An intuitive way to think tensors

Irregular Rhomboid

December 16, 2024

Multilinear algebra is a rich mathematical theory which is very important in physics, and yet often not very intuitive. This problem is made worse by the difficulty of fixing a notation that is both consistent and non-ambiguous. The goal of this document is to present a visual notation for tensors that renders without ambiguity the difference between contravariant and covariant indices.

Note: This document was initially written in French for an audience of freshman Math and Physics students who had already gone through a linear algebra course, and so this text was intended as synthesis of what they had learned so far that would insist on the fact that covariant and contravariant coordinates are fundamentally different objects – a common point of confusion among students. Likewise, the last section assumes the reader has already gone through a first year real analysis course. Further edits to this English version may introduce more basic definitions.

1 Vectors and covectors

1.1 Vectors

Let X be a finite dimensional vector space on a field $\mathbb{K}$, with dimension N. The elements of X are called *vectors*. If we have on hand a *basis* of X, i.e. a set of N linearly independent vectors $e = \{e_i\}_{i=1}^N$, we can express any vector $x \in X$ by a unique linear combination of basis elements,

$$x = \sum_{i=1}^{N} \alpha^i e_i,$$

where $\alpha^1, \ldots, \alpha^N \in \mathbb{K}$ are unique scalar coefficients. It is common to use Einstein notation for a more concise expression:

$$x = \alpha^i e_i.$$

In our notation, we will denote such vectors by *column vectors* :

$$[x]^e = \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix},$$

where $[x]^e$ denotes the *representation* of the vector x in the basis e. In the rest of this document, we will use this notation in parallel to Einstein notation.

At this point, it is useful to make a few remarks on the typographical conventions used in this text. By default, scalars will be denoted by greek letters, vectors by lowercase latin letters ($x, y, z, \dots$) and covectors will be the same as vectors, but with an additional $*$ (x^*, $y^*, \dots$), except for the letters i, j, k, l, m and n which denote indices.

1.2 Covectors

To the vector space X, we can associate its *dual space* X^*, whose elements are *linear forms* $y^* : X \to \mathbb{K}$. It is important to note that these objects are of a fundamentally different type from that of the elements of X. They are *functions* that take a vector as argument and return a scalar.

By linearity, each $y^* \in X^*$ is completely identified by the values it takes on basis elements :

$$y^*(x) = y^*(\alpha^i e_i) = \alpha^i y^*(e_i).$$

This suggests in particular the possibility of constructing a basis e^* of X^* associated to the basis e of X, by choosing basis elements e^{*i} such that $e^{*i}(e_i) = 1$ and $e^{*i}(e_j)$ for $j \neq i$. More concisely,

$$e^{*i}(e_j) = \delta^i_j,$$

with δ^i_j the Kronecker delta. A complete set of these elements can be computed in order to form a *dual basis* of X^*. Thus, any linear form $y^* \in X^*$ can be uniquely expressed as a linear combination of the elements e^{*i}:

$$y^* = \beta_i e^{*i},$$

where $\beta_i = y^*(e_i)$, and for $x = \alpha^j e_j \in X$, we have

$$\begin{aligned} y^*(x) &= \beta_i e^{*i}(x), \\ &= \beta_i e^{*i}(\alpha^j e_j), \\ &= \alpha^j \beta_i e^{*i}(e_j), \\ &= \alpha^j \beta_i \delta^i_j = \alpha^i \beta_i. \end{aligned}$$

In other words, applying y^* to the covector x is simply expressed by the sum of the products of their respective basis coefficients:

$$y^*(x) = \beta_i \alpha^i. \tag{1}$$

We will call the elements of the dual space *covectors*, and we will denote them in the following manner as *row vectors*:

$$[y^*]^{e^*} = [y^*]^1_e = \begin{pmatrix} \beta_1 & \dots & \beta_N \end{pmatrix},$$

where $[y^*]^1_e$ denotes the fact that we are representing y^* as a linear form from X with the basis e to $\mathbb{K}$ with the basis $\{1\}$, where '1' is the multiplicative identity of $\mathbb{K}$.

We can introduce an operation between covectors and vectors, which simply corresponds to equation (1), namely the *product* of a row vector (on the left) by a column vectors (on the right)

$$\begin{pmatrix} \beta_1 & \dots & \beta_N \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} = \beta_i \alpha^i. \tag{2}$$

2 Order 2 Tensors

2.1 Matrices

Given a linear map $f : X \to X$, we can completely characterize it with respect to the basis e by applying f to each basis vector, and decomposing the result in the basis e. We thereby for each e_i get coefficients γ_i^j such that

$$f(e_i) = \gamma_i^j e_j.$$

Then, for each $x = \alpha^i e_i$, we get

$$f(x) = f(\alpha^i e_i) = \alpha^i f(e_i) = \alpha^i \gamma_i^j e_j.$$

To the map $f : X \to X$, we can automatically associate the dual map $f^* : X^* \to X^*$, which is defined so that for $y^* \in X^*$ et $x \in X$,

$$(f^*(y^*))(x) = y^*(f(x)),$$

or, with respect to the bases e and e^*:

$$\begin{aligned}
(f^*(y^*))(x) &= y^*(f(x)), \\
&= y^*(\alpha^i \gamma_i^j e_j), \\
&= \alpha^i \gamma_i^j y^*(e_j), \\
&= \alpha^i \gamma_i^j \beta_j,
\end{aligned} \tag{3}$$

and also

$$f^*(y^*) = \beta_j \gamma_i^j e^{*i}.$$

We see that the coefficients γ_i^j appear in the representation of f as well as in the representation of f^*. Moreover, equation (3) suggests a third interpretation for γ_i^j, as the representation of a *bilinear form* $F : X^* \times X \to \mathbb{K}$, which takes as arguments a covector and a vector and returns a scalar. The maps f and f^* are in reality just facets of F. If we partially apply F to a vector x, we obtain

$$\begin{aligned}
F_x &: X^* \to \mathbb{K} \\
&\quad y^* \mapsto F_x(y^*) = F(y^*, x),
\end{aligned}$$

and if we partially apply F to $y^* \in X^*$,

$$\begin{aligned}
F_{y^*} &: X \to \mathbb{K} \\
&\quad x \mapsto F_{y^*}(x) = F(y^*, x).
\end{aligned}$$

By inspecting the type of these objects, we immediately see that $F_{y^*} = f^*(y^*)$ and F_x is an element of the *bidual space* of X, which corresponds to the vector $f(x) \in X$, by the Riesz isomorphism theorem. We are therefore tempted to write "$F_x \sim f(x)$", in the sense that these two objects are equivalent by isomorphism.

