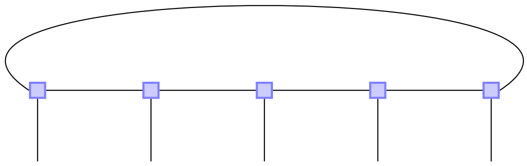
Tensor Calculus

Notes

Based on: many different sources



Sagnik Seth [1]

Dept of Physical Sciences
IISER Kolkata

[1] sag123nik456@gmail.com, ss22ms026@iiserkol.ac.in

Contents

1. Introduction

Hehe......😅 I am writing this as a way to understand tensors better and also to do something "productive". I don't know how much I will be able to complete but I intend to touch upon the basic aspects, albeit, in an extremely non-rigorous way....not going into heavy math perhaps (which I think is very bad 'coz math is great 😉). I will try to do some proofs (which I feel like doing) and skip others (since I don't care). I will definitely take $c = 1$ unless its necessary not to. I will use the convention $(+, -, -, -)$ for the metric tensor. I will also use some gen alpha slangs (which will indicate how chill I am!) and overall try to write in a fun way. A few references which I will use are listed below:

- *Tensor Calculus for Physics: A Concise Guide* by Dwight E Neuenschwander
- *The Poor Mans Introduction to Tensors* by Justin C. Feng
- *Geometry, Topology and Physics* by Mikio Nakahara
- *Spacetime and Geometry: An Introduction to General Relativity* by Sean M. Carroll
- *Modern Differential Geometry for Physicists* by Christopher Isham
- *Differential Geometry: A Theoretical Physics Approach* by Gerardo F. Torres del Castillo
- *An Introduction to Manifolds* by Loring W. Tu
- F.P. Schuller's Lecture notes of the course *Geometrical Anatomy of Theoretical Physics*
- Tensor Calculus YouTube video series by eigenchris, Andrew Dotson and ADG.

I will continue to add the resources as I progress. I am writing this as my own personal notes for the subject and if it helps anyone else, I will be super happy. If any mistake is there, let me know! ✌️

2. Indices: the ultimate rizzler!

Indices make our lives easier when writing abstract quantitites having multiple components, like vectors. If we have a three-dimensional vector, we can write it as v^i where i can take the values 1, 2, or 3.

Why are the indices written as superscript? Well, these are contravariant indices which will be discussed later. For now, let's just say that 'upstairs' indices are the 'normal thing'. Index placement is important and these are not powers...just the way we denote the components.

Consider the (in)famous equation:

$$\mathbf{F} = m\mathbf{a}$$

This can be written as $F^i = ma^i$, for each component i. Just remember that we should have the same kind of indices on both side of the equation finally. That is, if we have 'upstairs' index on the right, same should be on the left.

2.1. Einstein Convention

The OG rule...whenever you see two same indices, sum them. That's it! Let's make our hands dirty and look at some examples:

2.2. Examples

Matrix Multiplication:

Let us have the eigenvalue equation $M\mathbf{v} = \lambda\mathbf{v}$. We can write this as:

$$\sum_j M_{ij} v^j = \lambda v^i$$

Note two things here:

- The index j is summed over, so it does not come in the final expression (dummy index!).
- The index i occurs as a superscript on the right, so in the left also, the final expression should have the index i as a superscript.

Thus, using Einstein convention and correct index placement, the above equation can be written as:

$$M^i_j v^j = \lambda v^i$$

This can be visualised by treating each quantity as a 'box' with the indices as some 'hands' protruding out. When we sum, we just join these 'hands'. After taking the sum, the number of free hands decreases (index contraction)I A matrix has two hands and a vector has one hand. When we multiply a matrix with a vector, we obtain a vector, which should have one hand. This is represented in the diagram below:

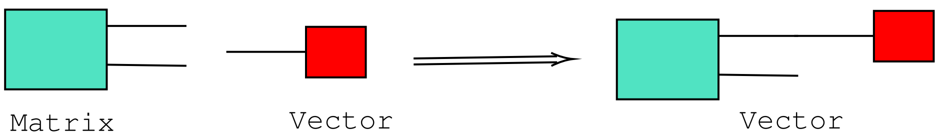


Figure 1: Matrix-vector multiplication. The final product has one free hand and is thus a vector.

The Scalar Product:

The dot product or scalar product of two vectors is a scalar (no hand). Then, there should be no free index in the expression. Thus, in the index notation:

$$\mathbf{v} \cdot \mathbf{v} = v^i v_i = v_i v^i$$

The last two expressions are same. The upstairs or downstairs indices do not matter, as these are summed over. Note that in this definition, we have used a *dual vector*, having a lower index. We can also define the dot product using the *regular* vector with an upper index but then a *metric* comes in.

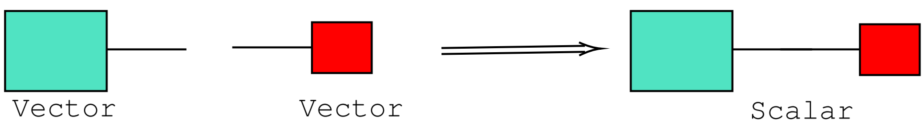


Figure 2: Dot product of two vectors. The final product is a scalar and has no free hands.

So basically a *scalar* is something that does not change under coordinate transformation, that is, if we go from a coordinate (x, y, z) to (x', y', z'), a scale $\lambda = \lambda'$.

Euclidean Vectors:

Any vector can be written as in terms of basis vectors: $\mathbf{A} = A^i \mathbf{e}_i$ where A^i are the components of the vector in the chosen basis. Now, we define

$$\mathbf{e}_n \cdot \mathbf{e}_m = g_{nm}$$

These g_{nm} are coefficients of metric tensor which will be discussed later. If these basis vectors are orthonormal, then the coefficients become the kronecker delta. Then we have the scalar product:

$$\begin{aligned} \mathbf{A} \cdot \mathbf{B} &= (A^m \mathbf{e}_m) \cdot B^n \mathbf{e}_n \\ &= (A^m B^n \mathbf{e}_m) \cdot \mathbf{e}_n \\ &= A^m B^n g_{nm} \end{aligned}$$

Note that we have used the *regular vector* with an upper index here, with the metric g. We can define a cross product of two vectors as:

$$(\mathbf{A} \times \mathbf{B})^i = \epsilon^{ijk} A^j B^k$$

where ϵ^{ijk} is the Levi-Civita symbol (cyclic permutation of i, j, k gives 1 and non-cyclic permutation gives -1 while repeated index in the symbol gives 0).

2.3. Some vector BS

Vectors as Directional Derivatives:

A vector can be thought of as a directional derivative. We define the directional derivative operator as:

$$\mathbf{v} \cdot \nabla = v^i \partial_i \equiv v^i \frac{\partial}{\partial x^i}$$

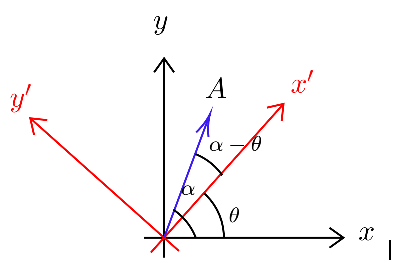
This is very similar to the vector expansion in terms of the basis vectors. Thus, the partial derivatives somewhat act like a basis. The basis of partial derivatives is indeed called a *coordinate basis*. Now let us calculate:

$$\mathbf{v} \cdot \nabla x^j = v^i \partial_i x^j = v^i \frac{\partial x^j}{\partial x^i} = v^i \delta_i^j = v^j$$

We have used the fact that coordinate components are independent of each other, that is, partial derivative of one component with respect to another gives a kronecker delta. Note that we have used the proper index placement here: $\partial_j \equiv \frac{\partial}{\partial x^j}$ has a lower index (by definition) while x^j has an upper index, thus kronecker delta has an upper as well as lower index.

Vector Transformation:

Let us suppose we have a coordinate system (x, y, z) and we rotate it about the z-axis by an angle θ. The new coordinates are given by:



$$x' = x \cos\theta - y \sin\theta$$

$$y' = x\sin\theta + y\cos\theta$$

$$z' = z$$

These relations can be readily found out using the following: Suppose we have a point A making an angle α with the original system. Then we have $x = r\cos(\alpha), y = r\sin(\alpha)$. After rotation, the new coordinates are given by:

$$x' = r\cos(\alpha - \theta) = r\cos\alpha\cos\theta + r\sin\alpha\sin\theta$$

$$y' = r\sin(\alpha - \theta) = r\sin\alpha\cos\theta - r\cos\alpha\sin\theta$$

which gives the previous result. Note that the z coordinate does not change as the rotation is about the z-axis. Now we consider the infinitesimal displacement in the new coordinate frame:

$$\begin{aligned}
(ds')^2 &= (dx')^2 + (dy')^2 + (dz')^2 \\
&= (dx\cos\theta - dy\sin\theta)^2 + (dx\sin\theta + dy\cos\theta)^2 + dz^2 \\
&= dx^2(\cos^2\theta + \sin^2\theta) + dy^2(\cos^2\theta + \sin^2\theta) - \cancel{2dxdy\sin\theta\cos\theta} + \cancel{2dxdy\sin\theta\cos\theta} + dz^2 \\
&= dx^2 + dy^2 + dz^2 \\
&= ds^2
\end{aligned}$$

Thus we see that the infinitesimal displacement is invariant under coordinate transformation and is thus a scalar.

