Chapter 5: Tensors

Author: Shadman Salam, PhD
PHY405 – Mathematical Physics
Fall 2024

Reference Texts: *An Introduction to Tensors and Group Theory for Physicists* by N. Jeevanjee; *Mathematical Methods for Physicists* by Arfken, Weber and Harris.

1 Poor Man's Introduction to Tensors

Traditionally tensors have been defined as a collection of objects which carry **indices** and which *transform* in a particular way specified by those indices. However, that definition does not give much insight or physical intuition about what a tensor actually is. An equivalent but possibly a more conceptual definition constitutes a more modern interpretation of a tensor. According to this modern view, a tensor can be defined as a *function* which "eats" a certain number of vectors (known as the **rank** ($\boldsymbol{r}$) of the tensor) and "spits" out a *scalar* number. The central principle of tensor analysis thus lies in the mathematical fact that scalars remain unaffected (or *invariant*) by coordinate changes. Hence, the main result of tensor analysis can be summarized in a few words: **an equation written in tensor form is valid in any coordinate system**.

Obviously, the section is aptly titled – I left out many key details about the mathematical object known as tensors. Hopefully, we will cover the important stuff in the following sections.

2 Properties of Tensors

The distinguishing feature of a tensor is a property known as ***multilinearity*** – that, a tensor T must be linear in each of its r number of arguments. For instance, for a tensor with a single argument (i.e. $r = 1$ for T) linearity implies that

$$T(v + cw) = T(v) + cT(w) \ , \tag{1}$$

for all vectors v, w and numbers c. Multilinearity enables us to express the value of the function on an *arbitrary* set of r vectors in terms of the values of the function on the r **basis vectors** such as $\hat{\mathbf{x}},\ \hat{\mathbf{y}},\ \hat{\mathbf{z}}$. The values of the function on basis vectors are the familiar notion (from previous knowledge of vectors) of **components** of the tensor.

Example 1. Consider a rank 2 tensor T. Note that the purpose of a rank 2 tensor is to "eat" 2 vectors v and w and "spit out" a number expressed as $T(v, w)$. Then, multilinearity of T implies that,

$$\begin{aligned} T(v_1 + cv_2, w) &= T(v_1, w) + cT(v_2, w) \\ T(v, w_1 + cw_2) &= T(v, w_1) + cT(v, w_2) \ , \end{aligned} \tag{2}$$

for any number c and all vectors v and w. Now, assuming that v and w lives in a vector space with coordinate basis $\hat{\mathbf{x}},\ \hat{\mathbf{y}},\ \hat{\mathbf{z}}$, then we can write,

$$v = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + \hat{\mathbf{z}}$$

$$w = w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}$$

which then plugged in to Eq.(2) leads to,

$$
\begin{aligned}
T(v,w) &= T(v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}},\ w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}) \\
&= v_x T(\hat{\mathbf{x}},\ w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}) + v_y T(\hat{\mathbf{y}},\ w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}) \\
&\quad + v_z T(\hat{\mathbf{z}},\ w_x\hat{\mathbf{x}} + w_y\hat{\mathbf{y}} + w_z\hat{\mathbf{z}}) \\
&= v_x w_x T(\hat{\mathbf{x}},\hat{\mathbf{x}}) + v_x w_y T(\hat{\mathbf{x}},\hat{\mathbf{y}}) + v_x w_z T(\hat{\mathbf{x}},\hat{\mathbf{z}}) \\
&\quad + v_y w_x T(\hat{\mathbf{y}},\hat{\mathbf{x}}) + v_y w_y T(\hat{\mathbf{y}},\hat{\mathbf{y}}) + v_z w_z T(\hat{\mathbf{y}},\hat{\mathbf{z}}) \\
&\quad + v_z w_x T(\hat{\mathbf{z}},\hat{\mathbf{x}}) + v_z w_y T(\hat{\mathbf{z}},\hat{\mathbf{y}}) + v_z w_z T(\hat{\mathbf{z}},\hat{\mathbf{z}})
\end{aligned} \tag{3}
$$

where we can make use of the notation,

$$
\begin{aligned}
T_{xx} &\equiv T(\hat{\mathbf{x}},\hat{\mathbf{x}}) \\
T_{xy} &\equiv T(\hat{\mathbf{x}},\hat{\mathbf{y}}) \\
T_{yx} &\equiv T(\hat{\mathbf{y}},\hat{\mathbf{x}})\ ,
\end{aligned} \tag{4}
$$

and so on which leads to a compact form of Eq.(3):

$$
\begin{aligned}
T(v,w) = v_x w_x T_{xx} &+ v_x w_y T_{xy} + v_x w_z T_{xz} \\
&+ v_y w_x T_{yx} + v_y w_y T_{yy} + v_y w_z T_{yz} \\
&+ v_z w_x T_{zx} + v_z w_y T_{zy} + v_z w_z T_{zz}\ .
\end{aligned} \tag{5}
$$

Usually, Eq.(5) is taken as the definition of a second rank tensor but from the modern perspective of looking at tensors, this equation is clearly just a consequence of multilinearity. In this setting, the **components** $\{T_{xx}, T_{xy}, T_{xz},\ ...\}$ of T are the **values of the tensor when it is evaluated on a given set of basis vectors.**

2.1 Tensor Transformation Law from Multilinearity

What we want to do now is show that looking at tensors as multilinear functions is completely equivalent to the traditional approach to treating tensors as objects that must obey certain transformation laws under coordinate changes. To that end, we will attempt to derive the transformation law for tensors from our standpoint.

First, we ask what happens to the tensor T from **Example 1** if we switch to a new set of basis $\{\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}'\}$ which is related to the old basis as,

$$
\begin{aligned}
\hat{\mathbf{x}}' &= A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}} \\
\hat{\mathbf{y}}' &= A_{y'x}\hat{\mathbf{x}} + A_{y'y}\hat{\mathbf{y}} + A_{y'z}\hat{\mathbf{z}} \\
\hat{\mathbf{z}}' &= A_{z'x}\hat{\mathbf{x}} + A_{z'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}\ .
\end{aligned} \tag{6}
$$

Note that, the mathematical construct we are calling a tensor, does not care about this shift in basis since T **will remain a tensor in the new basis**. What will change though is its values (or components). Think of vectors under basis changes if this confuses you: a vector remains a vector if you move to a new basis (in the primitive sense, it is still a pointy arrow of a certain length); however, the value of its components are subject to change. Thus, the new tensor components expressed in terms of the new basis are $\{T_{x'x'}, T_{x'y'}, ...\}$ For instance, working out one of these new components yields,

$$\begin{aligned}
T(\hat{\mathbf{x}}, \hat{\mathbf{x}}) \equiv T_{x'x'} &= T(A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}, A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}) \\
&= A_{x'x}T(\hat{\mathbf{x}}, A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}) + A_{x'y}T(\hat{\mathbf{y}}, A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}) \\
&\quad + A_{x'z}T(\hat{\mathbf{z}}, A_{x'x}\hat{\mathbf{x}} + A_{x'y}\hat{\mathbf{y}} + A_{x'z}\hat{\mathbf{z}}) \\
&= A_{x'x}(A_{x'x}T(\hat{\mathbf{x}}, \hat{\mathbf{x}}) + A_{x'y}T(\hat{\mathbf{x}}, \hat{\mathbf{y}}) + A_{x'z}T(\hat{\mathbf{x}}, \hat{\mathbf{z}})) \\
&\quad + A_{x'y}(A_{x'x}T(\hat{\mathbf{y}}, \hat{\mathbf{x}}) + A_{x'y}T(\hat{\mathbf{y}}, \hat{\mathbf{y}}) + A_{x'z}T(\hat{\mathbf{y}}, \hat{\mathbf{z}})) \\
&\quad + A_{x'x}(A_{x'x}T(\hat{\mathbf{z}}, \hat{\mathbf{x}}) + A_{x'y}T(\hat{\mathbf{z}}, \hat{\mathbf{y}}) + A_{x'z}T(\hat{\mathbf{z}}, \hat{\mathbf{z}})) \\
&= A_{x'x}A_{x'x}T_{xx} + A_{x'x}A_{x'y}T_{xy} + A_{x'x}A_{x'z}T_{xz} \\
&\quad + A_{x'y}A_{x'x}T_{yx} + A_{x'y}A_{x'y}T_{yy} + A_{x'y}A_{x'z}T_{yz} \\
&\quad + A_{x'z}A_{x'x}T_{zx} + A_{x'z}A_{x'y}T_{zy} + A_{x'z}A_{x'z}T_{zz} \ .
\end{aligned} \tag{7}$$

Quite frankly, this looks pretty messy and working out such a messy expression could easily lead to mistakes. As such, we never write out the components explicitly and rather resort to the aesthetic nature of tensors which enables to write (7) in the compact form using ***indices***

$$\boxed{T_{i'j'} = A^k_{i'} A^l_{j'} T_{kl}} \ , \tag{8}$$

which is the well-known **tensor transformation law**. If this looks confusing to you, we are on the right track! Soon, the tensor form of the transformation law will make the most sense, I hope.

