**Jad Nohra**, Private Note. Revised: October 4, 2016

# Covariance

pre-short

## I Hinterfragung I

**1.** Is there a correspondence (1-1) between m-lin maps and tensors? tensor products?

**2.** Tensors vs. tensor products.

**3.** Where is the proof that covar, contrav and mixed cover all possible 'kinds of tensors'?

**4.** The levi-civita pseudo-tensor is not a tensor, is a multilinear, or 'somehow' linear? This would answer [1].

**5.** How come we can add pseudo-tensor in 'tensor expressions', like the one expressing the cross product in terms of levi-civita?

**6.** What is 'wrong' with presenting tensors (exclusively) as m-dim arrays?

**7.** Why is variance not treated in lin-alg?

**8.** Is duality a (mathematical, algebraic) way to 'model' the extrinsic information of physical (geometric, coord-free) quantities?

**9.** Where exactly do 'partial derivatives' come in? Aren't they essentially unrelated to lin-alg and m-lin-alg?

**10.** Why exactly do we resort to 'coordinate basis vectors'? Are there alternatives? Is this mainly due to free us (automatically) from atlases? [@c10, (Chap4, p10), (Chap24, p10)]

## II Hinterfragung II

**11.** The 'tensor product' structural extension of 'cartesian product' is not (at least not obviously) a 'product of tensors', is 'basis product' or 'basis combination' a better name?

**12.** If the exterior product is a tensor, is then the 'operator of multiplication' a tensor? This does not seem to fit matrices being tensors, is the 'operator of multiplication of matrices' a tensor? Maybe index notation helps? In the sense that we write down the indices and keep the entries whose indices are to be taken as 'variables'?

Consider matrix multiplication in index notation for matrices $a, b$ as

$$a_{i,j} b_{j,k},$$

is the 'matrix multiplication tensor' then

$$(-)_{ij}(-)_{jk} \ ?$$

an 'object' obviously different from another like

$$(-)_{ij}(-)_{jk} \ ?$$

Does the above not contradict the usage of $\varepsilon$ as the index notation representative of the determinant 'tensor'? Maybe not since one can write the matrix multiplication as

$$1_j[(-)_{ij}(-)_{jk}]$$

where $1_j$ takes the place of the $\varepsilon$ symbol's function, in contrast to another 'matrix multiplication tensor':

$$2_j[(-)_{ij}(-)_{jk}].$$

All this in contrast with

$$\varepsilon_{ij} a_i b_j.$$

In fact this might just fit, with the determinant tensor being

$$\varepsilon_{ij}(-)_i(-)_j.$$

This is slowly leading us to the 'slot' interpretation, basically that fact that tensors are 'codings of algorithms (using vectors)'. Note that the 'problem' the slot notation solves is addressing the ambiguity between $\varepsilon$ds being part of the tensor, while $a$ being an 'argument'. the slots nevertheless look clumsy. What about

$$\varepsilon_{ij}{}^{\nearrow i}_{\searrow j}$$

## III Hinterfragung III

**13.** From [@c11, p2], one has the impression that a big amount of fuss is, just as the duality between vectors and linear functionals can be exploited to obtain a more-or-less canonical way to determine the coordinates of 'a' linear map, a tensor. Is this 'the' 'problem to solve'?. Finding a canonical basis for multi-linear maps? Is the solution the tensor product, which provides such a basis?

**14.** It seems that a path to the above is traced in [Nering], with bilinear forms being treated, which are simple forms of multi-linear maps. The relation between bilinear forms and tensors is per example shown in https://en.wikipedia.org/wiki/Bilinear_form, which nicely summarizes multiple concepts and ways to interpret multi-linear maps that we all plan to pin down. The relation is the following:

By the universal property of the tensor product, bilinear forms on $V$ are 1-1 with linear maps $V \otimes V \to K$. If $B$ is a bilinear form on $V$ the corresponding linear map is given by

$$v \otimes w \mapsto B(v, w).$$

The set of all linear maps $V \otimes V \to K$ is the dual space of $V \otimes V$, so bilinear forms may be thought of as elements of

$$(V \otimes V)^* \cong V^* \otimes V^*.$$

Likewise, symmetric bilinear forms may be thought of as elements of $\mathrm{Sym}^2(V^*)$ (the second symmetric power of $V^*$), and alternating bilinear forms as elements of $\sigma^2 V^*$ (the second exterior power of $V^*$).

So here we already have the 'inner product' and the 'determinant'. So we even get the scratch the differential forms, a concrete example using only alternating bilinear forms which can shed light and supports this Hinterfragung is this: How to identify $V \wedge V$ with the space of all alternating bilinear forms .

**15.** Also not that at least one author agrees that variance is an obfuscating and non-essential aspect of tensors. "A perhaps unconventional aspect of our approach is that for clarity we isolate the notion of covariance and contravariance (see Section 15.10) from our definition of a tensor. We do not view this as an essential part of the definition but a source of obfuscation." [@c12, sec.15-7].