Now, note one thing: If we consider the new coordinates as a function of the old coordinate that is $x' \equiv x'(x, y, z)$, we can write:

$$(dx')^i = \frac{\partial x'^i}{\partial x^j} dx^j$$

Using this analogy, we can define the transformation of a vector as:

$$(v')^i = \frac{\partial x'^i}{\partial x^j} v^j$$

Thus a vector is a quantity which transform like this. The terms $\frac{\partial x'^i}{\partial x^j}$ are the components of the transformation matrix Λ^i_j. As we defined the transformation from x to x', we can also define the reverse transformation from x' to x as:

$$\begin{aligned}
x^i &= \frac{\partial x^i}{\partial x'^j} x'^j \\
&= \left(\frac{\partial x^i}{\partial x'^j}\frac{\partial x'^j}{\partial x^k}\right) x^k
\end{aligned}$$

Now in the above sum, j and k indices are summed over. We must obtain x^i from the right hand side also. Thus by observation, we can see that the term $\frac{\partial x^i}{\partial x'^j}\frac{\partial x'^j}{\partial x^k}$ must be equal to the kronecker delta δ^i_k.

Now, let us consider we have a position vector written in the Cartesian coordinate system with basis $\{\mathbf{e}_i\}$, that is,

$$\mathbf{r} = x^i \mathbf{e}_i$$

Now, consider another coordinate system and write the cartesian coordinates as a function of these new coordinates, that is, $x^i \equiv x^i(x'^j)$. Since $\mathbf{r}$ depends on the coordinate $\{x'^i\}$, we can expand the differential displacement as:

$$\begin{aligned}
d\mathbf{r} &= \frac{\partial \mathbf{r}}{\partial x'^i} dx'^i \\
&= \frac{\partial (x^j \mathbf{e}_j)}{\partial x'^i} dx'^i \\
&= \left(\frac{\partial x^j}{\partial x'^i} \mathbf{e}_j + \frac{\partial \mathbf{e}_j}{\partial x'^i} x^j \right) dx'^i \\
&= \left(\frac{\partial x^j}{\partial x'^i} \mathbf{e}_j \right) dx'^i \\
&= \mathbf{e}'_j \; dx'^j
\end{aligned}$$

The second term in the third step is zero, as the Cartesian basis vectors are independent of the new coordinates. The final line is obtained from the previous step as this defines the transformation of the basis vectors. Comparing this with the first line of the previous expansion we have:

$$\boxed{\mathbf{e}'_i = \frac{\partial \mathbf{r}}{\partial x'^i}}$$

Thus any basis vector (in a coordinate system) can be obtained from the partial derivative of the position vector with respect to the coordinates (in that system).

3. Contravariant and Covariant: why the skibidi!

Let us suppose we have a vector space $\mathbf{V}$ and two bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$. We can the write the transformation of the basis into one another as:

$$\begin{aligned}
\mathbf{e}_i &= \Lambda^j_i \mathbf{e}'_j \\
\mathbf{e}'_i &= (\Lambda^{-1})^j_i \mathbf{e}_j
\end{aligned}$$

Now if we have a vector, we can write it in terms of the basis vectors as:

$$\mathbf{x} = (x')^j \mathbf{e}'_j = x^i \mathbf{e}_i = (x^i \Lambda^j_i) \mathbf{e}'_j$$

From this we get: $(x')^j = \Lambda^j_i x^i$. Well note that, in the transformation equation of the basis, if we have the primed basis in the left, then we had the inverse transformation matrix Λ^{-1} in the right, but here it is different (primed component in the left and Λ in the right). Thus, the basis vectors and the components transform in the "opposite" or "**contra**ry" way. Thus, these components are called the **contravariant** components of the vector.

Let us now consider the dual space $\mathbf{V}^*$[1] of the vector space $\mathbf{V}$. From the linearity property, we have:

$$f(x^i \mathbf{e}_i) = x^i f(\mathbf{e}_i) \equiv x^i f_i$$

Now, we use the basis transformation equation:

$$f_i = f(\mathbf{e}_i) = f(\Lambda^j_i \mathbf{e}'_j) = \Lambda^j_i f(\mathbf{e}'_j) = \Lambda^j_i f'_j$$

[1] The dual space is the set of all linear functionals, that is, linear maps $f : \mathbf{V} \to \mathbb{R}$. One example is say the *bra* vector which is dual to the *ket*. So basically a bra takes a ket and returns a real (complex) number: $\langle \psi | \psi \rangle$ (braket)

These f_i are the components of the "dual vector". Note that if we have unprimed things on the left, then we have the transformation matrix Λ on the right, which is similar to the transformation of the basis. Thus, we see that this transformation follows the same transformation as the basis vectors. Thus, these components are called the **covariant** components of the vector.

So, the components are named according to how the basis vectors transform. If they transform together, the are called **co**variant (and denoted by downstairs index) and if they transform in the opposite way, they are called **contra**variant (and denoted by upstairs index). The contravariant and the covariant components together form an 'invariance' like the scalar product (which do not change under coordinate transformation):

$$\mathbf{v} \cdot \mathbf{v} = v_i v^i$$

We had earlier seen another definition of the inner product, using both contravariant components and the matric tensor, which was $\mathbf{v} \cdot \mathbf{v} = v^i v^j g_{ij}$. Comparing both these definitions, we can see a relation:

$$v_i = g_{ij} v^j$$

Thus when changing from contravariant to covariant, we just need to invoke the holy metric tensor (to be dicussed later further).

Note: In the Cartesian coordinates, the metric tensor is the Kronecker delta, that is, $g_{nm} = \delta_{nm}$ and hence the components of the vectors and dual vectors are the same, that is, $x^i = x_i$.

Transformation of dual vectors: We had seen the transformation of the vectors. Now, generally we want to keep the inner product same in any basis that we choose. Then, we have:

$$\begin{aligned} A'^i A'_i &= A^j A_j \\ \implies \frac{\partial x'^i}{\partial x^j} \cancel{A^j} A'_i &= \cancel{A^j} A_j \\ \implies \frac{\partial x'^i}{\partial x^j} A'_i &= A_j \\ \implies A'_i &= \frac{\partial x^j}{\partial x'^i} A_j \end{aligned}$$

4. Why are tensors so sigma!

Nakahara defines tensor as:

> *A tensor T of type (p, q) is a multi-linear map that maps q vectors and p dual vectors to $\mathbb{R}$, that is:*
>
> $$T : \left(\bigotimes^p \mathbf{V}^* \right) \bigotimes \left(\bigotimes^q \mathbf{V} \right) \to \mathbb{R}$$

Dayummm!! 😑 Let us break this down. Consider a scalar which has no vector and no dual vector. Thus, it is a $(0, 0)$ type tensor. Now, let us consider a vector $\mathbf{v}$. This is a $(1, 0)$ tensor, that is, it maps a dual vector to a scalar. If we have a dual vector $\mathbf{f}$, then it is of type $(0, 1)$ and maps a vector to a scalar. This does not clear anything. Let us instead consider few examples:

Moment of Intertia Tensor:

Perhaps the first example of a tensor we had encountered during our classical mechanics course (which we had been told to understand just as a 'matrix').