We will represent F with respect to the bases e and e^* by the *matrix*

$$[F]^{e\otimes e^*} = \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \cdots & \gamma^N_N \end{pmatrix}.$$

However, we will interpret this object in two equivalent ways. Namely, as a *covector of vectors*

$$[F]^{e\otimes e^*} = \begin{pmatrix} \begin{pmatrix} \gamma^1_1 \\ \vdots \\ \gamma^N_1 \end{pmatrix} & \cdots & \begin{pmatrix} \gamma^1_N \\ \vdots \\ \gamma^N_N \end{pmatrix} \end{pmatrix},$$

or as a *vector de covectors*

$$[F]^{e\otimes e^*} = \begin{pmatrix} \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_N \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \gamma^N_1 & \cdots & \gamma^N_N \end{pmatrix} \end{pmatrix}.$$

We just extended our definitions of vectors and covectors, since we now admit that the elements of these objects no longer have to be scalars, but potentially other vectors/covectors. We will provide a formal definition of these objects in a few sections.

These notations let us naturally extend the product defined in (2) to covector-matrix products and matrix-vector products. For example, the classical definition of the product of a matrix by vector immediately comes out by using the "covector of vectors" form:

$$\begin{aligned} \begin{pmatrix} \begin{pmatrix} \gamma^1_1 \\ \vdots \\ \gamma^N_1 \end{pmatrix} & \cdots & \begin{pmatrix} \gamma^1_N \\ \vdots \\ \gamma^N_N \end{pmatrix} \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} &= \alpha^1 \begin{pmatrix} \gamma^1_1 \\ \vdots \\ \gamma^N_1 \end{pmatrix} + \cdots + \alpha^N \begin{pmatrix} \gamma^1_N \\ \vdots \\ \gamma^N_N \end{pmatrix}, \\ &= \begin{pmatrix} \alpha^i \gamma^1_i \\ \vdots \\ \alpha^i \gamma^N_i \end{pmatrix} = \alpha^i \gamma^j_i e_j. \end{aligned}$$

And the product of a covector by a matrix comes out similarly with the "vector of covectors" form,

$$\begin{aligned} \begin{pmatrix} \beta_1 & \ldots & \beta_N \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_N \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \gamma^N_1 & \cdots & \gamma^N_N \end{pmatrix} \end{pmatrix} &= \beta_1 \begin{pmatrix} \gamma^1_1 & \ldots & \gamma^1_N \end{pmatrix} + \cdots + \beta_N \begin{pmatrix} \gamma^N_1 & \ldots & \gamma^N_N \end{pmatrix}, \\ &= \begin{pmatrix} \beta_j \gamma^j_1 & \ldots & \beta_j \gamma^j_N \end{pmatrix} = \beta_j \gamma^j_i e^{*i}. \end{aligned}$$

By inspecting the result of these operations, we see that they correspond respectively to calculating $f(x)$ and $f^*(y^*)$. Similarly, we can compute $F(y^*, x)$ by

$$\begin{pmatrix} \beta_1 & \ldots & \beta_N \end{pmatrix} \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \cdots & \gamma^N_N \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} = \gamma^j_i \beta_j \alpha^i,$$

where to obtain the final result, we can either compute the covector-matrix product first, or the matrix vector product.

2.2 Tensor Products

In the preceding section, we defined the representation of F without expliciting the basis in which it was expressed. We will now formally construct this object.

For a vector $x = \alpha^i e_i$ and a covector $y^* = \beta_j e^{*j}$, we define the *tensor product* of x (on the left) and y^* (on the right) by

$$x \otimes y^* = \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} \begin{pmatrix} \beta_1 & \dots & \beta_N \end{pmatrix} = \begin{pmatrix} \alpha^1\beta_1 & \dots & \alpha^1\beta_N \\ \vdots & \ddots & \vdots \\ \alpha^N\beta_1 & \dots & \alpha^N\beta_N \end{pmatrix}.$$

In other words, contrary to the product of y^* with x, we now have a matrix whose coefficients are given by $\alpha^i\beta_j$. We can construct a basis of the space of matrices with the tensor products of basis vectors/covectors. For instance

$$e_i \otimes e^{*j} = \begin{pmatrix} \delta_i^1 \\ \vdots \\ \delta_i^N \end{pmatrix} \begin{pmatrix} \delta_1^j & \dots & \delta_N^j \end{pmatrix} = \begin{pmatrix} \delta_i^1\delta_1^j & \dots & \delta_i^1\delta_N^j \\ \vdots & \ddots & \vdots \\ \delta_i^N\delta_1^j & \dots & \delta_i^N\delta_N^j \end{pmatrix}.$$

We can easily convince ourselves using the definition of the Kronecker delta, that the only non-zero coefficient of this matrix is that of the i-th row and j-th column. Thus, a matrix of coefficients γ^i_j can be expressed as a linear combination of tensor products $e_i \otimes e^{*j}$:

$$\begin{pmatrix} \gamma_1^1 & \dots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \dots & \gamma_N^N \end{pmatrix} = \gamma^i_j(e_i \otimes e^{*j}),$$

where for a given bilinear form F, the coefficients γ^i_j are defined as

$$\gamma^i_j = F(e^{*i}, e_j).$$

We can now express the product of two matrices in our notation. Given two matrices with coefficients γ^i_j and η^i_j, we have

$$
\begin{aligned}
&\begin{pmatrix} \gamma^1 & \cdots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \cdots & \gamma^N_N \end{pmatrix}\begin{pmatrix} \eta^1_1 & \cdots & \eta^1_N \\ \vdots & \ddots & \vdots \\ \eta^N_1 & \cdots & \eta^N_N \end{pmatrix},\\
=&\left(\begin{pmatrix} \gamma^1_1 \\ \vdots \\ \gamma^N_1 \end{pmatrix} \quad \cdots \quad \begin{pmatrix} \gamma^1_N \\ \vdots \\ \gamma^N_N \end{pmatrix}\right)\begin{pmatrix} \begin{pmatrix} \eta^1_1 & \cdots & \eta^1_N \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \eta^N_1 & \cdots & \eta^N_N \end{pmatrix} \end{pmatrix},\\
=&\begin{pmatrix} \gamma^1_1 \\ \vdots \\ \gamma^N_1 \end{pmatrix}\begin{pmatrix} \eta^1_1 & \cdots & \eta^1_N \end{pmatrix} + \cdots + \begin{pmatrix} \gamma^1_N \\ \vdots \\ \gamma^N_N \end{pmatrix}\begin{pmatrix} \eta^N_1 & \cdots & \eta^N_N \end{pmatrix},\\
=&\begin{pmatrix} \gamma^1_1\eta^1_1 & \cdots & \gamma^1_1\eta^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1\eta^1_1 & \cdots & \gamma^N_1\eta^1_N \end{pmatrix} + \cdots + \begin{pmatrix} \gamma^1_N\eta^N_1 & \cdots & \gamma^1_N\eta^N_N \\ \vdots & \ddots & \vdots \\ \gamma^N_N\eta^N_1 & \cdots & \gamma^N_N\eta^N_N \end{pmatrix},\\
=&\begin{pmatrix} \gamma^1_k\eta^k_1 & \cdots & \gamma^1_k\eta^k_N \\ \vdots & \ddots & \vdots \\ \gamma^N_k\eta^k_1 & \cdots & \gamma^N_k\eta^k_N \end{pmatrix}.
\end{aligned}
$$

And we see that the element θ^i_j of the product of the two matrices is given by

$$\theta^i_j = \gamma^i_k\eta^k_j.$$

In the above calculation, we first applied the covector-vector product rule, then the vector-covector rule. The result is exactly the same if we first apply the vector-covector product and then the covector-vector product. The verification of this assertion is left as an exercice.