**16.** In [@c12], we read the following: "In this chapter we generalize our earlier discussion of bilinear forms, which leads in a natural manner to the concepts of tensors and tensor products. While we are aware that our approach is not the most general possible (which is to say it is not the most abstract), we feel that it is more intuitive, and hence the best way to approach the subject for the first time. In fact, our treatment is essentially all that is ever needed by physicists, engineers and applied mathematicians. More general treatments are discussed in advanced courses on abstract algebra." Now this is commendable but also a little bit confused. Commendable because it uses exactly our approach in this Hinterfragung, but confused because it calls this an inferior approach. One can understand the 'property based', 'coord-free' approach, and that is not a problem. One misses a part of it, which is as essential when one goes down from hoity-toity to actual calculations. Both parts should be treated as a hole. Which direction one starts with is irrelevant.

**17.** It could very well be that [@c12, p462-465] is golden.

**18.** It seems that being isomorphic, a linear space and its dual are, given no extra choices, not distinguishable. The (seeming) necessity to distinguish them must come through a choice of basis that is not easy to change. Is this what happens in diff-geom, in the case per example of coordinate functions giving rise to a coordinate basis? What then distinguishes variance, is it again a choice of relation between basis and coordinates that is not easy to change, or not desirable to change in order to keep a certain form of a problem? This choice should be subject to an explicit example.

**19.** Given $E^3$ one can use angles as coordinates, in that sense is there an infinity of possible relations between basis and vectors through coordinates? If so, why only var and covar? Are these the only 'linear' ones? It is probably useful to find this out (V).

**20.** We fall into (IV) and then fall out of it.

**21.** We fall into (VI) tbc

# IV Bemrekungen I

**22.** After falling out of (V), we note a latent misconception, its fix being: (Multi-)Linear algebra and vectors are more abstract than one initially thinks they are. This 'concreteness' misconception is responsible for many troubles.

**23.** A clearer but more verbose rewording of [@c14, p.134] is:

Consider the act of looking at the space of linear functionals. Staring with $V$, we pass to (its dual-iso, requiring a choice of basis, linked using bi-orthogonality, viz. coordinate picking) $V^*$ and then to $V^{**}$. But by a simple evaluational (or seemingly, maybe also category theoretically, natural) isomorphism between $V$ and $V^{**}$ (which does not require any choice of basis to be expressed), we can stop the ascent to $V^{***}$ and create a loop containing only $V$ and $V^*$. We call this symmetry of ' 'is' the space of linear functionals on' a 'duality' since there is no way to choose the starting point (us being in a loop). A good notation expression of this is simply dropping the parenthesis in $\phi(\alpha)$, and using $\phi\alpha$ instead, where $\alpha \in V, \phi \in V^*$.

**24.** With the help of [@c14, p.136], we find a slightly better exposition that allows us to treat on the spot the task of (39), which we do in (IX.3). Do not forget Nering's application to small oscillations, to Vector calculus, and to Spectral Decomposition

**25.** This ends this section.

# V Erfassung I

**26.** This Erfassung seems to uncover the fact that parametrization is also a topic in linear spaces.

**27.** Consider two vector spaces $V, W$ over $\mathcal{F}$ of equal finite dimension $n$. It follows that $V \cong W$. Now given no extra information, how could one find out if $W$ is the dual of $V$, that is, that $W$ is actually a (hidden) $V^*$.

**28.** Does this question make sense at all? How do we distinguish two isomorphic linear spaces? Obviously, by whatever distinguishes them.

**29.** What distinguishes $V$ and $V^*$ is the way that they are algebraically, formally, defined, described, 'from the top'. By the mention of functionals as set theoretical mappings on $V$, with certain algebraic (formal) constraints and proving they exist (usually done explicitly for infinite dimensional spaces). If we 'forget' this information, is there any other way? A way 'from the bottom'?

**30.** Without mentioning bases, there is no way. We call a linear space equipped with a basis a ***basis space***. In the following, we assume that $V$ and $W$ are basis spaces. When there is no ambiguity, we refer to both $V$ and its basis as $V$, the elements of which being $V_i$.

**31.** There is a hidden choice being sneaked in when one says that a point $v$ (vector) in $V$ is expressed by $v = v_i V_i$. Why exactly this choice of mapping? Could we have not chosen other mappings $\mathcal{F}^n \mapsto V$? why not $v = 2v_i V_i$? Clearly, this is similar to general coordinate functions.

**32.** The formalities of $V$ disallow us to do anything else with points (hence also with $V_i$) other than take scalar multiplied sums of them. This already starkly (quite beautifully) limits the possible mappings. Consider for $n = 2$ the mapping $v = v_1^3 V_1 + v_2 V_2$, or more generally, $v = f_1(v_1, v_2)V_1 + f_2(v_1, v_2)V_2$. As long as $f$ are bijective (we do not need continuity here?), we should have a valid way of relating $v$ to $V_i$ spanning $V$. I suppose this is ignored because it can be easily treated away and related to our chosen 'canonical' mapping.