2.2 Are Matrices Tensors?

Say, someone hands you a square matrix

$$T_{ij} = \begin{pmatrix} 4 & 8 & 1 \\ 1 & 4 & 2 \\ 12 & 9 & 5 \end{pmatrix}$$

and asks: "is this a tensor?" You may think, well it is denoted by indices and surely that looks like something that emerges from Eq.(5), and naturally be inclined to respond "yes it is a tensor". But the best answer here is: **"I do not have enough information to answer that!"**

Our interpretation of a tensor as a multilinear function on a vector space requires us to first choose a basis, say $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, which then allows us to express a tensor T in the chosen basis as,

$$[T] = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix} \tag{9}$$

where $\{T_{xx}, T_{xy}, T_{xz}, ...\}$ are the components of T. This then allows us to write, for example Eq.(5), in a more aesthetically pleasing form

$$T(v, w) = \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} \underbrace{\begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix}}_{[T]} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \ . \tag{10}$$

However, one must always remember that the association between a tensor and a matrix is dependent entirely on the choice of basis. Thus, it is best to think of a tensor

$$T: \text{ a multilinear function}$$

and the matrix $[T]$ as

$$[T]: \text{ a matrix representation of } T \text{ in a particular coordinate system.}$$

2.3 Tensors as Operators

Sometimes, matrices and tensors are considered to be **linear operators** which act on vectors and generates a new vector. Say, we have a linear operator R such that R can be turned into a second rank tensor T_R by

$$T_R(v, w) \equiv v \cdot Rw \ , \tag{11}$$

where the '$\cdot$' denotes usual vector dot product. This equation may be read as, "T takes in v and w and spits out the components of Rw parallel to v". For instance, if we compute the components of T_R as,

$$\begin{aligned} T_R(\hat{\mathbf{x}}, \hat{\mathbf{x}}) &\equiv (T_R)_{xx} \\ &= \hat{\mathbf{x}} \cdot R\hat{\mathbf{x}} \\ &= \hat{\mathbf{x}} \cdot (R_{xx}\hat{\mathbf{x}} + R_{xy}\hat{\mathbf{y}} + R_{xz}\hat{\mathbf{z}}) \\ \Rightarrow (T_R)_{xx} &= R_{xx} \ , \end{aligned} \tag{12}$$

and similarly for the other components, one notices that the components of T_R are the same as the components of the linear operator R! Thus the action of T_R (in the operator sense of the word) assumes the form,

$$T_R(v, w) = \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \ , \tag{13}$$

that is of course identical to Eq.(10). Hence, to turn a a linear operator R into a **second rank** tensor, just sandwich the component matrix in between two vectors (or more precisely, between a ***dual*** vector and a vector). The converse is also true, i.e. one can turn a second-rank tensor into a linear operator as,

$$(R_T(v))_x \equiv T(v, \hat{\mathbf{x}}) \ , \tag{14}$$

and similarly for other components. This shows that there is a one-to-one correspondence between linear operators and second rank tensors and thus one can indeed regard them as equivalent.

2.4 A Basic Physical Application of Tensors

Recall, from your introductory mechanics classes, the formula for the kinetic energy of rotating objects, usually written as,

$$KE = \frac{1}{2} I\omega^2 \tag{15}$$

where I is interpreted as the analogous quantity to m (mass) from translational motion, and ω of course denotes the angular speed. This interpretation leads one to think that, like m, I is also a scalar $-$ a big misconception. In reality, I is a function $I \equiv \mathcal{I}(\omega, \omega)$ in the sense of Eq.(3) that it "eats" two copies of ω ($v = w = \omega$) and "spits out" a scalar (kinetic energy), such that Eq.(15) can be recast as,

$$KE = \frac{1}{2}\mathcal{I}(\omega, \omega) \ , \tag{16}$$

which, upon choosing a basis, say $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$, can be expressed as

$$KE = \frac{1}{2} = \begin{pmatrix} \omega_x & \omega_y & \omega_z \end{pmatrix} \begin{pmatrix} \mathcal{I}_{xx} & \mathcal{I}_{xy} & \mathcal{I}_{xz} \\ \mathcal{I}_{yx} & \mathcal{I}_{yy} & \mathcal{I}_{yz} \\ \mathcal{I}_{zx} & \mathcal{I}_{zy} & \mathcal{I}_{zz} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \ . \tag{17}$$

Each component can then be calculated using the standard formulae for rigid objects.

The moment of inertia tensor can also be interpreted as a linear operator in the sense of Eq.(11) in that $\mathcal{I}$ is the linear operator that takes ω into the angular momentum L,

$$L = \begin{pmatrix} \mathcal{I}_{xx} & \mathcal{I}_{xy} & \mathcal{I}_{xz} \\ \mathcal{I}_{yx} & \mathcal{I}_{yy} & \mathcal{I}_{yz} \\ \mathcal{I}_{zx} & \mathcal{I}_{zy} & \mathcal{I}_{zz} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \ . \tag{18}$$

3 An Incomplete but Necessary Account of Vector Spaces

3.1 Abstract Vector Spaces

Whenever you think of objects termed as vectors, automatically the image of straight lines with pointy heads pop up in your mind. While that primitive idea is not to be completely discarded, it has to be acknowledged that the deeper one goes into physics (especially quantum physics) that simple notion of vectors becomes increasingly limited to work with. As such, one needs a more *abstract* notion of vectors and *vector spaces*.

Definition 1. An **abstract vector space** is a set V (whose elements are called *vectors*), together with a set of scalars C (we will always take C to be $\mathbb{R}$ or $\mathbb{C}$) and the operations of addition and multiplication that satisfy the following axioms:

1. $v + w = w + v$; $\forall\ v, w \in V$ (**Commutativity**).
2. $v + (w + x) = (v + w) + x$; $\forall\ v, w, x \in V$ (**Associativity**).
3. $\exists$ a vector $0 \in V$ s.t. $0 + v = v\ \forall\ v \in V$.
4. $\exists$ a vector $-v \in V$ s.t. $-v + v = 0\ \forall\ v \in V$.
5. $c(v + w) = cv + cw$; $\forall\ v, w \in V$& scalars c (**Distributivity**).
6. $1v = v\ \forall\ v \in V$.
7. $(c_1 + c_2)v = c_1 v + c_2 v\ \forall\ v \in V$ and all scalars c_1, c_2.
8. $(c_1 c_2)v = c_1(c_2 v)\ \forall\ v \in V$ and all scalars c_1, c_2.

Crucially, our understanding of vector spaces relies mostly on these essential properties of addition and multiplication, rather than properties that are particular to $\mathbb{R}^n$ or $\mathbb{C}^n$. This indicates that abstract vector spaces can be thought of *independently of any basis.* Some examples of vector spaces are:

1. $\mathbf{R^n}$: the **real** vector space. If $n = 3$ we have the well-known Euclidean 3−dimensional space $\mathbb{R}^3$. For $n = 4$ we have the $\mathbb{R}^4$ spacetime of special relativity.

2. $\mathbb{C}^n$: the **complex** vector space occurring mainly in finite-dimensional quantum systems. Additionally, any complex vector space is also a real vector space. Example: a spin−1/2 particle *fixed in space* has a *ket* space identifiable with $\mathbb{C}^2$.

3. $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$: $n \times n$ matrices with real or complex entries.

3.2 Subspace, Span, Linear Independence, and Bases

Definition 2. A **subspace** is a subset of a vector space which itself forms a vector space by satisfying the requirement that the subset is closed under addition and includes the 0 vector.

Example 2. $H_n(\mathbb{C})$ the set of $n \times n$ Hermitian matrices with complex entries is a subset and a subspace of $M_n(\mathbb{C})$. A matrix, A, is Hermitian if it satisfies,

$$A^\dagger \equiv (A^T)^* = A \tag{19}$$

where the superscripts T and $*$ denotes the transpose and the complex conjugation of the entries, respectively. Another interesting fact about Hermitian matrices is the fact that even though the entries of its matrices can be complex, $H_n(\mathbb{C})$ **does not** form a complex vector space; multiplying a Hermitian matrix by i yields *anti*-Hermitian matrix, so $H_n(\mathbb{C})$ is **not** closed under complex scalar multiplication.

In quantum mechanics, physical observables are necessarily represented by Hermitian operators and in finite dimensional ket spaces, the operators can be represented as elements of $H_n(\mathbb{C})$. Example: Pauli spin matrices σ_i which define the angular momentum operator for a spin−1/2 particle, are Hermitian matrices.

