Appendix A

Einstein summation notation

This little introduction to index notation has been prepared by **Andrew Stasiuk**, former GR student.

Mathematics is highly cumulative – the higher level concepts rest heavily on much simpler mathematics, and with that, comes old notation. A notation that worked well for simple problems can become untenable if it needs to be repeated often and/or takes up a lot of space. For instance, summing over multiple variables should formally be written as

$$\sum_{\alpha_1=0}^{N} \cdots \sum_{\alpha_k=0}^{N} (\cdots), \tag{A.0.1}$$

but this is tedious and is often compacted into a single sum,

$$\sum_{\alpha_1,\cdots,\alpha_k=0}^{N} (\cdots). \tag{A.0.2}$$

These types of notational simplifications permeate throughout all of mathematics and physics, like Dirac's bra-ket notation or Einstein's summation notation. Einstein's summation notation, often just called index notation, originates with Einsteins original work with the theory of general relativity. General relativity is a geometric theory, that is, it rest on the concepts and tools of differential geometry. This requires handling vectors, matrices, and *tensors*, which all must change under coordinate transformations in a very particular way. Informally, tensors are like higher dimensional matrices – while it isn't too hard to write out the components of column vector or a matrix, trying to write the components of a rank 3 tensor in a cube is hard, verging on impossible.

Faced with a notationally tedious theory, index notation is the proposed (and ubiquitously used) fix to the problem. Let us sketch out the concept before diving into the details.

We start simply – just drop all the sums, and when there is a repeated index on an object, understand that this index should be summed over. This is the most basic version of the notation, and we will see that it can be improved by making the following observation – given a change of coordinates, all objects will transform either like a *vector*, or a *dual vector*, and occasionally a mix of the two. The notation can then be improved by now making a choice to separate indices representing different transformation properties into superscripts and subscripts. Thus, a formula using this notation might look like the following,

$$R_{\mu\nu} - \frac{1}{2}R_{\alpha}{}^{\alpha}g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}\,, \tag{A.0.3}$$

which is in fact the Einstein field equations[1] and its solutions will fully describe the curvature of space and time. In the following sections, we will develop this notation by starting with a familiar linear algebra approach to change of basis, and concluding with developing the notions of covariance and contravariance along with some examples.

A.1 Linear Algebra and Change of Basis

First, we review the notion of a coordinate transformation in the usual linear algebra sense. Let V be a real vector space of dimension n with an ordered basis $\mathcal{B} = \{\boldsymbol{e}_i\}$. A vector, $\boldsymbol{x} \in V$ is then written as

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i, \tag{A.1.1}$$

which we often write as *column vector*, using a subscript to clearly denote the ordered basis used for the entries of the column vector,

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{B}}. \tag{A.1.2}$$

Then, given any other basis for V, say $\mathcal{C} = \{\boldsymbol{e}'_i\}$, then there is some *change of basis*, which is a linear map that takes basis vectors of $\mathcal{B}$ and maps them to basis vectors

[1] Actually, because the written objects are symmetric under interchange of indices there are only 10 unique differential equations, assuming space time is 3+1 dimensional.

in $\mathcal{C}$. We require that this operation leaves the vector $\boldsymbol{x}$ unchanged. To gain some intuition on the matter, think of a vector in $\mathbb{R}^3$ as an arrow floating in space. No matter how we orient the x, y, and z axes in this space, the 'arrow' itself will not change, but the projection of this arrow onto the axes might change. This is the essence of a coordinate transformation – given the coordinates of a vector in one set of axes, we can find a map that correctly represents the vector in a new set of axes. Now, since the vector is unchanged, we know that the map acting on the vector $\boldsymbol{x}$ should be the identity map:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{x} \quad \forall \boldsymbol{x} \implies \boldsymbol{T} = \mathbb{1}. \tag{A.1.3}$$