**33.** Let us then leave it at the choice $f_i(v_*) = v_i$, which looks suspiciously like the coordinate functions used to 'define' the dual basis. By $(x_*)$ we mean $(x_1, \ldots, x_n)$, $n$ being clear from the context.

**34.** Instead, let us look at what we can do about $V_i$. Can we not have $v = v_1(V_1 + V_2) + v_2 V_2$? Yes, but this is trivial, since it is equivalent to a non-canonical choice of $f_i$. In general, having $v = v_i g_i(V_*)$, what comes out of it? Can variance come out of it?

**35.** Let us take inspiration from the definition of dual basis as source for possible $g$ functions. First note that we will do something quite strange here. We will be working within $V$ totally unrelated to $W$. In contrast, in the definition of dual basis, one relates basis points from two isomorphic but otherwise unrelated spaces. How does a dual basis achieve this? Using functionals, this is done by choosing a property (of being a certain kind of coordinate function) to be imposed on $w_i$ based on the chosen $v_i$, in other works, workable only in basis spaces.

There is another construction, and that uses bilinear mappings. Bilinear mappings are allowed to swallow different vector spaces producing a third. For the dual basis, they must be chosen so as they turn out to be equivalent to coordinate functions, and hence they have their codomain as $\mathcal{F}$ and can be fully characterized by the Kronecker delta $\delta^i_j$, which is nothing other than canonical coordinate picking again. We say canonical because there is nothing wrong with using $2\delta^i_j$ or any other, as long as coordinates can be reversibly picked, viz. 'corresponded'. This kind of bilinear mapping causes the two bases to be called a 'biorthogonal system'.

**36.** Now back to $V$ and the basis relating functions $g$, using the same construction one uses for the dual basis. In other words we choose $g$ such that they produce $v'_i = g_i(v_*)$ such that $v_i$ and $v'_i$ are a biorthogonal system, using the obvious bilinear mapping, which unlike for the dual basis, has both domain spaces equal. Working in three dimensions, a 'geometric' characterization that is equivalent to the biorthogonal system is known: $v'_i = \dfrac{v_{i'} \times v_{i''}}{v_1 \cdot (v_2 \times v_3)}$, where $i' = (i + 1) \bmod 3$.

**37.** The question now is if the contravariance of points when using the 'canonical' functions of $v_i$ ($g = \mathrm{id}$) somehow becomes a covariance when switching to $v'_i$.

**38.** We fully treat this and obtain a positive answer in IX.1.

**39.** Having treated this, we turn back into [@c13] and study the chapters 9 and 11. Finally, they are now fully readable. We also finally have access to understanding the cotangent space. It becomes now important that we treat the 'example' at p.551 in full detail (also see (24)). We do that in IX.3.

**40.** In [@c13] p.550 we read: "We point out that these transformation laws are the origin of the terms "contravariant" and "covariant." This is because the ***components of a vector transform oppositely ("contravariant") to the basis vectors*** $e_i$, while the components of dual vectors transform the same as ("covariant") these basis vectors." The emphasized statement seems wrong at first if one is paying attention. We are talking about basis vectors, and not their representations. How can a vector (and not a representation) be changing through a matrix? This is answered in IX.2. The rest was already treated in IX.1.

# VI Erfassung II

**41.** It could very well be that [@c12, p462-465] is golden, we treat it here.

**42.** At this point, we have used our 'notation' to good use in sections IX.2 and IX.3. These two gave us a good grounding, and the second also gave us a solid understanding of 'toolification' and 'formal' abstract usage, also in relation to calculus. One would like to extend the kind of notation used there to also handle everything matrix notation handles, but this might be asking for too much at the moment. One could express inverses, but the summation notation might obfuscate their meaning and make the usual matrix manipulations look clumsy. In essence, this might be telling us that best is to learn to juggle switching of notation, each having its own focus area, dimming the lights on other aspects. Our original plan was to pass explicitly from coordinate based summation notation, to tensor notation to abstract notation. But this might be too big of a project. We could instead, since now the understanding is there, focus on standard notation. Per example, we already now fully understand the choice of index placement and the statement on [@c12, p.543] (note that the brackets are not inner product, but a symmetric way to express duality of evaluation, as explained at p.447):

Given a vector space $V$ with basis $\{e_i\}$ , we defined the dual space $V^*$ (with basis $\{w^i\}$) as the space of linear functionals on $V$. In other words, if $\phi = \sum_i \phi_i w^i \in V^*$ and $v = \sum_j v^j e_j \in V$, then

$$\phi(v) = \langle \phi, v\rangle = \langle \sum_i \phi_i w^i, \sum_i v^j e_j \rangle = \sum_{i,j} \phi_i v^j \delta^i_j$$
$$= \sum_i \phi_i v^i$$