2.3 Covariance and contravariance

Until now, we have only worked with a single basis to represent our vectors, covectors and matrices. Let us now look at how these representations are transformed when switching to a new basis.

Let $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_N\}$ be another basis of X and its dual basis $\varepsilon^* = \{\varepsilon^{*1}, \ldots, \varepsilon^{*N}\}$. The bases e and ε are related with one another by an invertible linear transformation. Let γ^i_j and η^i_j be the coefficients of this transformation and of its inverse, such that

$$\varepsilon_i = \gamma^j_i e_j,$$

and

$$e_i = \eta^j_i \varepsilon_j.$$

Since η^i_j represents the inverse of γ^i_j (and reciprocally), the composition of these two operations, that is, the product of the corresponding matrices must be equal to the identity matrix, i.e.

$$\gamma^i_k\eta^k_j = \delta^i_j.$$

We can visualize this by

$$\begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix} \begin{pmatrix} \eta_1^1 & \cdots & \eta_N^1 \\ \vdots & \ddots & \vdots \\ \eta_1^N & \cdots & \eta_N^N \end{pmatrix} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix},$$

where this last matrix must be interpreted as a diagonal matrix with ones on the diagonal.

Actually, if we are willing to further bend the rules, we can also visualize the change of basis, if we write the bases e and ε as row vectors whose "components" are basis vectors,

$$\begin{pmatrix} \varepsilon_1 & \ldots & \varepsilon_N \end{pmatrix} = \begin{pmatrix} e_1 & \ldots & e_N \end{pmatrix} \begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix}.$$

We have here covectors whose components are basis elements. While these objects are not exactly of the same type of those we have manipulated until now, we can go back to something more familiar by replacing the basis elements e_i and ε_i by their representation in a given basis. In the basis e, for example, we have

$$\begin{aligned} \begin{pmatrix} \begin{pmatrix} \gamma_1^1 \\ \vdots \\ \gamma_1^N \end{pmatrix} & \cdots & \begin{pmatrix} \gamma_N^1 \\ \vdots \\ \gamma_N^N \end{pmatrix} \end{pmatrix} &= \begin{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} & \cdots & \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix}, \\ &= \begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix}. \end{aligned}$$

We trivially recover the fact that the columns of the change of basis matrix correspond to the representation of the new basis vectors in the old basis. Reciprocally, every invertible matrix corresponds to a change of basis matrix, and its columns correspond to the new basis vectors.

Let us examine the relation between the coefficients of a vector in these two bases. Let $x = \alpha^i e_i = \tilde{\alpha}^i \varepsilon_i$. We have

$$\begin{aligned} x &= \alpha^i e_i, \\ &= \alpha^i (\eta_i^j \varepsilon_j), \\ &= (\alpha^i \eta_i^j) \varepsilon_j, \\ &= \tilde{\alpha}^j \varepsilon_j, \end{aligned}$$

and we see that if we move from the basis e to the basis ε, the basis coefficients α^i are multiplied by the *inverse* transformation η_j^i to obtain the new basis coefficients $\tilde{\alpha}^j = \eta_i^j \alpha^i$. For this reason, we say that the superscript indices are *contravariant* indices, because they change *in opposition* to the primal basis.

Similarly, we can look at the effect of the change of basis in the dual space. First, from the identity $\varepsilon^{*i}(\varepsilon_j) = \delta_j^i$, we can deduce that $\varepsilon^{*i} = \eta_j^i e^{*j}$. Indeed, suppose that $\varepsilon^{*i} = \theta_j^i e^{*j}$, then

$$\begin{aligned} \varepsilon^{*i}(\varepsilon_j) &= (e^{*k} \theta_k^i)(\gamma_j^k e_k), \\ &= e^{*k} (\theta_k^i \gamma_j^k) e_k, \\ &= \delta_j^i, \end{aligned}$$

where to go to the last line, we need that $\theta^i_k \gamma^k_j = \delta^i_j$, in other words, $\theta^i_j = \eta^i_j$ (the inverse of an invertible linear transformation is unique). By the same argument, we also have that $e^{*i} = \gamma^i_j \varepsilon^{*j}$. We can also visualize this change of basis by representing the bases e^* and ε^* with column vectors:

$$\begin{pmatrix} \varepsilon^{*1} \\ \vdots \\ \varepsilon^{*N} \end{pmatrix} = \begin{pmatrix} \eta^1_1 & \dots & \eta^1_N \\ \vdots & \ddots & \vdots \\ \eta^N_1 & \dots & \eta^N_N \end{pmatrix} \begin{pmatrix} e^{*1} \\ \vdots \\ e^{*N} \end{pmatrix}.$$

Having established the relation between the dual bases, it will not be surprising to find that for $y^* = \beta_i e^{*i} = \tilde{\beta}_i \varepsilon^{*i}$, we have

$$\begin{aligned} y^* &= \beta_i e^{*i}, \\ &= \beta_i (\gamma^i_j \varepsilon^{*j}), \\ &= (\beta_i \gamma^i_j) \varepsilon^{*j}, \\ &= \tilde{\beta}_j \varepsilon^{*j}. \end{aligned}$$

The coefficients of a covector therefore change by the same transformation as the primal basis vectors. For this reason, they called *covariant*, because they change *with* the primal basis.

We can also look at how the coefficients of a matrix change. Let F be a bilinear form whose coefficients with respect to the bases e and ε are respectively θ^i_j and $\tilde{\theta}^i_j$, that is,

$$F = \theta^i_j \, e_i \otimes e^{*j} = \tilde{\theta}^i_j \, \varepsilon_i \otimes \varepsilon^{*j}.$$