Let's say that we characterize the change of basis by the invertible linear map $\boldsymbol{P}$ that takes you from an element of basis $\mathcal{B}$ to an element of the basis $\mathcal{C}$. That is, $\boldsymbol{P}(\boldsymbol{e}_i) = \boldsymbol{e}'_i$. Since $\boldsymbol{x}$ cannot change under the change of basis, and at the same time the basis elements $\boldsymbol{e}_i$ change with $\boldsymbol{P}$, we will need a copy of $\boldsymbol{P}^{-1}$ somewhere changing the components $x_1, \ldots, x_n$ in order for $\boldsymbol{x}$ to be unchanged, i.e., for the change of basis to be a net identity map on $\boldsymbol{x}$. To be more concrete, from equation (A.1.1), we can find the related coordinate transformation as follows:

$$\boldsymbol{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \mathbb{1} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \boldsymbol{P}^{-1}\boldsymbol{P} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix}. \tag{A.1.4}$$

This gives us the following relationships,

$$\begin{bmatrix} \boldsymbol{e}'_1 \\ \vdots \\ \boldsymbol{e}'_n \end{bmatrix} = \boldsymbol{P} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix}, \tag{A.1.5}$$

and

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} = (\boldsymbol{P}^{-1})^T \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{A.1.6}$$

Let us stress that a change of basis maps any vector to itself $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{x}$, and we are only changing the 'collections of numbers' we use to represent the vector $\boldsymbol{x}$. We can write $\boldsymbol{x}$ as a column vector with entries in two different bases $\mathcal{B}$ and $\mathcal{C}$,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{B}} \equiv \boldsymbol{x} \equiv \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}_{\mathcal{C}}, \tag{A.1.7}$$

where the column vector representations will have different components for $\boldsymbol{P} \neq \mathbb{1}$, but represent the same vector in their respective bases.

Example A.1.1. Consider $V = \mathbb{R}^2$, with standard basis $\mathcal{B} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$. Now consider another basis $\mathcal{C}$, which we can write $\{\boldsymbol{e}_+, \boldsymbol{e}_-\}$ which is related to our original basis by the following change of basis matrix,

$$\boldsymbol{P} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{A.1.8}$$

Then we see that,

$$\begin{aligned} \boldsymbol{e}_+ &= \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 + \boldsymbol{e}_2) \\ \boldsymbol{e}_- &= \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 - \boldsymbol{e}_2). \end{aligned} \tag{A.1.9}$$

Further, consider the vector $\boldsymbol{x} = \boldsymbol{e}_1$. Then,

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = (\boldsymbol{P}^{-1})^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \tag{A.1.10}$$

We clearly see that

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\mathcal{B}} \equiv \boldsymbol{x} \equiv \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}_{\mathcal{C}}, \tag{A.1.11}$$

but

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \neq \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \tag{A.1.12}$$

In this scenario, we had the somewhat unusual property that $\boldsymbol{P} = (\boldsymbol{P}^{-1})^T$, which is characteristic of orthogonal matrices, that is, matrices whose inverses are their transposes.

Again, it is extremely important to stress here that the abstract vector $\boldsymbol{x}$ is left *completely unchanged*, it is just the column vector representation which changes. If we endow this space with an inner product, and take $\mathcal{B}, \mathcal{C}$ to be orthonormal bases, the matrix representation of the change of basis matrix (which is now an orthogonal matrix) becomes a bit more obvious,

$$\boldsymbol{P} = \begin{bmatrix} \langle \boldsymbol{e}_1', \boldsymbol{P}(\boldsymbol{e}_1) \rangle & \cdots & \langle \boldsymbol{e}_1', \boldsymbol{P}(\boldsymbol{e}_n) \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{e}_n', \boldsymbol{P}(\boldsymbol{e}_1) \rangle & \cdots & \langle \boldsymbol{e}_n', \boldsymbol{P}(\boldsymbol{e}_n) \rangle \end{bmatrix}. \tag{A.1.13}$$

Now, let's start introducing some shorthand notation, —call $(\boldsymbol{P}^{-1})^T \equiv \tilde{\boldsymbol{P}}$, and the components (or entries) of any map $\boldsymbol{A}$ to be $(\boldsymbol{A})_{ij} = A_{ij}$. This convention allows

us to write a little more compactly that,

$$\boldsymbol{e}'_i = \sum_{j=1}^{n} P_{ij}\boldsymbol{e}_j. \tag{A.1.14}$$

At this point, many undergraduate courses in applied mathematics and physics will start to introduce *index notation*, which involves dropping the summation sign and stating that if two indices are repeated in an equation, then it should be understood that it represents a sum. Using this, we might rewrite some of the above equations as

$$\begin{aligned} x'_i &= P_{ij}x_j\,, \\ \delta_{ij} &= P_{ik}\tilde{P}_{kj}\,. \end{aligned} \tag{A.1.15}$$

This convention as a shorthand, while used very often, is not optimal – consider the following issue. P_{ij} represents the component located at the ith row and jth column of the change of basis matrix. So then a diagonal element of the change of basis should be written as P_{ii}. Based on the convention, this should be a sum over i, giving us the trace of $\boldsymbol{P}$. We see that by taking this convention as a shorthand, we have introduced ambiguities where there were once none.