The long-winded version of the above is the following: By duality, we can drop the evaluation parenthesis from statements like $\phi(v)$. This is in line with our notational concept that all linearity can be turned into a notation of simple distributive multiplication. expressed fully in terms of bases, we then have that

$$\phi(v) = \phi v = v\phi = [\phi_i w_i][v_i e_i].$$

Let us turn our attention to representations. To have $v$ invariant:

$$v = [\tfrac{v}{e}]e = [\tfrac{v}{e'}]e',$$

and since it is a 'multiplication' of two things, they both either have to be invariant, or they have to cancel each other's variances. They do, since basis vectors are covar and vector representations are contrav. This is expressed, switching to matrix notation for $[\frac{v}{e'}]e'$, by

$$Cv' = [BL^{-1}][Lv] = B[LL^{-1]}]v = Bv.$$

Having this, we start injecting variance into the summation notation and write:

$$\begin{cases} \phi & = [\frac{\phi}{w}]^i w_i \\ v & = [\frac{v}{e}]^i e_i \end{cases}$$

But we will be using indices to indicate variance a bit later. To ascend further to standard notation, we have to relate everything to a single basis (see IX.1 and IX.3). We can 'multiply' vectors with covectors by evaluation in this sense:

$$\begin{aligned} \phi v &= [\phi_i w_i][v_j e_j] \\ &= (\phi_1 w_1 + \phi_2 w_2 + \dots)(v_1 e_1 + v_2 e_2 + \dots) \\ &= w_1 e_1 \underbrace{\phi_1 e_1}_{1} + w_1 e_2 \underbrace{\phi_1 e_2}_{0} + \dots \end{aligned}$$

By duality of bases $\{e_i\}$ and $\{w_i\}$(bi-orthogonality) we then have

$$\begin{aligned} \phi v &= [\phi_i v_j w_i e_j] \\ &= \phi_i v_i \end{aligned}$$

Let us keep in mind that what follows seems to erase the abstraction of the difference between linear functionals and vectors, which is 'very bad', but there is no way around this. In the end, we 'do' allow interaction between vectors and covectors in the abstract world, and this has to happen at some point, that is, both isomorphic but not equal structures have to collapse in some concrete sense into one (the story might not be so nice in the infinite dimensional case). In any case, this brings us close to thinking in these terms: Starting with some common interaction where both 'axes' for vecs and covecs are 'in the absolute' unity, and the result of their interaction being also unity (let us call it the situation of concrete absolute unity, which is a way to treat non-absolute equivalences), we wish the keep the result of the interaction unity when we pass to other 'axes'. To have this work, whenever the vec axis is scaled up, the covec axis must scale down, this being expressed by $w = \frac{1}{e}$, while the multi-dimensionality being separated out by bi-orthgonality. Additionally, for vec representations, whenever the 'axes' scale, the representations have to scale inversely, and so we have two kinds of 'scaling to invariance'.

All of this is 'expressed' by the tensor notation all at once. And it becomes trivial to see why the indices are the way they are, that is:

$$\phi v = \langle \phi_i w^i, v^j e_j \rangle.$$

If we continue to insist on 'multiplication', we can agree that square brackets can reinterpret types. In summation notation the indices in $\phi_i w^i$ cancel out and we obtain and 'index-less thing', in this case a 'vector' which is of some dimension equal to $i$. so we reinterpret it as something with an index of $i$ as far as its interactions with other things is concerned. If we again adjust detail, we get the very trivial looking (and not wrong in the sense of notation reinterpretation and omission of detail):

$$\begin{aligned}\phi v &= [\phi_i w^i]^i [v^j e_j]_i \\ &= [\phi_i w^i v^j e_j] \\ &= \phi_i v^i.\end{aligned}$$

Note that in the first step, we started with two separate 'things' but then we decided to rebase the left one in terms of $\{e_i\}$ so that they can properly interact.

Note that we found [@c15] which could save us a lot of trouble, indicating we might not be able to 'solve' the notation problem in a satisfactory manner, confirming the need to also focus on standard notation 'as it is'.

**43.** In any case, what we proved here seems to be that we should now focus on the standard way of doing things, continuing with [@c12. p543].

# VII Remarks

**44.** Is the approach of 'tensors as slots' a syntactic/formalist one? [1]

**45.** Is this not 'enough'? Is modern mathematics overly concerned with rigor and can hence be bypassed until needed? What is the problem with a 'formalist' approach in the sense of 'it does in symbols what it does in what we are trying to model'? The problem is the danger of an inconsistent theory. But is not the price to pay too high for almost all intents and purposes? Should not the retranslation into FOL/set theory be the 'last thing on one's mind'? The question of 'existence' of set theoretic entities leading to a proof of consistence? How close is consistence to 'dynamical determinism' in the sense that no approach to a predicate should lead to a difference outcome (T,F), in other words no predicate should be mapped 'dynamically with possible proofs as paths' to two values, so 'functional' in nature? Is this formalist approach a fully valid one to shortcut learning and understanding, leaving the 'modern rigorous path' for later, possible to machines?