We can apply the map F to x and y^* to obtain

$$\begin{aligned} F(y^*, x) = \theta^i_j \alpha^j \beta_i \; &= \; \tilde{\theta}^k_\ell \tilde{\alpha}^\ell \tilde{\beta}_k, \\ \Leftrightarrow \theta^i_j (\gamma^j_\ell \tilde{\alpha}^\ell)(\eta^k_i \tilde{\beta}_k) \; &= \; \tilde{\theta}^k_\ell \tilde{\alpha}^\ell \tilde{\beta}_k, \\ \Leftrightarrow (\theta^i_j \gamma^j_\ell \eta^k_i) \tilde{\alpha}^\ell \tilde{\beta}_k \; &= \; \tilde{\theta}^k_\ell \tilde{\alpha}^\ell \tilde{\beta}_k, \end{aligned}$$

where in the second line, we have used the fact $\alpha^j = \gamma^j_\ell \tilde{\alpha}^\ell$ and $\beta_i = \eta^k_i \tilde{\beta}_k$. We therefore obtain that,

$$\tilde{\theta}^k_\ell = \theta^i_j \gamma^j_\ell \eta^k_i.$$

Thus, the contravariant index of θ^i_j changes with the inverse transformation η^k_i and the covariant index changes with the direct transformation γ^j_ℓ. We can visualize this coordinate change in our notation in the following way,

$$\begin{pmatrix} \tilde{\theta}^1_1 & \dots & \tilde{\theta}^1_N \\ \vdots & \ddots & \vdots \\ \tilde{\theta}^N_1 & \dots & \tilde{\theta}^N_N \end{pmatrix} = \begin{pmatrix} \gamma^1_1 & \dots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \dots & \gamma^N_N \end{pmatrix} \begin{pmatrix} \theta^1_1 & \dots & \theta^1_N \\ \vdots & \ddots & \vdots \\ \theta^N_1 & \dots & \theta^N_N \end{pmatrix} \begin{pmatrix} \eta^1_1 & \dots & \eta^1_N \\ \vdots & \ddots & \vdots \\ \eta^N_1 & \dots & \eta^N_N \end{pmatrix}.$$

We now see that the matrices θ^i_j and $\tilde{\theta}^i_j$ are related by a similarity transform. Reciprocally, every pair of similar matrices corresponds to the representations of the same bilinear form with respect to different bases. This justifies considering similar matrices as equivalent, in the sense that they are the representation of the same object.

2.4 Metrics

We now arrive to the subject of spaces equipped with a *metric*, i.e. a symmetric positive definite bilinear product that takes two vectors and returns a scalar. The general form of such an object, with respect to two vectors $x = \alpha^i e_i$ and $y = \beta^i e_i$ is given by a sum of the form

$$\langle x, y\rangle_g = g_{ij}\alpha^i\beta^j,$$

where the coefficients g_{ij} represent what we will call a 2-*covariant tensor*, because it has two covariant indices. Based on the same principles that let us construct a basis for matrices, we can construct a basis for these new objects by defining the tensor product of two covectors y_1^* and y_2^* as

$$\begin{aligned}(y_1^* \otimes y_2^*) :&X \times X \to \mathbb{K}\\ &(x_1, x_2) \mapsto y_1^*(x_1)y_2^*(x_2).\end{aligned}$$

If $y_1^* = \beta_i e^{*i}$, $y_2^* = \beta'_i e^{*i}$ and $x_1 = \alpha^i e_i$, $x_2 = \alpha'^i e_i$, we can write the above equation as

$$(y_1^* \otimes y_2^*)(x_1, x_2) = (\beta_i\alpha^i)(\beta'_j\alpha'^j) = (\beta_i\beta'_j)\alpha^i\alpha'^j.$$

The tensor product of two covectors is therefore a 2-covariant tensor, whose coefficients are $\beta_i\beta'_j$. It is clear that products of basis covectors $e^{*i} \otimes e^{*j}$ form a basis for these objects, which finally lets us write the metric (or any 2-covariant tensor) as

$$g = g_{ij}e^{*i} \otimes e^{*j},$$

where the coefficients g_{ij} are defined as

$$g_{ij} = g(e_i, e_j).$$

In our notation, we will naturally represent such objects as *covectors of covectors* :

$$g = \Big(\begin{pmatrix}g_{11} & \dots & g_{1N}\end{pmatrix} \;\dots\; \begin{pmatrix}g_{N1} & \dots & g_{NN}\end{pmatrix}\Big).$$

A brief comment on the ordering of indices: The left index corresponds to the index of the outermost covector, while the right index corresponds the index of the innermost covector. We will keep to this convention to represent higher order tensors later down the line.

Just like for matrices, we can interpret the result of the partial application of this object to one vector. Let $x = \alpha^i e_i$. The tensor g partially applied with x as its first argument is an object of type $g(x, -) : X \to \mathbb{K}$, i.e. a linear form. It is therefore a covector whose components are $g_{ij}\alpha^i$. We can visualize this operation in our notation:

$$\begin{aligned}&\Big(\begin{pmatrix}g_{11} & \dots & g_{1N}\end{pmatrix} \;\dots\; \begin{pmatrix}g_{N1} & \dots & g_{NN}\end{pmatrix}\Big)\begin{pmatrix}\alpha^1\\ \vdots\\ \alpha^N\end{pmatrix},\\ &=\begin{pmatrix}g_{11} & \dots & g_{1N}\end{pmatrix}\alpha^1 + \dots + \begin{pmatrix}g_{N1} & \dots & g_{NN}\end{pmatrix}\alpha^N,\\ &=\begin{pmatrix}g_{i1}\alpha^i & \dots & g_{iN}\alpha^i\end{pmatrix},\end{aligned}$$

where to step to the second line, we used the covector-vector product rule.

Multipliying a vector by the covariant metric produces a covector. We can do the inverse operation with a 2-*contravariant tensor* g^{ij}, where the contravariant coefficients g^{ij} (which should not be confused with the covariant coefficients g_{ij}) are such that

$$g^{ik}g_{kj} = \delta^i_j.$$

Just like we defined the basis of covariant 2-tensors, we can define a basis for contravariant tensors with the tensor product of two vectors,

$$x_1 \otimes x_2 = (\alpha^i e_i) \otimes (\alpha'^j e_j) = \alpha^i \alpha'^j (e_i \otimes e_j).$$

By reflexivity, we can interpret these objects as bilinear forms taking two covectors as argument. Each 2-contravariant tensor can therefore be written as

$$h = h^{ij} e_i \otimes e_j,$$

where the coefficientes h^{ij} are defined by

$$h^{ij} = h(e^{*i}, e^{*j}).$$

We will represent these objects in our notation by *vectors of vectors*:

$$\begin{pmatrix} \begin{pmatrix} h^{11} \\ \vdots \\ h^{1N} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} h^{N1} \\ \vdots \\ h^{NN} \end{pmatrix} \end{pmatrix}.$$

The partial application of such an object to a covector can be computed analogously to the product of a covariant 2-tensor by a vector. It is however more interesting to observe that for a 2-covariant tensor g_{ij} and a 2-contravariant tensor h^{ij},

$$\begin{aligned} &\begin{pmatrix} \begin{pmatrix} g_{11} & \cdots & g_{1N} \end{pmatrix} & \cdots & \begin{pmatrix} g_{N1} & \cdots & g_{NN} \end{pmatrix} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} h^{11} \\ \vdots \\ h^{1N} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} h^{N1} \\ \vdots \\ h^{NN} \end{pmatrix} \end{pmatrix}, \\ &= \begin{pmatrix} g_{11} & \cdots & g_{1N} \end{pmatrix} \begin{pmatrix} h^{11} \\ \vdots \\ h^{N1} \end{pmatrix} + \cdots + \begin{pmatrix} g_{N1} & \cdots & g_{NN} \end{pmatrix} \begin{pmatrix} h^{1N} \\ \vdots \\ h^{NN} \end{pmatrix}, \\ &= g_{ij} h^{ij}. \end{aligned}$$

In particular, we can check[1] that for two vectors $x_1 = \alpha^i e_i$ and $x_2 = \alpha'^i e_i$, the following identity is satisfied

$$(gx_1)x_2 = g(x_1 \otimes x_2) = g(x_1, x_2).$$

[1]This is left as an exercise.