We can fix this by making the following observation: the basis elements transform with the map $\boldsymbol{P}$, while the coordinates transform with $\tilde{\boldsymbol{P}}$. Let us pick the convention that any collection of objects, be it an n-tuple of real numbers, basis vectors, etc., that transforms with $\boldsymbol{P}$ has its indices written as subscripts, and a collection that transform with $\tilde{\boldsymbol{P}}$ will have its indices written as superscripts. Then we only sum over repeated indices that are up and down, which we call a *contraction* (of indices). For example,

$$\boldsymbol{x} = \sum_{i=1}^{n} x^i\boldsymbol{e}_i \equiv x^i\boldsymbol{e}_i, \tag{A.1.16}$$

We can guess at the location of the indices of the change of basis. We want one index up to contract it with the basis elements, and one index down to be left over for consistency. Embedding as a matrix vector multiplication, we find,

$$\begin{bmatrix} \boldsymbol{e}'_1 \\ \vdots \\ \boldsymbol{e}'_n \end{bmatrix} = \begin{bmatrix} {P_1}^1 & \cdots & {P_1}^n \\ \vdots & \ddots & \vdots \\ {P_n}^1 & \cdots & {P_n}^n \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix}. \tag{A.1.17}$$

In the index notation introduced, we write

$$\boldsymbol{e}'_i = {P_i}^j\boldsymbol{e}_j. \tag{A.1.18}$$

Similarly, we can do the same for the coordinates,

$$x'^i = \tilde{P}^i_{\ j}\, x^j. \tag{A.1.19}$$

We can use the fact that the change of basis is an identity mapping, i.e. $\boldsymbol{P}\tilde{\boldsymbol{P}}^T = \mathbb{1}$, so that

$$P_i^{\ k}\tilde{P}_k^{\ j} = \delta_i^j. \tag{A.1.20}$$

Let us summarize our findings about changes of basis with the following definitions.

Definition A.1.1. For a collection of n objects that transform like the elements of a basis, that is, with $\boldsymbol{P}$, we write its index as a subscript, and refer to it as a **covariant index**. The collection, can be written as a_i and explicitly satisfies

$$a'_i = P_i^{\ j} a_j. \tag{A.1.21}$$

We say that a_i represents the **components of a rank 1 contravariant tensor**.

Definition A.1.2. For a collection of n objects that transform like the coordinates of a vector, that is, with $\tilde{\boldsymbol{P}}$, we write its index as a superscript, and refer to it as a **contravariant index**. The collection, can be written as a^i and explicitly satisfies

$$a'^i = \tilde{P}^i_{\ j}\, a^j. \tag{A.1.22}$$

We say that a^i represents the **components of a rank 1 covariant tensor**.

We see that $\boldsymbol{P}$ must be represented with two indices, and so we should also extend these definitions to cover objects with an arbitrary number of indices.

Definition A.1.3. For a collection, $\boldsymbol{v}$, of n^{r+s} objects that transforms r times like coordinates of a vector and s times like the elements of basis, we write the r indices transforming with $\boldsymbol{P}$ as superscripts, and the s indices transforming with $\tilde{\boldsymbol{P}}$ as subscripts. We say this collection has ***r* covariant and *s* contraviant indices**, or simply **rank (*r*,*s*)**. It satisfies,

$$v'^{\alpha_1\cdots\alpha_r}_{\quad\beta_1\cdots\beta_s} = \tilde{P}^{\alpha_1}_{\ \ \gamma_1}\cdots\tilde{P}^{\alpha_r}_{\ \ \gamma_r} P_{\beta_1}^{\ \ \delta_1}\cdots P_{\beta_s}^{\ \ \delta_s} v^{\gamma_1\cdots\gamma_r}_{\quad\delta_1\cdots\delta_s}. \tag{A.1.23}$$

We say that $v^{\alpha_1\cdots\alpha_r}_{\quad\beta_1\cdots\beta_s}$ represents the **components of an *r* times contravariant, *s* times covariant tensor**.

