W.

Returning to the example at hand, can we find a linear map that sends everything in E to zero and $[[v', w']]$ to something non-zero? Let us remind ourselves of our earlier proposition.

Proposition. *(repeated) A linear combination of functions of the form $[v, w]$ is zero if and only if it is generated by functions of the form $[av, w] - a[v, w]$, $[v, aw] - a[v, w]$, $[v, w + w'] - [v, w] - [v, w']$ and $[v + v', w] - [v, w] - [v', w]$.*

That gives us an obvious map that takes everything in E to zero: just map $[[v, w]]$ to $[v, w]$ and extend linearly. So we are then done if $[v', w']$ is non-zero. But for $[v', w']$ not to be zero, all we have to do is come up with a bilinear map f from $V \times W$ to $\mathbb{R}$ such that $f(v', w')$ is not zero. To do this, extend the singletons $\{v'\}$ and $\{w'\}$ to bases of V and W and for any pair of basis vectors (x, y) let $f(x, y) = 0$ unless $x = v'$ and $y = w'$ in which case let $f(x, y) = 1$. Then extend f bilinearly.

Well, we have solved the problem, but we didn't really do it directly from the quotient-space definition. Indeed, we got out of the quotient space as quickly as we could. How much simpler it would have been to start thinking immediately about bilinear functions. In order to show that $v \otimes w$ is non-zero, we could have regarded it as $[v, w]$ instead, and instantly known that $v \otimes w$ is non-zero if and only if there is a bilinear function f defined on $V \times W$ such that $f(v, w)$ does not vanish.

So, here is a piece of advice for interpreting a linear equation involving expressions of the form $v \otimes w$. *Do not* worry about what the objects themselves mean, and instead use the fact that

$$a_1 v_1 \otimes w_1 + a_2 v_2 \otimes w_2 + \ldots + a_n v_n \otimes w_n = 0 \tag{21}$$

if and only if

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) = 0 \tag{22}$$

for every bilinear function f defined on $V \times W$. (We proved this earlier, except that instead of $v_i \otimes w_i$ we wrote $[v_i, w_i]$.)

Now algebraists have a more grown-up way of saying this, which runs as follows. Here is a sentence from earlier in this page:

What we are saying is more like this: if f is an *arbitrary* bilinear function, then the above linear equation will *always* be satisfied.

It would be nice if there were a bilinear map g on $V \times W$ that was so 'generic' that we could regard it itself as an 'arbitrary' bilinear map. But there *is* such a map, and we have more or less defined it. It takes (v, w) to $v \otimes w$. The bilinearity of this map is obvious (if you don't find it obvious then you are forgetting to

use the fact I have just mentioned and recommended). As for its 'arbitrariness', the fact above can be translated as follows:

$$a_1 g(v_1, w_1) + a_2 g(v_2, w_2) + \ldots + a_n g(v_n, w_n) = 0 \tag{23}$$

if and only if

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) = 0 \tag{24}$$

for every bilinear function f defined on $V \times W$. In brief, no linear equation holds for g unless it holds for all bilinear functions.

How do algebraists express this 'arbitrariness'? They say that the tensor product has a *universal property*. The bilinear map g is in a certain sense 'as big as possible'. To see what this sense is, let us return to our main fact in its $v_i \otimes w_i$ formulation. Let $f : V \times W \to U$ be some bilinear map, and let us try to define a linear map $h : V \otimes W \to U$ by sending $v \otimes w$ to $f(v, w)$ and extending linearly. It is not quite obvious that this is well-defined, since we must check that if we write an element of $V \otimes W$ in two different ways as linear combinations of $v \otimes w$'s, then the corresponding linear combinations of $f(v, w)$'s are equal. But this is exactly what is guaranteed by the main fact. So h is a well-defined linear map, and $h \circ g = f$, since $h \circ g(v, w) = h(v \otimes w) = f(v, w)$, by the definition of h. Moreover, it is clear that h is the only linear map such that $h \circ g = f$, since $h(v \otimes w)$ is forced to be $f(v, w)$ and we are forced to extend linearly. We have therefore proved the following.

Proposition. *For every bilinear map $f : V \times W \to U$ there is a unique linear map $h : V \otimes W \to U$ such that $h \circ g = f$, where g is the bilinear map from $V \times W$ to $V \otimes W$ that takes (v, w) to $v \otimes w$.*

The map f is said to *factor uniquely through g*.