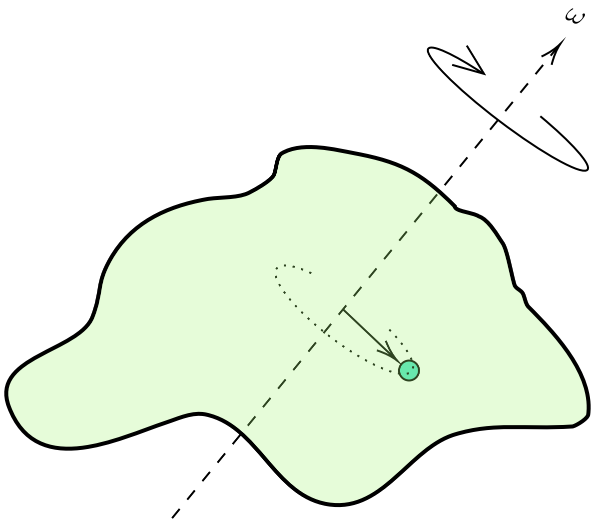


Figure 3: A rigid body rotating about an axis

Consider a rigid body made of tiny masses dm. Consider one such mass sitauted as a distance s from the fixed axis of rotation. It goes around a circle with speed $v = \omega s$. The angular momentum can be calculated as:

$$\mathbf{L} = \int (\mathbf{r} \times \mathbf{v}) dm = \int (\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r})) dm = \int (\mathbf{r} \cdot \mathbf{r})\boldsymbol{\omega} - \cancelto{0}{(\mathbf{r} \cdot \boldsymbol{\omega})} \mathbf{r} dm = \boldsymbol{\omega} \underbrace{\int s^2 dm}_{I}$$

The integral is called the moment of inertia. In a more general case, where there is no fixed axis of rotation, we write:

$$\begin{aligned} \mathbf{L} &= \int (\mathbf{r} \times \mathbf{v}) dm \\ &= \int (\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r})) dm \\ &= \int (\mathbf{r} \cdot \mathbf{r})\boldsymbol{\omega} - (\mathbf{r} \cdot \boldsymbol{\omega})\mathbf{r} dm \end{aligned}$$

We now write it in index notation, noting that $\omega^i = \delta^{ij}\omega_j$:

$$\begin{aligned} L^i &= \int (\mathbf{r} \cdot \mathbf{r})\delta^{ij}\omega_j - x^i(x^j\omega_j) dm \\ &= \omega_j \left(\int (\mathbf{r} \cdot \mathbf{r})\delta^{ij} - x^i x^j \ dm \right) \end{aligned}$$

The integral in the bracket is defined to be the inertia tensor:

$$I^{ij} = \int (\mathbf{r} \cdot \mathbf{r})\delta^{ij} - x^i x^j \ dm$$

Note that i and j goes from 1 to 3 and thus it has 9 components but since the expression is symmetric, we only have 6 independent components. This states that the angular momentum and the angular velocity are not necessarily parallel in some coordinate system where I have non-zero off-diagonal entries.

Electromagnetic Tensor:

The electromagnetic tensor is very useful in combining the electric field and magnetic field and finding their transformations. It is defined as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

where A_μ is the 4-potential. The indices μ and ν can take values from 0 to 3. The tensor has 16 components but only 6 of them are independent. The tensor is antisymmetric, that is, $F_{\mu\nu} = -F_{\nu\mu}$ and thus the diagonal entries are zero. We will discuss this later but Maxwell's equations can be written in a very compact form using the components of the electromagnetic tensor.

Electric-Susceptibility Tensor:

We had studied about polarisation in dielectrics in our classical electrodynamics course where we had often taken (for simplicity):

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}$$

Here we had taken the electric field to be parallel to the polarisation vector but in general, these are related by the susceptibility tensor as:

$$P^i = \epsilon_0 \chi^{ij} E^j$$

4.1. Gradient

The components of the gradient (basically partial derivative) are covariant. We denote it with a lower index explicitly to show that it is a covariant vector:

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$$

Then accordingly we will have:

$$\partial^\mu \equiv \frac{\partial}{\partial x_\mu} \qquad \partial^\mu = g^{\mu\nu}\partial_\nu$$

4.2. Tensor Transformations

As previously seen for vector transformation, the transformation of a tensor from one system to another is very similar. For simplicity, we show for a second rank tensor:

$$T'^{ij} = \frac{\partial x'^i}{\partial x^k}\frac{\partial x'^j}{\partial x^n} T^{kn}$$

This is simple: in the left everything is prime and contravariant, so in the right the numerator must have primes and the denominator must have unprimed indices and contraction should be done so as to match the contra and co indices on both sides in the final expansion. Well, this transformation law is such a nice thing that many people say that:

$$\boxed{\text{Tensors are defined in the way they transform}}$$

Let us see some more examples of tensors and their transformation from lower to upper indices (it's easy, just use the metric tensor appropriately):

$$T^{\mu\nu} = g^{\rho\mu} g^{\sigma\nu} T_{\rho\sigma}$$
$$T^{\mu}{}_{\nu} = g^{\mu}{}_{\rho} g^{\sigma}{}_{\nu} T^{\rho}{}_{\sigma}$$

Lastly we show an example of a mixed tensor transformation from one system to another 🔥:

$$T'^{\mu}{}_{\nu\rho} = \frac{\partial x^\beta}{\partial x'^\rho}\frac{\partial x^\alpha}{\partial x'^\nu}\frac{\partial x'^\mu}{\partial x^\sigma} T^{\sigma}{}_{\alpha\beta}$$

So basically, while changing the covariant indices, the prime is in the denominator and for contravariant, it is above. And then the positioning of the indices is trivial (I guess...)

4.3. Matrices vs. Tensor? same same but different....

Not all matrices are second-rank tensors 😑. Yes, the components of a second-rank tensor can be arranged in a matrix form but there are many matrices which do not transform according to the above equation. Take for example the following matrix, which we assume as a tensor:

$$[T^{lm}] \equiv \begin{pmatrix} (x^2)^2 & x^1x^2 \\ x^1x^2 & (x^2)^2 \end{pmatrix}$$

Note that the 2 outside the bracket is the power and inside the bracket is the index of the component. Then after rotation, we have the following relation:

$$\begin{aligned} x'^1 &= x^1 \cos\theta + x^2 \sin\theta \\ x'^2 &= -x^1 \sin\theta + x^2 \cos\theta \end{aligned}$$

Let us find $T^{'11}$ which according to the transformation rule, should be:

$$\begin{aligned} T^{'11} &= \frac{\partial x'^1}{\partial x^k}\frac{\partial x'^1}{\partial x^n}T^{kn} \\ &= \frac{\partial x'^1}{\partial x^1}\frac{\partial x'^1}{\partial x^1}T^{11} + \frac{\partial x'^1}{\partial x^1}\frac{\partial x'^1}{\partial x^2}T^{12} + \frac{\partial x'^1}{\partial x^2}\frac{\partial x'^1}{\partial x^1}T^{21} + \frac{\partial x'^1}{\partial x^2}\frac{\partial x'^1}{\partial x^2}T^{22} \\ &= \cos\theta\cos\theta T^{11} + \cos\theta\sin\theta T^{12} + \sin\theta\cos\theta T^{21} + \sin\theta\sin\theta T^{22} \\ &= \cos^2\theta(x^2)^2 + 2\sin\theta\cos\theta(x^1)(x^2) + \sin^2\theta(x^2)^2 \\ &= (x^2\cos\theta + x^1\sin\theta)^2 \end{aligned}$$

Well note that if T was indeed a tensor, then $T^{'11}$ should be equal to $(x'^2)^2$ but it is not. Thus, T is not a tensor and our assumption was wrong. **Thus all matrices are not tensors!**

5. Metric Tensor: how yo mama's fatness is quantified!

Let us consider the spherical polar coordinates (r, θ, ϕ). Note that the coordinate displacement $d\phi$ does not have the dimension of length. So, while considering the displacement vector, we write $d\mathbf{r} \sim r\sin\theta d\phi\hat{\boldsymbol{\phi}}$. Thus, in general, for any displacement we write it in terms of the "metric tensor" g^{ij} as:

$$ds^2 = g_{ij}x^ix^j$$

In rectangular coordinates, we have $g_{ij} = \delta_{ij}$, that is, the metric tensor is just the identity matrix. In cylindrical coordinates where $dx^1 = d\rho, dx^2 = d\phi, dx^3 = z$, we have:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Thus, the displacement is written as

$$ds^2 = g_{ij}dx^idx^j = d\rho^2 + \rho^2 d\phi^2 + dz^2$$

In spherical polar coordinates, we have:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}$$

Thus, the displacement is written as:

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2$$

If all of g_{ij} are non-negative we call that geometry "Riemannian" and if some of them are negative, it is termed "pseudo-Riemannian".

Now, how the hell do we calculate the components of the metric tensor?

Well we have previously seen how we could obtain the basis vectors using the partial derivatives of the coordinates. So, suppose we have to find the metric tensor for spherical polar coordinate system. For that, let us first write the position vector:

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = r\sin\theta\cos\phi\ \mathbf{e}_x + r\sin\theta\sin\phi\ \mathbf{e}_y + r\cos\theta\ \mathbf{e}_z$$

From this we obtain:

$$\begin{aligned}
\mathbf{e}_r &= \frac{\partial \mathbf{r}}{\partial r} = \sin\theta\cos\phi\ \mathbf{e}_x + \sin\theta\sin\phi\ \mathbf{e}_y + \cos\theta\ \mathbf{e}_z \\
\mathbf{e}_\theta &= \frac{\partial \mathbf{r}}{\partial \theta} = r\cos\theta\cos\phi\ \mathbf{e}_x + r\cos\theta\sin\phi\ \mathbf{e}_y - r\sin\theta\ \mathbf{e}_z \\
\mathbf{e}_\phi &= \frac{\partial \mathbf{r}}{\partial \phi} = -r\sin\theta\sin\phi\ \mathbf{e}_x + r\sin\theta\cos\phi\ \mathbf{e}_y
\end{aligned}$$

Now, we had defined the metric tensor components to be the scalar product of the basis vectors. Also note that since the basis vectors are orthogonal, there will be no cross terms, so the tensor is diagonal. Using this we have:

$$\begin{aligned}
g_{rr} &= \mathbf{e}_r \cdot \mathbf{e}_r = \sin^2\theta\cos^2\phi + \sin^2\theta\sin^2\phi + \cos^2\theta = 1 \\
g_{\theta\theta} &= \mathbf{e}_\theta \cdot \mathbf{e}_\theta = r^2\cos^2\theta\cos^2\phi + r^2\cos^2\theta\sin^2\phi + r^2\sin^2\theta = r^2 \\
g_{\phi\phi} &= \mathbf{e}_\phi \cdot \mathbf{e}_\phi = r^2\sin^2\theta\sin^2\phi + r^2\sin^2\theta\cos^2\phi = r^2\sin^2\theta
\end{aligned}$$

This is exactly what we had written before. Thus using this procedure, we can easily find the components of the metric tensor and then just chill!