Given V, we can also define the *dual space* to V, written V^*, which is the set of all linear functionals of V. That is, for any $\boldsymbol{\omega} \in V^*$, there we have a linear map $\boldsymbol{\omega} : V \longrightarrow \mathbb{R}$, which we can write suspiciously like an inner product, such that for $\boldsymbol{x} \in V$, we have that $\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle \in \mathbb{R}$. By linearity there must always exist a basis for the

dual space, called the *dual basis*, which we (suggestively) can write as $\{\boldsymbol{e}^1, \cdots, \boldsymbol{e}^n\}$, such that

$$\langle \boldsymbol{e}^j, \boldsymbol{e}_i \rangle = \delta_i^j, \tag{A.1.24}$$

and a general $\boldsymbol{\omega}$ can be written with dual coordinates,

$$\boldsymbol{\omega} = x_i \boldsymbol{e}^i \tag{A.1.25}$$

We can use this to calculate how elements of the dual basis transform under an allowable coordinate transformation.

$$\begin{aligned} \langle \boldsymbol{e}^j, \boldsymbol{e}_i \rangle &= \langle \boldsymbol{e}'^j, \boldsymbol{e}'_i \rangle \\ \delta^l_j \langle \boldsymbol{e}^j, \boldsymbol{e}_i \rangle &= \delta^l_j \langle \boldsymbol{e}'^j, \boldsymbol{e}'_i \rangle \\ \langle \tilde{P}^k{}_j \, \boldsymbol{e}^j, P_k{}^l \boldsymbol{e}_i \rangle &= \langle \boldsymbol{e}'^l, \boldsymbol{e}'_i \rangle, \end{aligned} \tag{A.1.26}$$

and if we take $i = l$, then we find that

$$\langle \tilde{P}^k{}_j \, \boldsymbol{e}^j, P_k{}^l \boldsymbol{e}'_l \rangle = 1 \implies \tilde{P}^k{}_j \, \boldsymbol{e}^j = \boldsymbol{e}'^k. \tag{A.1.27}$$

We have just concluded that elements of the dual basis transform with $\tilde{\boldsymbol{P}}$, which by a similar argument forces the dual coordinates to transform with $\boldsymbol{P}$. Our choice of location for the indices of the dual elements can be seen as an *a postieri* choice, since it was made with the prior knowledge that the coordinates of a dual vector will transform covariantly. Nothing in the mathematics expressly prevented us from putting the indices elsewhere, and if anyone was feeling particular masochistic, they might haphazardly throw around indices wherever they felt, and have to remember each case individually. But the point of this notation is to be more efficient and clear, so we continue to make choices consistent with those determined to be the 'best' after many years of refinement.

In order to move indices up and down, that is, map a contravariant object to a covariant object, we essentially need find an isomorphism between the vector space V to the dual space V^*, and the rest follows from linearity. It turns out that in 'flat spaces' (the kind we just considered), this mapping can always be taken to be the identity transformation[2], up to relabeling of bases. We often call this map the *metric tensor*, and write it as g_{ij} (or often η_{ij} as well), and its inverse as g^{ij}, such that

$$x^i g_{ij} = x_j. \tag{A.1.28}$$

[2]The fact that the relationship between V and V^* is the identity map is often used as an argument that contravariant and covariant indices do not need to be distinguished. In more general geometries, this will most definitely not be the case, and it is good practice to be careful when manipulating indices.

For a Euclidean vector space, this takes the form of $g_{ij} = \delta_{ij}$. We say this metric has a completely positive signature since we have that $g_{00} = \cdots = g_{nn} = 1$, all others equal to 0. In a Minkowski space-time, the metric has a 3+1 signature, such that $g_{00} = -1$, $g_{11} = g_{22} = g_{33} = 1$, all others equal to zero[3]. In a more general (possibly not flat) space, the metric can be more complicated. We will always call it g_{ij} to maintain consistency, but it should be understood that in a Euclidean space we are free to move indices up and down without any modification to coordinates. This seems to contradict the notion that covariant and contravariant indices are innately different, but we will see that in order to properly introduce the metric tensor, we need a sufficiently defined inner product, which will force our transformation to be orthogonal. We have already seen that orthogonal coordinate transformations imply that $\tilde{\boldsymbol{P}} = \boldsymbol{P}$, and as it turns out, the converse is also true.