Now let us see why this proposition says exactly the same as what I have called the main fact about $V \otimes W$. Since it followed from the main fact, all I have to do is show that reverse implication holds as well: assuming this proposition we can recover the main fact. Suppose therefore that there is a bilinear function f such that

$$a_1 f(v_1, w_1) + a_2 f(v_2, w_2) + \ldots + a_n f(v_n, w_n) \tag{25}$$

is not zero. Since we can write f as $h \circ g$, and since h is linear, it follows that

$$a_1 g(v_1, w_1) + a_2 g(v_2, w_2) + \ldots + a_n g(v_n, w_n), \tag{26}$$

which equals

$$a_1 v_1 \otimes w_1 + a_2 v_2 \otimes w_2 + \ldots + a_n v_n \otimes w_n, \tag{27}$$

is also non-zero. And that establishes the fact.

A useful lemma about the tensor product is that it is unique, in the following sense.

Lemma. *Let U and V be vector spaces, and let $b : U \times V \to X$ be a bilinear map from $U \times V$ to a vector space X. Suppose that for every bilinear map f defined on $U \times V$ there is a unique linear map c defined on X such that $f = c \circ b$. Then there is an isomorphism $i : X \to U \otimes V$ such that $u \otimes v = i \circ b(u,v)$ for every (u,v) in $U \otimes V$.*

We can avoid mentioning $u \otimes v$ if we use the map $g : U \times V \to U \otimes V$. Then the lemma says that $g = i \circ b$. Briefly, the point of the lemma is that any bilinear map $b : U \times V \to X$ satisfying the universal property is isomorphic to the map $g : U \times V \to U \otimes V$ in an obvious sense.

Proof. Applying the hypothesis about b to the bilinear map $g : U \times V \to U \otimes V$, we obtain a linear map $i :\to U \otimes V$ such that $g = i \circ b$. Similarly, applying the universal property of g to the bilinear map b, we obtain a linear map $j : U \otimes V \to X$ such that $b = j \circ g$. It follows that $b = j \circ g = j \circ i \circ b$. Now let c be the identity on X. Then $b = c \circ b$. So by the uniqueness part of the hypothesis on X (applied when $f = b$) we find that $j \circ i = c$. Similarly, $i \circ j$ is the identity on $U \otimes V$, which shows that i is an isomorphism. $\square$

The reason algebraists prefer to talk about the universal property of $V \otimes W$ and factorization of maps is that it enables them to avoid dirtying their hands by considering the actual *elements* of $V \otimes W$. It can be hard to get used to this spaces-rather-than-objects way of thinking, so let me prove that the tensor product is associative (in the sense that there is a natural isomorphism between $U \otimes (V \otimes W)$ and $(U \otimes V) \otimes W$), first by using the main fact and then by using the universal property.

9 The associativity of the tensor product

Since $V \otimes W$ is a vector space, it makes perfectly good sense to talk about $U \otimes (V \otimes W)$ when U is another vector space. A typical element of $U \otimes (V \otimes W)$ will be a linear combination of elements of the form $u \otimes x$, where x itself is a linear combination of elements of $V \otimes W$ of the form $v \otimes w$. Hence, we can write any element of $U \otimes (V \otimes W)$ as

$$u_1 \otimes (v_1 \otimes w_1) + \ldots + u_n \otimes (v_n \otimes w_n). \tag{28}$$

(Here I have used facts such as that $a(x \otimes y) = x \otimes ay = ax \otimes y$ and $x \otimes (y + z) = x \otimes y + x \otimes z$.) Since we can say something very similar about elements of $(U \otimes V) \otimes W$, there is a very tempting choice for the definition of a (potential) isomorphism between the two spaces, namely that the above vector should map to

$$(u_1 \otimes v_1) \otimes w1 + \ldots + (u_n \otimes v_n) \otimes w_n. \tag{29}$$

Indeed, this works, but I haven't proved it yet because I haven't demonstrated that it is well-defined. For this it is enough to prove that if the first vector is zero then the second must be as well. And now there is a slight

problem. We would like to make everything simple by converting the question into one about bilinear maps, but we find ourselves looking at bilinear maps on $U \times (V \otimes W)$, and $V \otimes W$ is itself an object that we want to try to avoid thinking about too directly.