5.1. Metric in relativity:

We now consider the case of Minkowski space, where the coordinate displacements between two events are described by four component vector (**4 vector**):

$$dx^\mu = (dt, dx, dy, dz) \equiv (dt, d\mathbf{r})$$

The spacetime interval can be written in terms of the metric tensor $g_{\mu\nu}$ as:

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 = g_{\mu\nu}dx^\mu dx^\nu$$

Thus in this case the metric tensor is:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Note the convention. Other conventions include the $(+,-,-,-)$ or the obnoxious $(+,+,+,+)$ with an imaginary time coordinate $x^0 = ict$. The overall thing is, spatial and temporal part should have some difference.

We also define the "proper time" as $d\tau = \frac{ds}{c}$. Since we take $c = 1$, then both are equivalent but let's take c to be c for once. Then,

$$c^2 d\tau^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 = c^2 dt^2 \left(1 - \frac{dx^2 + dy^2 + dz^2}{c^2 dt^2}\right) = c^2 dt^2 \left(1 - \frac{v^2}{c^2}\right)$$

From this we have

$$\boxed{\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - \frac{v^2}{c^2}} := \frac{1}{\gamma}}$$

5.2. Metric Tensor is a Tensor!

Okay, we so since beginning we had been calling *g* to be the metric *tensor*. But hey, why so? Let us show that it is indeed a tensor. We have said that the displacement ~~doesn't~~ shouldn't change. So we have:

$$ds^2 = ds'^2 \implies g_{\alpha\beta} x^\alpha x^\beta = g'_{\mu\nu} x'^\mu x'^\nu$$

We already know how the coordinates transform, so well let's put that:

$$g_{\alpha\beta} x^\alpha x^\beta = g'_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} x^\alpha x^\beta$$

Since the coordinates are arbitrary (lol, the OG reason we use everytime to compare both sides), we have finally that:

$$g_{\alpha\beta} = \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} g'_{\mu\nu}$$

It indeed transforms like a tensor and hence is a tensor.

5.3. Relating Ordinary and Co/Contra components

Okay, so for this section, forget the Einstein summation rule. We will explicitly use the Sigma 💪 symbol to denote the sum when needed.

So, we consider the 'ordinary vectors' whose components are just the coefficients of the unit basis vectors. For example:

$$\begin{aligned} \widetilde{\mathbf{A}} &= \widetilde{A}_x \hat{i} + \widetilde{A}_y \hat{j} + \widetilde{A}_z \hat{k} \\ &= \widetilde{A}_\rho \hat{\rho} + \widetilde{A}_\phi \hat{\phi} + \widetilde{A}_z \hat{k} \end{aligned}$$

The subscripts do not mean contravariant here, these are just index of the components. To distinguish this, we use tilde to denote the ordinary vector components. Now, note that we had written the displacement as:

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

In a diagonal metric tensor then, each component of the metric can be written as $g_{\mu\nu} = h_\mu^2 \delta_{\mu\nu}$ (no sum is there). For example, see the displacement in spherical polar coordinates:

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin\theta^2 d\phi^2$$

So, $h_1 = 1, h_2 = r, h_3 = r\sin\theta$. We then do this magic and define the contravariant component to be:

$$A^\mu = \frac{\widetilde{A}_\mu}{h_\mu}$$

So why does that make sense?

Note that for ordinary vectors, we define the dot product as:

$$\mathbf{A}\cdot\mathbf{B} = \widetilde{A}_1\widetilde{B}_1 + \widetilde{A}_2\widetilde{B}_2 + \widetilde{A}_3\widetilde{B}_3 = \sum_\mu h_\mu^2 A^\mu B^\mu = \sum_{\mu\nu} g_{\mu\nu} A^\mu B^\nu$$

We had used the definition of h using the delta function and metric tensor. Thus we see that using this kind of definition, the dot product for ordinary vectors can be made analogous to the definition of the inner product using the metric tensor. Once we have the contravariant components, we can find the covariant using the metric tensor:

$$A_\mu = \sum_\nu g_{\mu\nu} A^\nu = \sum_\nu h_\mu^2 \delta_{\mu\nu} A^\nu = h_\mu^2 A^\mu = h_\mu \widetilde{A}^\mu$$

Let us see an example for the spherical polar coordinates:

$$\begin{aligned}
A^1 &= \frac{\widetilde{A}_r}{1} && A_1 = 1\cdot\widetilde{A}_r \\
A^2 &= \frac{\widetilde{A}_\theta}{r} && A_2 = r\cdot\widetilde{A}_\theta \\
A^3 &= \frac{1}{\sin\theta}\frac{\widetilde{A}_\phi}{r} && A_3 = r\sin\theta\cdot\widetilde{A}_\phi
\end{aligned}$$

5.3.1. Case of the Gradient

We define the ordinary gradient in similar lines to the ordinary vectors. Given a function f, the ordinary gradient is just:

$$\nabla f \equiv \frac{\partial f}{\partial x}\hat{\imath} + \frac{\partial f}{\partial y}\hat{\jmath} + \frac{\partial f}{\partial z}\hat{k}$$

So basically we take derivatives with respect to the lengths x, y or z. On the other hand, we have 'covariant' gradient ∂_μ where we take derivatives with respect to a coordinate. So, in systems where the coordinates do not have the unit of length, like θ does not have length unit, both of these will differ. We can relate them by:

$$\nabla_\mu = \frac{1}{h_\mu}\partial_\mu f$$

Here ∇_μ means the ordinary derivative with respect to coordinate μ but with units of length. Let's clear this with an example of the dear spherical coordinates:

$$\begin{aligned}
\nabla_r &= \frac{1}{h_1}\partial_r f && = \partial_r f \\
\nabla_\theta &= \frac{1}{h_2}\partial_\theta f && = \frac{1}{r}\partial_\theta f \\
\nabla_\phi &= \frac{1}{h_3}\partial_\phi f && = \frac{1}{r\sin\theta}\partial_\phi f
\end{aligned}$$

Note that using this, we easily get the formula for gradient in spherical coordinate:

$$\nabla f = \frac{\partial f}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial f}{\partial\theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}\hat{\boldsymbol{\phi}}$$

5.3.2. Jacobian

We had often written (while doing some calculations in classical mechanics) $dm = \rho dV \equiv \rho dxdydz$. Now, since mass is a scalar (duh!), there is no reason to disbelief that dm is also a scalar. We know that dV is not a scalar (here by dV I mean the product of the differentials), I mean dV depends on the coordinate system, $dxdxydx \neq dx'dy'dz'$. And, since ρ and dV combine to give a scalar, ρ can also not be a scalar. So, let's see how to deal with this.

Let us take the example of a curve C enclosing some area. We have two coordinate systems, $\{x, y\}$ and $\{x', y'\}$ with transformations given by:

$$x' = f(x, y) \qquad y' = g(x, y)$$

Let a and b be the point on the x-axis where the tangent to the curve is vertical.