Definition 3. A set of vectors **spans** a vector space if all the vectors in the space can be written as **linear combinations** of the spanning set. For instance, if $S = \{(1,0,0),(0,1,0)\} \subset \mathbb{R}^3$, then the linear combination

$$c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix}$$

spans $\mathbb{R}^3 \ \forall \ c_1, c_2 \in \mathbb{R}$.

Definition 4. Following from the last definition, we can say the vector space S is **linearly independent** since there is no way to generate one of the elements from the other. On the other hand, $S' = \{(1,0,0),(0,1,0),(1,1,0)\} \subset \mathbb{R}^3$ is linearly *dependent.*

Definition 5. A **basis** for a vector space is an *ordered* set $\mathcal{B}\{v_1, ..., v_k\} \subset V$, of linearly independent vectors that spans all of V. Simply put, one can express all other vectors in V as linear combinations of the basis vectors. Basis vectors are usually denoted as $\{\hat{\mathbf{e}}_i\}_{i=1,...,k}$.

Definition 6. The **dimension** of a vector space V, denoted as dim V, is the number of elements of any finite basis. If no finite basis exists, we conclude that V is **infinite dimensional**. For instance, for three dimensional Euclidean space, $\mathcal{B} = \{(1,0,0),(0,1,0),(0,0,1)\}$.

Putting the above notions in use, one can express the elements of a vector space as $n-$tuples (as row or column vectors). Given $v \in V$ and a basis $\mathcal{B} = \{e_i\}_{i=1,..,n}$ for V, we can write

$$\boxed{v = \sum_{i=1}^{n} v^i e_i} \tag{20}$$

for some numbers v^i, which you can identify as the **components of** v **with respect to the chosen basis** $\mathcal{B}$. It follows then that v can be represented by the **column vector**,

$$[v]_{\mathcal{B}} \equiv \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} \ , \tag{21}$$

or the row vector

$$[v]^T_{\mathcal{B}} \equiv \begin{pmatrix} v^1 & v^2 & ... & v^n \end{pmatrix} \ , \tag{22}$$

where, in both expressions the subscript $\mathcal{B}$ refers to the choses basis and the superscript T in the second expression denotes the usual transpose of a vector. We use these expressions only to simplify computations and one must keep in mind that **vectors exist independent of any chosen basis**. However, how the vectors appear as row and column matrices **very much depends on** the choice of the basis.

3.3 Linear Operators

In physics, especially in quantum mechanics, the notion of **linear operator** is of great importance.

Definition 7. A **linear operator** on a vector space V is a function T from V to itself satisfying the **linearity** condition

$$T(cv + w) = cT(v) + T(w) \ . \tag{23}$$

The set of all linear operators on V forms a vector space (straightforward to check if all the axioms for a vector space is satisfied). The vector space of all linear operators on V is denoted as $\mathcal{L}(V)$. Examples of linear operators are plentiful: any $n \times n$ matrix can be interpreted as a linear operator acting on column vectors by matrix multiplication; thus, $\mathbb{M}_n(\mathbb{R})$ can be viewed as a vector space whose elements are all linear operators which means

$$\mathbb{M}_n(\mathbb{R}) \equiv \mathcal{L}(\mathbb{R}^n)$$

in our notation.

3.3.1 Invertible Linear Operators

A linear operator T is said to invertible if there exists a linear operator T^{-1} such that,

$$TT^{-1} = T^{-1}T = I \tag{24}$$

where I is the identity operator.

3.3.2 A Matrix is a Linear Operator?

The answer is: *may or may not be.* Generally, **a linear operator is *not* the same thing as a matrix**; that identification can be made *only after* a basis has been chosen. For operators on finite-dimensional spaces this is done as follows: choose a basis $\mathcal{B} = \{e_i\}_{i=1,..,n}$, then the action of T is determined by its action on the basis vectors:

$$T(v) = T\left(\sum_{j=1}^{n} v^i e_i\right) = \sum_{j=1}^{n} v^i T(e_i) = \sum_{j=1}^{n} v^i T_i{}^j e_j \ , \tag{25}$$

where the numbers $T_i{}^j$ are the **components of T relative to the basis $\mathcal{B}$** are defined by

$$\boxed{T(e_i) = \sum_{j=1}^{n} T_i{}^j e_j} \ . \tag{26}$$

Then, the action of T on a vector in a chosen basis becomes,

$$[v]_{\mathcal{B}} \equiv \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} \quad \text{and} \quad [T(v)]_{\mathcal{B}} = \begin{pmatrix} \sum_{i=1}^n v^i T_i{}^1 \\ \sum_{i=1}^n v^i T_i{}^2 \\ \vdots \\ \sum_{i=1}^n v^i T_i{}^n \end{pmatrix}$$

which leads us to the definition of the **matrix of T in the basis $\mathcal{B}$** by the matrix equation,

$$[T(v)]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}} \tag{27}$$

where the product on the RHS is just the usual matrix multiplication. This means that one can express the linear operator in matrix form,

$$[T]_{\mathcal{B}} = \begin{pmatrix} T_1{}^1 & T_2{}^1 & \dots & T_n{}^1 \\ T_1{}^2 & T_2{}^2 & \dots & T_n{}^2 \\ \vdots & \vdots & \vdots & \vdots \\ T_1{}^n & T_2{}^n & \dots & T_n{}^n \ , \end{pmatrix} \tag{28}$$

entries are the components of T. In this form, the action of the linear operator T just becomes ordinary matrix multiplication by $[T]_{\mathcal{B}}$.

3.4 Dual Spaces

Given a vector space V with scalars C, a ***dual vector*** (or ***linear functional***) on V is a $C-$valued linear function f on V, where linearity implies,

$$f(cv + w) = cf(v) + f(w) \ . \tag{29}$$

Of note here is that $f(v),\ f(w)$ are scalars and so the the sum on the LHS happens in V while the sum on the RHS happens in C. Thus, dual vectors eats a vector and spits out a scalar. The set of all dual vectors on V is called the **dual space** of V, denoted by V^*.

Aside: During conversations with you all, I came to realize that a lot of you are familiar with the concept of a *field* and wanted to include this aside. A *field* is an algebraic structure that allows you to add $(+,-)$ and multiply $(\cdot, :)$ with the usual rules intact. The elements of a field F should be considered as numbers. For instance, $\mathbb{R}, \mathbb{C}$ are fields but $\mathbb{Z}$ (set of integers) is not $-$ the multiplicative inverse of non-zero elements $\mathbb{Z}$ takes you out of $\mathbb{Z}$. A *vector space* V is defined **over** a field F, where, in addition to the usual vector space operations being defined, we can define *scaling* of vectors by elements λ in $F : \lambda v \in V\ \forall\ v \in V$. It should be emphasized that the two notions of a field and a vector space are different that can make sense in the same discipline. Thus, in Eq.(28), the LHS is defined strictly **in** V but the RHS takes one out of the vector space.

Example 3. let $\{e_i\}$ be a basis for V, so that an arbitrary vector v can be written as

$$v = \sum_{i=1}^{n} v^i e_i$$

then for each i we can define a **dual basis vector** e^i as

$$e^i(v) \equiv v^i \ , \tag{30}$$

which is interpreted as e^i "picks off" the $i-$th component of the vector v; another way to look at it is that the $i-$th component of a vector is the value of e^i on that vector. We sometimes refer to e^i as being ***dual to*** e_i and satisfy

$$e^i(e_j) = \delta^i{}_j \ , \tag{31}$$

where $\delta^i{}_j$ is the usual Kronecker delta defined as

$$\delta^i{}_j = \begin{cases} 1; i = j \\ 0; i \neq j \end{cases} \tag{32}$$

As such e^i span the dual space V^* and form the basis for V^*. If V is finite dimensional with dimension n, we can thus write

$$f = \sum_{i=1}^{n} f_i e^i \tag{33}$$

for any dual vector f with components f_i defined in the dual basis $\mathcal{B}^* \equiv \{e^i\}$ which form the dual space V^* with dim $V^* = n$.

