Index Notation

J.Pearson

July 18, 2008

Abstract

This is my own work, being a collection of methods & uses I have picked up. In this work, I gently introduce index notation, moving through using the summation convention. Then begin to use the notation, through relativistic Lorentz transformations and quantum mechanical 2^{nd} order perturbation theory. I then proceed with tensors.

Contents

1 Index Notation **1**

1.1 Matrices . . . 1

1.2 Kronecker-Delta . . . 4

1.2.1 The Scalar Product . . . 5

1.2.2 An Example . . . 6

2 Applications **6**

2.1 Differentiation . . . 6

2.2 Transformations . . . 9

2.3 Relativistic Notation . . . 9

2.3.1 Special Relativity . . . 12

2.4 Perturbation Theory . . . 16

3 Tensors **19**

3.1 Vectors . . . 19

3.1.1 Transformation of the Components . . . 19

3.1.2 Transformation of the Basis Vectors . . . 20

3.1.3 Inverse Transformations . . . 21

3.2 Tensors . . . 23

3.2.1 The $\binom{0}{1}$ Tensors . . . 24

3.2.2 Derivative Notation . . . 27

3.3 The Geodesic . . . 29

3.3.1 Christoffel Symbols . . . 32

4 More Physical Ideas **38**

4.1 Curvilinear Coordinates . . . 38

4.1.1 Plane Polar Coordinates . . . 39

1 Index Notation

Suppose we have some Cartesian vector:

$$\boldsymbol{v} = x\hat{\boldsymbol{\imath}} + y\hat{\boldsymbol{\jmath}} + z\hat{\boldsymbol{k}}$$

Let us just verbalise what the above expression actually means: it gives the components of each basis vector of the system; that is, how much the vector extends onto certain defined axes of the system. We can write the above in a more general manner:

$$\boldsymbol{v} = x_1\hat{\boldsymbol{e}}_1 + x_2\hat{\boldsymbol{e}}_2 + x_3\hat{\boldsymbol{e}}_3$$

Now, we have that some vector has certain projections onto some axes. From now on, I shall supress the 'hat' of the basis vectors $\boldsymbol{e}$ (it is understood that they have unit length). Notice here, for generality, we have not mentioned what coordinate system we are using. Now, the above expression may of course be written as a sum:

$$\boldsymbol{v} = \sum_{i=1}^{3} x_i \boldsymbol{e}_i$$

Now, the *Einstein summation convention* is to ignore the summation sign:

$$\boldsymbol{v} = x_i \boldsymbol{e}_i$$

So that now, the above expression *is understood* to be summed over the index, where the index runs over the available coordinate system. That is, in the above system, the system is 3-dimensional, hence, $i = 1 \rightarrow 3$. If, for some reason, the system was 8-dimensional, then $i = 1 \rightarrow 8$. Also, as we see in relativistic-notation, it may be the case that i starts from 0. But this is usually understood in specific cases.

1.1 Matrices

Now, suppose we have the matrix multiplication:

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Let us write this as:

$$B = \mathsf{A}C$$

Where is is understood how the above two equations are linked. Also, we say that the vector B has components b_1, b_2, b_3, and the matrix A has elements $a_{11}\ldots$. Now, if we were to carry out the multiplication, we would find:

$$\begin{aligned} b_1 &= a_{11}c_1 + a_{12}c_2 + a_{13}c_3 \\ b_2 &= a_{21}c_1 + a_{22}c_2 + a_{23}c_3 \\ b_3 &= a_{31}c_1 + a_{32}c_2 + a_{33}c_3 \end{aligned}$$

Now, if we glance over the above equations, we may notice that b_1 is found from $a_{1j}c_j$, where $j = 1 \rightarrow 3$ (a sum). That is:

$$b_1 = \sum_{j=1}^{3} a_{1j}c_j = a_{11}c_1 + a_{12}c_2 + a_{13}c_3$$

Which is indeed true. So, we may suppose that we can form any component of the vector B, b_i (say) by replacing the '1' above by i. Let us do just this:

$$b_i = \sum_{j=1}^{3} a_{ij}c_j \qquad b_i = a_{ij}c_j$$

Where the RHS expression has used the summation convention of implied summation. One can verify that the above expression does indeed hold true, by setting i equal to 1, 2, 3 in turn, and carrying out the summations. Of course, this also works for vectors/matrices in higher dimensions (i.e. more columns/rows/components/elements). As this obviously works in an arbitrary number of dimensions, one must specify, or at least *understand* how far to take the sum. Infact, in relativistic notation, one uses greek letters (rather than latin) to denote the indices, so that the following is automatically understood:

$$b_\nu = \sum_{\mu=0}^{3} a_{\nu\mu}c_\mu \qquad b_\nu = a_{\nu\mu}c_\mu$$

So, we say:
The i^{th} component of vector B, resulting from the multiplication of some matrix A having elements a_{ij}, with vector C, can be found via:

$$b_i = a_{ij}c_j$$

Now, notice that we may exchange any letter for another: $i \rightarrow k$, $j \rightarrow n$, so that:

$$b_k = a_{kn}c_n$$

But also note, we must to the same for everything. We may also cyclicly change the indices $i \rightarrow j \rightarrow i$, resulting in:

$$b_j = a_{ji}c_i$$

This may seem obvious, or pointless, but it is very useful.

Suppose we have the multiplication of two matrices, to give a third:

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

Let us explore how to 'get to' the index notation for such an operation. Now, let us carry out the multiplication, and write each element of the matricx B separately (rather than in its matrix form):

$$\begin{aligned} b_{11} &= a_{11}c_{11} + a_{12}c_{21} \\ b_{12} &= a_{11}c_{12} + a_{12}c_{22} \\ b_{21} &= a_{21}c_{11} + a_{22}c_{21} \\ b_{22} &= a_{21}c_{12} + a_{22}c_{22} \end{aligned}$$

Then, again, we may be able to notice that we could write:

$$b_{21} = \sum_{j=1}^{2} a_{2j}c_{j1} = a_{21}c_{11} + a_{22}c_{21}$$

Let us replace the '2' with i, and the '1' with k. Then, the above summation reads:

$$b_{ik} = \sum_{j=1}^{2} a_{ij}c_{jk} \qquad b_{ik} = a_{ij}c_{jk}$$

Again, if the summations are done, one can verify that one gets all the components of the matrix B. Also, writing in summation notation (leave in the summation sign, for now), one may take the upper-limit to be anything! The only reason a 2-dimensional matrix was used was for brevity. So, let us say that the result of multiplication of 2 N-dimensional *square*[1] matrices, has elements which may be found from:

$$b_{ik} = \sum_{j=1}^{N} a_{ij}c_{jk} \tag{1.1}$$

Now, let us re-lable k with j , as we are at complete liberty to do (as per our previous discussion):

$$b_{ij} = \sum_{k=1}^{N} a_{ik}c_{kj} \tag{1.2}$$

We can also cyclicly change the indices (from the above) $i \to j \to k \to i$ (this time using the summation convention):

$$b_{jk} = a_{ji}c_{ik}$$

One must notice that the order of the same index is the same. That is, whatever comes first on the b-index (say), appears at the same position in all expressions.
Also notice, the index being summed-over changes. Originally, we summed over j, in (1.1), then it was k in (1.2); (after re-labling k with j), and finally the summed-index was i.
Then notice, the only index which is not on both the left and right hand sides of an equation is being summed over.

Also, by way of notation, it is common to have an expression such as $b_{ij} = a_{ik}c_{kj}$, but refer to the matrix a_{ij}. If confusion occurs, one must rememeber the redundancy of labelling in expressions.

Recall the method for matrix addition:

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + c_{11} & a_{12} + c_{12} \\ a_{21} + c_{21} & a_{22} + c_{22} \end{pmatrix}$$

Now, notice that $b_{12} = a_{12} + c_{12}$, for example. Then, one may write:

$$b_{ij} = a_{ij} + c_{ij}$$

Now, the concept of *repeated indices* becomes very important to note here! Every index which appears on the left also appears on the right. We are finding the sum of two matrices one element at a time; not summing the matrices outright.

[1] A square matrix is one whose number of columns and rows are the same. This works for non-square matrices, but we shall not go into that here.

1.2 Kronecker-Delta

Suppose we have some matrix which only has diagonal elements: all other elements are zero. The diagonal components are also unity. Then, in 3-dimensions, this matrix would look like:

$$\mathsf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This is called the *identity matrix* (hence, the I). Now, suppose we form the matrix product of this with another matrix:

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

Obviously we are now in 2-dimensions. Let us do the multiplication:

$$\begin{aligned} b_{11} &= c_{11} \\ b_{12} &= c_{12} \\ b_{21} &= c_{21} \\ b_{22} &= c_{22} \end{aligned}$$

A curious pattern emerges! Let us write down the multiplication in summation convention index notation:

$$b_{ij} = \delta_{ik} c_{kj}$$

Where the δ-symbol denotes the matrix I (just as a denotes elements of the matrix A).

Now, let us take a look at the identity matrix again. Every component is zero, except those which lie on the diagonal. That is, $\delta_{11} = 1$, $\delta_{21} = 0$, etc. Note, the elements which lie on the diagonal are those for which $i = j$, corresponding to element δ_{ij}. Then, we may write that the identity matrix has the following properties:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.3}$$

Now then, let us re-examine $b_{ij} = \delta_{ik} c_{kj}$. Let us write out the sum, for $i = 1, j = 2$:

$$b_{12} = \sum_{k=1}^{2} \delta_{1k} c_{k2} = \delta_{11} c_{12} + \delta_{12} c_{22}$$

Now, by (1.3), $\delta_{11} = 1$ and $\delta_{12} = 0$. Hence, we see that $b_{12} = c_{12}$.
Which is what we had by direct matrix multiplication, except that we have arrived at it by considering the inherent meaning of the identity matrix, and its elements.

So, we may write that the only non-zero components of $b_{ij} = \delta_{ik} c_{kj}$ are those for which $i = k$. Hence:

$$b_{ij} = \delta_{ik} c_{kj} = c_{ij}$$

Again, a result we had before, from direct matrix multiplication; except that now we arrived at it from considering the inherent properties of the element δ_{ij}. In this way, we usually refrain from referring to δ_{ij} as an element of the identity matrix, more as an object in its own right: the *Kronecker-delta*.

We may think of the following equation:

$$b_{ij} = \sum_k \delta_{ik} c_{kj}$$

As sweeping over all values of k, but the only non-zero contributions picked up, are those for which $k = i$. Leaving the summation sign in, in this case, allows for this interpretation to become a little clearer.

1.2.1 The Scalar Product

Now, we can use the Kronecker-delta in the index-notation for a scalar-product. Suppose we have two vectors:

$$\boldsymbol{v} = v_1 \boldsymbol{e}_1 + v_2 \boldsymbol{e}_2 + v_3 \boldsymbol{e}_3 \qquad \boldsymbol{u} = u_1 \boldsymbol{e}_1 + u_2 \boldsymbol{e}_2 + u_3 \boldsymbol{e}_3$$

Then, their scalar product is:

$$\boldsymbol{v} \cdot \boldsymbol{u} = v_1 u_1 + v_2 u_2 + v_3 u_3$$

Now, inlight of previous discussions, this is calling out to be put into index notation! Now, something that has been used, but not stated, in doing the scalar-product, is the orthogonality of the basis vectors $\{\boldsymbol{e}_i\}$[2]. That is:

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij}$$

Now, let us write each vector in index notation:

$$\boldsymbol{v} = v_i \boldsymbol{e}_i \qquad \boldsymbol{u} = u_i \boldsymbol{e}_i$$

Now, to go ahead with the scalar product, one must realise something: there is no reason to use the same index for both vectors. That is, the following expressions are perfectly viable:

$$\boldsymbol{u} = u_i \boldsymbol{e}_i \qquad \boldsymbol{u} = u_j \boldsymbol{e}_j$$

Then, using the second expression (as it is infact, a more general way of doing things) in the scalar product:

$$\boldsymbol{v} \cdot \boldsymbol{u} = v_i u_j \boldsymbol{e}_i \cdot \boldsymbol{e}_j$$

Then, using the orthogonality relation:

$$v_i u_j \boldsymbol{e}_i \cdot \boldsymbol{e}_j = v_i u_j \delta_{ij}$$

Then, using the property of the Kronecker-delta: the only non-zero components are those for which $i = j$; thus, the above becomes:

$$v_i u_i$$

Note, we could just have easily chosen $v_j u_j$, and it would have been equivalent. Hence, we have that:

$$\boldsymbol{v} \cdot \boldsymbol{u} = v_i u_i = \sum_{i=1}^{3} v_i u_i = v_1 u_1 + v_2 u_2 + v_3 u_3$$

[2]This notation simply reads 'the set of basis vectors'.

Again, a result we had found by direct manipulation of the vectors, but now we arrived at it by considering the inherent property of the Kronecker-delta, as well as the orthogonality of the basis vectors.

1.2.2 An Example

Let us consider an example. Consider the equation:

$$b_{ij} = \delta_{ik}\delta_{jn}a_{kn}$$

Then, the task is to simplify it. Now, the delta on the far-right will 'filter' out only the component for which $j = n$. Then, we have:

$$b_{ij} = \delta_{ik}a_{kj}$$

The remaining delta will filter out the component for which $i = k$. Then:

$$b_{ij} = a_{ij}$$

Infact, if the elements of a matrix are the same in two matrices; then the matrices are identical. That is the above statement.

2 Applications

Thus far, we have discussed the notations of vectors and matrices. Let us consider some ways of using such notation.

2.1 Differentiation

Suppose we have some operator which does the same thing to different components of a vector. Consider the divergence of a cartesian vector $\boldsymbol{v} = (v_x, v_y, v_z)$:

$$\nabla \cdot \boldsymbol{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}$$

Then, let us write that:

$$x_1 = x \qquad x_2 = y \qquad x_3 = z$$

Then, the divergence of $\boldsymbol{v} = (v_1, v_2, v_3)$ is:

$$\nabla \cdot \boldsymbol{v} = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3}$$

Immediate inspection allows us to see that we may write this as a sum:

$$\nabla \cdot \boldsymbol{v} = \sum_{i=1}^{3} \frac{\partial v_i}{\partial x_i} \qquad \nabla \cdot \boldsymbol{v} = \frac{\partial v_i}{\partial x_i}$$

Similarly, let us write the nabla-operator itself:

$$\nabla = \boldsymbol{e}_1 \frac{\partial}{\partial x_1} + \boldsymbol{e}_2 \frac{\partial}{\partial x_2} + \boldsymbol{e}_3 \frac{\partial}{\partial x_3}$$

This is of course, under immediate summation-convention:

$$\nabla = \boldsymbol{e}_i \frac{\partial}{\partial x_i}$$

Infact, let us just recall the Kronecker-delta. In the above divergence, we may write:

$$\nabla \cdot \boldsymbol{v} = \frac{\partial v_i}{\partial x_i} = \frac{\partial v_j}{\partial x_i} \delta_{ij}$$

Let us leave this alone, for now.