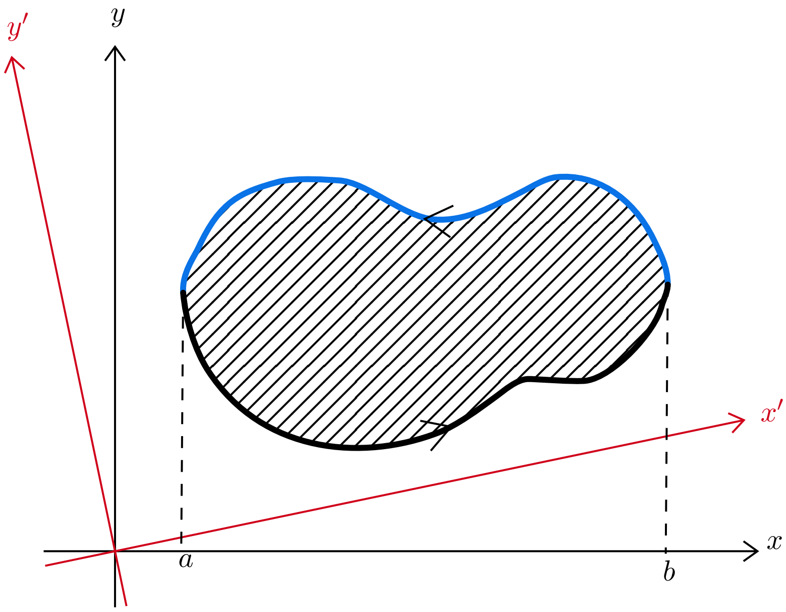


Figure 4: Measuring area in two coordinate systems

Now, the area enclosed by the curve in the $\{x, y\}$ system is given by:

$$A = \int_a^b y_{\text{blue}}(x)dx - \int_a^b y_{\text{black}}(x)dx = \int_a^b y_{\text{blue}}(x)dx + \int_b^a y_{\text{black}}(x)dx = -\oint_C ydx$$

The negative sign comes since the curve is taken anti-clockwise. Similarly for the $\{x', y'\}$ system, we have:

$$A' = -\oint_C y'dx'$$

Note that y' in the integrand is the y' coordinate. We are basically integrating the value ydx over the curve. Then we can write:

$$
\begin{aligned}
A' &= -\oint\limits_C y' dx' \\
&= -\oint\limits_C g(x,y)\left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy\right) \\
&= -\oint\limits_C (M_x dx + M_y dy) \\
&= -\oint\limits_C \mathbf{M}\cdot d\mathbf{r} \\
&= -\int\int\limits_S (\nabla\times\mathbf{M})\cdot\hat{\mathbf{n}}\ da
\end{aligned}
$$

where have defined $M_x = g(x,y)\frac{\partial f}{\partial x} \equiv gf_x$ and $M_y = g(x,y)\frac{\partial f}{\partial y} \equiv gf_y$ and used Stokes' Theorem. In our case, $\hat{\mathbf{n}}$ is along the z-direction and $da \equiv dxdy$, so only $(\nabla\times\mathbf{M})_z = \frac{\partial M_y}{\partial x} - \frac{\partial M_x}{\partial y}$ will survive. Thus, we finally have:

$$
\begin{aligned}
A' &= -\iint\limits_S \left(\frac{\partial(gf_y)}{\partial x} - \frac{\partial(gf_x)}{\partial y}\right) dxdy \\
&= -\iint\limits_S \left(\cancel{g\frac{\partial f_y}{\partial x}} + f_y\frac{\partial g}{\partial x} - \cancel{g\frac{\partial f_x}{\partial y}} - f_x\frac{\partial g}{\partial y}\right) dxdy \\
&= -\iint\limits_S \left(\frac{\partial f}{\partial y}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial g}{\partial y}\right) dxdy \\
&= \iint\limits_S \left(\frac{\partial x'}{\partial x}\frac{\partial y'}{\partial y} - \frac{\partial x'}{\partial y}\frac{\partial y'}{\partial x}\right) dxdy \\
&= \iint\limits_S \det\begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y}\end{bmatrix} dxdy
\end{aligned}
$$

Note that the terms in the second step cancelled because partial derivatives commute. Like we have terms like $\frac{\partial f_x}{\partial y} \equiv \frac{\partial^2 f}{\partial x\partial y}$ and $\frac{\partial f_y}{\partial x} \equiv \frac{\partial^2 f}{\partial y\partial x}$, both of which are equal. Now we define this weird determinant with a high-sounding name called Jacobian:

$$
\mathrm{J} := \det\begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y}\end{bmatrix}
$$

Then the integral becomes:

$$
A' = \iint\limits_S \mathrm{J}\ dxdy = \iint\limits_S dx'dy'
$$

In short notation, we write $\mathrm{J} = \left|\frac{\partial x'}{\partial x}\right|$. Well the thing is, technically we should call the Jacobian matrix as J but it's fine I guess. Just understand from the context. So this serves two purposes:

- Firstly, it tells where the prime and the unprimed things come in the matrix. Since unprimed things are in the denominator, we have the unprimed in the denominator in the matrix also.

- Secondly, it tells us where to put the Jacobian while doing the transformation. Like we can make the mathematicians crazy and write:

$$\iint \left|\frac{\partial x'}{\cancel{\partial x}}\right| \cancel{dx}dy \equiv \iint dx'dy'$$

So the unprimed denominator cancels with the unprimed differentials and the net result is the primed thing. So we can say that the from primed to unprimed frame, the Jacobian always comes with '$dxdy$'

Generalising this to a volume element, we would have obtained $dx'dy'dz' = \mathrm{J}dxdydz$

5.3.3. Relating Metric and Jacobian

We take the absolute value of the determinant of both sides of the metric tensor transformation equation:

$$\left|\det g'_{\mu\nu}\right| = \left|\det \frac{\partial x'^{\rho}}{\partial x^{\mu}}\right| \left|\det \frac{\partial x'^{\sigma}}{\partial x^{\nu}}\right| \left|\det g_{\rho\sigma}\right|$$

Note that the determinants written above are just representations, like they are actually determinants of the matrices whose components are written in the above expression. For simplicity, we will denote the absolute value of the determinant of something simply by $|\cdot|$. Thus, we have $|g'| = \left|\frac{\partial x}{\partial x'}\right|^2 |g|$ We take the absolute value since we had seen that for pseudo-Riemannian metric, the determinant of g is negative, like the Minkowski metric has determinant -1. Also note that since $\frac{\partial x}{\partial x'}$ and $\frac{\partial x'}{\partial x}$ are inverses of each other (again, their matrices), then we have

$$\left|\frac{\partial x}{\partial x'}\right| = \left|\frac{\partial x'}{\partial x}\right|^{-1}$$

Then we can write:

$$\left|\frac{\partial x}{\partial x'}\right| |g| = \left|\frac{\partial x'}{\partial x}\right| |g'| \implies \mathrm{J}^{-1}|g| = J|g'|$$

We used the fact that determinant of the inverse matrix is just the inverse (reciprocal) of the determinant. Then we have

$$\mathrm{J} = \sqrt{\frac{|g|}{|g'|}}$$

Then we can write the area transformation as:

$$\sqrt{|g'|}dx'dy' = \sqrt{|g|}dxdy$$

This quantity is a scalar and this is a *scalar transformation*. Generalising this, we can write for arbitrary dimension:

$$\sqrt{|g'|}d^n x' = \sqrt{|g|}d^n x$$

So we saw how volume element changes between coordinate systems. Now, since we have to keep dq or dm for example, to be scalars then we need the density to transform as $\rho' = \mathrm{J}^{-1}\rho$, so that the jacobian cancels from the volume element and it gives us a scalar.

In a nutshell...

Any quantity that transform in the following way:

$$Q' = J^w Q$$

is called a *tensor density* of weight w. So the volume element is a tensor density of weight $+1$ and the charge or mass density is of weight -1.

6. Tensor Derivatives: levelling up the rizz!

Till now, we had just seen the transformation of tensors and tensor densities. However, an important aspect of any calculation is the ability to take derivatives [1]. Derivatives occur everywhere in calculations and we need a way to tackle them. So let's start...

6.1. Velocity

Well velocity is a vector (we had been reminded many a times) and it should then transform as a vector. Now, we know the transformation:

$$dx'^i = \frac{\partial x'^i}{\partial x^l} x^l$$

To find velocity components, we have to differentiate with respect to time t.

$$\begin{aligned}
\frac{\mathrm{d}x'^i}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial x'^i}{\partial x^l} x^l\right) \\
&= \frac{\partial x'^i}{\partial x^l}\frac{\mathrm{d}x^l}{\mathrm{d}t} + x^l \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial x'^i}{\partial x^l}\right) \\
&= \frac{\partial x'^i}{\partial x^l}\frac{\mathrm{d}x^l}{\mathrm{d}t} + x^l \frac{\partial^2 x'^i}{\partial x^k \partial x^l}\frac{\mathrm{d}x^k}{\mathrm{d}t}
\end{aligned}$$

Now we define $v^k = \frac{\mathrm{d}x^k}{\mathrm{d}t}$. Then the above expression would give:

$$v'^i = \frac{\partial x'^i}{\partial x^l} v^l + \frac{\partial^2 x'^i}{\partial x^k \partial x^l} x^l v^k$$

The first term gives the proper thing for velocity to be a vector, like the correct transformation. The second term is the BAD term 😤. Let's see some examples of this term in some transformation:

Rotation:

$$\begin{aligned}
x' &= x\cos\theta + y\sin\theta \\
y' &= -x\sin\theta + y\cos\theta
\end{aligned}$$

So we have:

$$\frac{\partial x'}{\partial x} = \cos\theta \quad \frac{\partial x'}{\partial y} = \sin\theta \quad \frac{\partial y'}{\partial x} = -\sin\theta \quad \frac{\partial y'}{\partial y} = \cos\theta$$