Let us think, though, what a bilinear map f defined on $U \times (V \otimes W)$ is like. By definition it is linear in each variable when the other is held fixed. So every u in U defines for us a linear map f_u on $V \otimes W$ by the formula $f_u(x) = f(u, x)$. But linear maps on $V \otimes W$ correspond to *bilinear* maps on $V \times W$: given a bilinear map on $V \times W$ we have seen how to associate a linear map on $V \otimes W$, and the reverse process is trivial—just compose it with the bilinear map $g : V \times W \to V \otimes W$. So it looks as though bilinear maps on $U \times (V \otimes W)$ ought to correspond to *trilinear* maps on $U \times (V \times W) = U \times V \times W$. (That last equality is, strictly speaking, a very natural bijection rather than an exact equality, since $(u, (v, w))$ is not the same object as (u, v, w).)

This should interest us, because the definition of trilinear maps on $U \times V \times W$ makes no mention of how you bracket the product, and that ought to help if we are searching for an associativity proof. So let us try to make the connection precise.

First, then, let us take a bilinear map f defined on $U \times (V \otimes W)$ and try to associate with it a trilinear map e on $U \times V \times W$. There is an obvious candidate: $e(u, v, w) = f(u, v \otimes w)$, and it is easy to check that e is indeed trilinear.

As for the other direction, let e be a trilinear map defined on $U \times V \times W$. It isn't immediately obvious how to proceed, so let us remind ourselves what we do know: that bilinear maps on $V \times W$ correspond to linear maps on $V \otimes W$. Do we have any bilinear maps on $V \times W$? Yes we do: for each fixed u in U we can define $e_u(v, w)$ to be $e(u, v, w)$. This then gives us, again for each u, a linear map f_u defined on $V \otimes W$ by the formula $f_u(v \otimes w) = e_u(v, w) = e(u, v, w)$ (extended linearly). But now it follows from the trilinearity of e that the map $f(u, v \otimes w) = f_u(v \otimes w)$ is linear in u for fixed $v \otimes w$—and because we extended $f_u(v \otimes w)$ linearly, this is true of $f(u, x)$ for more general elements x of $V \otimes W$.

We have now shown that bilinear maps on $U \times (V \otimes W)$ are in a natural one-to-one correspondence with trilinear maps on $U \times V \times W$. But a very similar argument will clearly prove the same for bilinear maps on $(U \otimes V) \times W$. This observation has solved our problem, because

$$u_1 \otimes (v_1 \otimes w_1) + \ldots + u_n \otimes (v_n \otimes w_n) = 0 \tag{30}$$

if and only if

$$f(u_1, (v_1 \otimes w_1)) + \ldots + f(u_n, (v_n \otimes w_n)) = 0 \tag{31}$$

for every bilinear map f. By what we have just proved, this is true if and only if

$$e(u_1, v_1, w_1) + \ldots + e(u_n, v_n, w_n) = 0 \tag{32}$$

for every trilinear map e. But this, again by what we have just proved, is true if and only if

$$d((u_1 \otimes v_1), w_1) + \ldots + d((u_n \otimes v_n), w_n) = 0 \tag{33}$$

for every bilinear map d, this time defined on $(U \otimes V) \times W$. Finally, this is true if and only if

$$(u_1 \otimes v_1) \otimes w_1 + \ldots + (u_n \otimes v_n) \otimes w_n = 0. \tag{34}$$

10

Now let me give a slightly slicker, high-level proof of the same fact. We have just discovered the (not terribly surprising—none of these results are supposed to be surprising) relevance of trilinear functions on $U \times V \times W$. This strongly suggests that we should do to trilinear maps what we have already done to bilinear ones—that is, define a sort of triple tensor product $U \otimes V \otimes W$. What should this be? Well, presumably we would like a typical element to be a linear combination of elements of the form $u \otimes v \otimes w$ and for

$$u_1 \otimes v_1 \otimes w_1 + \ldots + u_n \otimes v_n \otimes w_n \tag{35}$$

to be zero if and only if

$$e(u_1, v_1, w_1) + \ldots + e(u_n, v_n, w_n) = 0 \tag{36}$$

for every trilinear map e. This can be done, and since the proofs are very closely analogous to those for products of two spaces, I shall not give them. Let me simply note that the universal property of $U \otimes V \otimes W$ is that every trilinear map $e : U \times V \times W$ decomposes uniquely as $h \circ g$, where g is the obvious map from $U \times V \times W$ to $U \otimes V \otimes W$ and h is linear. Moreover, any other space with this sort of universal property is naturally isomorphic to $U \otimes V \otimes W$.