3.5 Non-degenerate Hermitian Forms and Connection to Dual Spaces

A **non-degenerate Hermitian form** on a vector space V is a scalar valued function denoted as $(\cdot \mid \cdot)$, which assigns to an ordered pair of vectors $v,\ w \in V$ a scalar $(v|w)$ having the following properties:

1. $(v|w_1 + cw_2) = (v|w_1) + c(v|w_2)$ (**linearity in the second argument**).
2. $(v|w) = \overline{(w|v)}$ (**Hermiticity** - the bar denotes complex conjugation).
3. For each $v \neq 0 \in V$, there exists $w \in V$ such that $(v|w) \neq 0$ (**non-degeneracy**).
4. $(v|v) > 0$ for all $v \in V,\ v \neq 0$ (**positive-definiteness**). (*optional property*)

Note that 1 & 2 $\Rightarrow$ $(cv|w) = \bar{c}(v|w)$ which means that $(\cdot \mid \cdot)$ is **conjugate linear** in the first argument. Moreover, for real vector spaces, condition 2 $\Rightarrow$ $(\cdot \mid \cdot)$ is symmetric i.e. $(v|w) = (w|v)$ which means that $(\cdot \mid \cdot)$ is linear in both the first and the second arguments and in this case we refer to them as **bilinear**. When Condition 4 is satisfied, the Hermitian form is called an **inner product** and the vector space with such Hermitian forms are referred to as **inner product spaces**. You may recall the notion of *norm* or *length squared* of a vector v defined as $||v|| = \sqrt{(v|v)}$ - this is nothing but the square root of the Hermitian form. A widely applied aspect of Hermitian forms is to define **orthonormal bases** with the property $(e_i|e_j) = \delta_{ij}$ for bases $\mathcal{B} = \{e_i\}$, which are extremely computation friendly.

Example 4. The Hermitian scalar product on $\mathbb{C}^n$.

$$\text{Let } v = (v^1, .., v^n),\ w = (w^1, .., w^n) \in \mathbb{C}^n.$$

$$\text{Then: } (v|w) \equiv \sum_{i=1}^{n} \bar{v}^i w^i\ .$$

This is the *Hermitian scalar product* and is found everywhere on quantum mechanical vector spaces. Would this be a inner product if the conjugation in v was not there?

Example 5. The Minkowski metric on $4-D$ spacetime.

Consider two vectors (*events in spacetime*) $v_i = (t_i, x_i, y_i, z_i) \in \mathbb{R}^4,\ i = 1, 2$. One can then define the Minkowski metric as,

$$\eta(v_1, v_2) = t_1t_2 - x_1x_2 - y_1y_2 - z_1z_2$$

which is written in the form of a particle physicists metric - differs from that of a cosmologist! η is defined on a real vector space and thus is *symmetric*; it's clearly *bilinear* as well and furthermore one can always find a $w \neq v \in V \Rightarrow (v|w) \neq 0$ which satisfies non-degeneracy. However, note that, if $v = (1, 0, 0, 1)$, then $\eta(v, v) = 0$ (not positive definite) and hence $(v|v)$ is **not** an inner product. Thus, general non-degenerate Hermitian forms do not necessarily need to satisfy condition 4. Note that the Minkowski metric can be written in matrix form upon choosing a basis, say $\{e_i\}_{i=0,..,3} \in \mathbb{R}^4$, with its components η_{ij} defined as

$$\boxed{\eta_{ij} \equiv \eta(e_i, e_j)}\ . \tag{34}$$

3.5.1 Connection to Dual Spaces

For a finite dimensional vector space V, for any $v \in V$ there is an associated dual vector $\tilde{v} \in V^*$ defined by

$$\tilde{v} \equiv (v|w) \ , \tag{35}$$

which defines the map,

$$\begin{aligned} L &: V \to V^* \\ v &\mapsto \tilde{v} \ . \end{aligned} \tag{36}$$

Thus, $\tilde{v}$ is sometimes denoted as $L(v)$ or $(v|\cdot)$ and referred to as the **metric dual** of v. L is conjugate linear since for $v = cx + z, \ v, z, x \in V$,

$$\begin{aligned} &\tilde{v}(w) = (v|w) = (cx + z|w) = \bar{c}(x|w) + (z|w) = \bar{c}\tilde{x}(w) + \tilde{z}(w) \\ &\Rightarrow \tilde{v} = L(cx + z) = \bar{c}\tilde{x} + \tilde{z} = \bar{c}L(x) + L(z) \ . \end{aligned} \tag{37}$$

Furthermore, satisfying the non-degeneracy of the Hermitian form implies the map L is **one-to-one** and **onto** (surjective), so that L is an invertible map from V to V^*.

Aside: If we have a basis $\{e_i\}_{i=1,..,n}$ and the corresponding dual basis $\{e^i\}_{i=1,..,n}$, **it is not necessarily true that a dual vector e^i in the dual basis is the same as the metric dual** $L(e_i)$.

Example 6. Bras and kets in quantum mechanics.

Let $\mathcal{H}$ be a quantum mechanical Hilbert space with inner product $(\cdot \mid \cdot)$. If $\psi, \phi \in \mathcal{H}$ then

$$(\phi|\psi) = \langle\phi|\psi\rangle = \langle\psi|\phi\rangle$$

where $\langle\psi|$ is read as *bra vector* ψ which is just the **dual vector to** $|\psi\rangle$. In our notation then, $\langle\psi| = L(\psi) = (\psi| \ \cdot)$.

Example 7. Index gymnastics in relativity.

In $\mathbb{R}^4$ let us choose the standard basis $\mathcal{B} = \{e_\mu\}_{\mu=0,..,3}$ and its dual basis $\mathcal{B}' = \{e^\mu\}_{\mu=0,..,3}$ which allows us to write $v \in \mathbb{R}^4$ as

$$v = \sum_{\mu=0}^{3} = v^\mu e_\mu \ .$$

What then are the components of the dual vector $\tilde{v}$ in terms of the v^μ?

$$\tilde{v}_\mu = \tilde{v}(e_\mu) = (v|e_\mu) = \sum_{\nu=0}^{3}(v^\nu e_\nu|e_\mu) = \sum_{\nu=0}^{3} v^\nu \underbrace{(e_\nu|e_\mu)}_{=\eta(e_{\nu\mu})} = \sum_{\nu=0}^{3} v^\nu \eta_{\nu\mu} \equiv v_\mu$$

where v^ν are identified as the **contravariant** components of v while v_μ are known as the **covariant** components of v. In matrix form the above expression can be written as,

$$[\tilde{v}]_{\mathcal{B}'} = [\eta]_{\mathcal{B}}[v]_{\mathcal{B}} \ . \tag{38}$$

Conversely then, for a dual vector f, the invertibility of L implies

$$[\tilde{f}] = [\eta]^{-1}[f] \ , \tag{39}$$

where $\tilde{f} \equiv L^{-1}(f)$ and in component notation we have,

$$\tilde{f}^{\mu} = \sum_{n} \eta^{\nu\mu} f_{\nu} \ .$$

Thus, the process of *raising* and *lowering* indices that one encounters in relativity is just the application of the map L (and its inverse) in components.

Aside: From here on we will use **Einstein summation convention** which states that **whenever you see repeated indices, one raised and one lowered, a sum is implied**; e.g.

$$\sum_{n} \eta^{\nu\mu} f_{\nu} \equiv \eta^{\nu\mu} f_{\nu} = f_{\nu} \ .$$

When there are exceptions to this, we will explicitly mention that in context.

4 Tensors, Tensors, and more Tensors

Suppose V is a vector space and V^* is its corresponding dual space, then a **tensor of type (r, s)** is the **multilinear** map

$$T^r_s \ : \ \underbrace{V \times ... \times V}_{r \text{ times}} \times \underbrace{V^* \times ... \times V^*}_{s \text{ times}} \to \mathbb{R} \ . \tag{40}$$

The set of all such mappings for fixed r and s forms a vector space denoted by $T^r_s(V) \equiv T^r_s$. The number r is called the **covariant degree** of the tensor while the number s is called the **contravariant degree** of the tensor. In this setting, a scalar is a $(0, 0)$ tensor, an ordinary vector is a $(0, 1)$ tensor (hence called a ***contravariant vector***) while a dual vector is a $(1, 0)$ tensor (hence called a ***covariant vector***), while linear operators can be viewed as $(1, 1)$ tensors. Multilinearity of tensors implies that tensor components are just the values of the tensors on the basis vectors, once a basis is chosen:

$$\boxed{T_{i_1,..,i_r}{}^{j_1,..,j_s} \equiv T(e_{i_1}, ..., e_{i_r}, e^{j_1}, ..., e^{j_s})} \ . \tag{41}$$

Example 8. $(1, 1)$ tensors in quantum mechanics.

Given an operator H on a quantum mechanical Hilbert space spanned by the orthonormal basis $\{e_i\}$ (usually written as $\{|i\rangle\}$), we can write H as

$$\begin{aligned} H_i{}^j &= H(e_i, e^j) \\ &= e^j(He_i) \\ &= \langle j| \, |Hi\rangle \\ &= \langle j| \, H \, |i\rangle \ . \end{aligned}$$

in usual Dirac notations. $\langle j| \, H \, |i\rangle$ are often referred to as **matrix elements** of the operator H.