A.2 Tensor Manipulations and Calculations

There are a number of operations which can be done using tensors – since a tensor technically lives in a vector space, tensor spaces have both addition and scalar multiplication defined, with addition only being defined for tensors of the same rank. We can also take the tensor product of two tensors (just multiply the coordinates together), or contract along indices (using the metric tensor to raise and lower as needed). For example, let $a \in \mathbb{R}$ be a scalar, and b_{ij}, c_{ij} be two rank $(0, 2)$ tensors. The following operations are allowed:

$$b_{ij} + c_{ij}$$
$$ab_{ij}$$
$$b_{ij}c_{kl}$$
$$b_{ij}c^{ij} = b^{ij}c_{ij} = b_{ij}c_{lk}g^{li}g^{ki}.$$

From the top down, we see that we were able to create two new rank $(0, 2)$ tensors, a rank $(0, 4)$ tensor, and a scalar. In fact, by adding copies of b_{ij} and c_{ij} with different indices, or contracting indices by introducing copies of the metric tensor, we are able to build a tensor of any rank we choose. We are also allowed to swap indices, which equates to taking a transposition over those two indices. For rank two tensors, it is always clear which indices will be swapped, but for higher rank tensors, it should be indicated. For example,

$$(b^T)_{ij} = b_{ji}. \tag{A.2.1}$$

We can also symmetrize or anti-symmetrize the components of a tensor, which we indicate with a () for symmetrization and [] for anti-symmetrization. Doing this

[3]There is a competing standard used often by high energy physicists where the metric is defined as $g_{00} = 1$, $g_{11} = g_{22} = g_{33} = -1$, all others equal to 0

over two indices is relatively straightforward:

$$T_{(\alpha\beta)} = \frac{1}{2}(T_{\alpha\beta} + T_{\beta\alpha})$$
$$T_{[\alpha\beta]} = \frac{1}{2}(T_{\alpha\beta} - T_{\beta\alpha}). \tag{A.2.2}$$

To do these operations about more than two indices, we must sum over all possible permutations of the interchanged indices for symmetrization and introduce a (-1) factor dependent on the permutation parity for anti-symmetrization (recall the permutation group group S_n). For symmetrization, we have, for example

$$T^{\alpha_1\cdots\alpha_r}{}_{(\beta_1\cdots\beta_s)} = \frac{1}{s!}\sum_{\pi\in S_s} T^{\alpha_1\cdots\alpha_r}{}_{\pi(\beta_1)\cdots\pi(\beta_s)}\,, \tag{A.2.3}$$

as well as its anti-symmetrization,

$$T^{\alpha_1\cdots\alpha_r}{}_{[\beta_1\cdots\beta_s]} = \frac{1}{s!}\sum_{\pi\in S_s} (-1)^{\pi} T^{\alpha_1\cdots\alpha_r}{}_{\pi(\beta_1)\cdots\pi(\beta_s)}\,. \tag{A.2.4}$$

Let us now run through a series of examples with rank two tensors - these have direct matrix analogues and will help build intuition for the higher rank tensors.

Example A.2.1. Let $\boldsymbol{A}, \boldsymbol{B}$ be two rank $(1,1)$ tensors, which can be represented as matrices in the usual way with coordinates $\boldsymbol{A} = A_i{}^j \boldsymbol{e}^i \boldsymbol{e}_j$, $\boldsymbol{B} = B_i{}^j \boldsymbol{e}^i \boldsymbol{e}_j$. Then,

$$\begin{aligned}
(AB)_i{}^j &= A_i{}^l B_l{}^j \\
(BA)_i{}^j &= A_l{}^j B_i{}^l \\
(A^T B)_i{}^j &= (A^T)_i{}^l B_l{}^j = A^l{}_i B_l{}^j \\
(AB^T)_i{}^j &= A_i{}^l (B^T)_l{}^j = A_i{}^l B^j{}_l \\
(A^T B^T)_i{}^j &= (A^T)_i{}^l (B^T)_l{}^j = A^l{}_i B^j{}_l \\
(B^T A^T)_i{}^j &= (B^T)_i{}^l (A^T)_l{}^j = B^l{}_i A^j{}_l \\
\operatorname{tr} A &= A_i{}^i
\end{aligned}$$

