Notes on Tensor Products PhL 9.16.15

last update 9.19.15

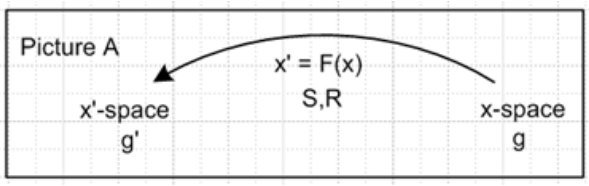
I will highlight questionable items in red and add comments in red.

1. Tensor Doc use of tensor product in Appendix E 1
E.1 Direct Product Notation 1
E.2 Tensor Expansions and Bases 3
E.4 Dyadic Products 4
2. Connection between tensor doc and my current tensor thread. 4
4. A Free Vector Space 5
4.1 Free Vector Space according to the wiki tensor product page 5
4.2 Free Vector Space according to Drexel 6
4.3 Free Vector Space according to Algebra pdf 8
4.4 Free Vector Space according to unknown source 9
4.5 Free Vector Space according to "trans-la1" 9
5. Digression: equivalence classes and relations and related notation 10
6. The Wiki Tensor Product Page (including notion of Equivalence Classes) 10
7. My summary of what I think wiki is saying 13
8. Some other sources have their say: 14
8.1 Lang's Algebra book 14
9. Kronecker Products 16
10. Axler book has nada 18
11. Ganatra's class notes 19
12. Roman book preview 19
13. Roman Notes, Starting on Page 355 which is Chapter 14 Tensor Products. 21
13.1 The Universal Pair Concept 21
Universality Example 1: the triangle with inclusion map j on top 27
Universality Example 2: The quotient of a vector space divided by a subspace. 29
13.2 Direct Sums, External and Internal 31
13.3 Bilinear Maps 33
13.4 Tensor Products 35
13.5 Review of Ganatra and Explanation of why τ is linear 39
13.6 Conclusions 43

1. Tensor Doc use of tensor product in Appendix E

E.1 Direct Product Notation

This entire Section uses the general Picture A context where x-space need not be Cartesian,



(E.1.1)

The Standard Notation of Chapter 7 is used throughout. Useful forms can be found in Section 7.18.

The key tool required for the expression of tensor expansions is the notion of a direct product of n tensorial vectors defined in this simple way,

$$(\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)^{abc\cdots} \equiv A^aB^bC^c.....$$
$$(\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)^a{}_b{}^c{}_{\cdots} \equiv A^aB_bC^c..... \quad \text{etc} \quad . \tag{E.1.2}$$

This is the Kronecker product, not a tensor product, and direct product is the wrong phrase!

The key tool required for the expression of tensor expansions is the notion of a direct product of n tensorial vectors defined in this simple way,

$$(\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)^{abc\cdots} \equiv A^aB^bC^c.....$$
$$(\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)^a{}_b{}^c{}_{\cdots} \equiv A^aB_bC^c..... \quad \text{etc} \quad . \tag{E.1.2}$$

Again, this is a Kronecker product = outer product. It is not a tensor product or direct product.

The tensor $\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots$ is nothing more than the outer product of vectors A,B,C as in (7.1.1) for contravariant vectors, but later extended to any mixture of vector types. One can define a **direct** or outer product of two rank-2 tensors in this way,

$$(M\otimes N)_{ab,AB} \equiv M_{aA}N_{bB} \qquad \text{// rank(M) = rank(N) = n = 2; number of tensors = I = 2}$$
$$(M\otimes N)^{ab}{}_{,AB} \equiv M^a{}_A N^b{}_B \quad \text{etc} \tag{E.1.3}$$

Here at least I use the correct phrase "outer product, but I again sneak in the word "direct".

Notice how the indices are arranged on the left side of each equation. The same idea can be applied to form a **direct product** of tensors of any rank, for example

$$(M\otimes N)_{ab,AB,\alpha\beta} = M_{aA\alpha}N_{bB\beta} \quad \text{etc} \quad \text{// rank(M) = rank(N) = n = 3; number of tensors = I = 2} \tag{E.1.4}$$

On the left side the number of groups of indices equals the tensor rank n of the tensors on the right, and the number of indices within each group matches the number I of tensors being direct-product-multiplied.

In what follows, only the (E.1.2) direct product of vectors shall be considered. One can define the dot product of two direct-product-space vectors in this obvious manner,

$$(\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)\bullet(\mathbf{A}'\otimes\mathbf{B}'\otimes\mathbf{C}'\ \ldots) \equiv (\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}\ \ldots)^{abc\cdots}\,(\mathbf{A}'\otimes\mathbf{B}'\otimes\mathbf{C}'\ \ldots)_{abc}$$

$$= A^{a}B^{b}C^{c}..... \; A'_{a}B'_{b}C'_{c}..... \; = \; \mathbf{A}\bullet\mathbf{A}' \; \mathbf{B}\bullet\mathbf{B}' \; \mathbf{C}\bullet\mathbf{C}' \; ... \qquad \text{(E.1.5)}$$

where of course the indices abc can be "tilted" in any way desired according to (7.11.3).

This seems like a reasonable dot product of two Kronecker products.

E.2 Tensor Expansions and Bases

Preamble: A seeming paradox and how it is resolved

This section does not mention tensor product objects.

Tensor Expansions

In what follows, we use the example of a rank-3 tensor and the reader can easily see how this applies to a rank-n tensor.

Let $\mathbf{b}_i$ be an arbitrary complete set of basis vectors in x-space. As shown in the notes following (6.2.8) there exists a unique set of *dual* ("reciprocal") basis vectors $\mathbf{b}^i$ (also in x-space) such that $\mathbf{b}^i\bullet\,\mathbf{b}_j = \delta^i{}_j$, where we now use the Standard Notation $\mathbf{b}_n$ equations of (7.18.6). Consider then the following expansion of rank-3 tensor A with contravariant components A^{abc},

$$A = \Sigma_{ijk}\,\alpha^{ijk}\,(\mathbf{b}_i\otimes\mathbf{b}_j\otimes\mathbf{b}_k) \quad \text{where} \quad (\mathbf{b}_i\otimes\mathbf{b}_j\otimes\mathbf{b}_k\,...)^{abc} \;=\; (\mathbf{b}_i)^a\,(\mathbf{b}_j)^b\,(\mathbf{b}_k)^c \qquad \textit{(E.1.2)}$$

$$A^{abc} = \; \Sigma_{ijk}\,\alpha^{ijk}\,(\mathbf{b}_i\otimes\mathbf{b}_j\otimes\mathbf{b}_k\,...)^{abc} \;=\; \Sigma_{ijk}\,\alpha^{ijk}\;(\mathbf{b}_i)^a\,(\mathbf{b}_j)^b\,(\mathbf{b}_k)^c \;. \qquad \text{(E.2.1)}$$

OK, so what is happening here? I define a tensor in terms of the Kronecker product of basis vectors, not the tensor product. This does cause A^{abc} to equal the real number on the right.

As we did in Example 3 (2.9.5) for $\mathbf{e}_n$, we declare the $\mathbf{b}_n$ vectors to be "contravariant by definition" by writing the rule $(\mathbf{b}'_n)^i \equiv R^i{}_j(\mathbf{b}^n)_j$. So then the vectors $\mathbf{b}_n$ are true rank-1 tensors under $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. According to Section 7.1, the outer product $(\mathbf{b}_i)^a\,(\mathbf{b}_j)^b\,(\mathbf{b}_k)^c$ is a true rank-3 contravariant tensor under **F**. Then (E.2.1) says that A^{abc} is a linear combination of these tensors, which we know is a tensor because the sum of two tensors of some type is a tensor of the same type.

The expansion coefficients α^{ijk} can be obtained by dotting both sides of (E.2.1) with $(\mathbf{b}^{i'}\otimes\mathbf{b}^{j'}\otimes\mathbf{b}^{k'})$ and using (E.1.5),

$$(\mathbf{b}^{i'}\otimes\mathbf{b}^{j'}\otimes\mathbf{b}^{k'})\bullet(\mathbf{b}_i\otimes\mathbf{b}_j\otimes\mathbf{b}_k) \;=\mathbf{b}^{i'}\bullet\mathbf{b}_i\;\;\mathbf{b}^{j'}\bullet\mathbf{b}_j\;\mathbf{b}^{k'}\bullet\mathbf{b}_k \;=\delta^{i'}{}_i\delta^{j'}{}_j\delta^{k'}{}_k \;. \qquad \text{(E.2.2)}$$

The result is then

$$A\bullet(\mathbf{b}^{i'}\otimes\mathbf{b}^{j'}\otimes\mathbf{b}^{k'}) = \{\Sigma_{ijk}\alpha^{ijk}\,(\mathbf{b}_i\otimes\mathbf{b}_j\otimes\mathbf{b}_k)\}\bullet\;(\mathbf{b}^{i'}\otimes\mathbf{b}^{j'}\otimes\mathbf{b}^{k'}) = \Sigma_{ijk}\,\alpha^{ijk}\,\delta^{i'}{}_i\delta^{j'}{}_j\delta^{k'}{}_k \;=\alpha^{i'j'k'}$$

or unpriming indices,

$$\alpha^{ijk} = A \bullet (\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k) \, .$$

But since both A and $(\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k)$ are elements of the triple direct-product space spanned by the vectors $(\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k)$, we use the dot product of (E.1.5) to claim that

$$A \bullet (\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k) \;\; = A^{abc} \, (\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k)_{abc} = A^{abc} \;\; (\mathbf{b}^i)_a \, (\mathbf{b}^j)_b \, (\mathbf{b}^k)_c \, .$$

The coefficients α^{ijk} may then be written in all these ways :

$$\alpha^{ijk} = A \bullet (\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k) = A^{abc} \, (\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k)_{abc} = A^{abc} \;\; (\mathbf{b}^i)_a \, (\mathbf{b}^j)_b \, (\mathbf{b}^k)_c \qquad \text{(E.2.3)}$$

This "space spanned by basis vectors $(\mathbf{b}^i \otimes \mathbf{b}^j \otimes \mathbf{b}^k)$ now sounds more like a tensor product. Ouch!

where A^{abc} are the contravariant components of tensor A in x-space, and $(b^i)_a$ are the covariant components of vector $\mathbf{b}^i$ in x-space. As noted in Example 3 of the previous subsection, the object α^{ijk} transforms as a *scalar*, since it is the dot product of two direct-product-space vectors. This fact is especially obvious from the last expression in (E.2.3) where all tensor indices are contracted so the result must be a scalar according to the neutralization rule (7.12.1). So α^{ijk} is a set of 3^3 =27 scalars.

OK, let's stop here. It seems clear to me that I am only talking about the Kronecker = outer product in this appendix.

Then we have

E.4 Dyadic Products

When two vectors **A** and **B** are combined in polyadic notation, the result is called a dyadic product (**AB**) [also known as a dyad or just a dyadic]

$$(\mathbf{AB})^{ij} \equiv A^i B^j \qquad // = (\mathbf{A} \otimes \mathbf{B})^{ij} = \text{component of the matrix } \mathbf{A} \otimes \mathbf{B} = \mathbf{AB} \, . \qquad \text{(E.4.1)}$$

So again I am talking outer product, not tensor product. Except for my use of the word "direct", everything seems to be self-consistent here, and a tensor really does seem to be an expansion on the outer product of the basis vectors. In particular, the vector is this

$$V = \Sigma_i v_i e_i \qquad V_k = \Sigma_i v_i (e_i)_k$$

so even this simple "tensor" can be written in terms of those real basis vector components.

---

2. Connection between tensor doc and my current tensor thread.

So how does the above self-consistent tensor doc stuff relate to my new tensor product discussion where I really seem to use the tensor product and not the Kronecker product? That is the $64 question! I write things like this:

$\{e_i\}$ = basis of V $v = \Sigma_{i=1}^{n} v_i e_i$ = general vector in V

$T \equiv \Sigma_{ij} T_{ij} e_i \otimes e_j$. = general object in V^2 $T \in V \times V = V^2$

These two objects exist in V^1 and V^2, for example. $\{e_i\}$ is an arbitrary basis like $\{b_i\}$ of tensor doc.