Note that the first derivatives do not depend on the coordinate anymore. For a fixed θ, the first derivatives are constants. And if we consider the transformation of the velocity components, then the second term will vanish, since these contain double derivatives. Thus, the BAD term vanishes and we happily see that velocity is a vector under rotation. In Galilean transformation ($x' = x - vt, y' = y, z' = z, t' = t$) too, the second term vanish. Even in Lorentz transformation ($t' = \gamma\left(t - vx\right), x = \gamma(x - vt), y' = y, z' = z$) the bad term vanishes. So in basic transformations, velocity is indeed a vector. Let us now see how derivative of a tensor component transforms. So we

[1] A person's intellectual prowess can be judged their ability to take derivatives 😎

have,

$$
\begin{aligned}
(\partial_\lambda T^\alpha)' = \partial_{\lambda'} T'^\lambda &= \frac{\partial}{\partial x'^\lambda}\left(\frac{\partial x'^\lambda}{\partial x^\sigma} T^\sigma\right) && \text{(contravariant transformation of tensor)} \\
&= \frac{\partial x^\rho}{\partial x'^\lambda}\frac{\partial}{\partial x^\rho}\left(\frac{\partial x'^\lambda}{\partial x^\sigma} T^\sigma\right) && \text{(covariant transformation of derivative)} \\
&= \frac{\partial x^\rho}{\partial x'^\lambda}\frac{\partial x'^\alpha}{\partial x^\sigma}\partial_\rho T^\sigma + \frac{\partial x^\rho}{\partial x'^\lambda}\frac{\partial^2 x'^\alpha}{\partial x^\rho \partial x^\sigma} T^\sigma && \text{(chain rule)}
\end{aligned}
$$

Again, the first term is the usual thing but the BAD term appears again! Notice how always the bad term contains a double derivative. Now we know that double derivatives have something to do with curvatures. Let us clarify a bit more.

Imagine the position vector $\mathbf{r}(t)$ on a flat space. Then the tangent vector to a point having position $\mathbf{r}(t)$, say $\mathbf{v}$, will lie entirely on the same space, right? But now, imagine the space being curved. Now if we draw the tangent, it will inevitable leave the space. Imagine the people living on the surface of a sphere. The velocity vector for a moving body in the space will be tangent to the sphere and points off of it. So, for the inhabitants, the velocity vector doesn't exist since it is not contained entirely on the space (Is this the same case with *God*? Hmm, something to think about 🤔). What they can do it, just take some small patch of the sphere (which is flat) where the tangent touches the surface and around it, locally, they can define the velocity vector. So, in general the velocity is not a tensor in curved space, where the second derivative is non-zero.

6.2. Affine Connection

Suppose in a locally inertial frame, for an object we have zero acceleration, that is:

$$\frac{\mathrm{d}^2 X^\alpha}{\mathrm{d}\tau^2} = 0$$

Suppose the coordinates $X^\alpha \equiv X^\alpha(x^\mu)$ where x^μ are coordinates of a ground-based inertial reference frame relative to which the object accelerates. Then using chain rule, we have from the previous relation:

$$\frac{d}{d\tau}\left(\frac{\partial X^\alpha}{\partial x^\mu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\right) = 0$$

Using product rule and chain rule again, we have:

$$\frac{\partial X^\alpha}{\partial x^\mu}\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \frac{\partial^2 X^\alpha}{\partial x^\mu \partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = 0$$

Now we do one nice thing: multiply the above with $\frac{\partial x^\lambda}{\partial X^\alpha}$:

$$\frac{\partial x^\lambda}{\partial X^\alpha}\left(\frac{\partial X^\alpha}{\partial x^\mu}\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2}\right) + \frac{\partial x^\lambda}{\partial X^\alpha}\left(\frac{\partial^2 X^\alpha}{\partial x^\mu \partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\right) = 0$$

Doing this, a nice thing occurs but for that we have to note that $\frac{\partial x^\lambda}{\partial X^\alpha}\frac{\partial X^\alpha}{\partial X^\mu} = \delta^\lambda{}_\mu$. Then from the first term, we get the kronecker delta and the expression reduces to:

$$\frac{\mathrm{d}^2 x^\lambda}{\mathrm{d}\tau^2} + \frac{\partial x^\lambda}{\partial X^\alpha}\frac{\partial^2 X^\alpha}{\partial x^\mu \partial x^\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = 0 \quad \implies \quad \frac{\mathrm{d}^2 x^\lambda}{\mathrm{d}\tau^2} + \Gamma^\lambda{}_{\mu\rho}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tau}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = 0$$

Here we have identified the scary blue term with a symbol with three indices, one upper and two lower (since in the blue term, there is one upper and two lower indices). This thing, we call the

affine connection. Note that since derivatives commute[1] generally, we have the lower indices of the affine connection to be symmetric, that is,

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

6.2.1. Definition in terms of basis vectors

We can define the affine connection using basis vectors also. Consider the derivative of the basis vector $\mathbf{e}_i$ with respect to some coordinates. This derivate is another vector which when expanded in terms of the basis, the coefficients are nothing but the affine connection.

$$\frac{\partial \mathbf{e}_i}{\partial x^j} = \Gamma^{k}{}_{ij}\mathbf{e}_k$$

6.2.2. Transformation of Affine Connection

Note that the affine connection contains both coordinates X^α and $x\mu$. Suppose we want to transform from x to x' coordinate system, then only the x things will be changed, not the X things. Leave the X alone because the transformation being studied is specifically about how the connection coefficients behave when you change coordinate systems on the ground. So we have:

$$\begin{aligned}
(\Gamma')^{\lambda}{}_{\mu\nu} &= \frac{\partial x'^\lambda}{\partial X^\alpha}\frac{\partial^2 X^\alpha}{\partial x'^\mu \partial x'^\nu} \\
&= \frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\frac{\partial^2 X^\alpha}{\partial x'^\mu \partial x'^\nu} \quad \text{(using chain rule)} \\
&= \left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\right)\frac{\partial}{\partial x'^\mu}\left(\frac{\partial X^\alpha}{\partial x^\sigma}\frac{\partial x^\sigma}{\partial x'^\nu}\right) \quad \text{(using chain rule again)} \\
&= \left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\right)\left(\frac{\partial x^\sigma}{\partial x'^\nu}\frac{\partial}{\partial x'^\mu}\left(\frac{\partial X^\alpha}{\partial x^\sigma}\right) + \frac{\partial X^\alpha}{\partial x^\sigma}\frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu}\right) \quad \text{(using product rule)} \\
&= \left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\right)\left(\frac{\partial x^\sigma}{\partial x'^\nu}\frac{\partial^2 X^\alpha}{\partial x^\kappa \partial x^\sigma}\frac{\partial x^\kappa}{\partial x'^\mu} + \frac{\partial X^\alpha}{\partial x^\sigma}\frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu}\right) \quad \text{(using chain rule)}
\end{aligned}$$

Well well, I know this was a shitty calculation but hey, sometimes shit is what relieves us! We now focus on the two terms separately in the above expression.

- **The First Term:** $\left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\right)\frac{\partial x^\sigma}{\partial x'^\nu}\frac{\partial^2 X^\alpha}{\partial x^\kappa \partial x^\sigma}\frac{\partial x^\kappa}{\partial x'^\mu}$

  This can be rearranged a bit and can be written as

$$\left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\sigma}{\partial x'^\nu}\frac{\partial x^\kappa}{\partial x'^\mu}\right)\left(\frac{\partial x^\rho}{\partial X^\alpha}\frac{\partial^2 X^\alpha}{\partial x^\kappa \partial x^\sigma}\right) \equiv \left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\sigma}{\partial x'^\nu}\frac{\partial x^\kappa}{\partial x'^\mu}\right)\Gamma^{\rho}{}_{\kappa\sigma}$$

- **The Second Term:** $\left(\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial x^\rho}{\partial X^\alpha}\right)\frac{\partial X^\alpha}{\partial x^\sigma}\frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu}$ The blue terms together gives $\delta^\rho{}_\sigma$ which reduces the expression to:

$$\frac{\partial x'^\lambda}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\nu}$$

[1] If a space has something called a 'torsion', then the derivatives no longer commute and the following property does not hold true

Thus, finally we obtain the expression for the transformation of the affine connection:

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

Note that the first term is the usual transformation rule for the affine connection but again the second BAD term emerges which, if non-zero, will lead to the affine connection not being a tensor.