Consider the differential of a vector, with respect to any component. Let me formulate this a little more clearly. Let us have the following vector[3]:

$$\boldsymbol{v} = \boldsymbol{e}_1 v_1 + \boldsymbol{e}_2 v_2 + \boldsymbol{e}_3 v_3$$

Then, let us differentiate the vector with respect to, say x_2. Then:

$$\frac{\partial \boldsymbol{v}}{\partial x_2} = \boldsymbol{e}_1 \frac{\partial v_1}{\partial x_2} + \boldsymbol{e}_2 \frac{\partial v_2}{\partial x_2} + \boldsymbol{e}_3 \frac{\partial v_3}{\partial x_2}$$

Before I put the differential-part into index notation, let me put the vector into index notation. The above is then:

$$\frac{\partial \boldsymbol{v}}{\partial x_2} = \boldsymbol{e}_i \frac{\partial v_i}{\partial x_2}$$

Then, the differential of $\boldsymbol{v}$, with respect to any component, is:

$$\frac{\partial \boldsymbol{v}}{\partial x_j} = \boldsymbol{e}_i \frac{\partial v_i}{\partial x_j}$$

Again, let us stop here with this.

Consider the derivative of x_1 with respect to x_2. It is clearly zero:

$$\frac{\partial x_1}{\partial x_2} = 0$$

That is, for $i \neq j$:

$$\frac{\partial x_i}{\partial x_j} = 0 \qquad i \neq j$$

Also, consider the derivative of x_1 with respect to x_1. That is clearly unity:

$$\frac{\partial x_1}{\partial x_1} = 1$$

[3]It is understood that we are using some basis, and that $v_1(x_1, x_2, x_3)$: the components are functions of x_1, x_2, x_3.

Hence, combining the above two results:

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} \tag{2.1}$$

Similarly, suppose we form the second-differential:

$$\frac{\partial^2 x_i}{\partial x_j \partial x_k} = \frac{\partial}{\partial x_j}\frac{\partial x_i}{\partial x_k} = \frac{\partial}{\partial x_j}\delta_{ik}$$

However, differentiation is commutative; that is:

$$\frac{\partial^2 x_i}{\partial x_j \partial x_k} = \frac{\partial^2 x_i}{\partial x_k \partial x_j}$$

Thus, the right hand side of the above is:

$$\frac{\partial^2 x_i}{\partial x_k \partial x_j} = \frac{\partial}{\partial x_k}\delta_{ij}$$

Hence:

$$\frac{\partial}{\partial x_j}\delta_{ik} = \frac{\partial}{\partial x_k}\delta_{ij}$$

Notice, this highlights the interchangeability of indices: notice what happens to the left hand side if j is changed for k.

Consider the following differential:

$$\frac{\partial x'_i}{\partial x'_j} = \delta'_{ij}$$

Where a prime here just denotes that the component is of a vector different to unprimed components. Now, we may of course multiply and divide by the same factor, x_k, say:

$$\frac{\partial x'_i}{\partial x'_j}\frac{\partial x_k}{\partial x_k} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_k}{\partial x'_j}$$

Now, the differential on the very-far RHS (i.e. $\frac{\partial x_k}{\partial x'_j}$), if we let the $k \to l$, and use a Kronecker delta δ_{kl}. That is:

$$\frac{\partial x'_i}{\partial x_k}\frac{\partial x_k}{\partial x'_j} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_j}\delta_{kl}$$

Now, let us just string together everything that we have done, in one big equality:

$$\delta'_{ij} = \frac{\partial x'_i}{\partial x'_j} = \frac{\partial x'_i}{\partial x'_j}\frac{\partial x_k}{\partial x_k} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_k}{\partial x'_j} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_j}\delta_{kl}$$

Hence, we have the rather curious result:

$$\delta'_{ij} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_j}\delta_{kl} \tag{2.2}$$

This is infact the statement that the Kronecker-delta transforms as a second-rank tensor. This is something we will come back to, in detail, later.

Let us consider the following:

$$\nabla(\nabla \cdot \boldsymbol{v})$$

Now, to put this into index notation, one must first notice that the expression is the grad of a scalar (the divergence). Hence:

$$\nabla(\nabla \cdot \boldsymbol{v}) = \boldsymbol{e}_i \frac{\partial}{\partial x_i} \frac{\partial v_j}{\partial x_j}$$

Which is of course the same as:

$$\nabla(\nabla \cdot \boldsymbol{v}) = \boldsymbol{e}_i \delta_{jk} \frac{\partial}{\partial x_i} \frac{\partial v_j}{\partial x_k}$$

This may seem like a pointless usage of the Kronecker-delta; and infact here, it is. However, it does serve the purpose of highlighthing how it may be used.

2.2 Transformations

Let us suppose of some transformation matrix A (components a_{ij}), which acts upon some vector (by matrix multiplication) B (components b_i), resulting in some new vector B' (components b'_i). Then, we have:

$$b'_i = a_{ij} b_j$$

Now then, here, the a_{ij} are some *transformation* from one frame (unprimed) to another (primed). Let us just recap what the above statement is saying: if we act upon a vector with a transformation, we get a new vector.

Now, rather than a vector, suppose we had something like b_{ij}, and that it transforms via:

$$b'_{ij} = a_{in} a_{jm} b_{nm}$$

Now, initially notice: we are implicitly implying that there are two summations on the right hand side: over n and m. Now, depending on the space in which the objects are embedded, the transformation matrices a_{ij} take on various forms. Again, depending on the system, b_{ij} will usually be called a tensor of second rank.

2.3 Relativistic Notation

Here, we shall consider the application of index notation to special relativity, and explore the subject in the process.

Let us first say that we shall use 4 coordinates: (ct, x, y, z). However, to make things more transparent, we shall denote them by:

$$x^\mu = (x^0, x^1, x^2, x^3) \tag{2.3}$$

So that such an infinitesimal is:

$$dx^\mu = (dx^0, dx^1, dx^2, dx^3)$$

Now, we shall define the subscript versions thus:

$$dx_0 \equiv -dx^0 \quad dx_1 \equiv dx^1 \quad dx_2 \equiv dx^2 \quad dx_3 \equiv dx^3 \tag{2.4}$$

Then:

$$dx_\mu = (dx_0, dx_1, dx_2, dx_3)$$

And, from their definitions, notice that we can write:

$$dx_\mu = (dx_0, dx_1, dx_2, dx_3) = (-dx^0, dx^1, dx^2, dx^3)$$

Now, an interval of space is such that:

$$-\, ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \tag{2.5}$$

Notice how such an interval is defined. Now notice that the above sum looks 'nice', except for the minus sign infront of the dx^0-term. Now If we recall from above, we have that $dx_0 = -dx^0$. Then, $dx_0 dx^0 = -(dx_0)^2$, which is the term above. Thus, by inspection we see that we may form the following:

$$-ds^2 = dx_0 dx^0 + dx_1 dx^1 + dx_2 dx^2 + dx_3 dx^3 = \sum_{\mu=0}^{3} dx_\mu dx^\mu$$

Hence, under this summation convention:

$$-\, ds^2 = dx_\mu dx^\mu \tag{2.6}$$

Now, let us introduce the concept of the *Minkowski metric*, $\eta_{\mu\nu}$. Let us say that the following is true:

$$-\, ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{2.7}$$

Notice, this expression seems a little more general than (2.6). Now, let us explore the properties of $\eta_{\mu\nu}$, in much the same way we explored the Kronecker-delta δ_{ij}.

Now, let us say that $\eta_{\mu\nu} = \eta_{\nu\mu}$; i.e. that it is symmetric. Infact, any anti-symmetric terms (that is, those for which $\xi_{\mu\nu} = -\xi_{\nu\mu}$) will drop out. This is not shown here.

Now, let us begin to write out the summation in (2.7):

$$\begin{aligned} -ds^2 &= \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} \eta_{\mu\nu} dx^\mu dx^\nu \\ &= \eta_{00} dx^0 dx^0 + \eta_{10} dx^1 dx^0 + \eta_{01} dx^0 dx^1 + \ldots + \eta_{11} dx^1 dx^1 + \ldots \end{aligned}$$

Now, compare this expression with (2.5). We see that the coefficient of $dx^0 dx^0$ is -1, and that of $dx^1 dx^1$ is 1; and that of $dx^1 dx^0$ is zero; etc. Then, we start to see that $\eta_{\mu\nu}$ only has non-zero entries for $\mu = \nu$, and that for $\mu = \nu = 0$ (note, indexing starts from 0, not 1), it has value -1; and that $\mu = \nu = 1, 2, 3$ it has value 1. Then, in analogue with the Kronecker delta, let us write a matrix which represents $\eta_{\mu\nu}$:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

Now, from (2.4), we are able to see that we can write a relation between dx^μ and dx_ν, in terms of $\eta_{\mu\nu}$:

$$dx_\mu = \eta_{\mu\nu} dx^\nu \tag{2.9}$$

Let us just write out the relevant matrices, just to see this 'in action':

$$\begin{aligned} \eta_{\mu\nu} dx^\nu &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dx^0 \\ dx^1 \\ dx^2 \\ dx^3 \end{pmatrix} \\ &= \left(-dx^0, dx^1, dx^2, dx^3\right) \\ &= dx_\mu \end{aligned}$$

So, in writing the relations in this way, we see that we have the *interpretation* that a vector with 'lower' indices is the transpose of the 'raised' index version. This interpretation is limited however, but may be useful.
So, we may write in general then, that a (4)vector $\boldsymbol{b}$, having components b^μ can have its indices 'lowered' by acting the Minkowsi metric upon it:

$$b_\mu = \eta_{\mu\nu} b^\nu$$

The *relativistic scalar product* of two vectors is:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a^\mu b_\mu$$

However, we can write that $b_\mu = \eta_{\mu\nu} b^\nu$; hence the above expression is:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a^\mu b_\mu = \eta_{\mu\nu} a^\mu b^\nu$$

And that is, of course, with reference to the metric itself:

$$\boldsymbol{a} \cdot \boldsymbol{b} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3$$

Now, if we have some matrix M, a product of M with its inverse gives the identity matrix I. So, if we write that the inverse of $\eta_{\mu\nu}$ is $\eta^{\mu\nu}$. Now, the identity matrix is infact the Kronecker delta symbol. Hence, we have that:

$$\eta^{\nu\rho} \eta_{\rho\mu} = \delta^\nu_\mu \tag{2.10}$$

Where we have used conventional notation for the RHS. Notice that the indices has been chosen in that order to be inaccord with matrix multiplication. Now, from the matrix form of $\eta_{\mu\nu}$, we may be able to spot that its inverse actually has the same form. Let us demonstrate this by confirming that we get to the identity matrix:

$$\begin{matrix} \eta^{\nu\rho} & \eta_{\rho\mu} & = \delta^\nu_\mu \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Hence confirmed.

Now, just as we had $b_\mu = \eta_{\mu\nu} b^\nu$, what about $\eta^{\rho\mu} b_\mu$? Let us consider it (notice that the choice of letter in labelling the indices is completely arbitrary, and is just done for clarity):

$$\begin{aligned} \eta^{\rho\mu} b_\mu &= \eta^{\rho\mu}(\eta_{\mu\nu} b^\nu) \\ &= (\eta^{\rho\mu}\eta_{\mu\nu}) b^\nu \\ &= \delta^\rho_\nu b^\nu \\ &= b^\rho \end{aligned}$$

It should be fairly obvious what we have done, in each step.

Thus, we have a way of raising, or lowering indices:

$$\eta^{\nu\mu} b_\mu = b^\nu \qquad \eta_{\mu\nu} b^\nu = b_\mu \tag{2.11}$$

Notice, the redundancy in labelling. In the first expression, we can exchange $\nu \to \mu \to \nu$, and then the indices appear in the same order as the second expression.

2.3.1 Special Relativity

I shall not attempt at a full discussion of SR, more the index notation involved!

Here, let us consider some 'boost' along the x-axis. In the language of SR, we have that two intertial frames are coincident at $t = 0$, and subsequently move along the x-axis with velcity v. The relationship between coordinates in the stationary frame (unprimed) and moving frame (primed) are given below:

$$\begin{aligned} ct' &= \gamma(ct - \beta x) \\ x' &= \gamma(x - \beta ct) \\ y' &= y \\ z' &= z \end{aligned}$$

Where:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \qquad \beta \equiv \frac{v}{c}$$

Then, using the standard $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$, we have that:

$$\begin{aligned} x'^0 &= \gamma(x^0 - \beta x^1) \\ x'^1 &= \gamma(-\beta x^0 + x^1) \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned}$$

Now, there appears to be symmetry in the first two terms. Let us write down a transformation matrix, then we shall discuss it:

$$\Lambda^\mu{}_\nu \equiv \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.12}$$

Now, we may see that we can form the components of x'^μ (i.e. in the moving frame), from those in the stationary frame, by multiplication of the above matrix, with the x^ν. Now, matrix multiplication cannot strictly be used in this case, as the indices of the x's are all -higher. So, let us write:

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \tag{2.13}$$

Consider this, as a matrix equation:

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

If 'matrix multiplication' is then carried out on the above, then the previous equations can be recovered. So, we write that the Lorentz transformation matrix is:

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.14}$$

One should also know that the inverse Lorentz transformations are the set of equations transferring from the primed frame to the unprimed frame:

$$\begin{aligned} ct &= \gamma(ct' - \beta x') \\ x &= \gamma(x' - \beta ct') \\ y &= y' \\ z &= z' \end{aligned}$$

Then, we see that we can write the inverse Lorentz transformation as a matrix (in notation we shall discuss shortly):

$$\Lambda_\nu{}^\mu = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.15}$$

Now, distances remain unchanged. That is, the value of the scalar product of two vectors is the same in both frames:

$$\boldsymbol{a} \cdot \boldsymbol{a} = \boldsymbol{a}' \cdot \boldsymbol{a}'$$

From previous discussions, this is:

$$\eta_{\alpha\beta} x^\alpha x^\beta = \eta_{\mu\nu} x'^\mu x'^\nu$$

Now, we have a relation which tells us what x'^μ is, in terms of x^μ; namely (2.13). So, the expressions on the RHS above:

$$x'^\mu = \Lambda^\mu{}_\nu x^\alpha \qquad x'^\nu = \Lambda^\nu{}_\beta x^\beta$$

Hence:

$$\eta_{\alpha\beta} x^\alpha x^\beta = \eta_{\mu\nu} x'^\mu x'^\nu = \eta_{\mu\nu} \Lambda^\mu{}_\alpha x^\alpha \Lambda^\nu{}_\beta x^\beta$$

Hence, as we notice that $x^\alpha x^\beta$ appears on both sides, we conclude that:

$$\eta_{\alpha\beta} = \eta_{\mu\nu}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta$$

Which we may of course rearange into something which looks more like matrix multiplication:

$$\eta_{\alpha\beta} = \Lambda^\mu{}_\alpha\eta_{\mu\nu}\Lambda^\nu{}_\beta \tag{2.16}$$

Let us write out each matrix, and multiply them out, just to see that this is indeed the case:

$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.17}$$

Multiplying the matrices on the far right, and putting next to that on the left:

$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} -\gamma & \gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Multplying these togther results in the familiar form of $\eta_{\alpha\beta}$.