I never talk about "components" of something like $e_i \otimes e_j$, though I do mention $(e_i)_k$ in a few places. In tensor doc, I would say

$$A = \Sigma_{ijk} \alpha^{ijk} (\mathbf{b}_i \otimes \mathbf{b}_j \otimes \mathbf{b}_k) \quad \text{where} \quad (\mathbf{b}_i \otimes \mathbf{b}_j \otimes \mathbf{b}_k \ldots)^{abc} = (\mathbf{b}_i)^a (\mathbf{b}_j)^b (\mathbf{b}_k)^c \quad \textit{(E.1.2)}$$

$$A^{abc} = \Sigma_{ijk} \alpha^{ijk} (\mathbf{b}_i \otimes \mathbf{b}_j \otimes \mathbf{b}_k \ldots)^{abc} = \Sigma_{ijk} \alpha^{ijk} (\mathbf{b}_i)^a (\mathbf{b}_j)^b (\mathbf{b}_k)^c \ . \qquad \text{(E.2.1)}$$

and I would claim that A was a tensor expanded on the $\{b_i\}$ basis having strange components as shown.

This is indeed very confusing right now!!

Can you have something that is both a tensor product and, within that, an outer product as a sort of subclass of tensor product? Look again at wiki on tensor product. Wiki uses (v,w) notation but writes the space as V⊗W. They have a very good comment on quotient space which I defer for the moment.

4. A Free Vector Space

Since this concept is used in the Wiki Tensor Product discussion, I will have a section right here to examine this sub-topic before moving into the Wiki stuff.

4.1 Free Vector Space according to the wiki tensor product page

Here is the rather confusing wiki paragraph on free vector spaces. Comments will come after:

Prerequisite: the free vector space [edit]

The definition of ⊗ requires the notion of the free vector space $F(S)$ on some set S, a vector space whose basis is parametrized by S. By definition, $F(S)$ is the set of all functions f from S to a given field K that have finite support; i.e., f is identically zero outside some finite subset of S. It is a vector space with the usual addition and scalar multiplication of functions. It has a basis parametrized by S. Indeed, for each s in S, let δ_s: $S \to K$ be the function given by $\delta_s(s) = 1$ and $\delta_s(t) = 0$ for any $t \neq s$ (in a fancier language, it is the Dirac's delta function with point mass at s when S is viewed as a discrete space.) Then $\{ \delta_s \mid s \in S \}$ is a basis of $F(S)$, since each element of $F(S)$ can be uniquely written as a linear combination of δ_s's. Because of this explicit expression, an element of $F(S)$ is often called a formal sum of symbols in S.

By construction, the (possibly infinite) dimension of the vector space $F(S)$ equals the cardinality of the set S.

In this following clip, wiki states its definition of "free vector space" $F(S)$ as f: S→ K(a field). Notice that the *F* and the K are not the same. The *F* means "Free".

The nature of the set S is not clear to me. The clip refers to a "finite subset of S". I take that as meaning a subset having a finite number of elements, not a subset that is bounded. The clip says that S, in addition to being a set, is a vector space, which means it has addition and scalar multiplication amongst its elements s. Could we select S = R, the reals? That set is a vector space, but how do you pick a finite subset of S? I guess I am wondering if S can be a "continuous set"? I think the answer is no. The idea of $\delta_{s}(s') = \delta_{s,s'}$ can be extended to continuous sets as the Dirac, but I am pretty sure S is a discrete set.

The clip says any element of F(S) can be expressed as a linear combination of those δ's. That seems to lead to the following rather uninteresting set of functions

$f(s') = \Sigma_s f_s \delta_s(s')$ // "linear combination of the δ's

$= \Sigma_s f_s \delta_{s.s'} = f_{s'}$

So, what is the meaning of the most general function of F(S) having the form $f(s') = f_{s'}$? This seems to say nothing at all, like 2 = 2.

If S cannot be continuous, that we cannot say for example that $S = R^3$ with basis elements e_i. Could we instead say that $S = \{e_i\}$ for i = 1,2,3. But such an S is not a vector space, it is the basis of a vector space.

Conclusion: Although wiki provides some info on the notion of a free vector space, the full picture of such an animal is extremely unclear. What sets S are really allowed? So we shall now look at some alternate web offerings on the subject of free vector space.

4.2 Free Vector Space according to Drexel

https://drexel28.wordpress.com/2011/04/19/free-vector-spaces/

Free Vector Space Definition

Let X be a finite set and F a field. We define the set $F[X]$ to be the set of all formal sums of the form $\sum_{x \in X} f_x x$ where $f_x \in F$ for every $x \in X$. We then define the *sum* of two of these formal sums as

$$\left(\sum_{x \in X} f_x x\right) + \left(\sum_{x \in X} g_x x\right) = \sum_{x \in X} (f_x + g_x) x$$

We define the *scalar product* of some $f \in F$ and a formal sum as

$$f \sum_{x \in X} f_x x = \sum_{x \in X} (f f_x) x$$

Instead of a finite subset of some possibly infinite set S, Drexel starts us off with a finite set X.
Instead of talking about F(X) = a set of functions f: X→ reals, Drexel talks about F[X] = a set of sums of the elements of set X weighted by reals. This is I guess a special case of a function f. So Drexel would have his most general F(x) element looking like this,

$f = \Sigma_{\mathbf{x}} f_{\mathbf{x}} x$ = element of F[X]

as an element of his F(X). In a later comment he suggests that you could label your x's in X as $x_{\mathbf{i}}$ and then the above becomes

$f = \Sigma_{\mathbf{i}} f_{\mathbf{i}} x_{\mathbf{i}}$ = element of $F[I_{\mathbf{x}}]$

So this is still a sum over X, but we have a certain way to label the x's (I is an "indexing set"), so here you could consider S = I or S = X, your choice of viewpoint.

He does NOT say that the x or $x_{\mathbf{i}}$ are real or in field F, though he might intend that.

He does NOT say that X is a vector space (whereas wiki says S is a vector space). But perhaps I could just make X be a vector space by having it be closed under addition etc.

The two "rules" he give don't really seem like rules. The first says,

$f + g = (\Sigma_{\mathbf{i}} f_{\mathbf{i}} x_{\mathbf{i}}) + (\Sigma_{\mathbf{i}} g_{\mathbf{i}} x_{\mathbf{i}}) = (f+g) = \Sigma_{\mathbf{i}} (f_{\mathbf{i}}+g_{\mathbf{i}})x_{\mathbf{i}}$

Since the sums are finite, the last item is just rearranging the summation terms, so this is not even a rule. The other "rule" says

$\alpha * f = \alpha * (\Sigma_{\mathbf{i}} f_{\mathbf{i}} x_{\mathbf{i}}) = \Sigma_{\mathbf{i}} (\alpha f_{\mathbf{i}}) x_{\mathbf{i}}$

Again, since f_i and α are reals, this does not seem like much of a rule!

Why is F[X] defined in this way a vector space? Well, the space of sums of the above form is indeed closed. There is a zero sum for use as the "identity" element. Inverses exist and so on.

So I will state his definition in my own words:

" A free vector space over a finite set X with elements x_i is the space of sums of the x_i with real coefficients. " OR " F[X] is the free vector space of X over R ".

So Drexel's definition allows me to think of the x_i as my basis vectors of R^3 which are e_i. These e_i are not reals, they are tuplets of reals, but Drexel does not require that the x_i be reals. I'll be you are supposed to think of the x_i as "inert carriers" like the x in a polynomial. Drexel goes on to say:

With this definition of addition and scalar multiplication of elements of $F[X]$ we call $F[X]$ the *free vector space of* X *over* F. Suppose that we are given an indexing set I. We can then choose a formal object x to 'append' the indices to, namely we can form the set $I_x = \{x_i : i \in I\}$ and we can then define the *free vector space of the indexing set* I *over* F to be $F[I_x]$. Evidently then $F[X]$ is a F-space with $\dim_F F[X] = \#(X)$.

This would then give dim = 3 if I choose X = e_1, e_2, e_3. Note that x is called a "formal object", like my idea of inert carrier.

4.3 Free Vector Space according to Algebra pdf

Definition 1.14. Let S be any set. Then the *free vector space* on S is the set of all those functions $f : S \to \mathbf{R}$ such that $\{s \in S | f(s) \neq 0\}$ is finite. Addition and scalar multiplication are defined via

$$(f + g)(s) = f(s) + g(s)$$

$$(\alpha f)(s) = \alpha f(s)$$

Notice that S is a basis for F, so, by (1.9), we can specify a linear transformation $F \to W$ by specifying the values for $f(s)$ as s ranges over S.

Note also that F is finite dimensional if and only if S is finite.

This is closer to wiki in that we have f : S → R with its same idea of finite support. Whereas Drexel had only sums for functions, here we have general functions, again like wiki. But now we have rules on these functions which are the same rules that Drexel gave for his finite sums. The first rule just says this free vector space is closed, the second shows that (αf) is a certain function, so you can scale functions. The elements of set S can be any objects you want, but the functions f(s) have to map to reals. You could I

suppose have S = $\{e_i\}$, but not clear what typical functions f would look like. Author claims that S is a basis for F, the free vector space. This offering fits roughly with the previous two in its outline, but the specifics are vague.

4.4 Free Vector Space according to unknown source

Defintion. *If X is a set, then we define the free vector space on X, $\mathbb{F}X$ to be the direct sum over X of copies of $\mathbb{F}$:*

$$\bigoplus_{x \in X} \mathbb{F}.$$

This vector space has a very nice basis.

Proposition. *Let f_x denote the function*

$$f_x(y) = \begin{cases} 1 & y = x \\ 0 & otherwise. \end{cases}$$

Then $\{f_x, x \in X\}$ forms a basis for $\mathbb{F}X$.

The set X is not specified to be finite.
The object F is not defined, so unclear what it means to have a direct sum of undefined objects!
The basis functions are similar to the δ_s ones of wike, where now s,s' replaced by x,y.
Maybe the direct sum means the space of sums like those of Drexel with coefficients from F.

$\Sigma_x f_x x$

Not very helpful.

4.5 Free Vector Space according to "trans-la1"

EXAMPLE 1.5. *Let I be an index set. Then $F^I = \{f : I \to F : ffunction\}$ is a vector space over F.*

So this is a mapping from integers to reals (let's say) f_i = f(i) a real number. It could be a sequence of reals. The values of i are in an index set I, fine. Why is this a vector space? What are the vectors of this vector space? I guess the vectors are the functions f. So we have to show that

$\alpha(f + f') = \alpha f + \alpha f'$ etc

but since the f's map to F, this all follows naturally. Now author continues,

EXAMPLE 1.8. *Let I be an index set. Consider $F(I) \subset F^I$ defined by*

$$F(I) = \{f : I \to F : \exists_{S \subset I finite}(i \notin S \Rightarrow f(i) = 0)\}.$$

$F(I)$ is called the free vector space over I.

Here the set S is finite set of integers from the integer set I.
The functions f in effect map from I to R. No "inert carrier" is specified so on x_i here.
Otherwise, this is little like the Drexel 4.2 approach.

So we have five different definitions of a "free vector space".

Roman mentions "free vector space" in his Preface, but that phrase never appears in Chapters 14 and 15. So he is not very helpful.

5. Digression: equivalence classes and relations and related notation

Stop. *What is the equivalence class of a set S?* OK, wiki says this is a concept of set theory where you just partition your set in to subsets in each of which elements are equivalent by some rule a ~ b . The rule is not specified.

The equivalence class of an element a is denoted $[a]$ and is defined as the set

$$[a] = \{x \in X \mid a \sim x\}$$

Sometimes this is written, where a ~ x is written a R x and R is the equivalence relation

$[a]_R = \{x \in X \mid a\ R\ x\}$

So far so good. Suppose you successfully partition your set using R so you get

S = { [a], [b], [c] } which has 3 equivalence classes for some relation R

The equivalence classes themselves form a set which is known as X/R " X modulo R" "quotient set". Another notation for this set is then X/~ which I have seen!

6. The Wiki Tensor Product Page (including notion of Equivalence Classes)

" Notation

Elements of V ⊗ W are often referred to as tensors, although this term refers to many other related concepts as well.[1] If v belongs to V and w belongs to W, then the equivalence class of (v, w) is denoted by v ⊗ w, which is called the tensor product of v with w. In physics and engineering, this use of the "⊗ " symbol refers specifically to the outer product operation; the result of the outer product v ⊗ w is one of the standard ways of representing the equivalence class v ⊗ w.[2] An element of V ⊗ W that can be written in the form v ⊗ w is called a pure or simple tensor. In general, an element of the tensor product space is not a pure tensor, but rather a finite linear combination of pure tensors. For example, if v1 and v2 are linearly independent, and w1 and w2 are also linearly independent, then v1 ⊗ w1 + v2 ⊗ w2 cannot be written as a pure tensor. The number of simple tensors required to express an element of a tensor product is called the tensor rank (not to be confused with tensor order, which is the number of spaces one

has taken the product of, in this case 2; in notation, the number of indices), and for linear operators or matrices, thought of as (1, 1) tensors (elements of the space V ⊗ V*), it agrees with matrix rank."