Example 9. Levi-Civita Symbol, ϵ, on $\mathbb{R}^3$ is defined by

$$\epsilon(u,v,w) \equiv (u \times v)\cdot w,\ \ u,v,w \in \mathbb{R}^3\ . \tag{42}$$

In the standard basis

$$\begin{aligned}\epsilon_{ijk} &= \epsilon(e_i,e_j,e_k)\\ = (e_i \times e_j)\cdot e_k &= \begin{cases} +1; \{i,j,k\} = \text{ even permutation of } ijk\\ -1; \{i,j,k\} = \text{ odd permutation of } ijk\\ \ 0; \qquad\quad \text{any two indices equal.}\end{cases}\end{aligned}$$

4.1 Change of Basis

Consider a vector space V and you are given two choices for basis vectors: $\mathcal{B} = \{e_i\}_{i=1,..,n}$ and $\mathcal{B}' = \{e_{i'}\}_{i'=1,..,n}$. Thus, each of $e_{i'}$ can be expressed in terms of e_i as

$$e_{i'} = A^j_{i'}e_j \ \ \text{(notice Einstein sum applied here)} \tag{43}$$

for some numbers $A^j_{i'}$ and likewise

$$e_i = A^{j'}_i e_{j'}\ , \tag{44}$$

for some numbers $A^{j'}_i$. This leads to

$$e_i = A^{j'}_i e_{j'} = A^{j'}_i A^k_{j'} e_k \tag{45}$$

where we used Eq.(43) and Eq.(44) and replaced the repeated index j in (43) by k $-$ this is a general procedure and you can replace repeated indices with any other letters and as such repeated indices are termed **dummy indices**. From Eq.(45) we can conclude that,

$$A^{j'}_i A^k_{j'} = \delta^k_i\ , \tag{46}$$

where the $A^{j'}_i$ and $A^i_{j'}$ are the components of the tensor in the chosen basis. In a way, $A^{j'}_i$ and $A^j_{i'}$ are inverses of each other. The next question to investigate now is **how do the corresponding dual bases transform?**

$$\text{Let } \{e^i\}_{i=1,..,n},\ \{e^{i'}\}_{i=1,..,n} \text{ be the bases } \textit{dual to } \mathcal{B},\mathcal{B}'\ .$$

Then the components of $\{e^{i'}\}$ expressed relative to $\{e^i\}$ become

$$\begin{aligned} e^{i'}(e_j) = e^{i'}(A^{k'}_j e_{k'}) &= A^{k'}_j e^{i'} e_{k'} = A^{k'}_j \delta^{i'}_{k'} = A^{i'}_j\\ \Rightarrow e^{i'} &= A^{i'}_j e^j\\ \text{similarly, } \Rightarrow e^i &= A^i_{j'} e^{j'}\ . \end{aligned} \tag{47}$$

This leads us to the *de facto* definition of tensors that concerns how the components of a tensor transform under a change in basis

$$\begin{aligned} T_{i'_1,..,i'_r}{}^{j'_1,..,j'_s} &= T(e_{i'_1},...,e_{i'_r},e^{j'_1},...,e^{j'_s})\\ &= T(A^{k_1}_{i'_1}e_{k_1},...,A^{k_r}_{i'_r}e_{k_r},A^{j'_1}_{l_1}e^{l_1},...,A^{j'_s}_{l_s}e^{l_s})\\ &= A^{k_1}_{i'_1}...A^{k_r}_{i'_r}A^{j'_1}_{l_1}...A^{j'_s}_{l_s}T(e_{k_1},..,e_{k_r},e^{l_1},...,e^{l_s})\\ \Rightarrow \boxed{T_{i'_1,..,i'_r}{}^{j'_1,..,j'_s} = A^{k_1}_{i'_1}...A^{k_r}_{i'_r}A^{j'_1}_{l_1}...A^{j'_s}_{l_s}T_{k_1...k_r}{}^{l_1...l_s}}&\ . \end{aligned} \tag{48}$$

Clearly, then the perspective from which we have approached the study of tensors, i.e. tensors as multilinear functions on V and V^*, is equivalent to the view that **tensors are objects that transform in a certain way - i.e. by Eq.**(48).

4.2 Examples of Tensor Transformations

Example 10. Consider the the standard basis $\mathcal{B}$ in $\mathbb{R}^2$ and another basis $\mathcal{B}'$ obtained by rotating $\mathcal{B}$ by an angle ϕ, as shown in the figure below. How do the components transform under this change of basis? Upon rotation $|e_1| = |e_{1'}|$ must hold. As such, by simple trigonometry we can arrive at

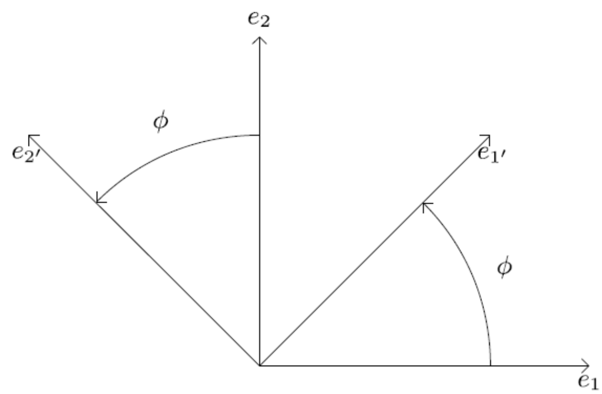

Figure 1: Change of basis under rotations.

the coordinates for $\mathcal{B}'$, written in matrix form as

$$R_i^{\ j} \equiv \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{49}$$

which is the famed **rotation matrix in** $2D$. This can be arranged into a system of equations as,

$$\begin{pmatrix} e_{1'} \\ e_{2'} \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \ . \end{pmatrix} \tag{50}$$

In general n dimensions, we can express this as

$$e_{i'} = R_{i'}^{\ j} e_j \ . \tag{51}$$

With the restriction that $e_{i'}$ are orthonormal such that we have,

$$e_i \cdot e_j = e_{i'} \cdot e_{j'} = \delta_{ij}$$

which implies

$$e_{i'} \cdot e_{j'} = \delta_{i'j'} = R_{i'}^{\ k} e_k \cdot R_{j'}^{\ l} e_l = R_{i'}^{\ k} R_{j'}^{\ l} \underbrace{e_k \cdot e_l}_{\delta_{kl}} = R_{i'}^{\ k} R_{j'}^{\ k}$$

which in matrix notation leads to

$$RR^T = \mathbb{I} \Rightarrow R^T = R^{-1} \ . \tag{52}$$

Matrices with this property are said to be **orthogonal** and we can write $R \in O(n)$. Additionally, the determinant of R, det $R = \pm 1$ and if we choose the positive unity for the determinant, the rotations are said to be **special orthogonal** and in that case we write $R \in SO(n)$. For instance, in $\mathbb{R}^3$, a rotation $R \in SO(3)$ takes a right-handed orthonormal basis into another right-handed orthonormal basis. What do you reckon happens if we choose the determinant to be negative unity?

Notice that Eq.(52) is a special case of the general tensor transformation law (Eq.(48)) if one replaces R with A.

Example 11. Vectors and dual vectors.

Remember a vector itself **remains unchanged** under a change of basis, but its components transform according to Eq.(48), as discussed below.

From Eq(48) we see that the components of a vector (or a $(0,1)$ tensor) transforms as

$$v^{i'} = A^{i'}_j v^j \tag{53}$$

while the components of a dual vector f transforms as,

$$f_{i'} = A^j_{i'} f_j \ . \tag{54}$$

Of note here is the realization that v transforms with the $A^{i'}_j$ (*opposite* of e_i basis vectors) while f transforms with $A^j_{i'}$ (*same* as e_i basis vectors). Thus, **vectors transform opposite (or *contrary*) to corresponding basis vectors and thus are referred to as *contravariant*** while the **dual vectors transform in the *same* way as the basis vectors and thus are referred to as *covariant***. As such, the opposite transformation laws for upper-indexed and lower-indexed objects mean that, for instance, the quantity $v \equiv v^i e_i$ represents an **invariant** − the vector itself is basis independent and blind to coordinate transformations.

Exercise 1. Using the discussion above, prove that

$$[f]^T_{\mathcal{B}'}[v]_{\mathcal{B}'} = [f]^T_{\mathcal{B}}[v]_{\mathcal{B}} \ .$$

Example 12. Linear Operators.

Linear operators can be viewed as $(1,1)$ tensors and as such transform according to Eq.(48) as

$$T_{i'}{}^{j'} = A_{i'}{}^k A_l{}^{j'} T_k{}^l \tag{55}$$

which in matrix form reads

$$[T]_{\mathcal{B}'} = A[T]_{\mathcal{B}}A^{-1} \tag{56}$$

which can be identified as the **similarity transformation** of matrices.