Now, let us look at the Lorentz transformation equations again:

$$\begin{aligned} x'^0 &= \gamma(x^0 - \beta x^1) \\ x'^1 &= \gamma(-\beta x^0 + x^1) \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned}$$

Let us do something, which will initially seem pointless (as do many of our 'new ideas', but they all end up giving a nice result!). Consider differentiating expressions thus:

$$\frac{\partial x'^0}{\partial x^0} = \gamma \qquad \frac{\partial x'^0}{\partial x^1} = -\beta\gamma \qquad \frac{\partial x'^0}{\partial x^2} = \frac{\partial x'^0}{\partial x^3} = 0$$

Also:

$$\frac{\partial x'^1}{\partial x^0} = -\beta\gamma \qquad \frac{\partial x'^1}{\partial x^1} = \gamma$$

Where we shall leave the others as trivial. Now, notice that we have generated the components of the Lorentz transformation matrix, by differentiating components in the primed frame, with respect to those in the unprimed frame. Let us write the Lorentz matrix in the following way:

$$\Lambda = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix}$$

That is, we see that we have generated its components by differentiating:

$$\Lambda^0{}_0 = \frac{\partial x'^0}{\partial x^0} \qquad \Lambda^1{}_0 = \frac{\partial x'^1}{\partial x^0}$$

Then, we can generalise to any element of the matrix:

$$\Lambda^\mu{}_\nu = \frac{\partial x'^\mu}{\partial x^\nu} \tag{2.18}$$

So that we could write the transformation between primed & unprimed frames in the two equivalent ways:

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \qquad x'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} x^\nu$$

Let us consider the inverse transformations, and its notation. Consider the inverse transformation equations:

$$\begin{aligned} x^0 &= \gamma(x'^0 + \beta x'^1) \\ x^1 &= \gamma(\beta x'^0 + x'^1) \\ x^2 &= x'^2 \\ x^3 &= x'^3 \end{aligned}$$

If we denote the inverse by matrix elements $\Lambda_\nu{}^\mu$, its not too hard to see that:

$$\Lambda_\nu{}^\mu = \frac{\partial x^\nu}{\partial x'^\mu}$$

We shall shortly justify this. Although this seems a rather convoluted way to write the transformations, it is actually a lot more transparent. If we consider that a set of coordinates are functions of another set, then the transformation is rather trivial. We will come to this later.

Let us justify the notation for the inverse Lorentz transformation of $\Lambda_\nu{}^\mu$, where the Lorentz transformation is $\Lambda^\mu{}_\nu$. Now, in the same way that the metric lowers an index from a contravariant vector:

$$\eta_{\mu\nu} x^\nu = x_\mu$$

We may also use the metric on mixed "tensors" (we shall discuss these in more detail later; just take this as read here):

$$\eta_{\mu\nu} A^{\nu\lambda} = A_\mu{}^\lambda \qquad \eta_{\mu\nu} A^\nu{}_\lambda = A_{\mu\lambda}$$

So, we see that in the LHS expression, the metric drops the ν and relabels it μ. One must note that the relative positions of the indices are kept: the column the index initially possesses is the column the index finally posseses. So, consider an amalgamation of two metric operations:

$$\eta_{\alpha\nu} \eta^{\beta\mu} A^\nu{}_\mu = \eta_{\alpha\nu} A^{\nu\beta} = A_\alpha{}^\beta$$

Now, consider replacing the symbol A with Λ:

$$\Lambda_\alpha{}^\beta = \eta_{\alpha\nu} \eta^{\beta\mu} \Lambda^\nu{}_\mu$$

Now, notice that $\eta^{\beta\mu} = (\eta^{\mu\beta})^T$, and that by using this, we get a form in which matrix multiplication is valid:

$$\Lambda_\alpha{}^\beta = \eta_{\alpha\nu}\Lambda^\nu{}_\mu(\eta^{\mu\beta})^T$$

Now, one can then carry out the three matrix multiplication:

$$\begin{aligned}
\eta_{\alpha\nu}\Lambda^\nu{}_\mu(\eta^{\mu\beta})^T &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \Lambda_\alpha{}^\beta
\end{aligned}$$

Hence, we see that $\Lambda_\nu{}^\mu$ has the components of the inverse Lorentz transformation, from (2.15). Hence our reasoning for using such notation.

2.4 Perturbation Theory

In quantum mechanics we are able to expand perturbed eigenstates & eigenenergies in terms of some known unperturbed states. So, by way of briefly setting up the problem:

$$\begin{aligned}
\hat{H} &= \hat{H}_0 + \hat{V} \\
\psi_k &= \psi_k^{(0)} + \psi_k^{(1)} + \psi_k^{(2)} + \dots \\
E_k &= E_k^{(0)} + E_k^{(1)} + E_k^{(2)} + \dots
\end{aligned}$$

Now, we have our following perturbed & underpurbed Schrodinger equations:

$$\hat{H}\psi_k = E_k\psi_k \qquad \hat{H}_0\phi_k = e_k\phi_k$$

So that, from the expansions, to zeroth order:

$$\psi_k^{(0)} = \phi_k \qquad E_k^{(0)} = e_k$$

Now, if we put our expansions into the perturbed Schrodinger equation, we will end up with equations correct to first and second order perturbations:

$$\begin{aligned}
\hat{H}_0\psi_k^{(1)} + \hat{V}\psi_k^{(0)} &= E_k^{(0)}\psi_k^{(1)} + E_k^{(1)}\psi_k^{(0)} \\
\hat{H}_0\psi_k^{(2)} + \hat{V}\psi_k^{(1)} &= E_k^{(0)}\psi_k^{(2)} + E_k^{(1)}\psi_k^{(1)} + E_k^{(2)}\psi_k^{(0)}
\end{aligned}$$

What we want to do, is to express the first and second order corrections to the eigenstates and eigenenergies of the perturbed Hamiltonian, in terms of the known unperturbed states.

Now, we shall use index notation heavily, but in a very different way to before, to solve this problem.

Now, we shall not solve for first order perturbation correction for the eigenstate, or energy; but we shall quote the results. We do so by the allowance of expanding any eiegenstate in terms of a complete set. The set we choose is the unperturbed states. Thus:

$$\psi_k^{(1)} = \sum_j c_{kj}^{(1)} \phi_j \qquad c_{kj}^{(1)} = \frac{V_{jk}}{e_k - e_j}$$

Where we have quoted the result[4]. Now, we will carry on & figure out what the $c_{kj}^{(2)}$ are:

$$\psi_k^{(2)} = \sum_j c_{kj}^{(2)} \phi_j$$

So, we proceed. We rewrite the second order Schrodinger equation thus:

$$E_k^{(2)} \psi_k^{(0)} = (\hat{V} - E_k^{(1)}) \psi_k^{(1)} + (\hat{H}_0 - E_k^{(0)}) \psi_k^{(2)}$$

Now, we insert, for the states, the expansions in terms of some coefficient & the unperturbed state[5]:

$$E_k^{(2)} \phi_k = \sum_j c_{kj}^{(1)} (\hat{V} - E_k^{(1)}) \phi_j + \sum_j c_{kj}^{(2)} (e_j - e_k) \phi_j$$

Now, using a standard quantum mechanics 'trick', we multiply everything by some state (we shall use ϕ_i^*), and integrate over all space:

$$E_k^{(2)} \int \phi_i^* \phi_k d\tau = \sum_j c_{kj}^{(1)} \left(\int \phi_i^* \hat{V} \phi_j d\tau - E_k^{(1)} \int \phi_i^* \phi_j d\tau \right) + \sum_j c_{kj}^{(2)} (e_j - e_k) \int \phi_i^* \phi_j d\tau$$

Using othornormality[6] of the eigenstates, as well as some notation[7], this cleans up to:

$$E_k^{(2)} \delta_{ik} = \sum_j c_{kj}^{(1)} \left(V_{ij} - E_k^{(1)} \delta_{ij} \right) + \sum_j c_{kj}^{(2)} (e_j - e_k) \delta_{ij}$$

Let us use the summation convention:

$$E_k^{(2)} \delta_{ik} = c_{kj}^{(1)} \left(V_{ij} - E_k^{(1)} \delta_{ij} \right) + c_{kj}^{(2)} (e_j - e_k) \delta_{ij}$$

Then, using the Kronecker-delta on the far RHS:

$$E_k^{(2)} \delta_{ik} = c_{kj}^{(1)} \left(V_{ij} - E_k^{(1)} \delta_{ij} \right) + c_{ki}^{(2)} (e_i - e_k) \tag{2.19}$$

Now, let us consider the case where $i = k$. Then, we have:

$$E_i^{(2)} = c_{ij}^{(1)} \left(V_{ij} - E_i^{(1)} \delta_{ij} \right) + c_{ii}^{(2)} (e_i - e_i)$$

[4]The full derivation of the above terms can be found in *Quantum Mechanics of Atoms & Molecules*. Notice however, that we must take $c_{ii}^{(1)} = 0$, to avoid infinities.

[5]We shall be using that $\psi_k^{(\alpha)} = \sum_j c_{kj}^{(\alpha)} \phi_j$, and that $\hat{H}_0 \phi_k = e_k \phi_k$, where we use that $E_k^{(0)} \equiv e_k$ and $\psi_k^{(0)} \equiv \phi_k$.

[6]That is, $\int \phi_i^* \phi_j d\tau = \delta_{ij}$

[7]We shall use that $\int \phi_i^* \hat{A} \phi_j d\tau \equiv A_{ij}$

The far RHS is obviously zero[8]. Thus:

$$\begin{aligned} E_i^{(2)} &= c_{ij}^{(1)}\left(V_{ij} - E_i^{(1)}\delta_{ij}\right) \\ &= c_{ij}^{(1)}V_{ij} - c_{ij}^{(1)}E_i^{(1)}\delta_{ij} \\ &= c_{ij}^{(1)}V_{ij} - c_{ii}^{(1)}E_i^{(1)} \end{aligned}$$

We then notice that $c_{ii}^{(1)} \equiv 0$, as has previously been discussed in a footnote. Hence:

$$E_i^{(2)} = c_{ij}^{(1)}V_{ij}$$

Hence, using our previously stated result for $c_{ij}^{(1)}$, and putting the summation sign in:

$$E_i^{(2)} = \sum_{j\neq i}\frac{V_{ji}V_{ij}}{e_i - e_j}$$

Hence, quoting that $E_k^{(1)} = V_{kk}$, we may write the energy of the perturbed state, up to second order correction:

$$E_k = e_k + V_{kk} + \sum_{j\neq k}\frac{V_{jk}V_{kj}}{e_k - e_j}$$

Now to compute the correction coefficient to the eigenstate.

If we look back to (2.19), we considered the case where $i = k$, and look at that. Now, let us consider $i \neq k$. Then, the LHS of (2.19) is zero, and the RHS is:

$$c_{kj}^{(1)}\left(V_{ij} - E_k^{(1)}\delta_{ij}\right) + c_{ki}^{(2)}(e_i - e_k) = 0$$

Now, using a previously quoted result $E_k^{(1)} = V_{kk}$, the above trivially becomes:

$$c_{kj}^{(1)}\left(V_{ij} - V_{kk}\delta_{ij}\right) + c_{ki}^{(2)}(e_i - e_k) = 0$$

Expanding out the first bracket results trivially in:

$$c_{kj}^{(1)}V_{ij} - c_{kj}^{(1)}V_{kk}\delta_{ij} + c_{ki}^{(2)}(e_i - e_k) = 0$$

Using the middle Kronecker-delta:

$$c_{kj}^{(1)}V_{ij} - c_{ki}^{(1)}V_{kk} + c_{ki}^{(2)}(e_i - e_k) = 0$$

Inserting our known expressions for $c_{ij}^{(1)}$ (i.e.[9]), being careful in using correct indices:

$$\frac{V_{jk}}{e_k - e_j}V_{ij} - \frac{V_{ik}}{e_k - e_i}V_{kk} + c_{ki}^{(2)}(e_i - e_k) = 0$$

[8] This is the case partly due to $e_i - e_i = 0$, and also as $c_{ii}^{(1)} \equiv 0$

[9] $c_{ij}^{(1)} = \frac{V_{ji}}{e_i - e_j}$

A trivial rearrangement[10]:

$$\frac{V_{jk}}{e_k - e_j}V_{ij} - \frac{V_{ik}}{e_k - e_i}V_{kk} = c_{ki}^{(2)}(e_k - e_i)$$

Thus:

$$c_{ki}^{(2)} = \frac{V_{jk}}{(e_k - e_j)(e_k - e_i)}V_{ij} - \frac{V_{ik}}{(e_k - e_i)^2}V_{kk}$$

Hence, putting the summation signs back in:

$$c_{ki}^{(2)} = \sum_{j\neq k} \frac{V_{jk}V_{ij}}{(e_k - e_j)(e_k - e_i)} - \frac{V_{ik}V_{kk}}{(e_k - e_i)^2}$$

Now, let us just put this into the form (i.e. exchange indices) $c_{kj}^{(2)}$:

$$c_{kj}^{(2)} = \sum_{n\neq k} \frac{V_{nk}V_{jn}}{(e_k - e_n)(e_k - e_j)} - \frac{V_{jk}V_{kk}}{(e_k - e_j)^2}$$

Now, if we recall:

$$\psi_k^{(2)} = \sum_j c_{kj}^{(2)}\phi_j$$

Then:

$$\psi_k^{(2)} = \sum_{j\neq k}\sum_{n\neq k} \frac{V_{nk}V_{jn}}{(e_k - e_n)(e_k - e_j)}\phi_j - \sum_{j\neq k} \frac{V_{jk}V_{kk}}{(e_k - e_j)^2}\phi_j$$

Where we must be careful to exclude infinities from the summations.

3 Tensors

Here we shall introduce tensors & a way of visualising the differences between *covariant* and *contravariant* tensors. This section will follow the structure of *Schutz*[11], but may well deviate in exact notation.

Let us start by looking at vectors, in a more formal manner. This is invariably repeat some of the other sections work, but this is to make this section more self-contained.