So this does suggest that the outer product might be a subclass of the tensor product, but I need then to know more about "equivalence class". I do like their claim that physics really does use the outer product, which is what tensor doc uses. Now look back at the opening wiki paragraph,

"In mathematics, the tensor product, denoted by ⊗ , may be applied in different contexts to vectors, matrices, tensors, vector spaces, algebras, topological vector spaces, and modules, among many other structures or objects. In each case the significance of the symbol is the same: the freest bilinear operation. In some contexts, this product is also referred to as outer product. The general concept of a "tensor product" is captured by monoidal categories; that is, the class of all things that have a tensor product is a monoidal category."

Notice the key word "bilinear" and "freest bilinear". I have not looked at all into "monoidal".

So let's back up and try to parse wiki's page on tensor products.

Tensor product of vector spaces [edit]

The tensor product of two vector spaces V and W over a field K is another vector space over K. It is denoted $V \otimes_K W$, or $V \otimes W$ when the underlying field K is understood.

I like using K for the field, all is well so far.

Now, here is where the "free vector space" buzzword appears, and this is why I tried to find someone who have a clean definition of such a space (see above).

Note that the "free vector space" F(VxW) is the set of functions f such that f : VxW → K (field) and which have finite support over VxW. Such a function would be written r = f((v,w)) where think of s = (v,w) = element of S.

Definition [edit]

Given two vector spaces V and W over a field K, the tensor product U of V and W, denoted as $U = V \otimes W$ is defined as the vector space whose elements and operations are constructed as follows:

From the cartesian product $V \times W$, the free vector space $F(V \times W)$ over K is formed. The vectors of $V \otimes W$ are then defined to be the equivalence classes of $F(V \times W)$ under the following equivalence relations:

$$v, v_1, v_2 \in V; w, w_1, w_2 \in W; c \in K;$$
$$(v_1, w) + (v_2, w) \sim (v_1 + v_2, w)$$
$$(v, w_1) + (v, w_2) \sim (v, w_1 + w_2)$$
$$c(v, w) \sim (cv, w) \sim (v, cw)$$

Note 1: I understand V x W as the Cartesian product with elements (v,w). Take the (v,w) elements to be the set S. This would seem to define a continuous set S. They talk about F(V x W) so are taking S = V x

W. Well, the elements of set S are things of the form (v,w) so there are a lot of such points in S if V and or W are continuous sets, yes.

Note 2: If we apply the Drexel idea, then F(VxW) is a set of functions which are sums of this form

$$f = \Sigma_i f_i (v_i, w_i) = \Sigma_{ij} f_{ij}(v_i, w_j)$$

where v_i is some arbitrary element of V, and so on. Here for example is a possible sum, for some v and w

$$f = c(v,w) - (cv,w) .$$

Again, sums are special cases of functions, so the wiki thing of 4.1 is OK too. This last function is going to get forced to zero soon because it is in an equivalence class with 0.

Now back to tensor product. I know what the equivalence classes of S means, so then if S = V x W I have to ask: what are the equivalence classes for the set elements which have the form (v,w) ? Whatever these classes are, they are used as the vectors of U = V⊗W. It seems that the scalar and distributive rules are imposed on the (v,w) as an extra feature.

Now here is another piece to chew on:

The operations of $V \otimes W$, i.e. the map of vector addition $+: U \times U \to U$ and scalar multiplication $\cdot: K \times U \to U$ are defined to be the respective operations $+_F$ and $\cdot_F$ from $F(V \times W)$, acting on *any* representatives

$$\tilde{u}_1, \tilde{u}_2$$

in the involved equivalence classes outputting the one equivalence class of the result.

$$\tilde{u}_1 \in u_1, \tilde{u}_2 \in u_2 \Rightarrow (+) : (u_1, u_2) \mapsto [\tilde{u}_1 +_F \tilde{u}_2]$$
$$\tilde{u}_1 \in u_1 \Rightarrow (\cdot) : (c, u_1) \mapsto [c \cdot_F \tilde{u}_1]$$

The result is independent of which representatives of the involved classes have been chosen. In other words, the operations are well-defined.

They are talking about the meaning of + and * for elements of U ≡ V⊗W where * is scalar mult of an element of U. I guess $\bar{u}_1$ is a "representative of an equivalence class" of S = V x W, and then maybe the class itself we can call u_1. The next lines say this:

For two representative elements of two equivalence classes (which classes are elements of U), you can do addition in this way: $\bar{u}_1 +_F \bar{u}_2$ where $+_F$ is the appropriate addition operator for U. It is what makes U be a vector space. Then to do scalar multiplication you have $c \bullet_F \bar{u}_1$ where $\bullet_F$ means scalar multiplication of an element in U. Any representatives will do.

So one idea is that you don't add and scalar-multiply the classes directly, you add and scalar multiply elements of those classes. I am fine so far. (familiar from Galois doc I think)

Note Added: This equivalence class stuff is just a fancy way to say that the + and • operations you have in the space V⊗W respect the "bilinear rules". It is I think better stated in the following clip:

In other words, the tensor product $V \otimes W$ is defined as the quotient space $F(V \times W)/N$, where N is the subspace of $F(V \times W)$ consisting of the equivalence class of the zero element, $N = [\varnothing]$, $\varnothing \in F(V \times W)$, under the equivalence relation of above. The following expression explicitly gives the subspace N:

$$\begin{aligned} N = \{n \in F(V \times W) \,|\, \exists v_1, v_2 \in V, \exists w_1, w_2 \in W, \exists c \in K : \\ n = (v_1, w_1) + (v_2, w_1) - (v_1 + v_2, w_1) \vee \\ n = (v_1, w_1) + (v_1, w_2) - (v_1, w_1 + w_2) \vee \\ n = c(v_1, w_1) - (cv_1, w_1) \vee \\ n = c(v_1, w_1) - (v_1, cw_1)\}. \end{aligned}$$

A whole new batch of stuff here. What is this little V symbol on the right? It is the OR operator in logic notation, fine. So at least one of the four equations has to be true.

What then is set N? First, what is F(VxW)? This is the set of functions f: VxW → K so these functions are K-valued. A typical function might be a(v,w) + b(v'w') where I guess + is addition in field K, and where a and b are elements of field K. So F(VxW) is the set of K-valued functions of (v,w) elements.

Note: Just maybe f: VxW → K somehow means that the (v,w) are my "inert carrier" objects and the → K just means that the coefficients are in K. It is then more like f: VxW → T where T is the space of linear combinations of the (v,w) inert (formal) objects with real coefficients. In this context, we are going to knock out from T certain elements.

Now four such functions appear on the right of the above 4 equations. Obviously if the scalar and distributive "rules hold", n = 0 for each of these four functions.

So I guess N is these specific four functions! Each of these functions is an element of F(VxW), I agree. I can see that N is a subset of F(VxW). There are really more than 4 functions because you get a function for each choice of v_1, v_2 etc. No doubt if you add two of these functions, you get a function in the same set, etc etc, so I suppose the set N could be a subspace of F(VxW).

Now using our rules, all these functions are equal to 0, even though they are all different functions. That is to say, every function in N equals 0 which is an element of field K I guess. Are such functions in the same equivalence class? Depends on how you define your relation ~.

I guess you could partition F(VxW) into classes where in each class the function takes some value in the set K. Then since 0 is an element of K, all the N functions above are in the equivalence class of functions which equal 0. I guess there might be more functions that those shown above which = 0. Maybe we define N to be an equivalence class and we label that class by the name [0]? Again, N is an equivalence class whose elements are n and such a class is written $[n]_{\mathbf{R}}$. Maybe there are only two classes, N and all other functions in F(VxW). The notation $\emptyset \in$ F(VxW) seems strange. It is VxW = set S which we partition into equivalence classes, so VxW = [0] + other classes. Well, I suppose if you were to partition S into classes so S = [α] + [β], you could talk about F = F([α]) + F([β]) and maybe these are then induced classes for F(S)?

OK, now comes the big line: V⊗W is the quotient space F(VxW)/N where N is the subspace of F defined above. I guess you can take a quotient in this manner.

7. My summary of what I think wiki is saying

1. The free vector space F(VxW) is the set of all finite linear combinations of (v,w) formal objects with real coefficients. These linear combinations are the Drexel sums above. If you were to single-index the (v,w) elements, you might call them (v_i,w_i) and then a most general element of the free vector space F(VxW) would be this

$$f = \Sigma_i \, f_i \, (v_i,w_i)$$

Here f_i are real coefficients, but the (v_i,w_i) is "formal" and not of any particular type. It need not be, for example, that (v_i,w_i) is a real number. It has no specified type IMHO.

2. Certain of these functions, such as f = 2(v,w) - (2v,w), are forced to be zero. This is done by saying for example that 2(v,w) - (2v,w) ~ 0, meaning this function "is equivalent to 0". Certain functions within F(VxW) are by fiat made to be equivalent to 0 in this manner.

3. Consider the following function which is a function in the free vector space F(VxW)

$$f = 3(v,w) + [\alpha(v,w) - (\alpha v,w)]$$

As one varies α, one has a lot of functions represented by this equation. These functions are elements of a row of the "chart" for the quotient F(VxW)/N . We can take the row representative to be 3(v,w). The row of the chart is an "equivalence class". All its many elements are represented by 3(v,w). So in the object that is F(VxW)/N, the only elements are the chart rows! This is a vector space whose elements are the rows of the chart. And 3(w,v) is a chart row, so is in F(VxW)/N . Then if we say

$$V \otimes W = F(VxW)/N$$

we are saying that the elements of the vector space V⊗W are those chart rows. In this V⊗W space, the quantity α(v,w) - (αv,w) is equal to 0.

4. In less formal language, V⊗W is the set of all functions in F(VxW) where we force certain of these functions to be 0. Just as an element of VxW is written (v,w), an element of V⊗W is written v⊗w. Then similar to the above

$$f = 3v \otimes w + [\alpha(v \otimes w) - (\alpha v) \otimes w] = 3v \otimes w$$

where we have set the function $\alpha(v \otimes w) - (\alpha v) \otimes w = 0$ by our definition of the V⊗W definition.

5. You could say that V⊗W starts out as VxW, but then we overlay certain "rules" on linear combinations of the (v,w) and when such rules are observed, we refer to (v,w) as v⊗w. There rules are I think exactly the "bilinear rules".

8. Some other sources have their say:

8.1 Lang's Algebra book

Regarding tensor product vs other things, Lang's "Algebra" supposedly has material. It is this book, there is one chapter of great interest

Chapter XVI The Tensor Product 601

1. Tensor product 601
2. Basic properties 607
3. Flat modules 612
4. Extension of the base 623
5. Some functorial isomorphisms 625
6. Tensor product of algebras 629
7. The tensor algebra of a module 632
8. Symmetric products 635

The book is 934 pages or so. I found a very slow Chinese server, trying to load it in. Have it!

But alas, it is **densepack jargon** of the worst kind for me! I would have to spend 6 months to be able to ready the chapter, and I don't really think it has what I want. But at least I have a nice book.

Note added: Well I reread it, and except for the fact that he uses the "module" word a lot (remember that a module is a vector space over a ring instead of over a field), he does exactly what I have just outlined above, concluding that $V \otimes W = F(V \times W)/N$. But Lang shows that

$$E_1 \otimes E_2 \otimes \ldots\ldots \otimes E_n = F(E_1 \times E_2 \times \ldots\ldots \times E_n)/N.$$

where he does more than two items. But his approach is more like the Roman one below where we use those triangle diagrams and the universal buzzward. Lang says:

"We use the words linear and homomorphism interchangeably".

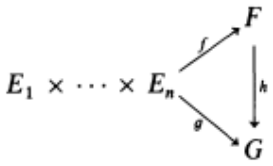
- A linear map is a homomorphism that preserves the vector space structure, namely the abelian group structure and scalar multiplication. The scalar type must further be specified to specify the homomorphism, e.g. every **R**-linear map is a **Z**-linear map, but not vice versa.

and here is his picture (notice the appearance of "free module" rather than "free vector space") And he suggests that the space N is a *subspace* of M, though that word did not appear in the above stuff.