**46.** Rigor, my prison, my castle. One day, a man thought up the most scenic path to walk. It was a great path and everyone who walked it agreed wholeheartedly, it took them nice places, real places, useful places. The man kept walking until he fell into a deep hole and died. There was a hole in this most scenic path. How did the man not know this fact about his praised path? It was because the path passed through slightly high grass, a most pleasurable height if you are inclined to ask. Everyone who enjoyed walking paths was terrified, no one was prepared to die, no matter how scenic their walk was going to be. What to do? They grumbled. They grumbled for years until one day, one of them proposed a possible and quite simple solution. Everyone likes simple solutions and it was swiftly declared very clever. From where everyone stands, where there was obviously no hole, burn, burn any grass around you and let the fire spread. Each fire had to be started with a match, which was, unfortunately for them, the only kind of fire starting tool they had. Then all places with no grass are provably good and safe paths. Much grass stayed, some grass was in islands to which one had to jump first to put on fire, but that was dangerous, since one could jump into a hole. Everyone was happy nevertheless, for there was enough for treading all the next hundred years. A few thought, this was a sad state of affairs but they were quickly banned to another group forever deemed inferior. In the end, left was much more space with grass than without, and actually, only few holes were really hidden out of sight, even fewer that would kill you. The most scenic paths had to stay unvisited until the men changed their minds, as a group, or convinced by a free spirit yet to come, and that was going to take a long long time. Some places had tiny grassy spots that were not possible to prove safe unless one actually jumped all the spots, there was no shortcut, and some of these formed infinite paths that were actually safe, but not provably safe, since no one could jump

[1] Actually, is this not the approach of Weintraub that we first appreciated and then abandoned?

all the infinitely many spots. Ah well.

**47.** Implicit definitions are strictly superior. What is the set of all Holzfeller? There is no reasonable explicit definition. The implicit definition is obvious, correct and useful. A Holzfeller is one who cuts would, and the set of all Holzfeller is then the set of all who cut wood. I cannot now go on a mission and name them all. I cannot define the set by its elements: Jack, Jim, etc. I say strictly because for sets where explicit definitions work, implicit ones do as well. The set X is "Jack, Jim", then I can say that the set is the set of everyone who is Jack and everyone who is Jim is in X and no one else.

**48.** Let us think about the isomorphism of the space of functional and forget that it is 'dual', why not leave it at that? Why call it dual? It is because the double dual is canonically isomorphic to itself and no more?

- " The work of Frigyes Riesz and others in the early 1900's considered concrete examples, and they spoke about linear functionals without feeling any need to gather them into a structured set (dual space). An analogue is perhaps Weierstrass, who discussed the convergence of sequences of functions in the 1870's without using the notion of a function space with a norm or a topology." http://math.stackexchange.com/questions/165389
- "Aside from treating the earlier problem strictly as one of extending linear functionals, Hahn also formally introduced the notion of dual space (polare Raum) for the first time, noted that X is embedded in its second dual X" and defined reflexivity (regularitaet). Duality theory had reached adolescence." [The Hahn-Banach Theorem: The Life and Times] [On The Hahn-Banach Theorem].
- The style of the answer in http://math.stackexchange.com/questions/111371/ is what we want to emulate with notation of restriction by properties, which it seems is what missing from our current approach.

# VIII Babelisms

**49.** When a physicist says 'tensor', he probably means 'tensor field'. [2]

**50.** When one says co(ntra)var tensor, one means tensor with co(ntra)var components.

# IX Treatments

## IX.1 Mathematical Rag: Elementary Covariance Through Uncanonical Vector Representation.

### IX.1.1 Motivation

**51.** Parametrization is a topic of geometry. What about parametrization within finite dimensional linear (vector) spaces? It is not a topic because it is trivial. However, we think it does deserve a short look.

**52.** Let us work within a linear space $V$ of dimension $n$, over $\mathcal{F}$. Given a chosen basis $B = \{b_i\}$, one chooses a standard bijective representation function

$$r_B : \mathcal{F}^n \to V,$$

such that it produces a point[3] $a \in V$ from a point $\underline{a} \in \mathcal{F}^n$ in the following manner:

$$a = \underline{a}_1 b_1 + \cdots + \underline{a}_n b_n.$$

Using a loose summation notation this is:

$$a = r(a_*) = \underline{a}_i b_i,$$

where $(x_*)$ stands for $(x_1, \ldots, x_n)$.

**53.** This is a useful and obvious choice, but a choice nevertheless. Without going into details of multilinear algebra, one can take any map $\phi$ that preserves linear independence and meaningfulness of representation

[2] https://en.wikipedia.org/wiki/Tensor#As_multidimensional_arrays

[3] We say point instead of vector to put ourselves into a more geometric mindset for the reason we explained.

of a point with respect to a basis (in short, bijective multilinear), and still have a valid unique representation of $a$:

$$a = \underline{a}_i \phi_i(b_*), \tag{1}$$

**54.** We can view all such $\phi$ as 'parametrizations' of $V$, and Equation 1 as a 'uncanonical vector representation'. This allows us to consider not only choices of basis, but also choices of parametrization. This is worthwhile because it gives a peek at the concept 'variance' (i.e covariance and contravariance), already within the realm of elementary linear algebra, without explicitly talking about duality, linear functionals or tensors, indicating that in some sense, the concept is quite elementary.