3.1 Vectors

3.1.1 Transformation of the Components

Suppose we have some vector, in a system Σ. Let it have 4 *components*:

$$\Delta\boldsymbol{x} = (\Delta t, \Delta x, \Delta y, \Delta z) \equiv \{\Delta x^\alpha\}$$

[10] We have only take the term to the RHS, we have not switched any indices over.

[11] *A First Course in General Relativity* - Schutz.

Now, in some other system Σ', the same vector will have different components:

$$\Delta \boldsymbol{x} = \{\Delta x'^{\alpha}\}$$

Where we denote components in Σ' with a 'prime'.
Now, as we have seen, we can transform *between the components* of the vectors, in the different systems (in equivalent notations):

$$\Delta x'^{\alpha} = \Lambda^{\alpha}{}_{\beta} \Delta x^{\beta} \qquad \Delta x'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}} \Delta x^{\beta} \qquad \beta \in [0,3]$$

We shall use the latter notation from now on. Now, of course, a general vector $\boldsymbol{A}$, in Σ, has *components* $\{A^{\alpha}\}$, which will transform as:

$$A'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}} A^{\beta} \tag{3.1}$$

Hence, just to stress the point once more: We have a transformation of the *components* of some vector $\boldsymbol{A}$ from coordinate system Σ to Σ'. So, for example, if the vector is $\boldsymbol{A} = (A^0, A^1, A^2, A^3)$, then its components in the frame Σ' are given by:

$$\begin{aligned} A'^0 &= \sum_{\beta=0}^{3} \Lambda^0{}_{\beta} A^{\beta} = \Lambda^0{}_0 A^0 + \Lambda^0{}_1 A^1 + \Lambda^0{}_2 A^2 + \Lambda^0{}_3 A^3 \\ &\vdots \\ A'^3 &= \sum_{\beta=0}^{3} \Lambda^3{}_{\beta} A^{\beta} = \Lambda^3{}_0 A^0 + \Lambda^3{}_1 A^1 + \Lambda^3{}_2 A^2 + \Lambda^3{}_3 A^3 \end{aligned}$$

3.1.2 Transformation of the Basis Vectors

Now, let us consider *basis vectors*. They are the set of vectors:

$$\begin{aligned} \boldsymbol{e}_0 &= (1,0,0,0) \\ \boldsymbol{e}_1 &= (0,1,0,0) \\ &\vdots \\ \boldsymbol{e}_3 &= (0,0,0,1) \end{aligned}$$

Notice: no 'prime', hence basis vectors in Σ. Now, notice that they may all be seen to satisfy:

$$(\boldsymbol{e}_{\alpha})^{\beta} = \delta_{\alpha}^{\beta} \tag{3.2}$$

Where we have denoted the different vectors themselves by the α-subscript, but the component by the superscript β. This is still in accord with the previous notation of A^{α} being the α^{th} component of $\boldsymbol{A}$. Now, as we know, a vector $\boldsymbol{A}$ is the sum of its components, and corresponding basis vectors:

$$\boldsymbol{A} = A^{\alpha} \boldsymbol{e}_{\alpha} \tag{3.3}$$

This highlights the use of our summation convention: a repeated index must appear both upper and lower, to be summed over. Now, we alluded to this earlier, with the discussion on the vector $\Delta\boldsymbol{x}$: a vector in Σ is the same as a vector in Σ'. That is:

$$A^{\alpha}\boldsymbol{e}_{\alpha} = A'^{\alpha}\boldsymbol{e}'_{\alpha}$$

Now, we have a transformation between A'^{α} and A^{α}, namely (3.1); hence, using this on the RHS of the above:

$$\begin{aligned} A^{\alpha}\boldsymbol{e}_{\alpha} &= A'^{\alpha}\boldsymbol{e}'_{\alpha} \\ &= \frac{\partial x'^{\alpha}}{\partial x^{\beta}} A^{\beta}\boldsymbol{e}'_{\alpha} \\ &= A^{\beta}\frac{\partial x'^{\alpha}}{\partial x^{\beta}}\boldsymbol{e}'_{\alpha} \end{aligned}$$

Where the last line just shuffled the expressions around. Now, as the RHS has no indices the same as the LHS (i.e. two summations on the RHS), we can change indices at will . Let us change $\beta \to \alpha$ and $\alpha' \to \beta'$. Then:

$$A^{\alpha}\boldsymbol{e}_{\alpha} = A^{\alpha}\frac{\partial x'^{\beta}}{\partial x^{\alpha}}\boldsymbol{e}'_{\beta}$$

Hence, taking the RHS to the LHS, and taking out the common factor:

$$A^{\alpha}\left(\boldsymbol{e}_{\alpha} - \frac{\partial x'^{\beta}}{\partial x^{\alpha}}\boldsymbol{e}'_{\beta}\right) = 0$$

Therefore:

$$\boldsymbol{e}_{\alpha} - \frac{\partial x'^{\beta}}{\partial x^{\alpha}}\boldsymbol{e}'_{\beta} = 0$$

Hence:

$$\boldsymbol{e}_{\alpha} = \frac{\partial x'^{\beta}}{\partial x^{\alpha}}\boldsymbol{e}_{\bar{\beta}} \qquad (3.4)$$

Hence, we have a transformation of the *basis vectors* in Σ' to those in Σ. Notice that it is different to the transformation of components. We write both transformations below, as way of revision:

$$\boldsymbol{e}_{\alpha} = \frac{\partial x'^{\beta}}{\partial x^{\alpha}}\boldsymbol{e}'_{\beta} \qquad\qquad A'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}}A^{\beta}$$

3.1.3 Inverse Transformations

Now, we see that the transformation matrix $\Lambda^{\beta}{}_{\alpha}$ is just a function of velocity, which we shall denote $\Lambda^{\beta}{}_{\alpha}(v)$. Hence, just writing down the basis vector transformation again, with this 'notation':

$$\boldsymbol{e}_{\alpha} = \Lambda^{\beta}{}_{\alpha}(v)\boldsymbol{e}'_{\beta} \qquad (3.5)$$

Now, let us conceive that the inverse transformation is given by swapping v for $-v$. Also, we must be careful of the position of the 'primes' in the above expression. Then[12]:

$$\boldsymbol{e}'_{\mu} = \Lambda^{\nu}{}_{\mu}(-v)\boldsymbol{e}_{\nu}$$

[12]Just to avoid any confusion that may occur: all sub-& super-scripts here are ν, and the argument of the transformation v.

Then, relabelling $\nu \to \beta$:

$$\boldsymbol{e'_\beta} = \Lambda^{\nu}{}_{\beta}(-v)\boldsymbol{e}_\nu$$

Hence, if we use this in the far RHS expression of (3.5):

$$\begin{aligned} \boldsymbol{e}_\alpha &= \Lambda^{\beta}{}_{\alpha}(v)\boldsymbol{e'_\beta} \\ &= \Lambda^{\beta}{}_{\alpha}(v)\Lambda^{\nu}{}_{\beta}(-v)\boldsymbol{e}_\nu \end{aligned}$$

We then see that we must have:

$$\Lambda^{\beta}{}_{\alpha}(v)\Lambda^{\nu}{}_{\beta}(-v) = \delta^{\nu}_{\alpha}$$

So that plugging it back in:

$$\boldsymbol{e}_\alpha = \delta^{\nu}{}_{\alpha}\boldsymbol{e}_\nu$$

Hence, writing our 'supposed inverse' identity another way (swap the expressions over):

$$\Lambda^{\nu}{}_{\beta}(-v)\Lambda^{\beta}{}_{\alpha}(v) = \delta^{\nu}{}_{\alpha}$$

We see that it is now of the form of matrix multiplication, where:

$$\Lambda^{-1}\Lambda = \mathsf{I}$$

Which is the standard rule for multiplication of a matrix with its inverse giving the identity matrix. Thus, leaving out the $v, -v$ notation, and using some more common notation:

$$\Lambda_{\beta}{}^{\nu}\Lambda^{\beta}{}_{\alpha} = \delta^{\nu}_{\alpha} \tag{3.6}$$

Let us write this in our other notation; as it will seem more transparent:

$$\frac{\partial x^\nu}{\partial x'^\beta}\frac{\partial x'^\beta}{\partial x^\alpha} = \delta^{\nu}_{\alpha} \qquad \Lambda_{\mu}{}^{\nu} = \frac{\partial x^\nu}{\partial x'^\mu}$$

That is, the transparency comes when we notice that the two factors of $\partial x'^\beta$ effectively cancel out, leaving something we know to be a Kronecker-delta.

Now, let us figure out the inverse transformation of components; that is, if we have the components in Σ', let us work them out in Σ. So, let us start with an expression we had before (3.1) $A'^\alpha = \Lambda^{\alpha}{}_{\beta}A^\beta$. Now, let us multiply both sides by $\Lambda_{\alpha}{}^{\nu}$ (i.e. the inverse of $\Lambda^{\alpha}{}_{\beta}$):

$$\begin{aligned} A'^\alpha &= \frac{\partial x'^\alpha}{\partial x^\beta}A^\beta \\ \Rightarrow \frac{\partial x^\nu}{\partial x'^\alpha}A'^\alpha &= \frac{\partial x^\nu}{\partial x'^\alpha}\frac{\partial x'^\alpha}{\partial x^\beta}A^\beta \end{aligned}$$

Now, we know that the RHS ends up being a Kronecker-delta. Thus, being careful with the indices:

$$\Rightarrow \frac{\partial x^\nu}{\partial x'^\alpha}A'^\alpha = \delta^{\nu}_{\beta}A^\beta = A^\nu$$

Where we have used the property of the Kronecker delta. Hence:

$$A^\nu = \frac{\partial x^\nu}{\partial x'^\alpha}A'^\alpha \tag{3.7}$$

So, we have a way of finding the components of a vector in some frame Σ', if they are known in Σ; and an inverse transformation in going from Σ' to Σ. Here they are:

$$A'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}} A^{\beta} \qquad A^{\beta} = \frac{\partial x^{\beta}}{\partial x'^{\alpha}} A'^{\alpha}$$

Let us finish by stating a few things.
The *magnitude* of a vector is:

$$\boldsymbol{A}^2 = -(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2$$

Where the quantity is frame invariant. Now, we have the following cases:

- If $\boldsymbol{A}^2 > 0$, then we call the vector *spacelike*;
- If $\boldsymbol{A}^2 = 0$, then the vector is called *null*;
- If $\boldsymbol{A}^2 < 0$, then we call the vector *timelike*.

We also have that the scalar product between two vectors is:

$$\boldsymbol{A} \cdot \boldsymbol{B} \equiv -A^0 B^0 + A^1 B^1 + A^2 B^2 + A^3 B^3$$

Which is also invariant. The basis vectors follow orthogonality, via the Minkowski metric:

$$\boldsymbol{e}_{\alpha} \cdot \boldsymbol{e}_{\beta} = \eta_{\alpha\beta} \qquad \boldsymbol{e}_{\bar{\alpha}} \cdot \boldsymbol{e}_{\bar{\beta}} = \eta_{\bar{\alpha}\bar{\beta}} \tag{3.8}$$

Where we have components as already discussed:

$$\eta_{00} = -1 \qquad \eta_{11} = \eta_{22} = \eta_{33} = 1$$

Infact, (3.8) describes a very general relation for any metric; the metric describes the way in which different basis vectors "interact". It is obviously the Kronecker-delta for Cartesian basis vectors.

This concludes our short discussion on vectors, and is mainly a way to introduce notation for use in the following section.

3.2 Tensors

Now, in Σ, if we have two vectors:

$$\boldsymbol{A} = A^{\alpha} \boldsymbol{e}_{\alpha} \qquad \boldsymbol{B} = B^{\beta} \boldsymbol{e}_{\beta}$$

Then their scalar product, via (3.8), is:

$$\begin{aligned} \boldsymbol{A} \cdot \boldsymbol{B} &= A^{\alpha} B^{\beta} \boldsymbol{e}_{\alpha} \cdot \boldsymbol{e}_{\beta} \\ &= A^{\alpha} B^{\beta} \eta_{\alpha\beta} \end{aligned}$$

So, we say that $\eta_{\alpha\beta}$ are *components of the metric tensor*. It gives a way of going from two vectors to a real number. Let us define the term tensor a bit more succinctly:

A tensor of type $\binom{0}{N}$ is a function of N vectors, into the real numbers; where the function is linear in each of its N arguments.

So, linearity means:

$$(\alpha \boldsymbol{A}) \cdot \boldsymbol{B} = \alpha(\boldsymbol{A} \cdot \boldsymbol{B}) \qquad (\boldsymbol{A} + \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{C} + \boldsymbol{B} \cdot \boldsymbol{C}$$

Each operation obviously also working on the $\boldsymbol{B}$ equivalently. This actually gives the definition of a vector space. To distinguish it from the 'vectors' we had before, let us call it the *dual vector space.*

Now, to see how some tensor g works as a function. Let it be a tensor of type $\binom{0}{2}$, so that it takes two vector arguments. Let us notate this just as we would a 'normal function' taking a set of scalar arguments. So:

$$\mathsf{g}(\boldsymbol{A}, \boldsymbol{B}) = a,\ a \in \mathbb{R} \qquad \mathsf{g}(\ ,\ )$$

The first expression shows that when two vectors are put into the function (the tensor), a real number 'comes out'. The second notation means that some function takes two arguments into the 'gaps'. Let the tensor be the metric tensor we met before:

$$\mathsf{g}(\boldsymbol{A}, \boldsymbol{B}) \equiv \boldsymbol{A} \cdot \boldsymbol{B} = A^\alpha B^\beta \eta_{\alpha\beta}$$

So, the function takes two vectors, and gives out a real number, and it does so by 'dotting' the vectors.

Note, the definition thus far has been regardless of reference frame: a real number is obtained irrespective of reference frame. In this way, we can think of tensors as being functions of vectors themselves, rather than their components.

As we just hinted at, a 'normal function', such as $f(t, x, y, z)$ takes scalar arguments (and therefore zero vector arguments), and gives a real number. Therefore, it is a tensor of type $\binom{0}{0}$.

Now, tensors have components, just as vectors did. So, we define:

In Σ, a tensor of type $\binom{0}{N}$ has components, whose values are of the function, when its arguments are the basis vectors $\{\boldsymbol{e}_\alpha\}$ in Σ.

Therefore, a tensors components are frame dependent, because basis vectors are frame dependent.

So, we can say that the metric tensor has elements given by the function acting upon the basis vectors, which are its arguments. That is:

$$\mathsf{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) = \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = \eta_{\alpha\beta}$$

Where we have used that the function is the dot-product.

Let us start to restrict ourselves to specific classes of tensors. The first being *covariant* vectors.