One may view the multilinear maps of a fixed set of modules $E_1, \ldots, E_n$ as the objects of a category. Indeed, if

$$f : E_1 \times \cdots \times E_n \to F \quad \text{and} \quad g : E_1 \times \cdots \times E_n \to G$$

are multilinear, we define a morphism $f \to g$ to be a homomorphism $h : F \to G$ which makes the following diagram commutative:



A universal object in this category is called a **tensor product** of $E_1, \ldots, E_n$ (over R).

Let M be the free module generated by the set of all n-tuples $(x_1, \ldots, x_n)$, $(x_i \in E_i)$, i.e. generated by the set $E_1 \times \cdots \times E_n$. Let N be the submodule generated by all the elements of the following type:

$$(x_1, \ldots, x_i + x_i', \ldots, x_n) - (x_1, \ldots, x_i, \ldots, x_n) - (x_1, \ldots, x_i', \ldots, x_n)$$
$$(x_1, \ldots, ax_i, \ldots, x_n) - a(x_1, \ldots, x_n)$$

for all $x_i \in E_i$, $x_i' \in E_i$, $a \in R$. We have the canonical injection

$$E_1 \times \cdots \times E_n \to M$$

of our set into the free module generated by it. We compose this map with the canonical map $M \to M/N$ on the factor module, to get a map

$$\varphi : E_1 \times \cdots \times E_n \to M/N.$$

We contend that φ is multilinear and is a tensor product.

The module M/N will be denoted by

$$E_1 \otimes \cdots \otimes E_n \quad \text{or also} \quad \bigotimes_{i=1}^{n} E_i.$$

I look more at the wiki refs and see Lang's Algebra sitting there, and **Birkhoff Mac Lane**. Another one is a book of **Bourbaki** which I just loaded in. Have it, all about rings and modules and fancy things relative to the tensor product. Lots of Hom operators, way too advanced for me.

Hom(E,F) = set of linear mappings E to F. // see Lang's comment above

I am looking for more of a primer on my topic, not the grand master work.

9. Kronecker Products

Here from https://en.wikipedia.org/wiki/Kronecker_product :

In mathematics, the **Kronecker product**, denoted by ⊗, is an operation on two matrices of arbitrary size resulting in a block matrix. It is a generalization of the outer product (which is denoted by the same symbol) from vectors to matrices, and gives the matrix of the tensor product with respect to a standard choice of basis. The Kronecker product should not be confused with the usual matrix multiplication, which is an entirely different operation.

From this definition, the Kronecker product is the generalization of the outer product of two vectors to the outer product of two matrices, as per my tensor doc statement that $(S\otimes\mathbf{T})_{\mathtt{ab,AB}} \equiv S_{\mathtt{aA}}T_{\mathtt{bB}}$

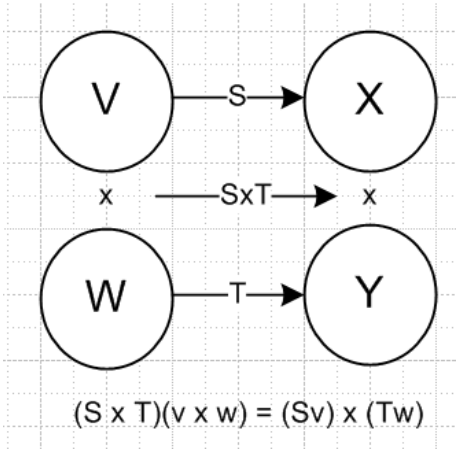
But here is more from the same wiki page:

3. **Relation to the abstract tensor product:**

   The Kronecker product of matrices corresponds to the abstract tensor product of linear maps. Specifically, if the vector spaces V, W, X, and Y have bases $\{v_1, \ldots, v_m\}$, $\{w_1, \ldots, w_n\}$, $\{x_1, \ldots, x_d\}$, and $\{y_1, \ldots, y_e\}$, respectively, and if the matrices A and B represent the linear transformations $S : V \to X$ and $T : W \to Y$, respectively in the appropriate bases, then the matrix $A \otimes B$ represents the tensor product of the two maps, $S \otimes T : V \otimes W \to X \otimes Y$ with respect to the basis $\{v_1 \otimes w_1, v_1 \otimes w_2, \ldots, v_2 \otimes w_1, \ldots, v_m \otimes w_n\}$ of $V \otimes W$ and the similarly defined basis of $X \otimes Y$ with the property that $A \otimes B(v_i \otimes w_j) = (Av_i)\otimes(Bw_j)$, where i and j are integers in the proper range.[3]

   When V and W are Lie algebras, and $S : V \to V$ and $T : W \to W$ are Lie algebra homomorphisms, the Kronecker sum of A and B represents the induced Lie algebra homomorphisms $V \otimes W \to V \otimes W$.

Well this is that QM notation of the cross product of two operators in different spaces



The claim then is that

$(S\otimes T)(v\otimes w) = (Sv)\otimes(Tw)$

$(S\otimes T) : V\otimes W \rightarrow (SV)\otimes(TW) = X\otimes Y$

Suppose I take $v = e_{\mathtt{i}}$ and $w = e'_{\mathtt{i}}$. Then the first line says,

$(S\otimes T)(e_{\mathtt{i}}\otimes e'_{\mathtt{j}}) = (Se_{\mathtt{i}})\otimes(Te'_{\mathtt{j}})$.

Now I know that $Se_{\mathtt{i}}$ is a vector, so lets add some subscripts to the action

$$[(S\otimes T)(e_i\otimes e'_j)]_{ab} = [(Se_i)\otimes(Te'_j)]_{ab} = (Se_i)_a(Te'_j)_b$$

Now consider

$$(Se_i)_a = S_{ar}(e_i)_r = S_{ar}\,\delta_{i,r} = S_{ai} \qquad \text{similarly } (Te'_j)_b = T_{bj}$$

Then we end up with

$$[(S\otimes T)(e_i\otimes e'_j)]_{ab} = S_{ai}T_{jb}$$

Then I would have to define

$$(e_a\otimes e'_b)^T(S\otimes T)(e_i\otimes e'_j) \equiv (S\otimes T)_{ai,bj} = S_{ai}T_{bj}$$

and that then produces the desired result. The left side could also be written

$$(e_a\otimes e'_b)^T(S\otimes T)(e_i\otimes e'_j) = (e_a^T\, S\, e_j)(e'^T_b\, T\, e_{j'}) = S_{aj}\,T_{bj}$$

which is another way to get the result.

A different take on this goes back to

$$(S\otimes T)(v\otimes w) = (Sv)\otimes(Tw)$$

where this gives the quantum mechanics Hilbert space notion that S acts on the first item and T acts on the second item. This is a tensor product operator acting on a tensor product vector, so is really a new concept from the regular tensor product idea.

Possible link

https://www.dpmms.cam.ac.uk/~wtg10/tensors3.html

it is very wordy and has poor graphics but it does talk in primer language.

I have looked pretty hard on the web under "tensor product" "outer product" and I just don't see anyone talking about the latter being a special case of the former.

Most relevant sources do talk about a "free vector space" defined on sets.

10. Axler book has nada

Sheldon Axler's *Linear Algebra Done Right*.

Try to get this item: got it, but it has nothing, it is just a basics LA book.

11. Ganatra's class notes

I just invested heavily in this guy **Ganatra**,

http://math.stanford.edu/~ganatra/math113/

In homework 7 you are supposed to show a certain **triangle fact**, and he even has a solution set, but I don't think the triangle fact is proved. You have to show that something is linear if two other things are bilinear. Basically, no proof is given in the homework. But the buzz word "universal bilinear map" is mentioned and maybe that would lead to another source on this.

https://books.google.com/books?id=bSyQr-wUys8C&pg=PA361&lpg=PA361&dq=%22universal+bilinear+map%22&source=bl&ots=Laf7MeHJWr&sig=LUDaYqVMPeG6jGW_3T2goLvUtuU&hl=en&sa=X&ved=0CDgQ6AEwA2oVChMI1MTigM38xwIV2A-SCh39CgCo#v=onepage&q=%22universal%20bilinear%20map%22&f=false

Here is some similar material: [this is where I found the Roman book probably in google books]

12. Roman book preview

Tensor Products

Let U and V be vector spaces. Our guide for the definition of the tensor product $U \otimes V$ will be the desire to have a universal property for bilinear functions, as measured by linearity. Referring to Figure 14.4, we want to define a vector space T and a bilinear map $t: U \times V \to T$ so that any bilinear map f with domain $U \times V$ can be factored through t. Intuitively speaking, t is the most "general" or "universal" bilinear map with domain $U \times V$: It is bilinear *and nothing more*.

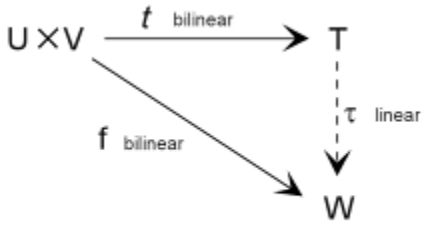


Figure 14.4

Now clear why this is a "guide for the definition of tensor product U ⊗ V". I think perhaps the idea is that there is really only one bilinear map off UxV. Here they show one going to T and another to W, but since T and W are linearly related, perhaps the bilinear maps are in some sense the same. Let's try do decode some of this guy's stuff, I doubt I will get very far.

Definition *Let $U \times V$ be the cartesian product of two vector spaces over F. Let $\mathcal{S} = \mathrm{Vect}(F)$. Let*

$$\mathcal{F} = \bigcup_{W} \{\hom_F(U, V; W) \mid W \in \mathcal{S}\}$$

be the family of all bilinear maps from $U \times V$ to any vector space W. The measuring family $\mathcal{H}$ is the family of all linear transformations.

Decoder activating: hom shown is the set of all *bilinear* maps from UxV to ANY vector space W, so I guess S is a set of candidate vector spaces for W. But I don't know what "measuring family" means. Maybe he is defining: "measuring family" ≡ all linear transformations in the universe??

Now how obscure can he make it:

A pair $(T, t: U \times V \to T)$ is **universal for bilinearity** *if it is universal for $(\mathcal{F}, \mathcal{H})$, that is, if for every bilinear map $f: U \times V \to W$, there is a* unique *linear transformation $\tau: T \to W$ for which*

$$f = \tau \circ t$$

The map τ is called the **mediating morphism** *for f.* □

All these glossy words just say that the τ link in the triangle is linear. I guess T and t here are two maps of the same type. The "pair" here is the space T and the mapping t. The *pair* is what is universal IF the rest of the statement is true. This is just a definition of universal, there is no proof that anything is linear.

We can now define the tensor product via this universal property.

Definition *Let U and V be vector spaces over a field F. Any universal pair $(T, t: U \times V \to T)$ for bilinearity is called a* **tensor product** *of U and V. The vector space T is denoted by $U \otimes V$ and sometimes referred to by itself as the tensor product. The map t is called the* **tensor map** *and the elements of $U \otimes V$ are called* **tensors**.

It is customary to use the symbol $\otimes$ to denote the image of any ordered pair (u, v) under the tensor map, that is,

$$u \otimes v = t(u, v)$$

for any $u \in U$ and $v \in V$. A tensor of the form $u \otimes v$ is said to be **decomposable**, *that is, the decomposable tensors are the images under the tensor map.* □

So T = U⊗V is a tensor product if {t,T} is a universal pair. Every universal pair defines a tensor product. Again: Any universal pair {T; t:UxV→T } defines a tensor product U⊗V ≡ T. So I guess if you want to test some specific U⊗V for being a tensor product, you have to identify the "universal pair" thing.

Since universal pairs are unique up to isomorphism, we may refer to "the" tensor product of vector spaces. Note also that the tensor product $\otimes$ is not a product in the sense of a binary operation on a set. In fact, even when $V = U$, the tensor product $u \otimes u$ is not in U, but rather in $U \otimes U$.

The text then skips two pages. The book here is

Advanced Linear Algebra, by Steven Roman (in google books). Can I find that book? I got the TOC at least. OK, found a slow copy from China at 5.23 MB, then I can continue with the above presentation. It would be good if the author would give AN EXAMPLE!!! Can you always find a universal pair? I guess you have to say something about the space T.