Now note that the BAD term contains the second derivative of the old coordinates with respect to the old coordinates, but generally we have the other way. So it would be a bit nice if we could change it. For that, note the identity and differentiate with respect to x'^{μ}:

$$\begin{aligned}
&\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\rho}}{\partial x'^{\nu}} = \delta^{\lambda}{}_{\nu}\\
\implies &\frac{\partial}{\partial x'^{\mu}}\left(\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\right) = 0\\
\implies &\left(\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\right)\left(\frac{\partial^2 x^{\rho}}{\partial x'^{\nu}\partial x'^{\mu}}\right) + \left(\frac{\partial x^{\rho}}{\partial x'^{\nu}}\right)\left(\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x'^{\mu}}\right) = 0\\
\implies &\left(\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\right)\left(\frac{\partial^2 x^{\rho}}{\partial x'^{\nu}\partial x'^{\mu}}\right) + \left(\frac{\partial x^{\rho}}{\partial x'^{\nu}}\right)\left(\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\right) = 0
\end{aligned}$$

The first term in the above is exactly the BAD term in the affine connection transformation and thus we replace this. Then we have:

$$\boxed{(\Gamma')^{\lambda}{}_{\mu\nu} = \left(\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\sigma}}{\partial x'^{\nu}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\right)\Gamma^{\rho}{}_{\kappa\sigma} - \left(\frac{\partial x^{\rho}}{\partial x'^{\nu}}\right)\left(\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\right)}$$

6.3. Covariant Derivatives

In every transformation seen so far, we had got a BAD term (containing a second derivative) which spoils the transformation. So wouldn't it be nice if we just redefine the definition of a derivative so that this BAD term gets cancelled from the definition only? This brings us to *covariant derivative*

A Notational Nightmare:

The symbol of covariant derivative is very confusing. Different people use different notation for it. Some use D for it, some use ∇ while some use ;. We will use ; for it, I guess but we may sometimes shift to D notation if ambiguity arises...

The covariant derivative of a vector component with respect to a scalar is defined as:

$$\frac{DA^{\lambda}}{D\tau} := \frac{dA^{\lambda}}{d\tau} + \Gamma^{\lambda}{}_{\mu\nu}\frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau}A^{\nu}$$

Here we used the D notation since the ; notation is mostly used when we differentiate with respect to some coordinate. The covariant derivative of a vector component with respect to a coordinate is then[1]:

$$A^{\lambda}{}_{;\mu} := A^{\lambda}{}_{,\mu} + \Gamma^{\lambda}{}_{\mu\nu}A^{\nu}$$

The covariant derivative of a covariant component is similarly:

$$A_{\lambda;\mu} := A_{\lambda,\mu} - \Gamma^{\alpha}{}_{\lambda\mu}A_{\alpha}$$

[1] $A^{\nu}{}_{,\mu}$ means the normal derivative, that is, $\partial_{\mu}A^{\nu}$

6.3.1. Transformation of Covariant Derivative

In the primed frame, we have:

$$
\begin{aligned}
{A'^{\lambda}}_{;\tau} &= {A'^{\lambda}}_{,\tau} + {\Gamma'^{\lambda}}_{\mu\nu}\frac{\mathrm{d}x'^{\mu}}{\mathrm{d}\tau}A'^{\nu} \\
&= \left(\frac{\partial x'^{\lambda}}{\partial x^{l}}\frac{\mathrm{d}A^{l}}{\mathrm{d}\tau} + A^{l}\frac{\partial^2 x'^{\lambda}}{\partial x^{k}\partial x^{l}}\frac{\mathrm{d}x^{k}}{\mathrm{d}\tau}\right) \\
&\quad + \left[\left(\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\sigma}}{\partial x'^{\nu}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\right){\Gamma^{\rho}}_{\kappa\sigma} - \left(\frac{\partial x^{\rho}}{\partial x'^{\nu}}\right)\left(\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\right)\right] \\
&\qquad \times \left(\frac{\partial x'^{\mu}}{\partial x^{\omega}}\frac{\mathrm{d}x^{\omega}}{\mathrm{d}\tau} + x^{\omega}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}\right)\frac{\partial x'^{\nu}}{\partial x^{\theta}}A^{\theta}
\end{aligned}
$$

Let's see this term by term.

- **The First Term:** Transformation of normal derivative

  $$\boxed{\frac{\partial x'^{\lambda}}{\partial x^{l}}\frac{\mathrm{d}A^{l}}{\mathrm{d}\tau}} + A^{l}\frac{\partial^2 x'^{\lambda}}{\partial x^{q}\partial x^{l}}\frac{\mathrm{d}x^{q}}{\mathrm{d}\tau}$$

  This is fine for now. Let's leave it here!

- **The Second Term:** Multiplication of three individual terms Well, this is the monster 👾 actually. When expanded, it will have four terms. Let us write them one by one:

  1. After reducing the Kronecker delta, the final expression becomes:

     $$\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\sigma}}{\partial x'^{\nu}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x'^{\nu}}{\partial x^{\theta}}\frac{\partial x'^{\mu}}{\partial x^{\omega}}\frac{\mathrm{d}x^{\omega}}{\mathrm{d}\tau}{\Gamma^{\rho}}_{\kappa\sigma}A^{\theta} = \boxed{\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\mathrm{d}x^{\kappa}}{\mathrm{d}\tau}{\Gamma^{\rho}}_{\kappa\sigma}A^{\sigma}}$$

  2. After reducing the Kronecker delta, we have:

     $$-\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\frac{\partial x'^{\nu}}{\partial x^{\theta}}\frac{\partial x'^{\mu}}{\partial x^{\omega}}\frac{\mathrm{d}x^{\omega}}{\mathrm{d}\tau}A^{\theta} = -\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\mathrm{d}x^{\sigma}}{\mathrm{d}\tau}A^{\rho}$$

     Note this final term and the **first term**, both of these differ only by the fact that $q \to \sigma$ and $l \to \rho$ but since these are dummy indices, we can just rename them and these terms are actually equal but we have a minus sign in this term, so this cancels with the first term.

  3. After reducing the Kronecker delta, we have:

     $$\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial x'^{\sigma}}{\partial x'^{\nu}}\frac{\partial x'^{\nu}}{\partial x^{\theta}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}x^{\omega}{\Gamma^{\rho}}_{\kappa\sigma}A^{\theta} = \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}x^{\omega}{\Gamma^{\rho}}_{\kappa\sigma}A^{\sigma}$$

  4. Same, after reducing the Kronecker delta, we have:

     $$-\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\frac{\partial x^{\rho}}{\partial x'^{\nu}}\frac{\partial x'^{\nu}}{\partial x^{\theta}}\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}x^{\omega}A^{\theta} = -\frac{\partial x^{\sigma}}{\partial x'^{\mu}}\frac{\partial^2 x'^{\lambda}}{\partial x^{\rho}\partial x^{\sigma}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}x^{\omega}A^{\rho}$$

Okay, so the green terms cancel and note the boxed terms, these are actually the terms which should have been in the transformation equation of covariant derivative, if it were a tensor. So, let us now focus on the remaining terms, that is, the last two terms. Note the third term:

$$\begin{aligned}\frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}x^{\omega}\left(\Gamma^{\rho}{}_{\kappa\sigma}\right)A^{\sigma} &= \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}\left(\frac{\partial x^{\rho}}{\partial x'^{\gamma}}\frac{\partial^2 x'^{\gamma}}{\partial x^{\kappa}\partial x^{\sigma}}\right)x^{\omega}A^{\sigma} \\ &= \frac{\partial x^{\kappa}}{\partial x'^{\mu}}\frac{\partial^2 x'^{\mu}}{\partial x^{\omega}\partial x^{\beta}}\frac{\mathrm{d}x^{\beta}}{\mathrm{d}\tau}\frac{\partial^2 x'^{\lambda}}{\partial x^{\kappa}\partial x^{\sigma}}x^{\omega}A^{\sigma}\end{aligned}$$

We had just replaced the affine connection coefficient and reduced the kronecker delta. Then this turns into the fourth term ($\rho \to \sigma$ and $\sigma \to \kappa$) and hence these two terms cancel. Then we remain only with the boxed terms and hence the covariant derivative with respect to a scalar actually transforms as a vector.

$$A'^{\lambda}{}_{;\tau} = \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\mathrm{d}A^{\rho}}{\mathrm{d}\tau} + \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\mathrm{d}x^{\kappa}}{\mathrm{d}\tau}\Gamma^{\rho}{}_{\kappa\sigma}A^{\sigma} = \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\left(\frac{\mathrm{d}A^{\rho}}{\mathrm{d}\tau} + \frac{\mathrm{d}x^{\kappa}}{\mathrm{d}\tau}\Gamma^{\rho}{}_{\kappa\sigma}A^{\sigma}\right) = \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\mathrm{d}A^{\rho}}{\mathrm{d}\tau}A^{\rho}{}_{;\tau}$$

We saw how *easy* it was to show the transformation of covariant derivative. It's going to get easier as we progress. Now, similar to the above, we can show that the covariant derivative with respect to a coordinate is a second-rank tensor.

$$A'^{\lambda}{}_{;\mu} = \frac{\partial x'^{\lambda}}{\partial x^{\rho}}\frac{\partial x^{\sigma}}{\partial x'^{\mu}}A^{\rho}{}_{;\mu}$$