Example 13. $(2,0)$ tensors

The Minkowski metric is an example of such tensors. Consider a $(2,0)$ tensor g_{ij} which, under a change of basis, transforms according to Eq.(48) as,

$$g_{i'j'} = A^k_{i'} A^l_{j'} g_{kl} \tag{57}$$

which in matrix form becomes,

$$[g]_{\mathcal{B}'} = A^{-1^T}[g]_{\mathcal{B}}A^{-1} = A[g]_{\mathcal{B}}A^{-1} \tag{58}$$

which closely resembles Eq.(56).

Fun fact: If one chooses orthonormal bases, then (56) and (58) are identical and there is no need to distinguish between a $(1,1)$ or a $(2,0)$ tensor.

4.3 Some Definitions for Tensor Analysis

4.3.1 Addition and Subtraction of Tensors

This is defined very similarly to vector addition in that for two rank 2 tensors A and B we have

$$A^{ij} + B^{ij} = C^{ij} \tag{59}$$

and in general, A and B must be tensors of the same rank and in the same space.

4.3.2 Symmetric and Antisymmetric Tensors

A tensor A with the property

$$A^{mn} = A^{nm} \tag{60}$$

is known as a **symmetric tensor**. On the other hand, if

$$A^{mn} = -A^{nm} \tag{61}$$

then A is an **antisymmetric tensor**. As such, every rank 2 tensor can be decomposed into a symmetric part and an antisymmetric part using the identity

$$A^{mn} = \underbrace{\frac{1}{2}(A^{mn} + A^{nm})}_{\text{symmetric}} + \underbrace{\frac{1}{2}(A^{mn} - A^{nm})}_{\text{antisymmetric}} \ . \tag{62}$$

4.3.3 Isotropic Tensors

An isotropic tensor is a tensor which has the same components in all rotated coordinate systems. ϵ_{ijk} and δ_i^j are examples of (Cartesian) isotropic tensors. To show that a tensor is isotropic, one needs to establish that the rotated tensor (*primed*) is the same as the un-rotated one.

Example 14. Using Eq.(48) let's investigate if the Kronecker delta tensor is indeed isotropic.

$$\delta_{i'}^{\ j'} = A_{i'}^{\ m}\ A_n^{\ j'}\ \delta_m^{\ n} = A_{i'}^{\ m}\ A_m^{\ j'} = \delta_{i'}^{\ j'} \ . \tag{63}$$

Note the Einstein summation convention in use above: in the second equality, $A_n^{\ j'} \to A_m^{\ j'}$ because $\delta_m^{\ n} \neq 0$ only when $m = n$. The next equality leads to summing over m's. You can also think of $A_{i'}^{\ m}\ A_m^{j'}$ as the dot product of rows i' and j' of the rotation matrix A which is just $\delta_{i'}^{\ j'}$. It would probably be helpful at this point to introduce the matrices,

$$A \equiv \begin{pmatrix} A_1^{1'} & A_2^{1'} & \dots & A_n^{1'} \\ A_1^{2'} & A_2^{2'} & \dots & A_n^{2'} \\ \vdots & \vdots & \vdots & \vdots \\ A_1^{n'} & A_2^{n'} & \dots & A_n^{n'} \end{pmatrix} \tag{64}$$

$$A^{-1} \equiv \begin{pmatrix} A_{1'}^{1} & A_{2'}^{1} & \dots & A_{n'}^{1} \\ A_{1'}^{2} & A_{2'}^{2} & \dots & A_{n'}^{2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{1'}^{n} & A_{2'}^{n} & \dots & A_{n'}^{n} \end{pmatrix} \tag{65}$$

4.3.4 Contraction

For ordinary vectors, we write the dot product as

$$\mathbf{A} \cdot \mathbf{B} = \sum_i A_i B_i \ .$$

The generalization of this expression in tensor analysis is known as *contraction* or summing over repeated upper and lower indices.

Example 15. Contract ${B^i}_j$.

$$\text{First, set } j = i \Rightarrow B^{i'}_{i'} = \underbrace{{A_k}^{i'} {A_{i'}}^{l}}_{\delta^l_k} B^k_l = B^k_k \ .$$

In general, the operation of contraction reduces the rank of a tensor by 2.

4.3.5 Direct Product

The **direct product** of two tensors of any rank and covariant/contravariant characters is the multiplication of the tensors, component by component. The result is a tensor whose rank is the *sum of the ranks of the factors*, for instance:

$$C^{ij}_{klm} = A^i_k B^j_{lm}, \qquad F^{ij}_{kl} = A^j B^i_{lk} \ . \tag{66}$$

Note: the index order in the direct product can be defined as desired, but the covariance/contravariance of the factors must be maintained in the direct product. Moreover, with the advent of direct products, one can now derive meaningful quantities like $\boldsymbol{\nabla}\mathbf{E}$, which was not defined as a suitable product in the realm of vector analysis. However, other than in Cartesian coordinates, one should be careful about using such quantities involving differential operators since their transformation rules are only simple in Cartesian coordinates.

Example 16. The direct product of a_i and b^j yields a rank 2 tensor $C^i_j = a_i b^j$. Justify that C^i_j is indeed a tensor.

To show that the rank 2 object is a tensor, we investigate its transformation under coordinate changes using (48).

$$ {C_{i'}}^{j'} = a_{i'} \, b^{j'} = A^k_{i'} \; A^{j'}_l \; a_k \, b^l = A^k_{i'} \; A^{j'}_l \; {C^l}_k$$

which confirms that ${C_i}^j$ is a mixed tensor of rank 2.

4.4 Yet Another Way to Interpret Tensor Transformations

Consider a function, f, of the coordinates that represents a physical quantity. Let us choose two generic coordinate systems, x_i and x'_i, where the primed coordinate indicates the transformed frame. If the components of the gradient of f in x_j, $\partial f/\partial x_j$, are known, then one can find the components of the gradient in the transformed coordinates by the chain rule as,

$$\frac{\partial f}{\partial x'_i} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial x'_i} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{x'_i} + ... + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial x'_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\frac{\partial x_j}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i}\frac{\partial f}{\partial x_j} \ . \tag{67}$$

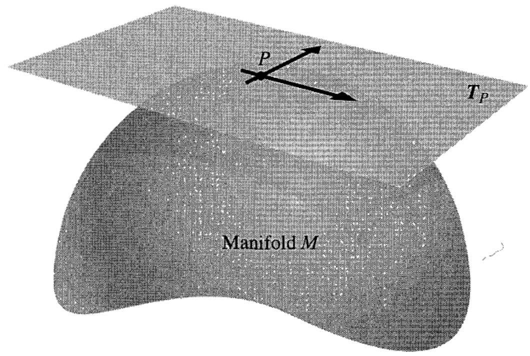

Figure 2: Tangent space - an abstract vector space which includes all vectors at the point p.

Note that the coordinate transformation information appears as partial derivatives of the *old* coordinates x_j with respect to the new coordinates x'_i. The next question we can raise is how a differential of one of the new coordinates, dx'_i, is related to differentials of the old coordinates, dx_i. Again, upon application of the chain rule we get,

$$dx'_i = dx_1 \frac{\partial x'_i}{\partial x_1} + dx_2 \frac{\partial x'_i}{\partial x_2} + ... + dx_n \frac{\partial x'_i}{\partial x_n} = \sum_{j=1}^{n} dx_j \frac{\partial x'_i}{\partial x_j} = dx_j \frac{\partial x'_i}{\partial x_j} \ . \tag{68}$$

This time we notice that the coordinate transformation information appears as partial derivatives of the *new* coordinates, x'_i, with respect to the *old* coordinates x_j. **Of note here is the fact that we were not very careful about index placements on the differentials to make things look similar to usual functions and calculus operations**.

Now, we are used to thinking of vectors as *arrows* stretching from one point in space to another, endowed with the ability to be freely relocated without worry. In general though, especially when dealing with curved spaces, vectors must be more carefully handled. What is done then is − to each point in space we associate the set of all possible vectors located at that point; this set is called the **tangent space at, say, point** P (denoted as T_p), which itself is considered to be embedded on a **manifold**. A suggestive image of this figure is given in Figure 2. T_p is spanned by the basis tangent vectors $\partial/\partial x_i$ at point P. Recall that covariant vectors transform like basis vectors while contravariant vectors transform like dual basis vectors. As such we can now redefine what we call covariant and contravariant vectors, with respect to T_p.