OK, I have the Roman book and it is very good! I need to study it a bit in my next session. I think it makes the most general definition of a tensor product, and I think I will be able to show that the outer product is a particular case of a tensor product.

13. Roman Notes, Starting on Page 355 which is Chapter 14 Tensor Products.

13.1 The Universal Pair Concept

"Universal Pair" is a concept in "Category Theory". Roman will use "simple terminology".

Referring to Figure 14.1, consider a set A and two functions f and g, with domain A.

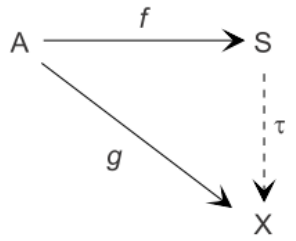


Figure 14.1

Suppose that there exists a function $\tau: S \to X$ for which this diagram *commutes*, that is,

$$g = \tau \circ f$$

This is sometimes expressed by saying that g can be **factored through** f. What does this say about the relationship between the functions f and g?

Notice that nothing is bilinear or linear yet, just a generic picture.

Let us think of the "information" about A contained in a function $h: A \to B$ as the way in which h *distinguishes* elements of A using *labels* from B. The relationship above implies that

$$g(a) \neq g(b) \Rightarrow f(a) \neq f(b)$$

and this can be phrased by saying that whatever ability g has to distinguish elements of A is also possessed by f. Put another way, except for labeling differences, any information about A that is contained in g is also contained in f.

So the ability to "distinguish" elements in A is important, and the above triangle for me says that f(a) = f(b) $\Rightarrow$ g(a) = g(b), so the contra shown above. So f and g have "something in common". Notice that no statement is made about the nature of mappings f,g,τ.

If τ happens to be injective, then the *only* difference between f and g is the values of the labels. That is, the two functions have the same information about A. However, in general, τ is not required to be injective and so f may contain more information than g.

Injective means 1-to-1, so then we can say $f = \tau^{-1} \circ g$ and then we get the other directions above, and then f and g exactly distinguish things in A in the same manner.

Now consider a family $\mathcal{S}$ of sets and a family

$$\mathcal{F} = \{g: A \to X \mid X \in \mathcal{S}\}$$

Assume that $S \in \mathcal{S}$ and $f: A \to S \in \mathcal{F}$. If the diagram in Figure 14.1 commutes for all $g \in \mathcal{F}$, then the information contained in every function in $\mathcal{F}$ is also contained in f. Moreover, since $f \in \mathcal{F}$, the function f cannot contain more information than is contained in the entire family and so we conclude that f contains exactly the same information as is contained in the entire family $\mathcal{F}$. In this sense, $f: A \to S$ is *universal* among all functions $g: A \to X$ in $\mathcal{F}$.

Set $\boldsymbol{\mathcal{S}}$ seems to be the set of all possible spaces X (no claim of vector space). Then S is one of those spaces so we have g: A→S. Note that "g: A→S" is an element of $\boldsymbol{\mathcal{F}}$. Imagine there is another element of $\boldsymbol{\mathcal{F}}$ which is "f: A→S" ($\boldsymbol{\mathcal{F}}$ is a very large set of mappings from set A). In some sense, f and g are both functions associated with this space $\boldsymbol{\mathcal{F}}$. We just said above that for any S, the information about A contained in g is also contained in f. By letting g and f wander in $\boldsymbol{\mathcal{F}}$, I guess all the functions in $\boldsymbol{\mathcal{F}}$ have the same information about A as does a particular function f, assuming the above triangle idea. I would perhaps say that f was "representative" of F, and as such, contains the same info about A as any other representative. Their word is to say that f is "universal" in its information content, all other g have the same information. The real phrase is that "f is universal among the functions in $\boldsymbol{\mathcal{F}}$.

In this way, a single function $f: A \to S$, or more precisely, a single pair (S, f), can capture a mathematical concept as described by a family of functions. Some examples from linear algebra are basis for a vector space, quotient space, direct sum and bilinearity (as we will see).

The pair here is (S, f), a particular range space for function f, and the function f. I guess we can apply this idea to any set A, so we don't need to indicate A in this notation. I would like to see some examples. Not obvious to me why a basis for a vector space is an example.

Let us make a formal definition.

Definition *Referring to Figure 14.2, let A be a set and let $\mathcal{S}$ be a family of sets. Let*

$$\mathcal{F} = \{g: A \to X \mid X \in \mathcal{S}\}$$

be a family of functions, all of which have domain A and range a member of $\mathcal{S}$. Let

$$\mathcal{H} = \{\tau: X \to Y \mid X, Y \in \mathcal{S}\}$$

be a family of functions with domain and range in $\mathcal{S}$. We assume that $\mathcal{H}$ has the following structure:

So $\mathcal{F}$ is exactly as before, but the new item is $\mathcal{H}$. In the first thing, X is some space in the set $\mathcal{S}$. In the second thing, this space X is the domain, and the range is some other space in $\mathcal{S}$. Also part of $\mathcal{H}$ is a new function τ. Basically, $\mathcal{H}$ is the space of all mappings between any two spaces in $\mathcal{S}$. But we now add requirement on the set of functions H:

1) *$\mathcal{H}$ contains the identity function ι_S for each member of $\mathcal{S}$.*

What does that mean exactly? Well, that applies only if X = Y, and then $i_X : X \to X$ could be a well defined identity function, so that is OK.

2) *$\mathcal{H}$ is closed under composition of functions, which is an associative operation.*

I think this means that if τ_1: X→Y and τ_2:Y→Z, then τ_1 o τ_2: X → Z "makes sense", where X,Y and Z are all in $\mathcal{S}$. I guess this would be "associative" because for more functions o is associative?? Unclear.

3) *For any $\tau \in \mathcal{H}$ and $f \in \mathcal{F}$, the composition $\tau \circ f$ is defined and belongs to $\mathcal{F}$.*

OK, this is just a straight assumption, fine by me. Here then is a picture:

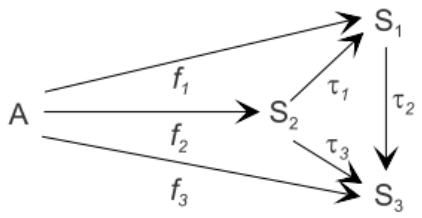


Figure 14.2

This fancy picture is one of many you could make. Three arbitrary spaces S_i as shown, and A. The three functions f_i are all obviously in the set $\mathcal{F}$. And the three τ_i functions are in $\mathcal{H}$. Notice how assumption 3 really says that $\tau_1 \circ f_2 = f_1$ and also $\tau_3 \circ f_2 = f_3$ and also $\tau_2 \circ f_1 = f_3$. Meanwhile. assumption 2 allows the fact that $\tau_2 \circ \tau_1 = \tau_3$. I write these equations using "commuting paths" in the picture.

We refer to $\mathcal{H}$ as the **measuring family** *and its members as* **measuring functions**.

But what is it that they are measuring?

Definition of a Universal Pair:

A pair $(S, f: A \to S)$, where $S \in \mathcal{S}$ and $f \in \mathcal{F}$ has the **universal property** *for the family $\mathcal{F}$* **as measured by** *$\mathcal{H}$, or is a* **universal pair** *for $(\mathcal{F}, \mathcal{H})$, if for every $g: A \to X$ in $\mathcal{F}$, there is a unique $\tau: S \to X$ in $\mathcal{H}$ for which the diagram in Figure 14.1 commutes, that is, for which*

$$g = \tau \circ f$$

or equivalently, any $g \in \mathcal{F}$ can be **factored through** *f. The unique measuring function τ is called the* **mediating morphism** *for g.*□

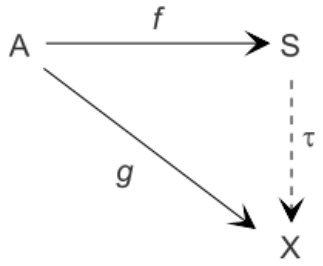


Figure 14.1

In this chunk, Roman is zeroing in that universal buzzword. Here is what I think the defining clause says starting with "if for every": for every (g,X) and (f,S) in $\mathcal{F}$, there exists a τ in $\mathcal{H}$ which is unique AND which allows the above picture to commute, which means you can write g = f o τ ("g can be factored through f"). If this situation is true, THEN

- The pair (S,f) has the universal property for set $\mathcal{F}$ as measured by $\mathcal{H}$
- The pair (S,f) is a universal pair for ($\mathcal{F}$,$\mathcal{H}$)

In the picture, the unique τ function is called the "mediating morphism for g".

OK, look at the above picture where f and g are in $\mathcal{F}$ **.** If the triangle picture can be drawn, then (S,f) is *almost* a universal pair. The only extra thing you need is that τ is unique.

Recall from earlier category notes and a morphism is an arrow in a category diagram like that above. It is also a map which "preserves some structure". Here, perhaps that is the information-bearing capability of the functions f and g. Perhaps it is just the commuting of the diagram is the structure being preserved.

Note the requirement that the mediating morphism τ be unique. Universal pairs are essentially unique, as the following describes.

Unclear. The universal pair is the thing (S,f), so how can this pair be unique? (X,g) is another pair that is a different pair. Yes, they are linked by a unique function τ. We continue on:

Theorem 14.1 *Let $(S, f: A \to S)$ and $(T, g: A \to T)$ be universal pairs for $(\mathcal{F}, \mathcal{H})$. Then there is a bijective measuring function $\mu \in \mathcal{H}$ for which $\mu S = T$. In fact, the mediating morphism of f with respect to g and the mediating morphism of g with respect to f are isomorphisms.*

Here we have two pairs (S,f) and (T,g). Each is a universal pair. I don't know what μS=T means. Perhaps for all elements s in S, you get μ(s) = t where t ϵ T. Also unclear: if τ_1 and τ_2 are the two unique functions as indicated, how can two functions be an isomorphism? Maybe if you run over all f and g in $\mathcal{F}$, the set of τ_1 mediating morphisms f to g is isomorphic to the set τ_2 for g to f. Hopefully the following proof of the theorem will help to clarify what it is trying to say. Proof starts off,

Proof. With reference to Figure 14.3, there are mediating morphisms $\tau: S \to T$ and $\sigma: T \to S$ for which

$$g = \tau \circ f$$
$$f = \sigma \circ g$$

Hence,

$$g = (\tau \circ \sigma) \circ g$$
$$f = (\sigma \circ \tau) \circ f$$

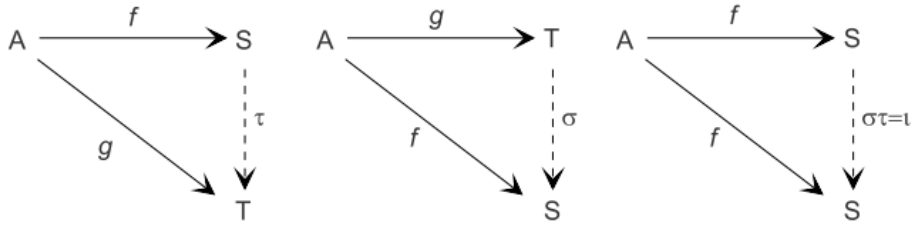


Figure 14.3

On his second line, he is saying f = σ o (τ o f) = (σ o τ) o f which quietly makes use of the assumed associativity property of the composition operator o. So far all is fine. We continue,

However, referring to the third diagram in Figure 14.3, both $\sigma \circ \tau: S \to S$ and the identity map $\iota: S \to S$ are mediating morphisms for f and so the uniqueness of mediating morphisms implies that $\sigma \circ \tau = \iota$. Similarly $\tau \circ \sigma = \iota$ and so τ and σ are inverses of one another, making τ the desired bijection.□

OK, in that third picture the unique dotted line function is (σ o τ) since we know f = (σ o τ) o f. On the other hand, we know that f = i_S f where i_S is the identity operator. Since things are unique, we know then that σ o τ = i_S. A similar argument would show τ o σ = i_S Thus, σ and τ are inverses of each other (left and right inverses). So the function τ is in fact 1-to-1. Not exactly clear why τ is "onto" but I guess it is defined for every element in S, and therefore it is a bijection as stated.

I still don't see the significance of the universal pair concept, nor do I see what is "unique" about it. I guess in the left picture, Roman is studying the universal pair (S,f), while in the middle (T,g). Since it turns out that τ and σ are inverses, maybe THAT is why one says that (S,f) and (T,g) are essentially the same, and so in that sense there is only one universal pair out of A, and so that pair is then unique.