### IX.1.2 Variance

**55.** In the abstract setting, one considers a point to be independent of its representation. Even though we forego the abstract treatment, we can still express this by remembering that any property that makes sense in the abstract setting can be mirrored in the coordinate setting.

**56.** Without repeating known details pertaining to change of basis in linear algebra, without mentioning choices of basis when they do not matter for our purpose, working directly with matrices, let us consider the following:

- A point $a$ which we identify with its representation in some basis.
- A basis $B$ which we identify with a matrix, such that
$$a = B\underline{a}_i.$$
- Another basis $C$, such that
$$a = C\underline{\alpha}_i,$$
$$C = BL,$$
$L$ being the change-of-basis matrix.

**57.** Our coordinate based version of representation independence requires the following:

$$Ia = B\underline{a}_i = C\underline{\alpha}_i,$$

we have:

$$\begin{cases} \underline{\alpha_i} & = L^{-1}\underline{a_i} \\ C & = BL \end{cases} \tag{2}$$

**58.** We can see that coordinates and bases change in 'opposite' ways. The first is called contravariant and the second covariant. In what follows, we make use of what we called paramterization in Equation 1 to show that one can choose it such that something different than the above happens.

**59.** Let us denote our parametrization function as $F$, a function acting multilinearly on a basis (its matrix). Instead of searching for a useful parametrization, we emulate the construction of dual basis and chose:

$$F : X \mapsto X^{-T},$$

that is, the inverse of the transpose, which is equal to the transpose of the inverse, and hence the use of the common shorthand where the order does not matter.

**60.** The motivation behind this choice is biorthogonality, roughly, a strong form of linear independence. In one dimension ($\mathbb{R}^1$), the only way for two points to be orthogonal is for one of them to be zero, by usual multiplication. In general, orthgonality takes a more relative face: two points (in $\mathbb{R}^n$) can be orthogonal relative to each other, without the constraint that at least one of them is orthogonal to all other points. Biorthogonality of a system of two sets of points simply says that each point from one set is orthogonal to all except one from the other set. This is a minimal and natural way to characterize a mapping that guarantees preservation of linear independence.

With basis points relating to matrix columns, and given the rules of matrix multiplication, expressing this requires taking the transpose of one of the matrices and writing:

$$XX^T = I.$$

This is reflected in $F$ which makes it so that a matrix is mapped to its biorthogonal.

**61.** Let us pickup our original change of basis example, but parametrizing through $F$. This means that we take out uncanonical representation in equation 1 and choose $\phi$ to be represented by our specified $F$. We then have

$$a = F(B)\underline{a'}_i = B^{-T}\underline{a'}_i = F(C)\underline{\alpha'}_i = (BL)^{-T}\underline{\alpha'}_i.$$

**62.** In contrast to equations 2 we now have:

$$\begin{cases} \underline{\alpha'_i} & = L^T \underline{a'_i} \\ C & = BL \end{cases}. \tag{3}$$

In this situation, the point is said to change covariantly, like the basis.

**63.** It is known that variance is invisible the concerned change-of-basis matrices are orthogonal, this is clear since in this case, $X^{-1} = X^T$ and the two cases become indistinguishable.

**64.** Can we obtain point covariance in a simpler way, foregoing the example of dual basis? The obvious idea that comes to mind is to use

$$F(X) = X^{-1}.$$

This does not work out as nicely, and one ends up with

$$\underline{\alpha'_i} = BLB^{-1}\underline{a'_i}.$$

This beautifully illustrates another 'reason' for the transposition in the biorthogonality matrix and motivates the proper and common abstract approach, which lifts us from the flattened state of affairs of working in matrices instead of linear maps. Nevertheless, it is useful to have seen a more 'computational' reason for it.

**65.** More ambitiously, can we go the other way round and, requiring that we obtain the simplest covariant relation

$$\underline{\alpha'_i} = L\underline{a'_i},$$

and working out what that imposes on $F$? Here, we meet another beautiful motivation for mutlilinear algebra proper. We used a transpose, but is that a 'linear' thing? Is that a linear map proper? Even a short look tells us that something is wrong.

If we say that a transpose is a linear map proper, we should be able to to find a matrix $T$ such that

$$X^T = TX.$$

If one examines what matrix multiplication constrains us with, one finds that there is no such $T$ in general. Nevertheless, transposition can be seen as a linear map proper if one flattens the transposed matrix $(n \times n)$ into a $(n^2 \times 1)$ matrix. But then, this 'flattened' matrix cannot be used to multiply the coordinates. This again motivates linear and multilinear algebra proper. In tensor notation, the transpose can be written as

$$M^T = \sum_{i=1}^{n} (e_i^T \otimes I_n) S_{n,n} (I_n \otimes M) \left( \sum_{i=1}^{n} (e_j \otimes e_j) \right) e_i^T.$$ [4]