Thus, partially applying g to x_1, then applying the resulting linear form to x_2 is equivalent to multiplying g by the tensor product of x_1 and x_2 and amounts to evaluating g on x_1 and x_2.

3 Higher Order Tensors

3.1 Formal Definition

Let us recap the objects we encountered so far:

Object	Basis	Components
scalar	1	α
vector $x \in X$	e_i	α^i
covector $y^* \in X^*$	e^{*i}	β_i
2-covariant tensor	$e^{*i} \otimes e^{*j}$	η_{ij}
matrix	$e_i \otimes e^{*j}$	γ^i_j
2-contravariant tensor	$e_i \otimes e_j$	θ^{ij}

Each of these objects can be interpreted as a function taking a certain number of vectors and covectors as arguments, and returns a scalar, while being linear with respect to each of its arguments. For example, a covector is a linear map $X \to \mathbb{K}$, a vector corresponds to a linear form $X^* \to \mathbb{K}$ by Riesz isomorphism, a matrix corresponds to a bilinear form taking a vector and a covector as arguments, ... As for scalars, we can interpret them as "functions" taking no arguments and returning a constant scalar.

Based on these considerations, we now define a *tensor* of shape (p, q) as a multilinear form

$$T : (X^*)^p \times (X)^q \to \mathbb{K},$$

which takes as arguments p covectors and q vectors. The *order* r of a tensor is its number of indices, i.e. $r = p + q$. A basis for these objects can be obtained with tensor products of the form

$$e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{*j_1} \otimes \cdots \otimes e^{*j_q},$$

and each tensor can be expressed in this basis by

$$T = T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{*j_1} \otimes \cdots \otimes e^{*j_q},$$

where

$$T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} = T(e^{*i_1}, \ldots, e^{*i_p}, e_{j_1}, \ldots, e_{j_q}).$$

When performing a change of basis given by $\varepsilon_i = \gamma^j_i e_j$ and $e_i = \eta^j_i \varepsilon_j$, the transformation rules seen previously for vectors, covectors and matrices generalize to

$$\tilde{T}^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} = T^{k_1,\ldots,k_p}_{\ell_1,\ldots,\ell_p} \eta^{i_1}_{k_1} \cdots \eta^{i_q}_{k_q} \gamma^{\ell_1}_{j_1} \cdots \gamma^{\ell_q}_{j_q}. \tag{4}$$

Each contravariant index changes with the inverse transform η^i_j and each covariant index changes with the direct transform γ^i_j.

We can also use the covariant or contravariant metric to convert a contravariant index to a covariant index, and vice-versa. For example,

$$U^{i_2,\ldots,i_p}_{k,j_1,\ldots,j_q} = g_{i_1,k} T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q}.$$

3.2 Recursive Definition

In our notation, we can represent arbitrary tensors as combinations of row vectors and column vectors. We can thereby define *recursively* a tensor of order r and shape (p, q) as one of the following alternatives:

- a scalar τ $(r = 0)$,
- a column vector
$$\begin{pmatrix} C^1 \\ \vdots \\ C^N \end{pmatrix},$$
whose components C^i are tensors of order $r-1$ and shape $(p-1, q)$,
- a row vector
$$\begin{pmatrix} L_1 & \dots & L_N \end{pmatrix},$$
whose components L_i are tensors of order $r-1$ and shape $(p, q-1)$.

It is important to note that within this notation, the same tensor can be represented in multiple different ways, when it has at least one covariant index and one contravariant index. As explained in the section on matrices, we consider these representations to be equivalent, and omit from specifying whether such a tensor is a vector of covectors or a covector of vectors when it is not necessary.

Here are a few examples of higher order tensors:

$$T = T^i_{jk} e_i \otimes e^{*j} \otimes e^{*k} = \begin{pmatrix} \begin{pmatrix} T^1_{11} & \dots & T^1_{1N} \\ \vdots & \ddots & \vdots \\ T^N_{11} & \dots & T^N_{1N} \end{pmatrix} & \dots & \begin{pmatrix} T^1_{N1} & \dots & T^1_{NN} \\ \vdots & \ddots & \vdots \\ T^N_{N1} & \dots & T^N_{NN} \end{pmatrix} \end{pmatrix},$$

$$U = U^{ij}_{kl} e_i \otimes e_j \otimes e^{*k} \otimes e^{*l} = \begin{pmatrix} \begin{pmatrix} U^{11}_{11} & \dots & U^{11}_{1N} \\ \vdots & \ddots & \vdots \\ U^{1N}_{11} & \dots & U^{1N}_{1N} \end{pmatrix} & \dots & \begin{pmatrix} U^{11}_{N1} & \dots & U^{11}_{NN} \\ \vdots & \ddots & \vdots \\ U^{1N}_{N1} & \dots & U^{1N}_{NN} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} U^{N1}_{11} & \dots & U^{N1}_{1N} \\ \vdots & \ddots & \vdots \\ U^{NN}_{11} & \dots & U^{NN}_{1N} \end{pmatrix} & \dots & \begin{pmatrix} U^{N1}_{N1} & \dots & U^{N1}_{NN} \\ \vdots & \ddots & \vdots \\ U^{NN}_{N1} & \dots & U^{NN}_{NN} \end{pmatrix} \end{pmatrix}.$$

Unfortunately, our notation quickly becomes cumbersome when representing tensors of large order. Nonetheless, it allows in principle to represent any tensor in a purely bidimensional manner, unlike most confused attempts to represent e.g. tensors of order 3 as three dimensional arrays, which have the problem of making ambiguous whether the third dimension should be interpreted as a covariant or contravariant index.

3.3 Generalized Tensor Product

We have also seen a few different rules to compute the product of two tensors. We can actually gather them all into a single generalized tensor product. Let us start by defining it with respect to our notation.