Universality Example 1: the triangle with inclusion map j on top

Examples of Universality

Now let us look at some examples of the universal property. Let Vect(F) denote the family of all vector spaces over the base field F. (We use the term *family* informally to represent what in set theory is formally referred to as a *class. A class* is a "collection" that is too large to be considered a set. For example, Vect(F) is a class.)

Vect(F) = all vector spaces over field F. Note terminology I have often wondered about

family = **class** = **collection** that is too large to be considered a **set**

There are certainly many infinite vector spaces on the reals, for example.

Example 14.1 (*Bases*) Let $\mathcal{B}$ be a nonempty set and let

1) $\mathcal{S} = \text{Vect}(F)$
2) $\mathcal{F} =$ set functions from $\mathcal{B}$ to members of $\mathcal{F}$
3) $\mathcal{H} =$ linear transformations

If $V_\mathcal{B}$ is a vector space with basis $\mathcal{B}$, then the pair $(V_\mathcal{B}, j\colon \mathcal{B} \to V_\mathcal{B})$, where j is the inclusion map $jv = v$, is universal for $(\mathcal{F}, \mathcal{H})$. To see this, note that the condition that $g \in \mathcal{F}$ can be factored through j,

$$g = \tau \circ j$$

is equivalent to the statement that $\tau v = gv$ for each basis vector $v \in \mathcal{B}$. But this uniquely defines a linear transformation τ.

Let's review the soup ingredients:

$\mathcal{B}$ = a set of at least one element (it will be the set of basis elements of a V)
$\mathcal{S}$ = all those vector spaces Vect(F) . Before, this was just a collection of sets.
$\mathcal{F}$ = the collection of "set functions" f such that f: B → $\mathcal{F}$
so this function f maps elements of the set B into other functions in $\mathcal{F}$, strange.
$\mathcal{H}$ = set of *linear* transformations I suppose from any space to any other space.

Now the idea is to take as the set $\mathcal{B}$ a set of basis elements of vector space V_B.

But what is an "inclusion map"?

Inclusion map

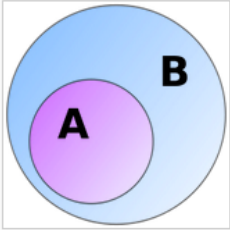
From Wikipedia, the free encyclopedia

In mathematics, if A is a subset of B, then the **inclusion map** (also **inclusion function**, **insertion**, or **canonical injection**) [1] is the function ι that sends each element, x of A to x, treated as an element of B:

$$\iota : A \to B, \qquad \iota(x) = x.$$

A "hooked arrow" $\hookrightarrow$ is sometimes used in place of the function arrow above to denote an inclusion map.

This and other analogous injective functions [2] from substructures are sometimes called ***natural injections***.



A is a subset of *B*, and *B* is a superset of *A*.

The inclusion map takes things in a subset and maps them into the larger space. So we can then for example map those basis elements from the smaller set $\mathcal{B}$ into the full vector space V_B. That is what the mapping j does. The pair under study here is then (V_B, j) or in more detail (V_B, j:$\mathcal{B}\to V_B$).

Now somehow Roman wants to say that this is an example of a "universal pair". We have to show that there is that triangle picture with a unique τ.

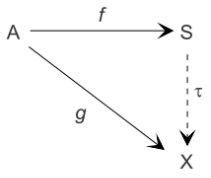
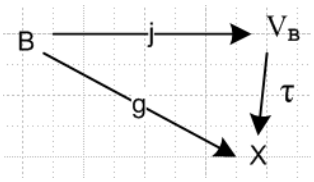


Figure 14.1



Repeat text above

If V_B is a vector space with basis $\mathcal{B}$, then the pair $(V_B, j{:}\mathcal{B} \to V_B)$, where j is the inclusion map $jv = v$, is universal for $(\mathcal{F}, \mathcal{H})$. To see this, note that the condition that $g \in \mathcal{F}$ can be factored through j,

$$g = \tau \circ j$$

is equivalent to the statement that $\tau v = gv$ for each basis vector $v \in \mathcal{B}$. But this uniquely defines a linear transformation τ.

What is object v here? It must be a basis element in B. We can map j(v) = v into V_B, then apply τ to get τ(v) which he writes as τv (a notation unclear to me). But start with the same v back in B and then g(v) = gv. Thus τv = gv. Write again as τ(v) = g(v) for all v in B. This seems to say that τ = g, and then certainly τ is unique, given g.

Now why is τ linear? You would have to show that τ(v + v') = τ(v) + τ(v'). That is the same then as g(v+v') = g(v) + g(v'), but Roman has not specified that g is a linear mapping. I keep losing the thread in

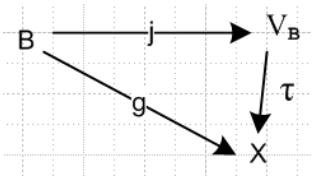
this kind of discussion when it comes to τ being linear! This is really the first appearance of the word linear, except he imposed that $\mathcal{H}$ = linear transformations.

The only way out I can see is this: Roman must have neglected to say that g ϵ $\mathcal{F}$ where $\mathcal{F}$ is a space of *linear* functions. Then when you end up with τ = g, then τ is also linear. He continues,

> In fact, the universality of the pair $(V_{\mathcal{B}}, j)$ is *precisely* the statement that a linear transformation τ is uniquely determined by assigning its values arbitrarily on a basis $\mathcal{B}$, the function g doing the arbitrary assignment in this context. Note also that Theorem 14.1 implies that if $(W, k\colon \mathcal{B} \to W)$ is also universal for $(\mathcal{F}, \mathcal{H})$, then there is a bijective mediating morphism from $V_{\mathcal{B}}$ to W, that is, W and $V_{\mathcal{B}}$ are isomorphic.□

I continue to think that g(v) must be linear. The above paragraph does not help me much. I hope this confusion won't wreck the rest of what Roman has to say.

[Note added 9.18.15 7 PM: Go back to the above picture



where we have

j(v) = v in V_B inclusion map
τ(v) = x
x = g(v)

Then we can define τ by

τ(v) = g(v)

Now suppose we simply declare that τ is linear, as I learned much later in these notes. Then we have

τ(v + v') = τ(v)+τ(v')

But this says

g(v+v') = g(v) + g(v')

Therefore, if you want τ to be linear, you must also have g be linear.]

Universality Example 2: The quotient of a vector space divided by a subspace.

Next example, here is the setup:

Example 14.2 (*Quotient spaces and canonical projections*) Let V be a vector space and let K be a subspace of V. Let

1) $\mathcal{S} = \text{Vect}(F)$
2) $\mathcal{F} =$ linear maps with domain V, whose kernels contain K
3) $\mathcal{H} =$ linear transformations

OK, now he says that $\mathcal{F}$ is *linear* maps. Also, K is a subspace of V and also this same K exists within the nullspace of any of those linear maps in $\mathcal{F}$. So this is a much more restricted collection $\mathcal{F}$. For example, if f is in $\mathcal{F}$ and if k is in K, then f(k) = 0.

Now I need to go access Theorem 3.4 from earlier in the book: I leave p 366 and go to p 104

Theorem 3.4 *Let S be a subspace of V and let $\tau \in \mathcal{L}(V, W)$ satisfy $S \subseteq \ker(\tau)$. Then, as pictured in Figure 3.2, there is a unique linear transformation $\tau' : V/S \to W$ with the property that*

$$\tau' \circ \pi_S = \tau$$

Moreover, $\ker(\tau') = \ker(\tau)/S$ *and* $\text{im}(\tau') = \text{im}(\tau)$.

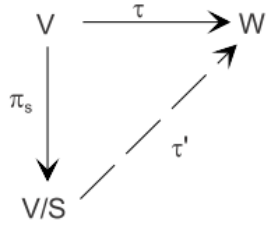


Figure 3.2: The universal property

Note that $\mathcal{L}$(V,W) is defined on page 73 as the obvious

Definition *Let V and W be vector spaces over a field F. A function $\tau : V \to W$ is a* **linear transformation** *if*

$$\tau(ru + sv) = r\tau(u) + s\tau(v)$$

for all scalars $r, s \in F$ and vectors $u, v \in V$. The set of all linear transformations from V to W is denoted by $\mathcal{L}(V, W)$.

So in Theorem 3.4 and its triangle picture, τ is a *linear* transformation τ:V→W. And in this theorem S is a subspace of V and all of S lies in the kernel of τ, so τ(s) = 0 for any s ϵ S. But now I need to know the meaning of the notation V/S. Well, in Galois doc I first discuss G/H being the factor group when H is a subgroup of G. Later I discuss R/I as the residue class ring for an ideal I in the ring R. But I never talk about the notion of A/B where B is a subspace of vector space A. That topic is addressed by Roman on page 100 and it is very similar to the group discussion. You write [v] = v + S, meaning that all elements

v+S are in the same row of the chart (v = representative) because here f(v+s) = f(v) since s is in the kernel of all functions f of interest. So the set of rows [v] form the quotient group which then is V/S.

So not dwelling too much on those V/S details, Theorem 3.4 above is saying that if π_S projects elements of vector space V onto those "chart rows", that is, into V/S, then for any linear τ:V→W, you can find a linear τ' such that τ' : V/S → W and you will then have $\tau' \circ \pi_S = \tau$.

This certainly seems a strange little world. Why would you want to carry out mapping τ by "factoring through π_S" and thus passing through the quotient space V/S of V relative to S on the way from V to W ?

Comment: since f(v+s) = f(v) for all s in the subgroup S, there is sort of only one f(v) here and you might as well talk about f([v]) and not bother with a list of all the s's. So you save effort by talking only about f([v]). You "mod out" all the equivalent arguments. So that is a good reason to work on V/S.

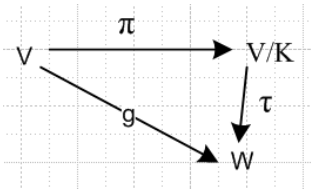
But let's just accept this rather obscure Theorem 3.4 and get back now to where we were.

Example 14.2 (*Quotient spaces and canonical projections*) Let V be a vector space and let K be a subspace of V. Let

1) $\mathcal{S} = \text{Vect}(F)$
2) $\mathcal{F}$ = linear maps with domain V, whose kernels contain K
3) $\mathcal{H}$ = linear transformations

Theorem 3.4 says precisely that the pair $(V/K, \pi: V \to V/K)$, where π is the canonical projection map, has the universal property for $\mathcal{F}$ as measured by $\mathcal{H}$.□

The pair here is (V/K, π) and we can then way in our local notation,



Since τ is a unique and linear transformation in $\mathcal{H}$, you can conclude that (V/K,π) is a universal pair, and that is what this is an example of. I guess π is called a canonical projection map in that earlier section.

[Why is (V/K,π) a universal pair: (1) because the above triangle exists; (2) τ is linear.]

I now skip example 14.3 regarding direct sums being an example of universal pairs. I hope it won't be needed for what I am trying to track down by reading in Roman.

13.2 Direct Sums, External and Internal

External Direct Sums

Definition *Let $V_1, \ldots, V_n$ be vector spaces over a field F. The* **external direct sum** *of $V_1, \ldots, V_n$, denoted by*

$$V = V_1 \boxplus \cdots \boxplus V_n$$

is the vector space V whose elements are ordered n-tuples:

$$V = \{(v_1, \ldots, v_n) \mid v_i \in V_i, i = 1, \ldots, n\}$$

with componentwise operations

$$(u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n)$$

and

$$r(v_1, \ldots, v_n) = (rv_1, \ldots, rv_n)$$

for all $r \in F$. □

What is this saying? When he says $v_i \in V_i$, I guess v_i is a vector which has a number of components equal to the dimension of V_i. So I think the element of the direct sum space is just a vertical stack of vectors of different sizes, and this is what I have always thought of as the direct sum. A matrix in block diagonal form would then map such an stacked element of V into V with no mixing. I have always used the symbol ⊕ for this direct sum idea, but he uses a boxed plus sign.