### IX.1.3 Nonlinear Parametrization

In our uncanonical vector representation, we constrained the map $\phi$ to be (multi)linear. What is wrong with a nonlinear continuous bijective one? This is almost never treated because it is not considered a useful avenue. Here is a simple reason for that. Let us take a one-dimensional vector space $V$ with a vector $a$, a basis $B = \{b\}$ and another $B' = \{b'\}$. Let us also take a non-linear function $f : V \to V$ acting in the following way: $f(x) = (\text{Rep}_B(x))^3 b$, clearly, $f$ is bijective so at least when applied to all vectors in $V$ one gets another sets that spans $V$. For contrast, Let $l$ be some linear bijection, $l : V \to V$. Let $y = l(a)$. Now $y$ is related to $a$, and this relation transfers nicely to their coordinates: $y = l(a) = l(a_1 b) = a_1 l(b)$. If $b' = l(b)$, then we can use $y = a_1 b'$. The non-usefulness comes from the fact that the coordinates 'break', become 'useless', when we consider the non-linear function. $z = f(a) = f(a_1 b) \neq a_1 f(b)$. If $b' = f(b)$, it gives us nothing we can ***use***, since $z \neq a_1 b'$. In short a non-linear change of basis takes us out of linear algebra and the crucial identification between vector and coordinates becomes unworkable in the linear algebraic context.

[4] http://math.stackexchange.com/questions/1143614, http://math.stackexchange.com/questions/125862

## IX.2 Vector Representations Transform Oppositely to Basis Vectors

**66.** This is a treatment of the question 40. This is possible to do quite elegantly once on realizes that one should use summation notation because it explicitly expresses the 'implementation' of matrix multiplication instead of hiding it. We add to this the introduction of a special notation for vector representation.

**67.** We shall work within a 'loose' but obvious summation notation that does not differentiate between upper and lower indices.

**68.** Notation:

| | |
|---|---|
| $[], ()$ | Separators |
| $x_j$ | A summation notation 'array' with $j$ elements. |
| $x_{j(i)}$ | The $i$-th element of $x_j$. |
| $x_{j(*)}$ | Any (all) element(s) of $x_j$. |
| $x_{jk}$ | A summation notation 'array' with $j \times k$ elements. |
| $\lfloor {}^i x \rfloor$ | Matrix layed out as such: ${}^1x, {}^2x, \ldots$ |
| $[\frac{x}{b}]$ | The representation of the vector $x$ relative to the basis whose elements are $b_k$. |
| $\mathbf{1}_j$ | The array of standard $\mathbb{R}^j$ basis vectors, ordered in the obvious way. |

**69.** With $j = k = q$ being the dimension of the vector space, let $l_{jk}$ be the representation of a 'forward' change of basis map, such that:

$$[\tfrac{v}{b}]_j = l_{jk}[\tfrac{v}{c}]_k \tag{4}$$

Since this holds for any $v$, it holds for all vectors of the basis $c$, so we have:

$$[\tfrac{c_q}{b}]_j = l_{jk}[\tfrac{c_q}{c}]_k$$

But

$$[\tfrac{c_q}{c}]_k = \mathbf{1}_k,$$

so passing to the individual elements, this means that

$$l_{jk(*,q)} = [\tfrac{c_q}{b}]_{j(*)} \tag{5}$$

In other words, the elements of $l$ are the elements of $[\frac{c}{b}]$. But by definition of vector representation, we have:

$$c_{q(i)} = [\tfrac{c_i}{b}]_k b_k, \tag{6}$$

where running over all $i$ exhausts all of $[\frac{c}{b}]$. Hence, dropping all indices, we find that:

$$\begin{cases} [\frac{v}{b}] = & l[\frac{v}{c}] \\ c = & lb \end{cases} \tag{7}$$

which demonstrates the opposite behavior.

**70.** In words, since the basis vectors change one way, the components have to offset that to keep representing the same vectors.

**71.** Given the above, the positions on indices and the whole statement itself from [@c13, p545] finally makes total sense:

$$v = v^i e_i.$$

The representation is contrav ($v^i$), the basis itself is covar ($e_i$), their summed multiplication produces an invar vector ($v$).

## IX.3 The Gradient is Covariant

**72.** This treatment is based on [@c14, p.136]. We have the following:

| | |
|---|---|
| $x_*$ | Variables |
| $w = f(x_*)$ | A function |
| $\nabla w = \left(\frac{\partial w}{\partial x_*}\right)$ | The gradient of $w$ |
| $\alpha, \beta$ | Two bases |
| $\xi$ | A vector such that $[\frac{\xi}{\alpha}]_i = x_i$ and $[\frac{\xi}{\beta}]_i = y_i$ |

**73.** Conceptually, we would like to show this very easily by explaining that a gradient is not 'directly' expressed w.r.t a basis, in effect mirroring exactly our thinking in (IX.1). We don't know why this is not the way things are explicitly explained since it almost trivializes the topic. In case of the gradient, the 'path' from gradient to its expression w.r.t a basis passes through coordinates, the rest is detail, and the exact relation leads to the gradient being covar.