A.3 General Coordinate Transformations

Following the work done in the previous section, we can generalize this notion to non-linear coordinate transformations. Now, let us consider as set V, such that every $\boldsymbol{x} \in V$ is described uniquely by a set of coordinates, usually written as a

tuple of real numbers, $(x^1, \cdots, x^n)$ with respect to some local coordinate basis, $\mathcal{B} = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$, which we will call the unprimed basis, or standard basis. Then, an arbitrary point, $\boldsymbol{x} \in V$ can be written as

$$\boldsymbol{x} = \sum_{i=1}^{n} x^i \boldsymbol{e}_i \equiv x^i \boldsymbol{e}_i. \tag{A.3.1}$$

Notice that we can take partial derivatives of a general vector with respect to its coordinates to recover the basis,

$$\boldsymbol{e}_i = \frac{\partial \boldsymbol{x}}{\partial x^i} \equiv \partial_i \boldsymbol{x} \tag{A.3.2}$$

Now, let us introduce another arbitrary coordinate system for V, which we call the primed coordinates, such that the primed coordinates depend smoothly on the unprimed coordinates,

$$x'^i = x'^i(x^1, \cdots, x^n), \tag{A.3.3}$$

and this map must be invertible, at least locally about any point, such that we can also write that the unprimed coordinates depend smoothly on the primed coordinates. That is,

$$x^i = x^i(x'^1, \cdots, x'^n) \tag{A.3.4}$$

Since an abstract element of V should be coordinate independent, we have the following,

$$\boldsymbol{e}'_i = \frac{\partial \boldsymbol{x}}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i} \frac{\partial \boldsymbol{x}}{\partial x^j} = \frac{\partial x^j}{\partial x'^i} \boldsymbol{e}_j, \tag{A.3.5}$$

which follows from the chain rule. Similarly, we have that

$$\boldsymbol{e}_i = \frac{\partial x'^j}{\partial x^i} \boldsymbol{e}'_j, \tag{A.3.6}$$

so it follows that,

$$\boldsymbol{e}_i = \frac{\partial x'^j}{\partial x^i} \frac{\partial x^l}{\partial x'^j} \boldsymbol{e}_l. \tag{A.3.7}$$

This means that,

$$\frac{\partial x'^j}{\partial x^i} \frac{\partial x^l}{\partial x'^j} = \begin{cases} 1 & i = l \\ 0 & \text{else} \end{cases} =: \delta_i^l, \tag{A.3.8}$$

where δ_i^l is called the Kronecker-delta. The object facilitating the change of coordinates is a map formed by the partial derivatives of the coordinates, and has two indices, which is commonly called the *Jacobian* of the coordinate transformation. Since this serves as our change of basis map for the coordinate basis, we can relabel for convenience:

$$\Lambda_i^{\ j} := \frac{\partial x^j}{\partial x'^i}. \tag{A.3.9}$$

Here, $\boldsymbol{\Lambda}$ serves the same purpose as $\boldsymbol{P}$ in the linear case. By chain rule, we immediately know that it's inverse is given by

$$(\Lambda^{-1})_i{}^j = \frac{\partial x'^j}{\partial x^i}, \tag{A.3.10}$$

since we have just shown that

$$(\Lambda^{-1})_i{}^l \Lambda_l{}^j = \delta_i^j. \tag{A.3.11}$$

Let's take a step back and write these products as matrix-vector multiplication, so as to utilize the intuition gained from the previous section. Since we have that $\boldsymbol{e}_i = \Lambda_i{}^j \boldsymbol{e}'_j$, we can write this symbolically as,

$$\begin{bmatrix} \boldsymbol{e}'_1 \\ \vdots \\ \boldsymbol{e}'_n \end{bmatrix} = \begin{bmatrix} \Lambda_1{}^1 & \cdots & \Lambda_1{}^n \\ \vdots & \ddots & \vdots \\ \Lambda_n{}^1 & \cdots & \Lambda_n{}^n \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix}. \tag{A.3.12}$$