3.2.1 The $\binom{0}{1}$ Tensors

These are commonly called (interchangeably): *one-forms, covariant vector* or *covector.*

Now, let $\tilde{p}$ be some one-form (we used the tilde-notation just as we did the arrow-notation for vectors). So, by definition, $\tilde{p}$ with one vector argument gives a real number:

$$\tilde{p}(\boldsymbol{A}) = a \qquad a \in \mathbb{R}$$

Now, suppose we also have another one-form $\tilde{q}$. Then, we are able to form two other one-forms via:

$$\tilde{s} = \tilde{p} + \tilde{q} \qquad \tilde{r} = \alpha\tilde{p}$$

That is, having their single vector arguments:

$$\tilde{s}(\boldsymbol{A}) = \tilde{p}(\boldsymbol{A}) + \tilde{q}(\boldsymbol{A}) \qquad \tilde{r}(\boldsymbol{A}) = \alpha\tilde{p}(\boldsymbol{A})$$

Thus, we have our dual vector space.

Now, the components of $\tilde{p}$ are p_α. Then, by definition:

$$p_\alpha = \tilde{p}(\boldsymbol{e}_\alpha) \tag{3.9}$$

Hence, giving the one-form a vector argument, and writing the vector as the sum of its components and basis vectors:

$$\tilde{p}(\boldsymbol{A}) = \tilde{p}(A^\alpha \boldsymbol{e}_\alpha)$$

But, using linearity, we may pull the 'number' (i.e. the component part) outside:

$$\tilde{p}(\boldsymbol{A}) = A^\alpha \tilde{p}(\boldsymbol{e}_\alpha)$$

However, we notice (3.9). Hence:

$$\tilde{p}(\boldsymbol{A}) = A^\alpha p_\alpha \tag{3.10}$$

This result may be thought about in the same way the dot-product *used* to be thought about: it has no metric. Hence, doing the above sum:

$$\tilde{p}(\boldsymbol{A}) = A^0 p_0 + A^1 p_1 + A^2 p_2 + A^3 p_3$$

Now, let us consider the transformation of $\tilde{p}$. Let us consider the following:

$$p'_\beta \equiv \tilde{p}(\boldsymbol{e'}_\beta)$$

This is the exact same thing as (3.9), except using the basis in Σ'. Hence, using the inverse transformation of the basis vectors:

$$\begin{aligned} p_{\bar{\beta}} &= \tilde{p}(\boldsymbol{e'}_\beta) \\ &= \tilde{p}({\Lambda_\beta}^\alpha \boldsymbol{e}_\alpha) \\ &= {\Lambda_\beta}^\alpha \tilde{p}(\boldsymbol{e}_\alpha) \\ &= {\Lambda_\beta}^\alpha p_\alpha \end{aligned}$$

That is, using differential notation for the transformation:

$$p'_\beta = \frac{\partial x^\alpha}{\partial x'^\beta} p_\alpha \tag{3.11}$$

Hence, we see that the components of the one-forms transform (if we look back) like the basis vectors, which is in the opposite way to components of vectors. That is, the components of a one-form (which is also known as a covariant vector) transform using the inverse transformation.

Now, as it is a very important result, we shall prove that $A^\alpha p_\alpha$ is frame invariant. So:

$$\begin{aligned} A'^\alpha p'_\alpha &= \frac{\partial x'^\alpha}{\partial x^\beta} A^\beta \frac{\partial x^\mu}{\partial x'^\alpha} p_\mu \\ &= \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x'^\alpha}{\partial x^\beta} A^\beta p_\mu \\ &= \delta^\mu_\beta A^\beta p_\mu \\ &= A^\mu p_\mu \end{aligned}$$

Hence proven. Just to recap how we did this proof: first write down the transformation of the components, then reshuffle the transformation matrices into a form which is recognisable by the Kronecker delta.

Hence, as we have seen, the components of one-forms transformed in the same way (or 'with') basis vectors did. Hence, we sometimes call them *covectors*.
And, in the same way, the components of vectors transformed opposite (or 'contrary') to basis vectors; and are hence sometimes called *contravariant vectors*.

One-forms Basis Now, as one forms form a vector space (the dual vector space), let us find a set of 4 linearly independent one-forms to use as a basis. Let us denote such a basis:

$$\{\tilde{\omega}^\alpha\} \qquad \alpha \in [0, 3]$$

So that a one-form can be written in the same form that a contravariant (as we shall now call 'vectors) vector was, in terms of its basis vectors:

$$\tilde{p} = p_\alpha \tilde{\omega}^\alpha \qquad \boldsymbol{A} = A^\alpha \boldsymbol{e}_\alpha$$

Now, let us find out what this basis is. Now, from above, let us give the one-form a vector argument:

$$\tilde{p}(\boldsymbol{A}) = p_\alpha \tilde{\omega}^\alpha(\boldsymbol{A})$$

And we do our usual thing of expanding the contravariant vector, and using linearity:

$$\begin{aligned} \tilde{p}(\boldsymbol{A}) &= p_\alpha \tilde{\omega}^\alpha(\boldsymbol{A}) \\ &= p_\alpha \tilde{\omega}^\alpha(A^\beta \boldsymbol{e}_\beta) \\ &= p_\alpha A^\beta \tilde{\omega}^\alpha(\boldsymbol{e}_\beta) \end{aligned}$$

From (3.10), which is that $\tilde{p}(\boldsymbol{A}) = A^\alpha p_\alpha$, we see that we must have:

$$\tilde{\omega}^\alpha(\boldsymbol{e}_\beta) = \delta^\alpha_\beta \tag{3.12}$$

Hence, we have a way of generating the β^{th} component of the α^{th} one-form basis. They are, not suprisingly:

$$\begin{aligned}\tilde{\omega}^0 &= (1,0,0,0)\\ \tilde{\omega}^1 &= (0,1,0,0)\\ \tilde{\omega}^2 &= (0,0,1,0)\\ \tilde{\omega}^3 &= (0,0,0,1)\end{aligned}$$

And, if we look carefully, we have a way of transforming the basis vectors for one-forms:

$$\tilde{\omega'}^\alpha = \frac{\partial x'^\alpha}{\partial x^\beta}\tilde{\omega}^\beta \tag{3.13}$$

Which was in the same way as transforming the components of a contravarient vector.

3.2.2 Derivative Notation

We shall use this a lot more in the later sections, but it is useful to see how this links in; and it is infact a more transparent notation as to what is going on.

Consider the derivative of some function $f(x,y)$:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = \frac{\partial f}{\partial x^i}dx^i$$

Then its not too hard to see that if we have more than one function, we have:

$$df^j = \frac{\partial f^j}{\partial x^i}dx^i$$

Then, going back to thinking about coordinates & transformations between frames, if a set of coordinates is dependant upon another set, then they will transform:

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu}A^\nu$$

This is infact the same as our defintion of the transformation of a contravaraint vector:

$$\Lambda^\mu{}_\nu \equiv \frac{\partial x'^\mu}{\partial x^\nu}$$

And the covariant vector transforms as:

$$A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}A_\nu$$

With:

$$\Lambda_\mu{}^\nu \equiv \frac{\partial x^\nu}{\partial x'^\mu}$$

And notice how the transformation and the inverse 'sit' next to each other:

$$\Lambda^{\mu}{}_{\rho}\Lambda_{\mu}{}^{\nu} = \frac{\partial x'^{\mu}}{\partial x^{\rho}}\frac{\partial x^{\nu}}{\partial x'^{\mu}} = \frac{\partial x^{\nu}}{\partial x^{\rho}} = \delta^{\nu}_{\rho}$$

Which is what we had before. We see that writing as differentials is a little more transparent, as it highlights the 'cancelling' nature of the differentials.

We shall now use index notation in a much more useful way: in the beginnings of the general theory of relativity.

3.3 The Geodesic

This will initially require knowledge of the calculus of variations result for minimising the functional:

$$I[y] = \int F(y(x), y'(x), x)dx \qquad y' = \frac{dy}{dx}$$

That is, to minimise I, one computes:

$$\frac{d}{dx}\frac{\partial F}{\partial y'} - \frac{\partial F}{\partial x} = 0$$

Now, the problem at hand is to compute the path between two points, which minimises that path length. That is, the total path is:

$$l = \int_0^s ds$$

Now, the line element is given by:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu$$

Multiplying & dividing by dt^2:

$$ds^2 = g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}dt^2$$

Square-rooting & putting back into the integral:

$$l = \int_0^s ds = \int \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}dt$$

After using the following standard notation:

$$\frac{dx^\mu}{dt} = \dot{x}^\mu$$

Where an over-dot denotes temporal derivative. So, we have that to minimise our functional length l, is to use:

$$F(x, \dot{x}, t) = \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} \qquad \dot{x^\mu} \equiv \frac{dx^\mu}{dt}$$

In the Euler-Lagrange equation:

$$\frac{d}{dt}\frac{\partial F}{\partial \dot{x}^\mu} - \frac{\partial F}{\partial x^\mu} = 0 \tag{3.14}$$

Now, the only things we must state about the form of our metric $g_{\mu\nu}$ is that it is a function of x only (i.e. not of $\dot{x}$).

Let us begin by differentiating F with respect to a temporal derivative. That is, compute:

$$\frac{\partial F}{\partial \dot{x}^\mu}$$

So:

$$\begin{aligned}
\frac{\partial F}{\partial \dot{x}^\mu} &= \frac{\partial}{\partial \dot{x}^\rho}\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} \\
&= \tfrac{1}{2}\left(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu\right)^{-1/2}\frac{\partial}{\partial \dot{x}^\rho}\left(g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta\right) \\
&= \frac{1}{2\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}}g_{\alpha\beta}\left(\dot{x}^\alpha\frac{\partial \dot{x}^\beta}{\partial \dot{x}^\rho} + \dot{x}^\beta\frac{\partial \dot{x}^\alpha}{\partial \dot{x}^\rho}\right)
\end{aligned}$$

Where in the first line we used the standard rule for differentiating a function. Then we used the product rule. Notice that we must be careful in relabeling indices. We continue:

$$\begin{aligned}\frac{\partial F}{\partial \dot{x}^\mu} &= \frac{1}{2\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}} g_{\alpha\beta}\left(\dot{x}^\alpha \frac{\partial \dot{x}^\beta}{\partial \dot{x}^\rho} + \dot{x}^\beta \frac{\partial \dot{x}^\alpha}{\partial \dot{x}^\rho}\right)\\ &= \frac{1}{2\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}} g_{\alpha\beta}\left(\dot{x}^\alpha \delta^\beta_\rho + \dot{x}^\beta \delta^\alpha_\rho\right)\\ &= \frac{1}{2\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}} \left(g_{\alpha\beta}\dot{x}^\alpha \delta^\beta_\rho + g_{\alpha\beta}\dot{x}^\beta \delta^\alpha_\rho\right)\\ &= \frac{1}{2\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}} \left(g_{\alpha\rho}\dot{x}^\alpha + g_{\rho\beta}\dot{x}^\beta\right)\end{aligned}$$

Finally, we use the symmetry of the metric: $g_{\mu\nu} = g_{\nu\mu}$. Then, the above becomes:

$$\frac{\partial F}{\partial \dot{x}^\mu} = \frac{1}{\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}} g_{\rho\alpha}\dot{x}^\alpha$$

Finally, we notice that:

$$\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} = \frac{ds}{dt} \equiv \dot{s}$$

And therefore:

$$\frac{\partial F}{\partial \dot{x}^\mu} = \frac{g_{\mu\nu}\dot{x}^\nu}{\dot{s}} \tag{3.15}$$

Let us compute the second term in the Euler-Lagrange equation:

$$\frac{\partial F}{\partial x^\mu}$$

Thus:

$$\begin{aligned}\frac{\partial F}{\partial x^\rho} &= \frac{1}{2\dot{s}}\frac{\partial}{\partial x^\rho}\left(g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta\right)\\ &= \frac{1}{2\dot{s}}\dot{x}^\alpha\dot{x}^\beta\frac{\partial g_{\alpha\beta}}{\partial x^\rho}\end{aligned}$$

After noting that $\frac{\partial \dot{x}^\mu}{\partial x^\nu} = 0$ in the relevant product-rule stage. So, changing indices, the above is just:

$$\frac{\partial F}{\partial x^\rho} = \frac{\dot{x}^\mu\dot{x}^\nu}{2\dot{s}}\frac{\partial g_{\mu\nu}}{\partial x^\rho} \tag{3.16}$$

And therefore, we have the two components of the Euler-Lagrange equation, which we then substitute in. So, putting (3.15) and (3.16) into (3.14):

$$\frac{d}{dt}\left(\frac{g_{\mu\nu}\dot{x}^\nu}{\dot{s}}\right) - \frac{\dot{x}^\mu\dot{x}^\nu}{2\dot{s}}\frac{\partial g_{\mu\nu}}{\partial x^\rho} = 0 \tag{3.17}$$

Let us evaluate the time derivative on the far LHS:

$$\begin{aligned}\frac{d}{dt}\left(\frac{g_{\mu\nu}\dot{x}^\nu}{\dot{s}}\right) &= \frac{1}{\dot{s}}\left(\dot{x}^\nu\frac{d}{dt}(g_{\mu\nu}) + g_{\mu\nu}\ddot{x}^\nu\right) - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}^2}\\ &= \frac{g_{\mu\nu}\ddot{x}^\nu}{\dot{s}} + \frac{\dot{x}^\nu}{\dot{s}^2}\left(\dot{s}\frac{d}{dt}(g_{\mu\nu}) - g_{\mu\nu}\ddot{s}\right)\end{aligned}$$

Where in the final step we just took out a common factor. Now, we shall proceed by doing things which seem pointless, but will end up giving a nice result. So just bear with it! Now, notice that we can do the following:

$$\frac{d}{dt}g_{\mu\nu} = \frac{dg_{\mu\nu}}{dt}\frac{dx^\rho}{dx^\rho} = \frac{dg_{\mu\nu}}{dx^\rho}\frac{dx^\rho}{dt} = \frac{dg_{\mu\nu}}{dx^\rho}\dot{x}^\rho$$

Hence, we shall use this to give:

$$\frac{d}{dt}\left(\frac{g_{\mu\nu}\dot{x}^\nu}{\dot{s}}\right) = \frac{g_{\mu\nu}\ddot{x}^\nu}{\dot{s}} + \frac{\dot{x}^\nu}{\dot{s}^2}\left(\dot{s}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\rho - g_{\mu\nu}\ddot{s}\right)$$

Expanding out:

$$\frac{d}{dt}\left(\frac{g_{\mu\nu}\dot{x}^\nu}{\dot{s}}\right) = \frac{g_{\mu\nu}\ddot{x}^\nu}{\dot{s}} + \frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{\dot{x}^\nu\dot{x}^\rho}{\dot{s}} - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}^2}$$