**74.** First let us note that we want to relate coordinates to 'velocities', and we can only do this using calculus. So we will have to go through calculus to determine the variance of the gradient. A first step towards this is using partial derivatives to describe the change-of-basis. This works out beautifully because of the linearity of the process. As in equation 4, we have

$$[\tfrac{\xi}{\beta}]_j = l_{jk}[\tfrac{\xi}{\alpha}]_k$$

By examination, we clearly have that

$$\frac{\partial\left([\frac{\xi}{\beta}]_{j(p)}\right)}{\partial\left([\frac{\xi}{\alpha}]_{k(q)}\right)} = l_{jk}(p,q) \tag{8}$$

We can continue with this notation as a sort of 'anti-summation', in the following way, removing some detail:

$$\frac{[\frac{\xi}{\beta}]_j}{[\frac{\xi}{\alpha}]_k} = l_{jk}$$

Removing even more detail, we obtain, (thanks to the separation powers of calculus):

$$\frac{\partial[\frac{\xi}{\beta}]}{\partial[\frac{\xi}{\alpha}]} = l \tag{9}$$

So then, all in all we have:

$$[\tfrac{\xi}{\beta}]_j = \left[\frac{[\frac{\xi}{\beta}]_j}{[\frac{\xi}{\alpha}]_k}\right]_{jk} [\tfrac{\xi}{\alpha}]_k \tag{10}$$

**75.** It is both silly and enlightening that we cannot simply work with the gradient directly w.r.t the bases by saying:

$$\frac{\partial w}{\partial \alpha} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial \alpha}$$

Clearly, this is gibberish since the partial derivatives are defined in calculus in terms of 'variables', not abstract vectors. But it seems this approach although illegal, can be seen as a shortcut notation to something that is valid. But until then, we need to pass through coordinates to relate general quantities (e.g, nonlinear functions,gradients, etc.) to linear ones (e.g, bases)

**76.** In any case, by the chain rule, we have that:

$$\frac{\partial w}{\partial x_j} = \sum_i \frac{\partial w}{\partial y_j}\frac{\partial y_i}{\partial x_j}$$

Adjusting detail we have,

$$\left[\partial^w_{x_j}\right]_j = \left[\frac{[\frac{\xi}{\beta}]_j}{[\frac{\xi}{\alpha}]_k}\right]_{jk} \left[\partial^w_{y_k}\right]_k \tag{11}$$

Clearly this varies oppositely to equation 10, and since vectors are contrav, gradients are then covar.

**77.** Note: A more elegant path is to consider the function as a scalar field, which is already a coordinate-free equivalent to 'function'. This fixes the lack of elegance due to talking about coordinate-free gradients, while describing them in terms of coordinate-based functions.

## IX.4 Reading the Dual Basis Independence Proof

**78.** We need to prove the linear independence of a dual basis without using coordinates, but only the definitional property that

$$w^i(e_j) = \delta^i_j.$$

**79.** We first note that we initially know nothing about the zero functional except that it is equal to any functional once that is multiplied by the scalar zero. In other words, we do not know that the zero functional is such that it produces zero when applied to any vector. Let us denote the zero functional by $\zeta$, it is defined by

$$\zeta = 0\phi \quad \Big|\ \forall \phi \in V^*$$

**80.** The subtly notational transformation that bridges us both ways to evaluational properties of functionals is set theoretical. If two set theoretical functions are equal, it means nothing else other than that (expansionally) their actions are equal on all elements.

$$\begin{aligned} &\phi = \varphi \\ \Leftrightarrow\quad &\phi(v) = \varphi(v) \quad \forall v \end{aligned}$$

**81.** This formal/syntactic/evaluational bridge is what we use as we proceed to show that the zero functional 'is' (evaluates to) zero when applied to any vector. For two functionals $\phi, \varphi$ we have that

$$\begin{array}{r|c} [\zeta] = [0\phi] & \\ [\zeta](v) = [0\phi](v) & \forall \phi \in V^*, \forall v \in V \\ [\zeta](v) = [\phi](0v) & " \\ [\zeta](v) = 0[\phi(v)] & " \\ \zeta(v) = 0 & \forall v \in V \end{array}$$

**82.** Finally, to prove the linear independence of $w^i$, we need to show that, in any way that we choose our $a_i$, we have that $a_i w^i = \zeta$ implies $a_i = 0$, let us do that:

$$\begin{array}{r|c} [a_i w^i] = [\zeta] & \\ [a_i w^i](v) = [\zeta](v) = 0 & \forall v \in V \\ [a_i w^i](e_j) = 0 & \text{for each } j \in \dim(V) \\ a_i \delta^i_j = 0 & " \\ a_j = 0 & " \end{array}$$

Note that we mean in line 3 that line 2 holds for all $v$, including for all $e_j$. Similarly, in line 4 we use 'for each' instead of 'for all' to emphasize that conceptually, we repeat the procedure for each of the $j$ (exhausting them), which is logically obvious, but a distinction that the mind naturally makes.

# X Sources

**83.** http://physics.stackexchange.com/questions/87775/is-the-covariance-or-contravariance-of-vectors-tensors-something-that-can-be-vi

1. Tangent spaces as paths and not as vandal arrows.
2. Tensors as slots, pointing to [@c10]