Definition 8. The components of a **covariant vector** (or a dual vector) transforms similar to a gradient (since the basis vectors are differentials) and obey the transformation law:

$$\boxed{V'_i = \frac{\partial x^j}{\partial x^{i'}} V_j} \ . \tag{69}$$

Definition 9. The components of a **contravariant vector** (or simply a vector) transform in the opposite manner to those of a covariant vector. Thus, a vector transforms like a coordinate differential with the following transformation law:

$$\boxed{V^{i'} = \frac{\partial x^{i'}}{\partial x^j} V^j} \ . \tag{70}$$

Note: a *contravariant* index in the denominator is equivalent to a *covariant* index in the numerator, and vice versa. So, in the construct $\partial x^j/\partial x^{i'}$, j is contravariant while i is taken to be covariant. Tensors of higher rank can then be analogously defined as objects which transform according to

$$\boxed{T_{i'_1,\dots,i'_p}{}^{k'_1,\dots,k'_q} = \frac{\partial x^{j_1}}{\partial x^{i'_1}}\cdots\frac{\partial x^{j_p}}{\partial x^{i'_p}}\frac{\partial x^{k'_1}}{\partial x^{l_1}}\cdots\frac{\partial x^{k'_q}}{\partial x^{l_q}}T_{j_1,\dots,j_p}{}^{l_1,\dots,l_q}}\,, \tag{71}$$

whose rank is $n = p + q$.

4.4.1 Invariance of Equations in Tensor Form

Consider an equation of the form

$$K_i^j A_j = B_i \tag{72}$$

which under a coordinate change can be written as,

$$K_{i'}^{j'} A_{j'} = B'_i\;. \tag{73}$$

We want to investigate if, in fact, the above equation is a valid "tensored" equation. To that end, let's investigate the transformation of $B_i \to B'_i$ by using (69):

$$B'_i = \frac{\partial x^m}{\partial x^{i'}}B_m = \frac{\partial x^m}{\partial x^{i'}}K_m^j A_j = \frac{\partial x^m}{\partial x^{i'}}\;K_m^j\;\frac{\partial x^{n'}}{\partial x^j}\;A'_n \tag{74}$$

where the last equality was obtained by applying an **inverse transformation** (primed to un-primed coordinates),

$$\boxed{A_i = \frac{\partial x^{j'}}{\partial x^i}A'_j}\;. \tag{75}$$

Getting back to (74), for aesthetic purposes we interchange the *dummy* indices j and n to get,

$$B'_i = \frac{\partial x^m}{\partial x^{i'}}\;\frac{\partial x^{j'}}{\partial x^n}\;K_m^n\;A'_j \tag{76}$$

which can be substituted in (73) to give,

$$\left[K_{i'}^{j'} - \frac{\partial x^m}{\partial x^{i'}}\;\frac{\partial x^{j'}}{\partial x^n}\;K_m^n\right]A_{j'} = 0 \tag{77}$$

and since A'_j is arbitrary, then the coefficient must vanish, i.e.

$$\begin{aligned} K_{i'}^{j'} - \frac{\partial x^m}{\partial x^{i'}}\;\frac{\partial x^{j'}}{\partial x^n}\;K_m^n &= 0 \\ \Rightarrow K_{i'}^{j'} &= \frac{\partial x^m}{\partial x^{i'}}\;\frac{\partial x^{j'}}{\partial x^n}\;K_m^n \end{aligned} \tag{78}$$

which is consistent with the transformation property of a $(1,1)$ tensor and we can safely say that Eq.(72) holds in all transformed coordinate systems. Conversely, this leads us to the **quotient rule** which states: *if the equation of interest holds in all transformed coordinate systems, then K is a tensor of the indicated rank and covariant/contravariant character.*

4.5 Tensor Product of Vector Spaces

Given two vector spaces V and W (over the same set of scalars C), we can construct a product vector space, denoted as $V \otimes W$, whose elements are, in some sense, 'products' of vectors $v \in V$ and $w \in W$ which can be denoted as $v \otimes w$. This product should be bilinear such that

$$\begin{aligned}&(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \ , \\ &v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2 \ , \\ &c(v \otimes w) = (cv) \otimes w = v \otimes (cw) \ , \quad c \in C \ .\end{aligned} \tag{79}$$

Any two arbitrary vectors can then be expanded in terms of bases $\{e_i\}_{i=1,..,n}$ and $\{f_i\}_{i=1,..,m}$ for V and W respectively as,

$$v \otimes w = (v^i e_i) \otimes (w^j f_j) = v^i w^j e_i \otimes f_j \ , \tag{80}$$

so that, $\{e_i \otimes f_j\}, \ i = 1, ..., n, \ j = 1, ..., m$ should be the basis for $V \otimes W$ with dimension nm.

Precisely then, we define the **tensor product** $V \otimes W$ to be the set of all $C-$valued bilinear functions on $V^* \times W^*$, which form a vector space themselves. Furthermore, given two vectors $v \in V$ and $w \in W$, we define their ***tensor product*** $v \otimes w$ to be the element of $V \otimes W$ defined as,

$$(v \otimes w)(h, g) \equiv v(h)w(g) \quad \forall h \in V^*, g \in W^* \ . \tag{81}$$

The tensor product has a couple of important properties besides bilinearity:

1. $(V \otimes W)^* = V^* \otimes W^*$ (**commutes**) with taking duals.
2. $(V_1 \otimes V_2) \otimes V_3 = V_1 \otimes (V_2 \otimes V_3)$ (**associative**) for tensor product of vector spaces.

4.5.1 Pseudotensors

Thus far, whenever we mentioned a transformation, we implicitly assumed a **passive transformation**, i.e. a change in basis with the vector or tensor orientations fixed. An **active transformation**, on the other hand, is a transformation which may leave the basis unchanged but change the vector or tensor itself. Another important transformation is the effect of **reflections** of the coordinate system, referred to as **inversions**. We know that a coordinate rotation on a fixed vector can be described by a transformation of its components according to,

$$A' = SA$$

where S is an orthogonal matrix with determinant $+1$. If reflections are included, then det $S = -1$. Quantities that transform like vectors under rotation but under reflections, produce the wrong sign, are known as **pseudovectors**. Examples of pseudovectors are angular velocity, angular momentum, magnetic field, etc. Pseudovectors thus obey the transformation rule,

$$A' = \det(S)SA \ . \tag{82}$$

The generalization of this to higher rank tensors yield objects called **pseudotensors**. If T is a tensor and P is a pseudotensor, then the direct product satisfies the rule

$$T \otimes T = P \otimes P = T \text{ and } T \otimes P = P \otimes T = P \ . \tag{83}$$

Example 17. Levi-Civita symbol revisited.

Recall that the three-indexed Levi-Civita symbol has the values,

$$\begin{aligned} &\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = +1 \ , \\ &\epsilon_{132} = \epsilon_{213} = \epsilon_{321} = -1 \ , \\ &\text{all other } \epsilon_{ijk} = 0 \ . \end{aligned} \tag{84}$$

Suppose we are given another rank 3 pseudotensor η_{ijk}, which in some Cartesian coordinate system is equal to ϵ_{ijk}. Then, by (82) we have

$$\eta'_{ijk} = \det(A) a_i^p \ a_j^q \ a_k^r \ \epsilon_{pqr} \tag{85}$$

where A is the matrix of coefficients in an orthogonal transformation of $\mathbb{R}^3$ and Einstein summation is implied. The presence of the $\det(A)$ factor means that η'_{ijk} **only** transforms as a tensor under rotations when $\det(A) = +1$ but if one includes reflection, i.e. $\det(A) = -1$, then η'_{ijk} **does not** transform as a tensor. The presence of the -1 gives η'_{ijk}, and hence ϵ'_{ijk}, the characteristics of a ***pseudotensor***. Eq.(85) also shows that ϵ_{ijk} is an ***isotropic*** pseudotensor since it has the same components in all rotated coordinate systems and -1 times those components upon inclusion of coordinate inversions.