But in the above, the V_i are ANY vector spaces. In Roman's "internal" direct sum, it is required that each of the V_i be a *subspace* of V. So then he writes $V = V_1 \oplus V_2 \oplus \ldots$ using the ⊕ symbol. So this then would be a special case of the box-plus symbol direct sum. Now back to where we were:

In fact, if V is a vector space, there are two classes of functions from $V \times V$ to W: the linear maps $\mathcal{L}(V \times V, W)$, where $V \times V = V \boxplus V$ is the direct product of vector spaces, and the bilinear maps $\hom(V, V; W)$, where $V \times V$ is just the cartesian product of sets. We leave it as an exercise to show that these two classes have only the zero map in common. In other words, the only map that is both linear and bilinear is the zero map.

Consider the meaning of $\mathcal{L}$(V x V, W) whose representative reads w= f((v_1,v_2)). The domain element is simply the pair (v_1,v_2), an element of V x V. Linearity (implied by $\mathcal{L}$) would mean

f((v_1,v_2) + (v'_1,v'_2)) = f((v_1,v_2)) + f ((v'_1,v'_2))

Note that there is no sense of "bilinearity" here. That is, we do not have,

f(v_1+ v_2,v_3) = f(v_1,v_3) + f(v_2,v_3). for example (part of bilinearity)

We cannot extract either argument of a pair in the first line and do something with it.

Now go back to the first concept with f((v_1,v_2)) . Think of (v_1,v_2) as a stacked vector, then you would say (v_1,v_2) = element of both V x V and V ⊞ V. You have a direct sum structure (external type). Probably V is not a subspace of V x V, so you do not have the type of direct sum V ⊕ V here. Remember that V ⊞ V is the rawest most generic direct sum.

I think the above quoted says: start with the Cartesian thing V x V and you have as yet no algebra. But then add the set of linear maps $\mathcal{L}$(V x V, W) defined on V x V, then you have some linear functions going on, f: V x V → W. This is the first of two possible structures you could have. The second is that you define a bilinear map w = f(v_1,v_2) and then your structure is this set of bilinear functions on V x V. Recall that this set of bilinear maps is called hom(V,V;W) whereas the first is sort of V ⊞ V.

13.3 Bilinear Maps

New Roman Topic

He then defines **bilinear** maps and I am fine with that. Here then are some definitions

The set of all bilinear functions from $U \times V$ *to* W *is denoted by* $\hom_F(U, V; W)$. *A bilinear function* $f: U \times V \to F$ *with values in the base field* F *is called a* **bilinear form** *on* $U \times V$.□

Notice these definitions and notations:

• set of all *bilinear* functions f such that f:UxV→W is called **$\hom_F$(U,V; W).**

I think the F refers to the field used for scalars in the linearity stuff, called "the base field".
This set is in fact itself a vector space. Probably there is some reason for the name "hom" relating to homomorphism, but no one wants to explain it on line. // Well above I quoted something that said with vector spaces, homomorphism and linear mean the same thing.

• If f:UxV→F, that base field, then f is called a **bilinear form**.

• If f:UxV→R, I have been referring to this as a **bilinear functional**.

It is important to emphasize that, in the definition of bilinear function, $U \times V$ is the *cartesian product of sets*, not the direct product of vector spaces. In other words, we do not consider any algebraic structure on $U \times V$ when defining bilinear functions, so expressions like

$$(x, y) + (z, w) \quad \text{and} \quad r(x, y)$$

are meaningless.

So elements of U x V are just (u,v) and a priori they are not part of any "algebra". I agree.

Now here is an example of the bilinear case.

Example 14.4 (**Multiplication is bilinear**) If A is an algebra, the product map $\mu: A \times A \to A$ defined by

$$\mu(a, b) = ab$$

is bilinear, that is, multiplication is linear in each position.□

I think of A as a set for which multiplication is defined. Then f: A x A → A says μ(a,b) = ab and since this is a product, it is obviously bilinear. Reminds me of the **outer product idea** yet to come.

Again: this Example 14.4 is an example of f : V x V → W with the bilinear structure. It happens that W = V so we have μ : V x V → V with w = μ(a,b) = ab. Manifestly bilinear!

I skip Example 14.5, and go to

Example 14.6 If V and W are vector spaces, and $f \in V^*$ and $g \in W^*$, then the product map $\phi: V \times W \to F$ defined by

$$\phi(v, w) = f(v)g(w)$$

is bilinear. Dually, if $v \in V$ and $w \in W$, then the map $\lambda: V^* \times W^* \to F$ defined by

$$\lambda(f, g) = f(v)g(w)$$

is bilinear.□

Here f and g are *linear* functionals, and this causes φ(u,v) defined above to be bilinear!! The Dually part above is certainly just a rehash of the first case with V* and W*. The functionals appear in the same way in both examples. I am happy with this example, both parts.

It is precisely the tensor product that will allow us to generalize the previous example. In particular, if $\tau \in \mathcal{L}(U, W)$ and $\sigma \in \mathcal{L}(V, W)$, then we would like to consider a "product" map $\phi: U \times V \to W$ defined by

$$\phi(u, v) = \tau(u) \,?\, \sigma(v)$$

The tensor product $\otimes$ is just the thing to replace the question mark, because it has the desired bilinearity property, as we will see. In fact, the tensor product is bilinear and nothing else, so it is *exactly* what we need!

Notice this statement that your basic skeletal tensor product is bilinear and nothing else!

I have now done all the background work and am ready to go over again the upcoming section on tensor products. On my previous pass, the background was not there yet.

(Resuming here at 3 PM on 9.18.15)

13.4 Tensor Products

Tensor Products

Let U and V be vector spaces. Our guide for the definition of the tensor product $U \otimes V$ will be the desire to have a universal property for bilinear functions, as measured by linearity. Referring to Figure 14.4, we want to define a vector space T and a bilinear map $t: U \times V \to T$ so that any bilinear map f with domain $U \times V$ can be factored through t. Intuitively speaking, t is the most "general" or "universal" bilinear map with domain $U \times V$: It is bilinear *and nothing more*.

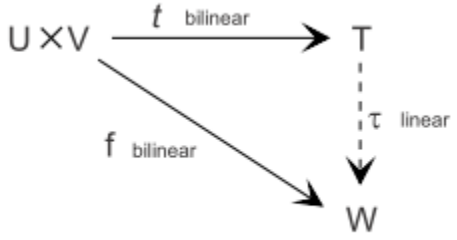


Figure 14.4

I am just quoting text as it appears, now on Roman p 362. Nothing more is said at this point about the above picture. For example, no proof is given that if f and t are bilinear, then τ is linear. But I will take a quick shot right here:

Plan A:

$t_1 = t(u.v)$ is bilinear

$w = f(u.v)$ is bilinear

what then do we know about τ which appears in $w = \tau(t_1)$?

To show that τ was linear, I would have to show that

$\tau(t_1 + t_2) = \tau(t_1) + \tau(t_2)$

or

$\tau(\ t(u1,v1) + t(u2,v2)) = \ \tau(\ t(u1,v1)) + \tau(\ t(u2,v2))$

or

$\tau \circ (\ t(u1,v1) + \ t(u2,v2)) = \tau \circ t(u1,v1) \ + \ \tau \circ t(u2,v2)$

I know that

RHS = $f(u1,v1) + f(u2,v2) = w_1 + w_2$

So I would have to show that

$\tau(\, t(u1,v1) + t(u2,v2)) = w_1 + w_2$

But if I know nothing at all about τ, then I cannot write $\tau(\, t(u1,v1) + t(u2,v2))$ in any way!

Plan B:

Try writing $\tau \circ t = f$ as $\tau = f \circ t^{-1}$ where t^{-1}: T → U x V. But it seems to be very unlikely that t is invertible, so this really goes nowhere.

Comment: Having tried this about three times now, I am getting the idea that maybe linear τ is by fiat and not something you prove. It is part of that universal buzzword definition. If you have τ linear, then maybe you can show also unique and then you have universal pair (T,t). Let's just hack onwards in hopes of getting some answers in a few days.

[Note added 9.18.15 7 PM.: I finally show below that you do in fact just *declare* τ to be linear as part of its definition. When you do this, you check to make sure everything in the triangle is consistent! You can never prove that τ is linear by Plans such as the above!]

Roman continues (no gap)

Definition *Let $U \times V$ be the cartesian product of two vector spaces over F. Let $\mathcal{S} = \mathrm{Vect}(F)$. Let*

$$\mathcal{F} = \bigcup_W \{\hom_F(U, V; W) \mid W \in \mathcal{S}\}$$

be the family of all bilinear maps from $U \times V$ to any vector space W. The measuring family $\mathcal{H}$ is the family of all linear transformations.

A pair $(T, t: U \times V \to T)$ is **universal for bilinearity** *if it is universal for $(\mathcal{F}, \mathcal{H})$, that is, if for every bilinear map $f: U \times V \to W$, there is a* unique *linear transformation $\tau: T \to W$ for which*

$$f = \tau \circ t$$

The map τ is called the **mediating morphism** *for f.* □

It seems that τ linear really is being imposed or assumed here by saying $\mathcal{H}$ is linear [correct]. But then all he really says here is that the pair (T,t) is universal for functions in set $\mathcal{F}$ IF you can find a unique and linear τ. So Roman is just repeating earlier material in this bilinear space context. Again, the above entire quote is just a definition of what you would have to show to show that pair (T,t) was universal for $\mathcal{F}$. Nothing at all has been proved or even claimed. We continue on some more:

Definition *Let U and V be vector spaces over a field F. Any universal pair $(T, t: U \times V \to T)$ for bilinearity is called a* **tensor product** *of U and V. The vector space T is denoted by $U \otimes V$ and sometimes referred to by itself as the tensor product. The map t is called the* **tensor map** *and the elements of $U \otimes V$ are called* **tensors**.

It is customary to use the symbol $\otimes$ to denote the image of any ordered pair (u, v) under the tensor map, that is,

$$u \otimes v = t(u, v)$$

for any $u \in U$ and $v \in V$. A tensor of the form $u \otimes v$ is said to be **decomposable**, *that is, the decomposable tensors are the images under the tensor map.* □

OK, this defines a tensor product *in terms of* (T,t) being a universal pair for the space $\mathfrak{F}$ of bilinear functions. So if you can somehow show that a particular (T,t) pair is *universal*, THEN you have a tensor product. So how are we to show such a thing if we are handed a candidate tensor product?

Since universal pairs are unique up to isomorphism, we may refer to "the" tensor product of vector spaces. Note also that the tensor product $\otimes$ is not a product in the sense of a binary operation on a set. In fact, even when $V = U$, the tensor product $u \otimes u$ is not in U, but rather in $U \otimes U$.

OK, we continue onto page 363

As we will see, there are other, more constructive ways to define the tensor product. Since we have adopted the universal pair definition, the other ways to define tensor product are, for us, constructions rather than definitions. Let us examine some of these constructions.

OK, I am all for "construction". Continue on:

Construction I: Intuitive but Not Coordinate Free

The universal property for bilinearity captures the essence of bilinearity and the tensor map is the most "general" bilinear function on $U \times V$. To see how this universality can be achieved in a constructive manner, let $\{e_i \mid i \in I\}$ be a basis for U and let $\{f_j \mid j \in J\}$ be a basis for V. Then a bilinear map t on $U \times V$ is uniquely determined by assigning arbitrary values to the "basis" pairs (e_i, f_j) and extending by bilinearity, that is, if $u = \sum \alpha_i e_i$ and $v = \sum \beta_j f_j$, then

$$t(u, v) = t\left(\sum \alpha_i e_i, \sum \beta_j f_j\right) = \sum \alpha_i \beta_j t(e_i, f_j)$$

Now, the tensor map t, being the most general bilinear map, must do this *and nothing more*. To achieve this goal, we define the tensor map t on the pairs (e_i, f_j) in such a way that the images $t(e_i, f_j)$ *do not interact*, and then extend by bilinearity.

OK, above he is writing

$u \otimes v = \Sigma_{ij}\, u_i\, v_j (e_i \otimes f_j)$ $\qquad u = \Sigma_i u_i e_i \qquad v = \Sigma_j\, v_j f_j$

which is something I write all the time. If you defined $u' = \Sigma_i u'_i e_i$ you could say

$(u + u') \otimes v = \Sigma_{ij}\, (u_i + u_i')\, v_j (e_i \otimes f_j) = u \otimes v + u' \otimes v$

So I agree, this "construction" using basis functions does produce bilinearity, either in terms of the operator ⊗ or since t(u+u',v) = t(u,v) + t(u',v) which is to say, in terms of a function t.

I agree that u ⊗ v and t(u,v) are each uniquely determined by this construction. He again states the idea of bilinear only in that "that is all there is".