In order to prove the associativity of $\otimes$, it is enough to show that $U \otimes (V \otimes W)$ is naturally isomorphic to $U \otimes V \otimes W$. By the uniqueness result just mentioned, all we have to do in order for this is to prove that every trilinear map $e : U \times V \times W \to X$ factors uniquely through $U \otimes (V \otimes W)$. (More precisely, $f = c \circ g$, where c is linear and $g(u, v, w) = u \otimes (v \otimes w)$.) How might that work? Well, for every u we have a bilinear map $e_u : V \times W \to X$ given by $e_u(v, w) = e(u, v, w)$. The universal property of ordinary tensor products gives us for each u a unique linear map $f_u : V \otimes W \to X$ such that $e(u, v, w) = f_u(v \otimes w)$. Defining $f(u, v \otimes w)$ to be $f_u(v \otimes w)$ gives us a map f which is unique, clearly bilinear (we have seen this already) and defined on $U \times (V \otimes W)$. Hence, it factors uniquely through $U \otimes (V \otimes W)$.

11   Yet another way of regarding tensor products

There is one further way, which is useful as a psychological crutch, but has disadvantages which I shall mention below. Suppose that V and W are finite-dimensional vector spaces with given bases, so that we can think of a typical vector in V in coordinate form as $v = (a_1, \ldots, a_m)$, and a typical vector in

W as $w = (b_1, \ldots, b_n)$. Then $v \otimes w$ can be thought of as the m-by-n matrix $A_{ij} = a_i b_j$. Notice that matrices of the form $v \otimes w$ are only a small subset of all m-by-n matrices. In fact, they are the matrices of rank one. For exactly this reason, an element of $V \otimes W$ of the form $v \otimes w$ is called a rank-one tensor, and in general the *rank* of an element x of $V \otimes W$ is defined to be the smallest number of $v \otimes w$'s you need to add together to make x. This coincides with the usual definition of the rank of a matrix. (Exercise: prove this.)

Why does this 'construct' the tensor product? To answer this question, let us prove the main fact. Just to avoid ambiguity, I will invent yet another piece of notation—let $\{v, w\}$ stand for the rank-one matrix defined above, and let us show that

$$\{v_1, w_1\} + \ldots + \{v_n, w_n\} = 0 \tag{37}$$

if and only if

$$f(v, w) + \ldots + f(v_n, w_n) = 0 \tag{38}$$

for every bilinear map f defined on $V \times W$. One direction is trivial, since the map taking (v, w) to $\{v, w\}$ is bilinear. As for the other, if the first sum is a non-zero matrix A, then pick (i, j) such that A_{ij} is non-zero and define a bilinear map f by $f(v, w) = v_i w_j$. It is clear that the second sum is non-zero for this choice of f.

An advantage of this way of thinking of tensor products is that it makes them easier to visualize. Its main disadvantage is that it relies on a particular choice of basis. Not only is there no canonical choice of basis for a general vector space (even if it is finite-dimensional), but also the tensor product can be defined for other algebraic structures where you don't have a basis at all.

Rather than define these other algebraic structures, let me give one example of a type of object which is not a vector space, but on which a tensor product can be defined. Let G, H and K be abelian groups, and write the operations on them as $+$. Let us call a map $f : G \times H \to K$ a *bihomomorphism* if $f(g, h)$ is a homomorphism in h for each fixed g and in g for each fixed h. We can ask ourselves what the most general sort of bihomomorphism is, and we will find that we can answer this question exactly as we did for vector spaces—either by considering all bihomomorphisms at once, or by taking the free abelian group generated by all expressions of the form $[[g, h]]$ and quotienting out by the subgroup generated by expressions of two obvious forms. This will construct for us a tensor product $G \otimes H$, which will have the property that

$$g_1 \otimes h + \ldots + g_n \otimes h_n = 0 \tag{39}$$

if and only if

$$r(g_1, h_1) + \ldots + r(g_n, h_n) = 0 \tag{40}$$

for every bihomomorphism r. As a quick (and surprising if you have not seen it before) example, let $\mathbb{Z}_p$ be the group of integers mod p under addition and let us work out $\mathbb{Z}_2 \otimes \mathbb{Z}_3$. To do this, we must look at what bihomomorphisms there are. Well, it is not hard to see that a bihomomorphism r will be determined by what it does to $(1, 1)$. But $r(1, 1) + r(1, 1) = r(0, 1) = 0$, as r is a homomorphism in the

first variable, and $r(1,1)+r(1,1)+r(1,1) = r(1,0) = 0$, as r is a homomorphism in the second variable. Subtracting, it follows that $r(1,1) = 0$ and hence that r is identically 0. From this it follows that $\mathbb{Z}_2 \otimes \mathbb{Z}_3 = 0$.