4.6 Metric Tensor

Thus far we worked mostly with Cartesian coordinate systems. Now we move on to treating more general **metric spaces**, also known as **Riemannian spaces**. Essentially, we can think of the **metric tensor** as a mathematical object that describes the geometry of a coordinate system or manifold. The components of the metric describes lengths and angles between basis vectors. It is important to note that the metric tensor itself is defined as a *bilinear form at a point on a tangent space.*

Letting q^i denote a set of generalized coordinates, we can define **covariant basis vectors** ε_i as,

$$\varepsilon_i = \frac{\partial x}{\partial q^i} e_x + \frac{\partial y}{\partial q^i} e_y + \frac{\partial z}{\partial q^i} e_z \ , \tag{86}$$

that describes the displacement per unit change in q^i, keeping the other q^j fixed. Note that, the direction and the magnitude of the ε_i may be functions of position. Any arbitrary vector v can then be written as the linear combinations of ε_i,

$$v = v^1 \varepsilon_1 + v^2 \varepsilon_2 + v^3 \varepsilon_3 \ . \tag{87}$$

Given this basis, we can compute the displacement associated with changes in the q^i and since the basis vectors depend on position, our computation needs to be infinitesimal displacements ds. Then,

$$\begin{aligned} &(ds)^2 = (\varepsilon_i \ dq^i) \cdot (\varepsilon_j \ dq^j) \\ \Rightarrow &(ds)^2 = \underbrace{\varepsilon_i \cdot \varepsilon_j}_{g_{ij}} \ dq^i \ dq^j \end{aligned} \tag{88}$$

is an invariant under rotations (and reflections) and thus is a scalar. Above, we have defined the ***covariant*** **metric tensor**

$$\boxed{g_{ij} = \varepsilon_i \cdot \varepsilon_j = \frac{\partial x^k}{\partial q^i} \frac{\partial x_k}{\partial q^j}} \ . \tag{89}$$

Note that the basis vectors ε_i can be defined by their Cartesian components, but generally they are **neither** unit vectors **nor** mutually orthogonal and furthermore are not required to be diagonal. Conveniently we can then define an **inverse** or **contravariant** metric tensor g^{ij} that satisfies,

$$g^{ik}g_{kj} = g_{jk}g^{ki} = \delta^i_j \ . \tag{90}$$

The metric tensor and its inverse are then employed to *raise* and *lower* indices in relativity such as

$$g_{ij}F^j = F_i \quad \text{and} \quad g^{ij}F_j = F^i \ . \tag{91}$$

In this setting Eq.(87) gives,

$$v = v^i\varepsilon_i = v^i\delta^k_i\varepsilon_k = \underbrace{v^i(g_{ij}}_{v_j}\ \underbrace{g^{jk})\varepsilon_k}_{\varepsilon^j} = v_j\varepsilon^j \ , \tag{92}$$

showing that the vector can be represented either by contravariant or covariant components, with the two sets of components related by the transformation equation (91). Thus, using similar lines of arguments, we can define the **contravariant basis vectors**

$$\varepsilon^i = \frac{\partial q^i}{\partial x}e_x + \frac{\partial q^i}{\partial y}e_y + \frac{\partial q^i}{\partial z}e_z \ . \tag{93}$$

4.7 Covariant Derivatives

Since the basis vectors ε_i are in general not constant, the derivatives of a vector will not be a tensor whose components transform as the derivatives of the vector components. Recall the transformation rule for a contravariant vector V^i,

$$V^{i'} = \frac{\partial x^{i'}}{\partial q^k}V^k \tag{94}$$

with respect to the generalized coordinate q^k. Upon differentiation with respect to q^j (94) gives,

$$\frac{\partial V^{i'}}{\partial q^j} = \frac{\partial x^{i'}}{\partial q^k}\ \frac{\partial V^k}{\partial q^j} + \frac{\partial^2 x^{i'}}{\partial q^k \partial q^j}\ V^k \ . \tag{95}$$

The first term on the right is what one would expect from a tensor transformation but the term in red certainly is not! This proves that, in general, tensor derivatives do not transform as tensors.

To get rid of the offending term in (95), one must add a *correction term* to the partial derivative $\partial V^k/\partial q^j$, in a way that the new term includes a factor of V^k. These considerations lead us to define a new **derivative operator** $V^k_{;j}$, called the **covariant derivative**,

$$\boxed{V^k{}_{;j} = \frac{\partial V^k}{\partial q^j} + \Gamma^k_{ij}V^j} \ , \tag{96}$$

identified with a very awkward notation, where Γ^i_{jk} are known as the **affine connections**, or the **Christoffel symbols**, or sometimes also as **connection coefficients**. We define the affine connections as,

$$\boxed{\Gamma^i_{jk} = \frac{\partial q^i}{\partial x^\alpha}\ \frac{\partial^2 x^\alpha}{\partial q^j \partial q^k}} \ , \tag{97}$$

It is important to note that even though it may appear as a mixed tensor, with all the various indices, it is actually a very important **non-tensor** that shows up everywhere in physical laws. How do I know it is a non-tensor? Well, the Γ^i_{jk}'s transform according to the law,

$$(\Gamma^i_{jk})' = \frac{\partial q^{i'}}{\partial x^\rho} \frac{\partial x^\tau}{\partial q^{j'}} \frac{\partial x^\sigma}{\partial q^{k'}} \Gamma^\rho_{\tau\sigma} + \frac{\partial q^{i'}}{\partial q^\rho} \frac{\partial^2 x^\rho}{\partial q^{j'} q^{k'}} \ , \tag{98}$$

where again the term in red is what makes it a non-tensor.

It is good to note that one can also define the covariant derivative of a *covariant vector* V_μ as,

$$\boxed{V_{\mu;\nu} = \frac{\partial V_\mu}{\partial x^\nu} - \Gamma^{\mu\nu}_\lambda V_\lambda} \ . \tag{99}$$

It is then up to the reader to verify the claim that: **both $V_{\mu;\nu}$ and $V^k_{;j}$ are tensors**.

4.7.1 An Attempt at Making Things Physically Intuitive

For the last part of this chapter, let us dip our tows into the realm of relativity such that we are now in Minkowski spacetime. At every point in spacetime, with arbitrary gravitational field, it is possible to choose a "locally inertial coordinate system" such that within a sufficiently small region in the neighbourhood of the point of interest, the laws of nature takes the same form as un-accelerated Cartesian coordinate frame in the absence of gravitation. This means that the equation of motion of a particle freely falling becomes,

$$\frac{d^2\xi^\alpha}{d\tau^2} = 0 \ , \tag{100}$$

with ξ^α denoting the freely falling coordinates ($\alpha = 0, .., 3$) and

$$d\tau^2 = \eta_{\alpha\beta} d\xi^\alpha d\xi^\beta \tag{101}$$

being the proper time and the Minkowski metric is defined as, $\eta_{\alpha\beta} = (1, -1, -1, -1)$. If one now moves to a different inertial frame of reference, say the laboratory frame with coordinates x^μ, the equation of motion for the same particle becomes,

$$\frac{d^2 x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0 \ , \tag{102}$$

where you should recognize the Γ as the affine connection defined before. Additionally, one can express proper time (100) in an arbitrary coordinate system, in terms of $g_{\mu\nu}$ (arbitrary coordinate metric tensor) as,

$$d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu \ \text{ with } \ g_{\mu\nu} \equiv \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta} \ . \tag{103}$$

Of course, (100) and (102) must be the same in all inertial frames - fundamental postulate of relativity! As such, it must be the case that

- $g_{\mu\nu}$ determines the proper time interval between two events in spacetime, in the case of the events being infinitesimally separated, while
- $\Gamma^\lambda_{\mu\nu}$ determines the gravitational force.

In that case, in all locally inertial reference frames, $\Gamma^{\lambda}_{\mu\nu} = 0$ leading to identical equations of motion and one recovers physics from breaking. Therefore it should not be surprising that $g_{\mu\nu}$ and $\Gamma^{\lambda}_{\mu\nu}$ are related as,

$$\boxed{\Gamma^{\sigma}_{\lambda\mu} = \frac{1}{2} g^{\nu\sigma} \left(\frac{\partial g_{\mu\nu}}{\partial x^{\lambda}} + \frac{\partial g_{\lambda\nu}}{\partial x^{\mu}} - \frac{\partial g_{\mu\lambda}}{\partial x^{\nu}} \right)} \ , \tag{104}$$

which hints at the fact that $g_{\mu\nu}$ is in fact a gravitational potential!

How does defining covariant derivatives in terms of Christoffel symbols help physics?

It allows to reaffirm the famed phrase: *equations of motion must be the same in all coordinate systems.* What this essentially means is that coordinates do not really mean anything by themselves when describing the physics. As such, it must always be true that an equation describing a law of nature must be the same in all coordinates. One achieves this **general covariance** by writing down the known equations in special relativity (no gravity), for instance the Maxwell equations:

$$\partial_{\alpha} F^{\alpha\beta} = j^{\beta} \ , \tag{105}$$

where $\partial_{\alpha} \equiv \frac{\partial}{\partial x^{\alpha}}$ and $F^{\alpha\beta}$ is the EM tensor and j^{ν} is the current density. But when relocated in a general coordinate space with curvature, this equation does not work because derivatives, as we saw, do not transform as tensors! Then, the covariant derivative formulation saves us; in a general coordinate, change $\eta_{\alpha\beta} \to g_{\alpha\beta}$ and replace $\partial_{\alpha} \to D_{\alpha}$ where D_{α} is the covariant derivate in a much friendlier notation. Then (105) becomes,

$$\boxed{D_{\alpha} F^{\alpha\beta} = j^{\beta}} \ , \tag{106}$$

which is in the same form as (105), and Einstein will rest easy! We say then that (106) is **manifestly covariant**.