Hence, we put this result back into (3.17):

$$\frac{g_{\mu\nu}\ddot{x}^\nu}{\dot{s}} + \frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{\dot{x}^\nu\dot{x}^\rho}{\dot{s}} - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}^2} - \frac{\dot{x}^\mu\dot{x}^\nu}{2\dot{s}}\frac{\partial g_{\mu\nu}}{\partial x^\rho} = 0$$

Let us cancel off a single factor of $\dot{s}$:

$$g_{\mu\nu}\ddot{x}^\nu + \frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\nu\dot{x}^\rho - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}} - \frac{\dot{x}^\mu\dot{x}^\nu}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho} = 0$$

Just so we dont get lost in the mathematics, we shall just recap what we are doing: the above is an expression which will minimise the length of a path between two points, in a space with a metric. Thus far, this is an incredibly general expression.
Now, in the second-from-left expression, change the indices thus: $\mu \to \rho, \rho \to \mu$. Giving:

$$\frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\nu\dot{x}^\rho \longmapsto \frac{\partial g_{\rho\nu}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\mu$$

Then, putting this in:

$$g_{\mu\nu}\ddot{x}^\nu + \frac{\partial g_{\rho\nu}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\mu - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}} - \frac{\dot{x}^\mu\dot{x}^\nu}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho} = 0$$

Collecting terms:

$$g_{\mu\nu}\ddot{x}^\nu + \left(\frac{\partial g_{\rho\nu}}{\partial x^\mu} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\right)\dot{x}^\nu\dot{x}^\mu - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}} = 0 \tag{3.18}$$

Now, consider the simple expression that $a = b$. Then, it is obvious that $a = \frac{1}{2}(a+b)$. So, if we have that:

$$\frac{\partial g_{\rho\nu}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\mu = \frac{\partial g_{\rho\mu}}{\partial x^\nu}\dot{x}^\mu\dot{x}^\nu$$

Which is true just by interchange of the indices $\mu \leftrightarrow \nu$. Then, by our simple expression, we have that:

$$\frac{\partial g_{\rho\nu}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\mu = \frac{1}{2}\left(\frac{\partial g_{\rho\nu}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\mu + \frac{\partial g_{\rho\mu}}{\partial x^\nu}\dot{x}^\mu\dot{x}^\nu\right)$$

Hence, using this in the first of the bracketed-terms in (3.18):

$$g_{\mu\nu}\ddot{x}^\nu + \left(\frac{1}{2}\left(\frac{\partial g_{\rho\nu}}{\partial x^\mu} + \frac{\partial g_{\rho\mu}}{\partial x^\nu}\right) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\right)\dot{x}^\nu\dot{x}^\mu - \frac{g_{\mu\nu}\dot{x}^\nu\ddot{s}}{\dot{s}} = 0$$

Which is of course just:

$$g_{\mu\nu}\ddot{x}^{\nu} + \frac{1}{2}\left(\frac{\partial g_{\rho\nu}}{\partial x^{\mu}} + \frac{\partial g_{\rho\mu}}{\partial x^{\nu}} - \frac{\partial g_{\mu\nu}}{\partial x^{\rho}}\right)\dot{x}^{\nu}\dot{x}^{\mu} - \frac{g_{\mu\nu}\dot{x}^{\nu}\ddot{s}}{\dot{s}} = 0$$

Now, let us use some notation. Let us define the bracketed-quantity:

$$[\mu\nu, \rho] \equiv \frac{1}{2}\left(\frac{\partial g_{\rho\nu}}{\partial x^{\mu}} + \frac{\partial g_{\rho\mu}}{\partial x^{\nu}} - \frac{\partial g_{\mu\nu}}{\partial x^{\rho}}\right)$$

Then we have:

$$g_{\mu\nu}\ddot{x}^{\nu} + [\mu\nu, \rho]\dot{x}^{\nu}\dot{x}^{\mu} - \frac{g_{\mu\nu}\dot{x}^{\nu}\ddot{s}}{\dot{s}} = 0$$

Now, let us choose $\ddot{s} = 0$. This gives:

$$g_{\nu\mu}\ddot{x}^{\mu} + [\mu\sigma, \nu]\dot{x}^{\mu}\dot{x}^{\sigma} = 0$$

Multiply through by $g^{\nu\lambda}$:

$$g^{\nu\lambda}g_{\nu\mu}\ddot{x}^{\mu} + g^{\nu\lambda}[\mu\sigma, \nu]\dot{x}^{\mu}\dot{x}^{\sigma} = 0$$

We also have the relation:

$$g^{\nu\lambda}g_{\nu\mu} = \delta^{\lambda}_{\mu}$$

Giving:

$$\delta^{\lambda}_{\mu}\ddot{x}^{\mu} + g^{\nu\lambda}[\mu\sigma, \nu]\dot{x}^{\mu}\dot{x}^{\sigma} = 0$$

Which is just:

$$\ddot{x}^{\lambda} + g^{\nu\lambda}[\mu\sigma, \nu]\dot{x}^{\mu}\dot{x}^{\sigma} = 0$$

Let us use another piece of notation:

$$\Gamma^{\lambda}{}_{\mu\sigma} \equiv g^{\nu\lambda}[\mu\sigma, \nu]$$

Then:

$$\ddot{x}^{\lambda} + \Gamma^{\lambda}{}_{\mu\sigma}\dot{x}^{\mu}\dot{x}^{\sigma} = 0$$

Changing indices around a bit:

$$\ddot{x}^{\mu} + \Gamma^{\mu}{}_{\nu\sigma}\dot{x}^{\nu}\dot{x}^{\sigma} = 0 \tag{3.19}$$

We have thus computed the equation of a *geodesic*, under the condition that $\ddot{s} = 0$.

3.3.1 Christoffel Symbols

Along the derivation of the geodesic, we introduced the notation of the *Christoffel symbol*; tensorial properties (infact, lack thereof) we shall soon discuss. They are defined:

$$\Gamma^{\lambda}{}_{\mu\nu} \equiv g^{\rho\lambda}[\mu\nu, \rho] \tag{3.20}$$

That is, expanding out the "bracket notation":

$$\Gamma^{\lambda}{}_{\mu\nu} = g^{\rho\lambda}\frac{1}{2}\left(\frac{\partial g_{\rho\nu}}{\partial x^{\mu}} + \frac{\partial g_{\rho\mu}}{\partial x^{\nu}} - \frac{\partial g_{\mu\nu}}{\partial x^{\rho}}\right) \tag{3.21}$$

Let us introduce another piece of notation, the "commma-derivative" notation:

$$y_{\mu\nu,\rho} \equiv \frac{\partial y_{\mu\nu}}{\partial x^\rho}$$

So that the Christoffel symbol looks like:

$$\Gamma^\lambda{}_{\mu\nu} = g^{\rho\lambda}\frac{1}{2}\left(g_{\rho\nu,\mu} + g_{\rho\mu,\nu} - g_{\mu\nu,\rho}\right)$$

We have been introducing a lot of notation that essentially just makes things a lot more compact; allowing us to write more in a smaller space, essentially.

Notice that from the definitions, we can see that the following interchange 'dont do anything':

$$\Gamma^\lambda{}_{\mu\nu} = \Gamma^\lambda{}_{\nu\mu}$$

Let us consider the transformation of the equation of the geodesic.

Let us initially consider the transformation of $\dot{x}^\mu$:

$$\begin{aligned}
\dot{x}^\mu &= \frac{\partial x^\mu}{\partial t} \\
&= \frac{\partial x^\mu}{\partial x'^\nu}\frac{\partial x'^\nu}{\partial t} \\
&= \frac{\partial x^\mu}{\partial x'^\nu}\dot{x}'^\nu
\end{aligned}$$

Which is, under no suprise, the standard rule for transformation of a contravariant tensor, of first rank. Infact, we assumed that $t = t'$ (in actual fact, what we have done is noted the invariance of s, and used this as the 'time derivative). Now let us consider the transformation of $\ddot{x}^\mu$:

$$\begin{aligned}
\ddot{x}^\mu &= \frac{d}{dt}\dot{x}^\mu \\
&= \frac{d}{dt}\left(\frac{\partial x^\mu}{\partial x'^\nu}\dot{x}'^\nu\right) \\
&= \ddot{x}'^\nu\frac{\partial x^\mu}{\partial x'^\nu} + \dot{x}'^\nu\frac{d}{dt}\frac{\partial x^\mu}{\partial x'^\nu} \\
&= \ddot{x}'^\nu\frac{\partial x^\mu}{\partial x'^\nu} + \dot{x}'^\nu\frac{\partial}{\partial t}\frac{\partial x^\mu}{\partial x'^\nu}\frac{\partial x'^\sigma}{\partial x'^\sigma} \\
&= \ddot{x}'^\nu\frac{\partial x^\mu}{\partial x'^\nu} + \frac{\partial^2 x^\mu}{\partial x'^\nu\partial x'^\sigma}\dot{x}'^\sigma\dot{x}'^\nu
\end{aligned}$$

Where we used the same technique for derivatives, that we used in computing the derivative of $g_{\mu\nu}$. Therefore, what we have is:

$$\ddot{x}^\mu = \ddot{x}'^\nu\frac{\partial x^\mu}{\partial x'^\nu} + \frac{\partial^2 x^\mu}{\partial x'^\nu\partial x'^\sigma}\dot{x}'^\sigma\dot{x}'^\nu$$

Now, let us substitute these into the equation of the geodesic (3.19), being careful about indices:

$$\ddot{x}'^\nu\frac{\partial x^\mu}{\partial x'^\nu} + \frac{\partial^2 x^\mu}{\partial x'^\nu\partial x'^\sigma}\dot{x}'^\sigma\dot{x}'^\nu + \Gamma^\mu{}_{\nu\sigma}\frac{\partial x^\nu}{\partial x'^\alpha}\frac{\partial x^\sigma}{\partial x'^\beta}\dot{x}'^\alpha\dot{x}'^\beta = 0$$

We see that we can change the indices in the middle expression $\nu \to \alpha, \sigma \to \beta$, so that we can take out a common factor:

$$\ddot{x}'^{\nu} \frac{\partial x^\mu}{\partial x'^\nu} + \left(\frac{\partial^2 x^\mu}{\partial x'^\alpha \partial x'^\beta} + \Gamma^\mu{}_{\nu\sigma} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} \right) \dot{x}'^\alpha \dot{x}'^\beta = 0$$

Now let us multiply by a factor of $\frac{\partial x'^\rho}{\partial x^\mu}$:

$$\frac{\partial x'^\rho}{\partial x^\mu} \ddot{x}'^{\nu} \frac{\partial x^\mu}{\partial x'^\nu} + \frac{\partial x'^\rho}{\partial x^\mu} \left(\frac{\partial^2 x^\mu}{\partial x'^\alpha \partial x'^\beta} + \Gamma^\mu{}_{\nu\sigma} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} \right) \dot{x}'^\alpha \dot{x}'^\beta = 0$$

The first expression reduces slightly:

$$\ddot{x}'^{\nu} \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial x^\mu}{\partial x'^\nu} = \ddot{x}'^{\nu} \delta^\rho_\nu = \ddot{x}'^{\rho}$$

Thus, using this:

$$\ddot{x}'^{\rho} + \frac{\partial x'^\rho}{\partial x^\mu} \left(\frac{\partial^2 x^\mu}{\partial x'^\alpha \partial x'^\beta} + \Gamma^\mu{}_{\nu\sigma} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} \right) \dot{x}'^\alpha \dot{x}'^\beta = 0$$

Now, consider the equation of the geodesic, in the primed frame:

$$\ddot{x}'^{\mu} + \Gamma'^\mu{}_{\nu\sigma} \dot{x}'^\nu \dot{x}'^\sigma = 0$$

Upon (a very careful) comparison, we see that:

$$\Gamma'^\rho{}_{\alpha\beta} = \frac{\partial x'^\rho}{\partial x^\mu} \left(\frac{\partial^2 x^\mu}{\partial x'^\alpha \partial x'^\beta} + \Gamma^\mu{}_{\nu\sigma} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} \right) \tag{3.22}$$

Expanding out:

$$\Gamma'^\rho{}_{\alpha\beta} = \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} \Gamma^\mu{}_{\nu\sigma} + \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial^2 x^\mu}{\partial x'^\alpha \partial x'^\beta}$$

Now, this is *almost* the expression for the transformation of a mixed tensor. A mixed first rank contravariant, second rank covariant tensor would transform as:

$$A'^\rho{}_{\alpha\beta} = \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial x^\nu}{\partial x'^\alpha} \frac{\partial x^\sigma}{\partial x'^\beta} A^\mu{}_{\nu\sigma}$$

But, as we saw, $\Gamma^\rho{}_{\alpha\beta}$ does not conform to this (due to the presence of the second additive term). Therefore, the Christoffel symbol does not transform as a tensor. Thus the reason we did not call it a tensor! Just to reiterate the point: $\Gamma'^\rho{}_{\alpha\beta}$ are not components of a tensor.

Contravariant Differentiation We shall now use the Christoffel symbols, but to do so will take some algebra!