Let T and U be two arbitrary tensors. The *product* $T * U$ is a tensor defined recursively by

scalar-scalar :

$$\tau\sigma,$$

if $T = \tau$ and $U = \sigma$ are scalars

scalar-vector :

$$\tau * \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} = \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} * \tau = \begin{pmatrix} \tau * U^1 \\ \vdots \\ \tau * U^N \end{pmatrix},$$

if $T = \tau$ is a scalar, and U is a column vector.

scalar-covector :

$$\tau * \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} = \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} * \tau = \begin{pmatrix} \tau * U_1 & \dots & \tau * U_N \end{pmatrix},$$

if $T = \tau$ is a scalar and U is a row vector.

covector-vector :

$$\begin{pmatrix} T_1 & \dots & T_N \end{pmatrix} * \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} = T_i * U^i,$$

if T is a row vector and U is a column vector.

vector-covector :

$$\begin{pmatrix} T^1 \\ \vdots \\ T^N \end{pmatrix} * \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} = \begin{pmatrix} T^1 * U_1 & \dots & T^1 * U_N \\ \vdots & \ddots & \vdots \\ T^N * U_1 & \dots & T^N * U_N \end{pmatrix},$$

if T is a column vector and U is a row vector. The result can be interpreted arbitrarily as a vector of covectors or a covector of vectors.

covector-covector :

$$\begin{aligned} &\begin{pmatrix} T_1 & \dots & T_N \end{pmatrix} * \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} \\ &= \begin{pmatrix} T_1 * \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} & \dots & T_N * \begin{pmatrix} U_1 & \dots & U_N \end{pmatrix} \end{pmatrix}, \end{aligned}$$

if T and U are both row vectors.

vector-vector :

$$\begin{pmatrix} T^1 \\ \vdots \\ T^N \end{pmatrix} * \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} = \begin{pmatrix} T^1 * \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} \\ \vdots \\ T^N * \begin{pmatrix} U^1 \\ \vdots \\ U^N \end{pmatrix} \end{pmatrix},$$

if T and U are both column vectors.

Remarkably, the result of this product is independent of the way tensors are represented as combinations of vectors and covectors. Let us illustrate this by revisiting the matrix-vector product

$$\begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \\ \vdots & \ddots & \vdots \\ \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix} * \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix} \end{pmatrix} * \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix},$$

$$= \begin{pmatrix} \begin{pmatrix} \gamma_1^1 & \cdots & \gamma_N^1 \end{pmatrix} * \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \gamma_1^N & \cdots & \gamma_N^N \end{pmatrix} * \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \gamma_j^1 \alpha^j \\ \vdots \\ \gamma_j^N \alpha^j \end{pmatrix} = \gamma_j^i \alpha^j e_i.$$

In the end, we recover the correct expression for the coefficients of the matrix-vector product. We leave as an exercise the care of checking the calculations for the covector-matrix product.

Let us finish by defining the product in a purely algebraic fashion. Let T and U be tensors with respective shapes (p, q) and (r, s). If the computation of $T * U$ does not require the covector-vector rule, the result is of shape $(p + r, q + s)$. Otherwise, we will have to use the covector-vector rule $t = \min\{q, r\}$ times, which decrements each time the number of contravariant and covariant indices of the result, whose shape is then $(p + r - t, q + s - t)$. This last formula is in fact always valid, and we can therefore express the product $V = T*U$ as a tensor of shape $(p + r - t, q + s - t)$ whose coefficients are given by

$$V^{i_1,\ldots,i_{p+r-t}}_{j_1,\ldots,j_{q+s-t}} = T^{i_1,\ldots,i_p}_{k_1,\ldots,k_t,j_1,\ldots,j_{q-t}} U^{k_1,\ldots,k_t,i_{p+1},\ldots,i_{p+r-t}}_{j_{q-t+1},\ldots,j_{q-t+s}}.$$

3.4 Other operations on tensors

The product defined in the previous section is elegant, but does not let us express some important tensor operations. For instance, swapping two covariant (or contravariant) indices

$$U^{i_1,\ldots,i_p}_{j_1,j_2,\ldots,j_q} = T^{i_1,\ldots,i_p}_{j_2,j_1,\ldots,j_q},$$

cannot be expressed in our notation. That being said, some operations like the trace of a matrix

$$\gamma_i^i,$$

can be expressed in the following manner by multiplying the matrix on the left and right by appropriate tensors

$$\delta_{ik} \gamma_j^i \delta^{j\ell},$$

where δ_{ik} and $\delta^{j\ell}$ should respectively be interpreted as the purely covariant/contravariant equivalent of the Kronecker delta δ^i_j. Visually, these correspond to

$$\begin{pmatrix}\begin{pmatrix}1 & \cdots & 0\end{pmatrix} & \cdots & \begin{pmatrix}0 & \cdots & 1\end{pmatrix}\end{pmatrix}\begin{pmatrix}\gamma^1_1 & \cdots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \cdots & \gamma^N_N\end{pmatrix}\begin{pmatrix}\begin{pmatrix}1\\ \vdots \\ 0\end{pmatrix} \\ \vdots \\ \begin{pmatrix}0 \\ \vdots \\ 1\end{pmatrix}\end{pmatrix}.$$

We leave checking that this expression evaluates to the trace γ^i_i as an exercise.

This last example, although elegant, also illustrates the limits of our notation compared with the purely symbolic notation, which can express complex operations with economy thanks to Einstein summation. The visual notation developed in this document is mostly useful to manipulate tensors of order up to 2 and is more cumbersome than not beyond that.

4 Application : the correct way to think derivatives

To close this document, we will use tensors as defined previously to clarify the notion of the derivative of a map between two vector spaces. We will start by revisiting the derivative of a map $f : \mathbb{R} \to \mathbb{R}$. It is defined as the following limit, if it exists

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

If we unpack the formal definition of the limit, this boils down to saying that for all $\varepsilon > 0$, there is some $\delta >$ such that if $0 < |x - x_0| < \delta$, then

$$\left| \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \right| < \varepsilon,$$

that is,

$$\frac{|f(x) - f(x_0) - f'(x_0)(x - x_0)|}{|x - x_0|} < \varepsilon, \tag{5}$$

where the expression in the numerator corresponds to the difference between $f(x)$ and its first order Taylor approximation around x_0.

Suppose now that we wished to extend this definition to a map $f : X \to Y$, where X and Y are now vector spaces equipped with their respective norms $||.||_X$ and $||.||_Y$. For equation (4) to make sense in this context, we need to replace the absolute values with the appropriate norms:

$$\frac{||f(x) - f(x_0) - f'(x_0)(x - x_0)||_Y}{||x - x_0||_X} < \varepsilon, \tag{6}$$

but what should be the type of "$f'(x_0)$" in this context? If we wish to preserve the idea of taking the first order Taylor approximation, this must be a linear map taking $x - x_0 \in X$ as argument and returning an element of Y (otherwise the expression in the numerator would not make sense). The derivative $f'(x_0)$ must therefore be a *linear map from X to Y*. By manipulating the equation above, we obtain

$$||f(x) - f(x_0) - f'(x_0)(x - x_0)||_Y < \varepsilon ||x - x_0||_X,$$

so for x sufficiently close to x_0, we have that

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

We then recover the first order Taylor approximation.