For convenience, let us formally call this Jacobian matrix $\boldsymbol{\Lambda}$ to save some space – in general, when we want to discuss the entire structure, we write the symbol boldfaced without any indices, the reason for this will become more clear soon. Now, we can also write $\boldsymbol{x}$ in this manner,

$$\boldsymbol{x} = x^i \boldsymbol{e}_i = \begin{bmatrix} x^1 & \cdots & x^n \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix}. \tag{A.3.13}$$

Let's 'insert the identity matrix' in between this row-column product, giving us,

$$\boldsymbol{x} = \begin{bmatrix} x^1 & \cdots & x^n \end{bmatrix} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix} = \left((\boldsymbol{\Lambda}^{-1})^T \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} \right)^T \boldsymbol{\Lambda} \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix} \tag{A.3.14}$$

Then then, since we know that the product on the right is the primed coordinate basis, we must have that

$$\left((\boldsymbol{\Lambda}^{-1})^T \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} \right)^T = \begin{bmatrix} x'^1 & \cdots & x'^n \end{bmatrix}, \tag{A.3.15}$$

and by moving back into index notation, we find that we have just proven

$$\tilde{\Lambda}^i{}_j x^j = x'^i, \tag{A.3.16}$$

where $\tilde{\Lambda}^i{}_j = (\Lambda^{-1})_j{}^i$, or rather, $\tilde{\mathbf{\Lambda}} = (\mathbf{\Lambda}^{-1})^T$. It's important to note that it is entirely equivalent to write

$$x'^i = (\Lambda^{-1})_j{}^i x^j, \tag{A.3.17}$$

but this contraction of indices is not directly seen as a formal matrix vector product, and so is considered 'less intuitive'. We have just found the transformation rules for coordinates of a vector, and the basis elements of a vector, and unsurprisingly it matches the transformation rules in the linear case. We say (for now) that the coordinates "transform with lambda tilde" and the basis elements "transform with lambda".

Often, a tensor defined by its transformation properties, as we have just done above, is the standard physics approach to covariance and contravariance. In reality, the object, $v^{\alpha_1\cdots\alpha_r}{}_{\beta_1\cdots\beta_s}$ is *not* the tensor, but represents the coordinates of a tensor! Like a vector, the entire object is a sum of the coordinates over the basis elements. We should properly write,

$$\boldsymbol{v} = v^{\alpha_1\cdots\alpha_r}{}_{\beta_1\cdots\beta_s} \boldsymbol{e}_{\alpha_1} \otimes \cdots \otimes \boldsymbol{e}_{\alpha_r} \otimes \boldsymbol{e}^{\beta_1} \otimes \cdots \otimes \boldsymbol{e}^{\beta_s}, \tag{A.3.18}$$

which is ubiquitously shortened to

$$v^{\alpha_1\cdots\alpha_r}{}_{\beta_1\cdots\beta_s} \boldsymbol{e}_{\alpha_1} \cdots \boldsymbol{e}_{\alpha_r} \boldsymbol{e}^{\beta_1} \cdots \boldsymbol{e}^{\beta_s}. \tag{A.3.19}$$

It should be clear that the coordinates are real numbers, that is, $v^{\alpha_1\cdots\alpha_r}{}_{\beta_1\cdots\beta_s} \in \mathbb{R}$, while the tensor itself is an element of the vector space $V^{*\otimes s} \otimes V^{\otimes r}$. Using the $\langle , \rangle$ map, an element in the dual space 'eats' an element of the vector space and spits out a real number, and the same can be said of a vector 'eating' a covector and spitting out a real number. Thus it should be understood that tensor of rank (r, s) is a *multi-linear map* that takes in elements of $V^{\otimes s} \otimes V^{*\otimes r}$ and spits out scalars in $\mathbb{R}$.

As it currently stands, V and V^* are only connected by the fact elements in these two spaces can be combined to make a real number, we wish to introduce a map giving us the generalized metric tensor alluded to previously. Now, recall that we can define an inner product, denoted by '$\cdot$', on V which will similarly give a bi-linear map, $\cdot : V \times V \longrightarrow \mathbb{R}$. Then for an arbitrary basis (possibly not orthonormal), we have that,

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = g_{ij} \in \mathbb{R}. \tag{A.3.20}$$

By linearity of the inner product, we can show using the transformation properties of vectors that g_{ij} defines the coordinates of a rank $(0, 2)$ tensor, which we have previously called **metric tensor**.