Consider the transformation of the contravariant vector:

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu$$

Let us differentiate this, with respect to a contravariant component in the primed frame:

$$\frac{\partial}{\partial x'^\rho} A'^\mu = \frac{\partial}{\partial x'^\rho} \left(\frac{\partial x'^\mu}{\partial x^\nu} A^\nu \right)$$

Let us multiply and divide by the same factor, on the RHS:

$$\frac{\partial}{\partial x'^{\rho}} A'^{\mu} = \frac{\partial x^{\alpha}}{\partial x^{\alpha}} \frac{\partial}{\partial x'^{\rho}} \left(\frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu} \right)$$

Which is of course:

$$\frac{\partial}{\partial x'^{\rho}} A'^{\mu} = \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \frac{\partial}{\partial x^{\alpha}} \left(\frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu} \right)$$

Using the product rule:

$$\begin{aligned} \frac{\partial A'^{\mu}}{\partial x'^{\rho}} &= \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \left(\frac{\partial^2 x'^{\mu}}{\partial x^{\alpha} \partial x^{\nu}} A^{\nu} + \frac{\partial x'^{\mu}}{\partial x^{\nu}} \frac{\partial A^{\nu}}{\partial x^{\alpha}} \right) \\ &= \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \frac{\partial x'^{\mu}}{\partial x^{\nu}} \frac{\partial A^{\nu}}{\partial x^{\alpha}} + \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \frac{\partial^2 x'^{\mu}}{\partial x^{\alpha} \partial x^{\nu}} A^{\nu} \end{aligned}$$

Using some (simple) notation:

$$B'^{\mu}{}_{\rho} \equiv \frac{\partial A'^{\mu}}{\partial x'^{\rho}} \qquad B^{\mu}{}_{\rho} \equiv \frac{\partial A^{\mu}}{\partial x^{\rho}}$$

We therefore have:

$$B'^{\mu}{}_{\rho} = \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \frac{\partial x'^{\mu}}{\partial x^{\nu}} B^{\nu}{}_{\alpha} + \frac{\partial x^{\alpha}}{\partial x'^{\rho}} \frac{\partial^2 x'^{\mu}}{\partial x^{\alpha} \partial x^{\nu}} A^{\nu} \tag{3.23}$$

Thus, we see that $B'^{\mu}{}_{\rho}$, defined in this way, is not a tensor. Now, to proceed, we shall find an expression for the final expression above, the second derivative. To do so, consider (3.22):

$$\Gamma'^{\rho}{}_{\alpha\beta} = \frac{\partial x'^{\rho}}{\partial x^{\mu}} \left(\frac{\partial^2 x^{\mu}}{\partial x'^{\alpha} \partial x'^{\beta}} + \Gamma^{\mu}{}_{\nu\sigma} \frac{\partial x^{\nu}}{\partial x'^{\alpha}} \frac{\partial x^{\sigma}}{\partial x'^{\beta}} \right)$$

Let us multiply this by $\frac{\partial x^{\kappa}}{\partial x'^{\rho}}$:

$$\frac{\partial x^{\kappa}}{\partial x'^{\rho}} \Gamma'^{\rho}{}_{\alpha\beta} = \frac{\partial x^{\kappa}}{\partial x'^{\rho}} \frac{\partial x'^{\rho}}{\partial x^{\mu}} \left(\frac{\partial^2 x^{\mu}}{\partial x'^{\alpha} \partial x'^{\beta}} + \Gamma^{\mu}{}_{\nu\sigma} \frac{\partial x^{\nu}}{\partial x'^{\alpha}} \frac{\partial x^{\sigma}}{\partial x'^{\beta}} \right)$$

Which is, noting the Kronecker-delta multiplying the bracket:

$$\begin{aligned} \frac{\partial x^{\kappa}}{\partial x'^{\rho}} \Gamma'^{\rho}{}_{\alpha\beta} &= \delta^{\kappa}_{\mu} \left(\frac{\partial^2 x^{\mu}}{\partial x'^{\alpha} \partial x'^{\beta}} + \Gamma^{\mu}{}_{\nu\sigma} \frac{\partial x^{\nu}}{\partial x'^{\alpha}} \frac{\partial x^{\sigma}}{\partial x'^{\beta}} \right) \\ &= \frac{\partial^2 x^{\kappa}}{\partial x'^{\alpha} \partial x'^{\beta}} + \Gamma^{\kappa}{}_{\nu\sigma} \frac{\partial x^{\nu}}{\partial x'^{\alpha}} \frac{\partial x^{\sigma}}{\partial x'^{\beta}} \end{aligned}$$

And thus, trivially rearanging:

$$\frac{\partial^2 x^{\kappa}}{\partial x'^{\alpha} \partial x'^{\beta}} = \frac{\partial x^{\kappa}}{\partial x'^{\rho}} \Gamma'^{\rho}{}_{\alpha\beta} - \Gamma^{\kappa}{}_{\nu\sigma} \frac{\partial x^{\nu}}{\partial x'^{\alpha}} \frac{\partial x^{\sigma}}{\partial x'^{\beta}}$$

We can of course dash un-dashed quantities; and undash dashed quantities:

$$\frac{\partial^2 x'^{\kappa}}{\partial x^{\alpha} \partial x^{\beta}} = \frac{\partial x'^{\kappa}}{\partial x^{\rho}} \Gamma^{\rho}{}_{\alpha\beta} - \Gamma'^{\kappa}{}_{\nu\sigma} \frac{\partial x'^{\nu}}{\partial x^{\alpha}} \frac{\partial x'^{\sigma}}{\partial x^{\beta}}$$

So, to use this in (3.23), we shall change the indices carefully, to:

$$\frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\nu} = \frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu} - \Gamma'^\mu{}_{\lambda\pi}\frac{\partial x'^\lambda}{\partial x^\alpha}\frac{\partial x'^\pi}{\partial x^\nu}$$

Thus, (3.23) can be written:

$$\begin{aligned}
B'^\mu{}_\rho &= \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\nu}B^\nu{}_\alpha + \frac{\partial x^\alpha}{\partial x'^\rho}\left(\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu} - \Gamma'^\mu{}_{\lambda\pi}\frac{\partial x'^\lambda}{\partial x^\alpha}\frac{\partial x'^\pi}{\partial x^\nu}\right)A^\nu \\
&= \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\nu}B^\nu{}_\alpha + \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu}A^\nu - \frac{\partial x^\alpha}{\partial x'^\rho}\Gamma'^\mu{}_{\lambda\pi}\frac{\partial x'^\lambda}{\partial x^\alpha}\frac{\partial x'^\pi}{\partial x^\nu}A^\nu \\
&= \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\nu}B^\nu{}_\alpha + \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu}A^\nu - \delta^\lambda_\rho\Gamma'^\mu{}_{\lambda\pi}\frac{\partial x'^\pi}{\partial x^\nu}A^\nu \\
&= \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\nu}B^\nu{}_\alpha + \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu}A^\nu - \Gamma'^\mu{}_{\rho\pi}A'^\pi
\end{aligned}$$

Therefore, taking over to the other side & removing our trivial notation:

$$\frac{\partial A'^\mu}{\partial x'^\rho} + \Gamma'^\mu{}_{\rho\pi}A'^\pi = \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\nu}\frac{\partial A^\nu}{\partial x^\alpha} + \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu}A^\nu$$

In the middle expression, changing $\nu \to \lambda$:

$$\frac{\partial A'^\mu}{\partial x'^\rho} + \Gamma'^\mu{}_{\rho\pi}A'^\pi = \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\frac{\partial A^\lambda}{\partial x^\alpha} + \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\Gamma^\lambda{}_{\alpha\nu}A^\nu$$

So, taking out the common factor:

$$\frac{\partial A'^\mu}{\partial x'^\rho} + \Gamma'^\mu{}_{\rho\pi}A'^\pi = \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}\left(\frac{\partial A^\lambda}{\partial x^\alpha} + \Gamma^\lambda{}_{\alpha\nu}A^\nu\right)$$

Using the notation that:

$$C'^\mu{}_\rho \equiv \frac{\partial A'^\mu}{\partial x'^\rho} + \Gamma'^\mu{}_{\rho\pi}A'^\pi \qquad C^\lambda{}_\alpha \equiv \frac{\partial A^\lambda}{\partial x^\alpha} + \Gamma^\lambda{}_{\alpha\nu}A^\nu$$

We therefore have:

$$C'^\mu{}_\rho = \frac{\partial x^\alpha}{\partial x'^\rho}\frac{\partial x'^\mu}{\partial x^\lambda}C^\lambda{}_\alpha$$

And hence we have shown that $C^\lambda{}_\alpha$ transforms as a mixed tensor. We write:

$$C^\lambda{}_\alpha = A^\lambda{}_{;\alpha}$$

So that we have an expression for the *covariant derivative of the contravariant vector*:

$$A^\lambda{}_{;\alpha} \equiv \frac{\partial A^\lambda}{\partial x^\alpha} + \Gamma^\lambda{}_{\alpha\nu}A^\nu \tag{3.24}$$

Other notation for this is:

$$A^\lambda{}_{;\alpha} = \partial_\alpha A^\lambda + \Gamma^\lambda{}_{\alpha\nu}A^\nu \equiv A^\lambda{}_{,\alpha} + \Gamma^\lambda{}_{\alpha\nu}A^\nu$$

Notice the difference between the comma and semi-colon derivative notation. Also, due to the afore mentioned symmetry of the 'bottom two indices' of the Christoffel symbol, this is the same as:

$$A^{\lambda}{}_{;\alpha} = \frac{\partial A^{\lambda}}{\partial x^{\alpha}} + \Gamma^{\lambda}{}_{\nu\alpha} A^{\nu}$$

If we were to repeat the whole derivation (which we wont) for the *covariant derivative of covariant vectors*, we would find:

$$A_{\lambda\ ;\alpha} \equiv \frac{\partial A_{\lambda}}{\partial x^{\alpha}} - \Gamma^{\nu}{}_{\alpha\lambda} A_{\nu} \tag{3.25}$$

Let us consider some examples, of higher order tensor derivatives:

- Second rank covariant:
$$A_{\mu\nu;\alpha} = \partial_{\alpha} A_{\mu\nu} - \Gamma^{\beta}{}_{\mu\alpha} A_{\beta\nu} - \Gamma^{\beta}{}_{\nu\alpha} A_{\mu\beta}$$

- Second rank contravariant:

$$A^{\mu\nu}{}_{;\alpha} = \partial_{\alpha} A^{\mu\nu} + \Gamma^{\mu}{}_{\beta\alpha} A^{\beta\nu} + \Gamma^{\nu}{}_{\alpha\beta} A^{\mu\beta}$$

4 More Physical Ideas

Let us now look at coordinate systems, and transformations between them, by considering some less abstract ideas.

4.1 Curvilinear Coordinates

Consider the standard set of transformations, between the 2D polar & Cartesian coordinates:

$$x = r\cos\theta \quad y = r\sin\theta \qquad r = \sqrt{x^2+y^2} \quad \theta = \tan^{-1}\tfrac{y}{x}$$

So that one coordinate system is represented by (x, y), and the other (r, θ). We see that $r(x, y)$ and $\theta(x, y)$. That is, each has its own differential:

$$dr = \frac{\partial r}{\partial x}dx + \frac{\partial r}{\partial y}dy \qquad d\theta = \frac{\partial \theta}{\partial x}dx + \frac{\partial \theta}{\partial y}dy$$

Now, consider some "abstract" set of coordinates (ξ, η), each of which is a function of the set (x, y). Then, each has differential:

$$\begin{aligned} d\xi &= \frac{\partial \xi}{\partial x}dx + \frac{\partial \xi}{\partial y}dy \\ d\eta &= \frac{\partial \eta}{\partial x}dx + \frac{\partial \eta}{\partial y}dy \end{aligned}$$

It should be fairly easy to see that this set of equations can be put into a matrix equation:

$$\begin{pmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} d\xi \\ d\eta \end{pmatrix}$$

Now, consider two distinct points $(x_1, y_1), (x_2, y_2)$. For the transformation to be "good", there must be distinct images of these points, $(\xi_1, \eta_1), (\xi_2, \eta_2)$. So, if the two points are the same, then the difference between the two points is zero. Then, that means that $dx = dy = 0$, as well as $d\xi = d\eta = 0$. That is, we must also have that the determinant of the matrix is non-zero. The determinant is known as the *Jacobian*. If the Jacobian vanishes at a point, then that point is called *singular*.

This should be fairly easy to see how to generalise to any number of dimensions.

Now, one of the ideas we introduced previously, was that the Lorentz transform, multiplied by its inverse, gave the identity matrix. Now, let us show that this is true for any transformation matrix. We shall do it for a 2D one; but generality is easy to see.
In analogue with the above discussion with $\xi(x, y), \eta(x, y)$; we also see that $x(\xi, \eta), y(\xi, \eta)$. So that we have a set of inverse transformations:

$$dx = \frac{\partial x}{\partial \xi}d\xi + \frac{\partial x}{\partial \eta}d\eta \qquad dy = \frac{\partial y}{\partial \xi}d\xi + \frac{\partial y}{\partial \eta}d\eta$$

So that we can see that the inverse transformation matrix is:

$$\begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{pmatrix}$$

Then, multiplying the transformation by its inverse, we have:

$$\begin{aligned} \begin{pmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{pmatrix} &= \begin{pmatrix} \frac{\partial \xi}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial \xi}{\partial y}\frac{\partial y}{\partial \xi} & \frac{\partial \xi}{\partial x}\frac{\partial x}{\partial \eta} + \frac{\partial \xi}{\partial y}\frac{\partial y}{\partial \eta} \\ \frac{\partial \eta}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial \eta}{\partial x}\frac{\partial y}{\partial \xi} & \frac{\partial \eta}{\partial x}\frac{\partial x}{\partial \eta} + \frac{\partial \eta}{\partial y}\frac{\partial y}{\partial \eta} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial \xi}{\partial \xi} & \frac{\partial \xi}{\partial \eta} \\ \frac{\partial \eta}{\partial \xi} & \frac{\partial \eta}{\partial \eta} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Thus we have shown it.

4.1.1 Plane Polar Coordinates

Let us now consider plane polars, and their relation to the cartesian coordinate system. We have, as standard relations:

$$x = r\cos\theta \qquad y = r\sin\theta$$

Now, consider the transformation of the basis:

$$\boldsymbol{e}'_\alpha = \Lambda^\beta{}_\alpha \boldsymbol{e}_\beta$$

Now, consider that the two sets of coordinates are the plane polar & the Cartesian:

$$\{\boldsymbol{e}'_\alpha\} = (\boldsymbol{e}_r, \boldsymbol{e}_\theta) \qquad \{\boldsymbol{e}_\alpha\} = (\boldsymbol{e}_x, \boldsymbol{e}_y)$$

Then, for the radial unit vector, as an example:

$$\boldsymbol{e}_r = \Lambda^x{}_r \boldsymbol{e}_x + \Lambda^y{}_r \boldsymbol{e}_y$$

Which, by using our differential notation for the Lorentz transformation, is just

$$\begin{aligned} \boldsymbol{e}_r &= \frac{\partial x}{\partial r}\boldsymbol{e}_x + \frac{\partial y}{\partial r}\boldsymbol{e}_y \\ &= \cos\theta \boldsymbol{e}_x + \sin\theta \boldsymbol{e}_y \end{aligned}$$

In a similar fashion, we can compute $\boldsymbol{e}_\theta$:

$$\begin{aligned} \boldsymbol{e}_\theta &= \Lambda^x{}_\theta \boldsymbol{e}_x + \Lambda^y{}_\theta \boldsymbol{e}_y \\ &= -r\sin\theta \boldsymbol{e}_x + r\cos\theta \boldsymbol{e}_y \end{aligned}$$

Thus, pulling the two results together:

$$\boldsymbol{e}_r = \cos\theta \boldsymbol{e}_x + \sin\theta \boldsymbol{e}_y \qquad \boldsymbol{e}_\theta = -r\sin\theta \boldsymbol{e}_x + r\cos\theta \boldsymbol{e}_y \tag{4.1}$$

Therefore, what we have computed is a way of transferring between two coordinate systems (the plane polars & Cartesian), at a given point in space. Now, these relations are something we "always knew", but we used to derive them geometrically. Here we have derived them purely by considering coordinate transformations. Notice that the polar basis vectors change, depending on where one is in the plane.