4.1 Gradient

Let us put this definition into context by an example. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable multivariate function. The derivative of f at x_0 is a linear map $f'(x_0) : \mathbb{R}^n \to \mathbb{R}$. In other words, $f'(x_0)$ is a *covector*, such that

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0),$$

but freshman real analysis tells us that

$$f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0).$$

The derivative $f'(x_0)$ is in fact the *gradient* of f! We conclude from all of this that the *gradient is a covector*, and not a vector as is commonly confused.

We can then compute the coefficients of the gradient with respect to some basis. Given a primal basis $\{e_i\}_{i=1}^N$ and its associated dual basis $\{e^{*i}\}_{i=1}^N$, we have $\nabla f(x_0) = \beta_i e^{*i}$, where $\beta_i = \nabla f(x_0) e_i$ and moreover, as a special case of equation (6)

$$\begin{aligned}
&\lim_{h\to 0} \frac{|f(x_0 + he_i) - f(x_0) - \nabla f(x_0)(x_0 + he_i - x_0)|}{||x_0 + he_i - x_0||} = 0, \\
\Leftrightarrow &\lim_{h\to 0} \frac{|f(x_0 + he_i) - f(x_0) - \nabla f(x_0) he_i|}{||he_i||} = 0, \\
\Leftrightarrow &\lim_{h\to 0} \frac{|f(x_0 + he_i) - f(x_0) - \beta_i h|}{||he_i||} = 0,
\end{aligned}$$

but this last equation is nothing more than the definition of the derivative at zero of the function $\eta_i : \mathbb{R} \to \mathbb{R}; h \mapsto f(x_0 + he_i)$, in other words

$$\nabla f(x_0) e_i = \beta_i = \frac{d\eta_i}{dh}(0).$$

In the special case that $\{e_i\}_{i=1}^N$ is the canonical basis, we have

$$\begin{aligned}
\beta_i &= \left[\frac{d}{dh} f(x_0 + he_i)\right](0) \\
&= \lim_{h\to 0} \frac{f(x_0 + he_i)}{h}, \\
&= \lim_{\Delta x^i \to 0} \frac{f(x_0 + \Delta x^i e_i)}{\Delta x^i}, \\
&= \frac{\partial f}{\partial x^i}(x_0).
\end{aligned}$$

In other words, the components of the gradient of f with respect to the canonical basis are the *partial derivatives* of f.

More generally, we will call the derivative

$$\left[\frac{d}{dh} f(x_0 + hv)\right](0),$$

the *directional derivative* of f at x_0 in the direction $v \in X$. By the above discussion, it is clear that partial derivatives are just a special case of the directional derivative.

4.2 Jacobian

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable. The derivative of f at $x_0 \in \mathbb{R}^n$ must now be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. If we take a basis $e = \{e_i\}_{i=1}^n$ of $\mathbb{R}^n$ and a basis $\{\epsilon_j\}_{j=1}^m$ of $\mathbb{R}^m$, we can now represent f' in the usual way

$$f'(x_0) = \gamma^i_j \epsilon_i \otimes e^{*j},$$

where the components γ^i_j are characterized by

$$f'(x_0) e_j = \gamma^i_j \epsilon_i.$$

And if we note with $f^i(x)$ the i-th component of $f(x)$ (in the ϵ basis) so that $f(x) = f^i(x)\epsilon_i \in \mathbb{R}^m$, we have

$$\begin{aligned} &\lim_{h\to 0} \frac{\|f(x_0 + he_j) - f(x_0) - f'(x_0)(he_j)\|}{\|he_j\|} = 0, \\ \Rightarrow &\lim_{h\to 0} \frac{\left|f^i(x_0 + he_j) - f^i(x_0) - h\gamma^i_j\right|}{\|he_j\|} = 0, \end{aligned}$$

and we conclude like for the gradient that the coefficient γ^i_j must be equal to

$$\gamma^i_j = \left[\frac{d}{dh} f^i(x_0 + he_j)\right](0).$$

In particular, if e and ϵ are the respective canonical bases of $\mathbb{R}^n$ et $\mathbb{R}^m$, we recover

$$\gamma^i_j = \frac{\partial f^i}{\partial x^j}(x_0).$$

The derivative of $f : \mathbb{R}^n \to \mathbb{R}^m$ therefore corresponds to the *Jacobian matrix*.

4.3 Higher order derivatives and Hessian

We have seen with the previous examples that our general definition of the derivative admits well-known notions like the gradient or the Jacobian matrix as special cases. What about higher order derivatives, like the hessian matrix? To answer this question, let us contemplate the *type* of the derivative f'.

If f is a map between the vector spaces X and Y, its derivative at a point $f'(x_0)$ is a linear map from X to Y. If we note the set of all linear maps from X to Y as $L(X, Y)$, we can express the type of f' as

$$f' : X \to L(X, Y).$$

We can now make an important observation. $L(X, Y)$ is itself a vector space. If the correct limits converge, we can express the second derivative as

$$f'' = (f')' : X \to L(X, L(X, Y)),$$

such that for $x_0 \in X$,

$$\lim_{x \to x_0} \frac{||f'(x) - f'(x_0) - f''(x_0)(x - x_0)||_{L(X,Y)}}{||x - x_0||_X} = 0,$$

where $||.||_{L(X,Y)}$ is a norm on $L(X, Y)$. The norm can be defined uniquely using the norms on X and Y, but we will not detail this point. It is clear that the same procedure can be repeated to obtain higher order derivatives.

Let us illustrate the second derivative in the case of a twice differentiable multivariate function $f : \mathbb{R}^n \to \mathbb{R}$. We already know that its derivative is none other than the gradient ∇f. The second derivative of f is of type

$$f'' = (\nabla f)' : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^{n*}).$$

The second derivative evaluated at a point x_0 is then a linear map that takes a vector $x \in \mathbb{R}^n$ and returns a covector $f''(x_0)(x) \in \mathbb{R}^{n*}$. We have already encountered this type of objects. It is a covariant tensor of order 2!

If we take a basis $e = \{e_i\}_{i=1}^n$ of $\mathbb{R}^n$ and its associated dual basis e^*, we can compute the coefficients of this tensor. First, the gradient is expressed as

$$\nabla f(x_0) = \beta_i(x_0) e^{*i},$$

where we explicitly signal that the coefficients β_i are functions of x_0 defined as the directional derivatives of f at x_0 in the directions e_i. The second derivative $f''(x_0)$ can be expressed as

$$f''(x_0) = \eta_{ij} e^{*i} \otimes e^{*j},$$

whose components are characterized by

$$f''(x_0) e_i = \eta_{ij} e^{*j}$$

and by definition, we have that

$$\lim_{h \to 0} \frac{||\nabla f(x_0 + he_i) - \nabla f(x_0) - f''(x_0)(he_i)||}{||he_i||} = 0,$$

which implies in particular for the j-th coefficient

$$\lim_{h \to 0} \frac{\left|\beta_j(x_0 + he_i) - \beta_j(x_0) - h\eta_{ij}\right|}{||he_i||} = 0.$$

We deduce from this that the coefficient η_{ij} is the directional derivative

$$\eta_{ij} = \frac{d}{dt} \left[\beta_j(x_0 + te_i)\right]\Big|_{t=0} = \frac{d}{dt} \left[\frac{d}{ds} f(x_0 + te_i + se_j)\Big|_{s=0}\right]\Big|_{t=0}.$$

In particular, if e is the canonical basis, we recover

$$\eta_{ij} = \frac{\partial}{\partial x^i}\frac{\partial f}{\partial x^j}(x_0).$$

The second derivative of f therefore corresponds to the hessian "matrix"[2] of f, which we will denote with $\nabla^2 f$.