In the above paragraph, he does not say that u_i are scalars or in some base field or anything like that. Perhaps they are in something like R where multiplication like $u_i v_j$ is at least defined (the destination space T is "an algebra".)

So the above construction defines t : U x V → T. Since he is saying U ⊗ V is a tensor product, it must be that the pair (W,t) is a universal pair, and he mentions that, but I see no argument for WHY (W,t) constructed as above would be a universal pair. Where is τ? Where is the proof that τ is unique and linear? [see below!] Let's hack a way on the next chunk,

In particular, for each ordered pair (e_i, f_j), we invent a new formal symbol, written $e_i \otimes f_j$, and define T to be the vector space with basis

$$\mathcal{D} = \{e_i \otimes f_j \mid i \in I, j \in J\}$$

The tensor map is defined by setting $t(e_i, f_j) = e_i \otimes f_j$ and extending by bilinearity. Thus,

$$t(u, v) = t\Big(\sum \alpha_i e_i, \sum \beta_j f_j\Big) = \sum \alpha_i \beta_j (e_i \otimes f_j)$$

Nothing new in that chunk, so keep going:

To see that the pair (T, t) is the tensor product of U and V, if $g: U \times V \to W$ is bilinear, the universality condition $g = \tau \circ t$ is equivalent to

$$\tau(e_i \otimes f_j) = g(e_i, f_j)$$

which does indeed uniquely define a *linear* map $\tau: T \to W$. Hence, (T, t) has the universal property for bilinearity and so we can write $T = U \otimes V$ and refer to t as the tensor map.

OK, he is finally making the claim. I agree with the above equation. But how does that equation show that the function τ is linear?? This is the point I keep missing. I have some kind of misconception right at this point that keeps causing trouble. Here is the triangle from above (with f in place of g)

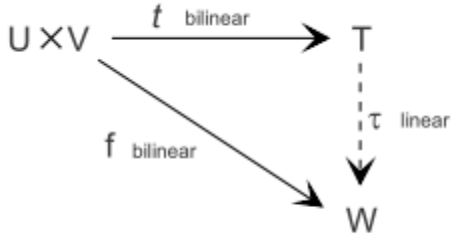


Figure 14.4

In my "book" showing that τ is a linear function requires showing that (at least)

$$\tau(e_i \otimes f_j + e_{i'} \otimes f_{j'}) = \tau(e_i \otimes f_j) + \tau(e_{i'} \otimes f_{j'})$$

At least now I know something about τ (above I knew nothing whatsoever). Again I see that the right side of this equation is the same as

$$g(e_i, f_j) + g(e_{i'}, f_{j'})$$

but as usual in all my earlier attempts, I see no way to show that

$$\tau(e_i \otimes f_j + e_{i'} \otimes f_{j'}) = g(e_i, f_j) + g(e_{i'}, f_{j'})$$

So this is the Kiss of Death in Roman's presentation, for me as a reader. I will try and fail as usual. For example,

$$\tau(e_i \otimes f_j + e_{i'} \otimes f_{j'}) = \tau [t(e_i, f_j) + t(e_{i'}, f_{j'})] = \text{what is the next step.}$$

The time has now come to pause in Roman until this "misconception" is cleared up. I had the same problem with Ganatra above. I will do some clips from his Problem and Solution on this matter, just to give myself another possible shot that maybe his "solution" really had a solution. $\varphi \; V \otimes W$

13.5 Review of Ganatra and Explanation of why τ is linear

Ganatra Problem Statement in HW 7: (this is Problem 3)

(a) There is a natural map

$$\phi : V \times W \to V \otimes W$$
$$(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \otimes \mathbf{w}.$$

Prove that this is a *bilinear* map, where *bilinear maps* were defined on HW6.

(b) In fact ϕ is the *universal bilinear map*, as we will see in this exercise. Namely, prove the following: if

$$T : V \times W \to X$$

is a *bilinear map*, in the sense of homework 6, prove that there exists a unique *linear map*

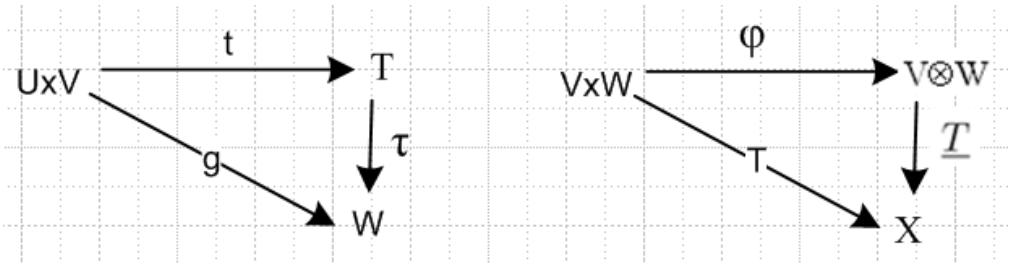
$$\underline{T} : V \otimes W \to X$$

such that $T = \underline{T} \circ \phi$. That is, T factors uniquely as

$$V \times W \xrightarrow{\phi} V \otimes W \xrightarrow{\underline{T}} X$$

where $\underline{T}$ is a linear map.

Here is a comparison of the Roman and Ganatra triangles,



So Ganatra's student has to show that $\underline{T}$ is linear (and maybe also unique). Now the homework assignment gives some "hints":

Hint: What does it mean for a map $\underline{T} : V \otimes W \to X$ to be a linear? Firstly, it means that on a formal sum

$$\sum_k a_k \mathbf{v}_k \otimes \mathbf{w}_k$$

$\underline{T}$ is additive and homogenous, so

$$\underline{T}(\sum_k a_k \mathbf{v}_k \otimes \mathbf{w}_k) = \sum_k a_k \underline{T}(\mathbf{v}_k \otimes \mathbf{w}_k).$$

It also means that $\underline{T}$ should not be affected by applying the relations (2); that is, $\underline{T}((\mathbf{v} + \mathbf{v}') \otimes \mathbf{w})$ should be equal to $\underline{T}(\mathbf{v} \otimes \mathbf{w} + \mathbf{v}' \otimes \mathbf{w})$ and so on.

OK, this really is not what I think of as being linear for τ. This is some whole new definition of linear that I do not understand! I will use τ and not $\underline{T}$. So his meaning of linear would be, for a simple case

$\tau(\, v_1 \otimes w_1 + v_2 \otimes w_2) = \tau(v_1 \otimes w_1 + v_2 \otimes w_2)$

so OK, apart from the two labels being the same, I guess that agrees with my definition of linear. So fine, now the student doing the homework is supposed to use this Hint to show that τ is linear.

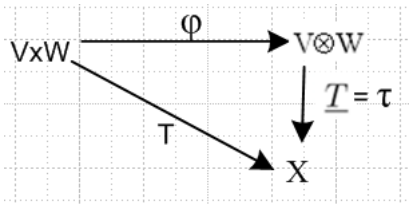
The homework assignment has nothing more to say, no more Hints. So I now turn to the homework solution handed out by some TA:

Ganatra Solution Statement for HW 7: (this is Problem 3)

The solution begins like so:

(b) We define $\underline{T}(v \otimes w) = T(v, w)$ for each $v \in V$ and each $w \in W$, and extend this definition to all of $V \otimes W$ linearly. The function $\underline{T}$ is well-defined because T is bilinear. (Could you write down the calculations that would verify this?). By construction, we have that $T = \underline{T} \circ \phi$. Therefore there exists at least one map $\underline{T}$ with the required properties.

Here again is the Ganatra triangle,



The solver proposes a *definition* of the function τ

τ (v⊗w) = T(v,w)

I like that, fine. We are then told to "*extend* this definition linearly to all of V⊗W." So I think τ is linear because it is constructed to be linear. So in his "extension" of the meaning of τ he would say

τ (v⊗w + v'⊗w') = T(v,w) + T(v',w')

In other words, τ is linear because we say it is linear. There is nothing to prove.

The triangle picture just shows τ, and we are supposed to construct a τ which makes the triangle commute. The commuting triangle tells us this, for all v and w:

v⊗w = φ(v,w)

x = τ (v⊗w)

x = T(v,w) bilinear

So it tells us that, for all v and w

$\tau\ (v \otimes w) = T(v,w)$

I guess the question is then: if we declare that τ is linear, does that contradict anything in the triangle? The triangle says nothing more than $\tau\ (v \otimes w) = T(v,w)$, so forcing τ to be linear does not contradict anything. We are *looking for* a linear and unique τ, so why not construct it in this manner

(1) $\tau\ (v \otimes w) = T(v,w)$
(2) τ is linear

So I guess we could ask: does this cause any triangle trouble in other ways? Well consider

$\tau\ (v \otimes w + v' \otimes w') = \tau(v,w) + \tau(v',w')$ // due to our declaration τ is linear

This then says

$\tau(\ v \otimes w + v' \otimes w'\ ) = x + x'$

Now we can write, since T is bilinear

$x + x' = T(v,w) + T(v',w')$

I cannot combine the T's in this form. But suppose we had a special case that

$x' = T(v',w)$ // no prime on w

Then we would have

$x + x' = T(v,w) + T(v',w) = T(v+v',w) = \tau\ ([v+v'] \otimes w) = \tau\ (\varphi(v+v',w))$

$= \tau\ [\ \varphi(v,w) + \varphi(v',w)] = \tau\ [\ \varphi(v,w)] + \tau\ [\ \varphi(v',w)] = x + x'$

The triangle itself tells us that

$\tau \circ \varphi = T$

Conclusion: Defining τ to be linear **"by fiat"** is consistent with the triangle diagram.

So OK, if we do this, why is τ unique? Suppose we had τ and τ' both linear. then

$\tau \circ \varphi = T$
$\tau' \circ \varphi = T$ $\quad (\tau - \tau') \circ \varphi = 0$

If this is true for ANY φ which defines $v \otimes w$, then I guess $\tau = \tau'$. Here is what the solver says:

Assume there were two linear maps $\underline{T}, \underline{T}'$ with $T = \underline{T} \circ \phi$ and $T = \underline{T}' \circ \phi$. Then $\underline{T}$ and $\underline{T}'$ would be equal on the image of ϕ. But this means that $\underline{T}$ and $\underline{T}'$ are equal on any pure tensor, and any element of $V \otimes W$ is a linear combination of pure tensors. Therefore $\underline{T} = \underline{T}'$, so there is a unique linear map with this property.

OK, I am getting happier now. You simply define τ by τ (v⊗w) = T(v,w) according the triangle, and you then declare that τ is linear, and this linear τ then satisfies the triangle and causes no problems, so you can do this declaration as part of the definition of τ. This is what Roman does not make clear.

13.6 Conclusions

1. The tensor product construction which I always make using the basis elements can be associated with a triangle diagram involving two bilinear functions and a linear function τ

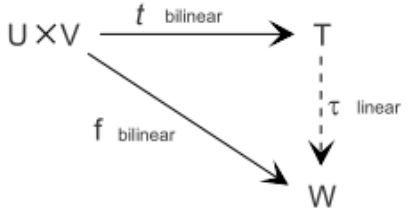


Figure 14.4

There is a trivial way to define τ as τ(u⊗v) = f(u,v), then you "extend" τ by declaring it to be linear, and this is then consistent with the picture. τ is unique as well, and therefore this picture describes (T,t) as a "universal pair". That in turn says that T = U ⊗ V is a "tensor product"!

2. There is no requirement to do "more". You *have* a tensor product.

3. The underlying category idea of the triangle is this: the set of bilinear maps t: U x V → T is more or less unique and completely well defined, up to isomorphism, so T can be referred to as U ⊗ V. For example, in the above triangle picture, if we consider another set of bilinear maps U x V → W where W is a different space from T, there is a 1:1 relationship between maps t: U x V → T and f: U x V → W as embodied in the unique linear function τ which is 1:1, known as a "mediating morphism" (morphism = arrow). In this sense, the set of maps t: U x V → T is "universal" and (T,t) form a "universal pair".

4. You could, however, if you wanted, continue on and define $[a \otimes b]_{ij} = a_i b_j$. This extra add-on fact does not conflict with the triangle or with the definition of tensor product *because* this add-on declaration is bilinear! That makes it consistent with the earlier stuff.

5. Therefore, an "outer product" is a special case of a "tensor product".

I will try to write up this conclusion more clearly. Ready to re-scan earlier sources maybe!