Example A.3.1. The metric tensor is indeed a tensor of rank $(0, 2)$. Basis vectors are covariant, so we recall that

$$\boldsymbol{e}'_i = \Lambda_i{}^j \boldsymbol{e}_j. \tag{A.3.21}$$

Then,

$$\begin{aligned} g'_{ij} &= (\boldsymbol{e}'_i) \cdot (\boldsymbol{e}'_j) \\ &= (\Lambda_i{}^l \boldsymbol{e}_l) \cdot (\Lambda_j{}^k \boldsymbol{e}_k) \\ &= \Lambda_i{}^l \Lambda_j{}^k (\boldsymbol{e}_l \cdot \boldsymbol{e}_k) \\ &= \Lambda_i{}^l \Lambda_j{}^k g_{lk}. \end{aligned} \tag{A.3.22}$$

Thus, by definition, the metric tensor is a twice covariant tensor. Further, by symmetry of the inner product, it is symmetric, that is,

$$g_{ij} = \boldsymbol{e}_i \cdot \boldsymbol{e}_j = \boldsymbol{e}_j \cdot \boldsymbol{e}_i = g_{ji}. \tag{A.3.23}$$

Given $\boldsymbol{a} = a^i \boldsymbol{e}_i, \boldsymbol{b} = b^i \boldsymbol{e}_i \in V$, we can write

$$\boldsymbol{a} \cdot \boldsymbol{b} = (a^i \boldsymbol{e}_i) \cdot (b^j \boldsymbol{e}_j) = a^i b^j g_{ij}. \tag{A.3.24}$$

The metric tensor allows us to map elements from V to V^* and vice-versa. Said a bit more obviously, the metric tensor defines a natural way to lower indices,

$$b^j g_{ij} = b_i. \tag{A.3.25}$$

Because of the properties of the inner product, in particular, the property of positive definite-ness, we have that g_{ij} is invertible, and we will call its inverse g^{ij} which satisfies the following property,

$$g_{il} g^{lj} = \delta_i^j. \tag{A.3.26}$$

This allows us to raise indices, such that

$$b_i g^{ij} = b^j, \tag{A.3.27}$$

which is a completely consistent definition, notice,

$$b^i = b_j g^{ji} = b^l g_{jl} g^{ji} = b^l \delta_l^i = b^i. \tag{A.3.28}$$

This can also be seen as consistent with the induced dual basis,

$$\begin{aligned} \langle \boldsymbol{e}^i, \boldsymbol{e}_j \rangle &= \langle \boldsymbol{e}_k g^{ki}, \boldsymbol{e}_j \rangle \\ &:= g^{ki} (\boldsymbol{e}_k \cdot \boldsymbol{e}_j) \\ &= g^{ki} g_{kj} = \delta_j^i. \end{aligned} \tag{A.3.29}$$

The action of the metric tensor can be applied to tensors of any rank, for example,

$$v^{\alpha_1\cdots\alpha_i\cdots\alpha_r}{}_{\beta_1\cdots\beta_s} g_{\gamma\alpha_i} = v^{\alpha_1\cdots\alpha_{i-1}}{}_{\gamma}{}^{\alpha_{i+1}\cdots\alpha_r}{}_{\beta_1\cdots\beta_s}. \tag{A.3.30}$$

Now that we have the ability to move to and from the dual space, there is little need to discuss the basis elements anymore. All computations can be computed using the components of the tensor, as the relationship between the basis elements is entirely encoded in the metric tensor. For example, it is common to refer to the collections a^i and b^i as vectors, even though they are only really the components of a vector. Their inner product, or contraction can be written as

$$a^i b_i = a^i b^j g_{ij}, \tag{A.3.31}$$

and their outer (or tensor) product as

$$a^i b^j. \tag{A.3.32}$$

The diagonal elements of the resulting rank 2 tensor is written,

$$a^i b^i, \tag{A.3.33}$$

where since the repeated indices are both superscripts, no sum is performed. The trace of their outer product is readily seen as equivalent to their inner product.