Before moving to the last section, let us briefly comment on derivatives beyond order two. The derivative of order three, for example will be of type

$$f''' : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^{n*})),$$

i.e. $f'''(x_0)$ is a linear map that takes a vector as argument and returns a covariant tensor of order two. It is therefore a covariant tensor of order three whose coefficients in the canonical basis are

$$\theta_{ijk} = \frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}\frac{\partial f}{\partial x^k}(x_0).$$

By pursuing this logic, it is clear that the derivative of order k of f will be a covariant tensor of order k whose coefficients will be given by derivatives of order k.

4.4 Taylor series

To conclude this section, we will show how to generalize the Taylor series to multivariate functions. As a reminder, the Taylor series of an infinitely differentiable function $f : \mathbb{R} \to \mathbb{R}$ around x_0 is given by

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k, \\ &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \dots \end{aligned}$$

If f is now an infinitely differentiable function from $\mathbb{R}^n$ to $\mathbb{R}$, the derivatives are covariant tensors and we need to make sense of the expressions $(x-x_0)^k$. Since these are multiplied to covariant tensors of order k and the result must be a scalar, it seems natural to interpret $(x-x_0)^k$ as the tensor product of the vector $(x - x_0)$ k times with itself, which is a contravariant tensor of order k, which will note by $(x - x_0)^{\otimes k}$.

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^{\otimes k}, \\ &= f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2} \nabla^2 f(x_0)(x - x_0) \otimes (x - x_0) + \dots. \end{aligned}$$

In the above expression, we find the second order Taylor approximation around x_0 with the gradient and hessian.

Finally, let us show that this series is valid. For it to make sense, it needs to converge at every point independently of the basis used to represent the tensors. In other words, for any point x within a neighbourhood of x_0 and for any basis $e = \{e_i\}_{i=1}^n$, the series

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{k!} D^{(k)}_{i_1,\dots,i_k} X^{(k),i_1,\dots,i_k},$$

[2] Calling it a "matrix" is of course a mistake, and we should instead call it the hessian 2-covariant tensor, but apparently it doesn't sound as good.

converges, where $D^{(k)}_{i_1,\ldots,i_k}$ and $X^{(k),i_1,\ldots,i_k}$ are respectively the coefficients of $f^{(k)}(x_0)$ and $(x-x_0)^{\otimes k}$ with respect to the basis e. If we now take another basis ε such that $\varepsilon_i = \gamma^j_i e_i$ and $e_i = \eta^j_i \varepsilon_j$ and the coefficients of $f^{(k)}(x_0)$ and $(x-x_0)^{\otimes k}$ with respect to this new basis are respectively $\tilde{D}^{(k)}_{i_1,\ldots,i_k}$ and $\tilde{X}^{(k),i_1,\ldots,i_k}$, the tensor transformation formula (4) lets us write

$$\begin{aligned}
&\sum_{k=0}^{\infty} \tilde{D}^{(k)}_{i_1,\ldots,i_k} \tilde{X}^{(k),i_1,\ldots,i_k},\\
&= \sum_{k=0}^{\infty} \frac{1}{k!} D^{(k)}_{i_1,\ldots,i_k} \gamma^{i_1}_{j_1} \ldots \gamma^{i_k}_{j_k} X^{(k),i_1,\ldots,i_k} \eta^{j_k}_{i_1} \ldots \eta^{j_k}_{i_k},\\
&= \sum_{k=0}^{\infty} \frac{1}{k!} D^{(k)}_{j_1,\ldots,j_k} X^{(k),j_1,\ldots,j_k},
\end{aligned}$$

where to go to the last line, we used the fact that $\gamma^i_k \eta^k_j = \delta^i_j$, which lets us simplify all pairs of terms $\gamma^{i_p}_{j_q} \eta^{j_q}_{i_p}$. The convergence of this series is therefore independent of the basis used, we just need to show that it converges in a basis of our choice.

For some x different from x_0, we can construct a basis by taking as the first vector $e_1 = x - x_0$ and choosing the other basis vectors arbitrarily (we can always do this in finite dimensions). We then have

$$\begin{aligned}
f(x) &= \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)(e_1)^{\otimes k},\\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d}{dh^k} \left[f(x_0 + he_1) \right] \Big|_{h=0},
\end{aligned}$$

where to step to the second line, we used the definition of the coefficients of the derivative tensors, with respect to the basis we have constructed. We just reduced our initial series to the Taylor series of $h \mapsto f(x_0 + he_1)$. If this series is well-defined for all x in a neighbourhood of x_0 (i.e. for any choice of e_1), we can then consider that the Taylor series of f is well-defined.

Conclusion

We developped in this document a visual notation that lets us express and visualize more clearly (we hope) the different types of tensors and their operations. That being said, we have omitted certain points for the sake of clarity and concision.

In particular, we have almost exclusively considered transformations between the same vector space. It is of course possible to apply our notation simultaneously to multiple different spaces, provided of course one pays attention to the bases used.

It is also possible to push further the developments initiated here. For instance, we could construct a vector space containing *all* tensors of all orders on the same base space X, whose elements would be *formal linear combinations* of scalars, vectors, covectors, ..., e.g.

$$\alpha + \alpha^i e_i + \beta_j e^{*j} + \gamma_{ij} e^{*i} \otimes e^{*j} + \theta^i_j e_i \otimes e^{*j} + \dots.$$

This space is infinite dimensional, and it may be hard to wrap one's head around it. Why it is interesting to consider such a space is that the product $*$, if we impose that it is distributive, then becomes everywhere-defined and associative on the space of tensors, which makes this space an *algebra*.

Finally, there seems to be a missing type of objects to the ones we already defined. The tensor product $y^* \otimes x$, which would be interpreted as a linear form taking a matrix as argument, i.e.

$$(y^* \otimes x)(\gamma^i_j e_i \otimes e^{*j}) = \beta_i \gamma^i_j \alpha^j,$$

or visually as

$$(y^* \otimes x) * (\gamma^i_j e_i \otimes e^{*j}) = \begin{pmatrix} \beta_1 & \dots & \beta_N \end{pmatrix} \begin{pmatrix} \gamma^1_1 & \dots & \gamma^1_N \\ \vdots & \ddots & \vdots \\ \gamma^N_1 & \dots & \gamma^N_N \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^N \end{pmatrix}.$$

It is not clear to us at the moment how to represent such an object, whether symbolically or visually.