Let us consider computing the metric for the polar coordinate system. Recall that the metric tensor is computed via:

$$g_{\mu\nu} = \mathsf{g}(\boldsymbol{e}_\mu, \boldsymbol{e}_\nu) = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu$$

Consider the lengths of the basis vectors:

$$|\boldsymbol{e}_\theta|^2 = \boldsymbol{e}_\theta \cdot \boldsymbol{e}_\theta$$

This is just:

$$\boldsymbol{e}_\theta \cdot \boldsymbol{e}_\theta = (-r\sin\theta\boldsymbol{e}_x + r\cos\theta\boldsymbol{e}_y) \cdot (-r\sin\theta\boldsymbol{e}_x + r\cos\theta\boldsymbol{e}_y)$$

Writing this out, in painful detail:

$$\boldsymbol{e}_\theta \cdot \boldsymbol{e}_\theta = (r^2\sin^2\theta)\boldsymbol{e}_x \cdot \boldsymbol{e}_x - (r^2\sin\theta\cos\theta)\boldsymbol{e}_x \cdot \boldsymbol{e}_y - (r^2\cos\theta\sin\theta)\boldsymbol{e}_y \cdot \boldsymbol{e}_x + (r^2\cos^2\theta)\boldsymbol{e}_y \cdot \boldsymbol{e}_y$$

Now, we use the orthonormality of the Cartesian basis vectors:

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij} \qquad \{\boldsymbol{e}_i\} = (\boldsymbol{e}_x, \boldsymbol{e}_y)$$

Hence, we see that this reduces our expression to:

$$\boldsymbol{e}_\theta \cdot \boldsymbol{e}_\theta = r^2\sin^2\theta + r^2\cos^2\theta = r^2$$

And therefore:

$$\boldsymbol{e}_\theta \cdot \boldsymbol{e}_\theta = |\boldsymbol{e}_\theta|^2 = r^2$$

Therefore, we see that the angular basis vector increases in magnitude as one gets further from the origin. In a similar fashion, one finds:

$$\boldsymbol{e}_r \cdot \boldsymbol{e}_r = |\boldsymbol{e}_r|^2 = 1$$

Let us compute the following:

$$\boldsymbol{e}_r \cdot \boldsymbol{e}_\theta = (\cos\theta\boldsymbol{e}_x + \sin\theta\boldsymbol{e}_y) \cdot (-r\sin\theta\boldsymbol{e}_x + r\cos\theta\boldsymbol{e}_y)$$

This clearly gives zero. Thus:

$$\boldsymbol{e}_r \cdot \boldsymbol{e}_\theta = 0$$

Now, we have enough information to write down the metric for plane polars. From the above definition, $g_{\mu\nu} = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu$. And therefore:

$$g_{rr} = 1 \qquad g_{\theta\theta} = r^2 \qquad g_{r\theta} = g_{\theta r} = 0$$

Therefore, arranging these components into a matrix:

$$(g_{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \tag{4.2}$$

We have therefore computed all the elements of the metric for polar space. Let us use it to compute the interval length:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \qquad \{dx^\mu\} = (dr, d\theta)$$

Hence, let us do the sum:

$$\begin{aligned} ds^2 &= \sum_{\mu=1}^{2}\sum_{\nu=1}^{2} g_{\mu\nu}dx^\mu dx^\nu \\ &= g_{11}dx^1dx^1 + g_{12}dx^1dx^2 + g_{21}dx^2dx^1 + g_{22}dx^2dx^2 \\ &= dr^2 + r^2d\theta^2 \end{aligned}$$

So here we see a line element "we always knew", but we have (again) derived it in a very different way.

Its not to hard to "intuit" the inverse matrix, by thinking of a matrix which when multiplied by (4.2) gives the identity matrix. This is seen to be:

$$(g^{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}$$

So that the product of the metric and its inverse gives the identity:

$$g_{\mu\nu}g^{\nu\rho} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

From here, we shall leave off the brackets denoting the matrix of elements. This shall be clear.

Covariant Derivative Let us consider the derivatives of the basis vectors. Consider:

$$\frac{\partial}{\partial r}\boldsymbol{e}_r = \frac{\partial}{\partial r}(\cos\theta\boldsymbol{e}_x + \sin\theta\boldsymbol{e}_y) = 0$$

Where we have used the appropriate result in (4.1). Consider also:

$$\begin{aligned} \frac{\partial}{\partial\theta}\boldsymbol{e}_r &= \frac{\partial}{\partial\theta}(\cos\theta\boldsymbol{e}_x + \sin\theta\boldsymbol{e}_y) \\ &= -\sin\theta\boldsymbol{e}_x + \cos\theta\boldsymbol{e}_y \\ &= \frac{1}{r}\boldsymbol{e}_\theta \end{aligned}$$

In a similar way, we can easily compute:

$$\frac{\partial}{\partial r}\boldsymbol{e}_\theta = \frac{1}{r}\boldsymbol{e}_\theta \qquad \frac{\partial}{\partial\theta}\boldsymbol{e}_r = -r\boldsymbol{e}_r$$

So, we see that if we were to differentiate some polar vector, we must also consider the differential of the basis vectors. So, consider some vector $\boldsymbol{V} = (V^r, V^\theta)$. Then, doing some differentiation:

$$\begin{aligned} \frac{\partial\boldsymbol{V}}{\partial r} &= \frac{\partial}{\partial r}(V^r\boldsymbol{e}_r + V^\theta\boldsymbol{e}_\theta) \\ &= \frac{\partial V^r}{\partial r}\boldsymbol{e}_r + V^r\frac{\partial\boldsymbol{e}_r}{\partial r} + \frac{\partial V^\theta}{\partial r}\boldsymbol{e}_\theta + V^\theta\frac{\partial\boldsymbol{e}_\theta}{\partial r} \end{aligned}$$

Where we have used the product rule. Notice that we do not have this to deal with when we use Cartesian basis vectors: they are constant. The "problem" arises when we note that the polar basis vectors are not constant. So, we may generalise the above by using the summation convention on the vector:

$$\frac{\partial \boldsymbol{V}}{\partial r} = \frac{\partial}{\partial r}(V^\alpha \boldsymbol{e}_\alpha) = \frac{\partial V^\alpha}{\partial r}\boldsymbol{e}_\alpha + V^\alpha \frac{\partial \boldsymbol{e}_\alpha}{\partial r}$$

Notice that the final term is zero for the Cartesian basis vectors. Let us further generalise this by choosing an arbitrary coordinate by which to differentiate:

$$\frac{\partial \boldsymbol{V}}{\partial x^\beta} = \frac{\partial V^\alpha}{\partial x^\beta}\boldsymbol{e}_\alpha + V^\alpha \frac{\partial \boldsymbol{e}_\alpha}{\partial x^\beta}$$

Now, instead of leaving the final expression as the differential of basis vectors, let us write it as a linear combination of basis vectors:

$$\frac{\partial \boldsymbol{e}_\alpha}{\partial x^\beta} = \Gamma^\mu{}_{\alpha\beta}\boldsymbol{e}_\mu \tag{4.3}$$

Notice that we are using the Christoffel symbols as the coefficient in the expansion. We shall link back to these in a moment. This is something which may seem wrong, or illegal, but recall that it is a standard procedure in quantum mechanics: expressing a vector as a sum over basis states. So that we have:

$$\frac{\partial \boldsymbol{V}}{\partial x^\beta} = \frac{\partial V^\alpha}{\partial x^\beta}\boldsymbol{e}_\alpha + V^\alpha \Gamma^\mu{}_{\alpha\beta}\boldsymbol{e}_\mu$$

In the final expression, let us relabel the indices $\alpha \leftrightarrow \mu$, giving:

$$\frac{\partial \boldsymbol{V}}{\partial x^\beta} = \frac{\partial V^\alpha}{\partial x^\beta}\boldsymbol{e}_\alpha + V^\mu \Gamma^\alpha{}_{\mu\beta}\boldsymbol{e}_\alpha$$

This of course lets us take out a "common factor":

$$\frac{\partial \boldsymbol{V}}{\partial x^\beta} = \left(\frac{\partial V^\alpha}{\partial x^\beta} + V^\mu \Gamma^\alpha{}_{\mu\beta}\right)\boldsymbol{e}_\alpha$$

Therefore, we see that we have created a new vector $\partial \boldsymbol{V}/\partial x^\beta$, having components:

$$\frac{\partial V^\alpha}{\partial x^\beta} + V^\mu \Gamma^\alpha{}_{\mu\beta}$$

Let us create the notation:

$$\frac{\partial V^\alpha}{\partial x^\beta} \equiv V^\alpha_{\ ,\beta} \qquad V^\alpha_{\ ;\beta} \equiv V^\alpha_{\ ,\beta} + V^\mu \Gamma^\alpha{}_{\mu\beta} \tag{4.4}$$

Then, we have an expression for the *covariant derivative*:

$$\frac{\partial \boldsymbol{V}}{\partial x^\beta} = V^\alpha_{\ ;\beta}\boldsymbol{e}_\alpha$$

Now, this will give a vector field, which is generally denoted by $\nabla \boldsymbol{V}$. This does indeed have a striking resemblance to the grad operator we "usually" worked with! We have components:

$$(\nabla \boldsymbol{V})^\alpha{}_\beta = (\nabla_\beta \boldsymbol{V})^\alpha = V^\alpha_{\ ;\beta}$$

The middle line highlights what is going on: the β^{th} derivative of the α^{th} component of the vector $\boldsymbol{V}$. Consider briefly, a scalar field ϕ. It does not depend on the basis vectors, so its covariant derivative will not take into account the α^{th} component; thus:

$$\nabla_\beta \phi = \frac{\partial \phi}{\partial x^\beta}$$

Which is just the gradient of a scalar! So, it is possible, and useful (but not necessarily correct) to think of the covariant derivative as the gradient of a vector.

Christoffel Symbols for Plane Polars Let us link the particular Christoffel symbols with what we know they must be, from differentiating the polar basis vectors. We saw that we had the following derivatives of the basis vectors:

$$\frac{\partial \boldsymbol{e}_r}{\partial r} = 0 \qquad \frac{\partial \boldsymbol{e}_r}{\partial \theta} = \frac{1}{r}\boldsymbol{e}_\theta$$
$$\frac{\partial \boldsymbol{e}_\theta}{\partial r} = \frac{1}{r}\boldsymbol{e}_\theta \qquad \frac{\partial \boldsymbol{e}_\theta}{\partial \theta} - r\boldsymbol{e}_r$$

Now, from (4.3), we see that when differentiating the basis vector $\boldsymbol{e}_\alpha$ with respect to the component x^β, we get some number ${\Gamma^\mu}_{\alpha\beta}$ infront of the basis vector $\boldsymbol{e}_\mu$. That is, taking an example:

$$\begin{aligned}\frac{\partial \boldsymbol{e}_\theta}{\partial r} &= \sum_{\mu=1}^{2} {\Gamma^\mu}_{r\theta}\boldsymbol{e}_\mu \\ &= {\Gamma^r}_{r\theta}\boldsymbol{e}_r + {\Gamma^\theta}_{r\theta}\boldsymbol{e}_\theta \\ &= \frac{1}{r}\boldsymbol{e}_\theta\end{aligned}$$

And therefore, we see that by inspection:

$${\Gamma^r}_{r\theta} = 0 \qquad {\Gamma^\theta}_{r\theta} = \frac{1}{r}$$

If we do a similar analysis on all the other basis vectors, we find that we can write down a whole set of Christoffel symbols (note that this has all only been for the polar coordinate system!):

$${\Gamma^r}_{rr} = {\Gamma^\theta}_{rr} = 0 \qquad {\Gamma^r}_{r\theta} = {\Gamma^r}_{\theta r} = 0 \qquad {\Gamma^\theta}_{r\theta} = {\Gamma^\theta}_{\theta r} = \frac{1}{r} \qquad {\Gamma^r}_{\theta\theta} = -r \qquad {\Gamma^\theta}_{\theta\theta} = 0$$

One may notice that the symbols have a certain symmetry:

$${\Gamma^\mu}_{\nu\rho} = {\Gamma^\mu}_{\rho\nu}$$

Although to derive this particular set of Christoffel symbols we used the fact that the Cartesian basis are constant, they do not appear in the final expressions. We used the symbols to write down the derivative of a vector, in some curvilinear space, using only polar coordinates. We have seen in previous sections how the symbols link to the metric. We saw:

$${\Gamma^\lambda}_{\mu\nu} = \frac{1}{2}g^{\rho\lambda}(g_{\rho\nu,\mu} + g_{\rho\mu,\nu} - g_{\mu\nu,\rho})$$

It would be extremely tedious to compute all the symbols again using this, but let us show one or two. Let us consider:

$$\begin{aligned}
\Gamma^r{}_{\theta\theta} &= \sum_{\rho=1}^{2} \frac{1}{2} g^{\rho r}(g_{\rho\theta,\theta} + g_{\rho\theta,\theta} - g_{\theta\theta,\rho}) \\
&= \frac{1}{2} g^{rr}(g_{r\theta,\theta} + g_{r\theta,\theta} - g_{\theta\theta,r}) + \frac{1}{2} g^{\theta r}(g_{\theta\theta,\theta} + g_{\theta\theta,\theta} - g_{\theta\theta,\theta})
\end{aligned}$$

With reference to the previously computed metrics:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \qquad g^{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}$$

We see that:

$$g_{r\theta} = 0 = g^{r\theta}$$

Also notice that:

$$g_{\theta\theta,r} = \frac{\partial g_{\theta\theta}}{\partial r} = \frac{\partial r^2}{\partial r} = 2r$$

Therefore, the only non-zero term gives $\Gamma^r{}_{\theta\theta} = -r$. Which we found before. In deriving the Christoffel symbols this way, the only thing we are considering is the metric of the space. This should highlight the inherent power of this notation & formalism! Let us compute another symbol. Consider $\Gamma^\theta{}_{r\theta}$:

$$\begin{aligned}
\Gamma^\theta{}_{r\theta} &= \sum_{\rho} \frac{1}{2} g^{\rho\theta}(g_{\rho\theta,r} + g_{\rho r,\theta} - g_{r\theta,\rho}) \\
&= \frac{1}{2} g^{r\theta}(g_{r\theta,r} + g_{rr,\theta} - g_{r\theta,r}) + \frac{1}{2} g^{\theta\theta}(g_{\theta\theta,r} + g_{\theta r,\theta} - g_{r\theta,\theta}) \\
&= \frac{1}{2} g^{\theta\theta} g_{\theta\theta,r} \\
&= \frac{1}{2} r^{-2} \frac{\partial r^2}{\partial r} \\
&= \frac{1}{r}
\end{aligned}$$

Where in the second step we just reduced everything back to what is non-zero. Again, it is a result we derived using a different method.