The covariant derivative of a general tensor is given by the following formula:

$$A^{i_1\ldots i_r}{}_{j_1\ldots j_s;p} = A^{i_1\ldots i_r}{}_{j_1\ldots j_s,p} + \sum_{u=1}^{r}\Gamma^{i_u}{}_{h_u p}A^{i_1\ldots i_{u-1}h_u i_{u+1}\ldots i_r}{}_{j_1\ldots j_s} - \sum_{u=1}^{s}A^{i_1\ldots i_r}{}_{j_1\ldots j_{u-1}h_u j_{u+1}\ldots j_s}\Gamma^{h_u}{}_{j_u p}$$

Basically, the first term is the conventional derivative. For all the contravariant indices, we have $+$ sign while for covariant indices, we have $-$ sign. And in each sum, remove the u^{th} index and replace it with an arbitrary variable in the tensor and then accordingly adjust the affine connection. Let us see some examples perhaps, using this formula:

$$\begin{aligned}T^{\mu\nu}{}_{;\beta} &= T^{\mu\nu}{}_{,\beta} + \Gamma^{\mu}{}_{\kappa\beta}T^{\kappa\nu} + \Gamma^{\nu}{}_{\kappa\beta}T^{\mu\kappa} \\ T^{\mu\nu}{}_{\sigma;\beta} &= T^{\mu\nu}{}_{\sigma,\beta} + \Gamma^{\mu}{}_{\kappa\beta}T^{\kappa\nu}{}_{\sigma} + \Gamma^{\nu}{}_{\kappa\beta}T^{\mu\kappa}{}_{\sigma} - \Gamma^{\kappa}{}_{\sigma\beta}A^{\mu\nu}{}_{\kappa}\end{aligned}$$

6.3.2. Covariant Derivatives using Basis Vectors

Theorem 1 (Product Rule):
The covariant derivative satisfies a kind of product rule like:

$$(A^{\nu}B^{\mu})_{;\alpha} = A^{\nu}B^{\mu}{}_{;\alpha} + B^{\mu}A^{\nu}{}_{;\alpha}$$

Proof. We show it for a rank two contravariant tensor. We begin with the usual expansion of the covariant derivative:

$$\begin{aligned}(A^{\nu}B^{\mu})_{;\alpha} &= \partial_{\alpha}(A^{\nu}B^{\mu}) + \Gamma^{\nu}{}_{\kappa\alpha}A^{\kappa}B^{\mu} + \Gamma^{\mu}{}_{\kappa\alpha}A^{\nu}B^{\kappa} \\ &= B^{\mu}\partial_{\alpha}A^{\nu} + A^{\nu}\partial_{\alpha}B^{\mu} + \Gamma^{\nu}{}_{\kappa\alpha}A^{\kappa}B^{\mu} + \Gamma^{\mu}{}_{\kappa\alpha}A^{\nu}B^{\kappa} \\ &= B^{\mu}\left(\partial_{\alpha}A^{\nu} + \Gamma^{\nu}{}_{\kappa\alpha}A^{\kappa}\right) + A^{\nu}\left(\partial_{\alpha}B^{\mu} + \Gamma^{\mu}{}_{\kappa\alpha}B^{\kappa}\right) \\ &= B^{\mu}A^{\nu}{}_{;\alpha} + A^{\nu}B^{\mu}{}_{;\alpha}\end{aligned}$$

This can be generalised to higher rank tensor with mixed indices as well.

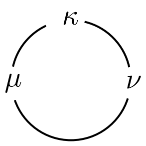
6.4. Relating metric and affine connection

We use the basis definition of the affine connection.

$$\begin{aligned}
\Gamma^{\lambda}{}_{\kappa\mu} g_{\lambda\nu} &= \Gamma^{\lambda}{}_{\kappa\mu} (\mathbf{e}_\lambda \cdot \mathbf{e}_\nu) \\
&= (\partial_\mu \mathbf{e}_\kappa) \cdot \mathbf{e}_\nu \\
&= \partial_\mu (\mathbf{e}_\kappa \cdot \mathbf{e}_\nu) - \mathbf{e}_\kappa \cdot \partial_\mu \mathbf{e}_\nu \\
&= \partial_\mu g_{\kappa\nu} - \Gamma^{\lambda}{}_{\nu\mu} \mathbf{e}_\kappa \cdot \mathbf{e}_\lambda
\end{aligned}$$

Thus we get:

$$\Gamma^{\lambda}{}_{\kappa\mu} g_{\lambda\nu} + \Gamma^{\lambda}{}_{\nu\mu} g_{\kappa\lambda} = \partial_\mu g_{\kappa\nu}$$



We now use a cyclic permutation of μ, κ, ν to obtain the following two equations[1]:

$$\begin{aligned}
\Gamma^{\lambda}{}_{\mu\nu} g_{\kappa\lambda} + \Gamma^{\lambda}{}_{\kappa\nu} g_{\lambda\mu} &= \partial_\nu g_{\mu\kappa} \\
\Gamma^{\lambda}{}_{\nu\kappa} g_{\lambda\mu} + \Gamma^{\lambda}{}_{\mu\kappa} g_{\lambda\nu} &= \partial_\kappa g_{\nu\mu}
\end{aligned}$$

Let us now add the first and last equations and subtract the middle one:

$$\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa} = \cancel{\Gamma^{\lambda}{}_{\nu\kappa} g_{\lambda\mu}} + \Gamma^{\lambda}{}_{\mu\kappa} g_{\lambda\nu} + \Gamma^{\lambda}{}_{\kappa\mu} g_{\lambda\nu} + \cancel{\Gamma^{\lambda}{}_{\nu\mu} g_{\lambda\kappa}} - \cancel{\Gamma^{\lambda}{}_{\mu\nu} g_{\lambda\kappa}} - \cancel{\Gamma^{\lambda}{}_{\kappa\nu} g_{\lambda\mu}}$$

We had assumed a torsion-free space and hence the affine connections commute in their lower indices. Hence those terms get cancelled and we finally have the expression as:

$$\Gamma^{\lambda}{}_{\mu\kappa} g_{\lambda\nu} = \frac{1}{2} (\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa})$$

Multiply the above by $g^{\nu\alpha}$, then we have:

$$\begin{aligned}
\Gamma^{\lambda}{}_{\mu\kappa} g_{\lambda\nu} &= \frac{1}{2} (\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa}) \\
\implies \Gamma^{\lambda}{}_{\mu\kappa} g_{\lambda\nu} g^{\nu\alpha} &= \frac{1}{2} g^{\nu\alpha} (\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa}) \\
\implies \Gamma^{\lambda}{}_{\mu\kappa} \delta^{\alpha}{}_{\lambda} &= \frac{1}{2} g^{\nu\alpha} (\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa}) \\
\implies \Gamma^{\alpha}{}_{\mu\kappa} &= \frac{1}{2} g^{\nu\alpha} (\partial_\kappa g_{\nu\mu} + \partial_\mu g_{\kappa\nu} - \partial_\nu g_{\mu\kappa})
\end{aligned}$$

Whew!!😌 Let us now see an example to find the connection coefficient for 2D polar coordinates.The metric and inverse metric tensor are:

$$g_{ij} \equiv \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \qquad g^{ij} \equiv \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2} \end{pmatrix}$$

[1] λ is summed over in the expression, so we do not consider it in permutation

Then taking the derivative of the metric tensor, we have:

$$\partial_r g_{ij} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 2r \end{pmatrix} \qquad \partial_\theta g_{ij} \equiv \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The only non-zero element is at $\partial_r g_{\theta\theta}$. Then we only have two non-zero connection coefficients:

$$\begin{aligned} {\Gamma^\theta}_{r\theta} &= \frac{1}{2} g^{\theta\theta} \left(\cancel{\partial_\theta g_{\theta r}} + \partial_r g_{\theta\theta} - \cancel{\partial_\theta g_{r\theta}} \right) \\ &= \frac{1}{2r^2} \cdot 2r \\ &= \frac{1}{r} \end{aligned}$$

$$\begin{aligned} {\Gamma^r}_{\theta\theta} &= \frac{1}{2} g^{rr} \left(\cancel{\partial_\theta g_{r\theta}} + \cancel{\partial_\theta g_{\theta r}} - \partial_r g_{\theta\theta} \right) \\ &= -\frac{1}{2} \cdot 2r \\ &= -r \end{aligned}$$

6.5. Curls, divergences and other craps

6.5.1. Covariant Divergence

The 'ordinary' gradient that we had written earlier, will be generalised to the covariant derivative. Then we will have the divergence modified as:

$$\nabla \cdot \tilde{A} \quad \to \quad {A^\mu}_{;\mu} = \partial_\mu A^\mu + {\Gamma^\mu}_{\mu\lambda}$$

From the above relation between metric tensor and the Christoffel symbol, we have:

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

Notice the two coloured terms. Since μ and ν are being summed over, we can just interchange them in the, say, red term. Then we will have that both the terms cancel. Thus, we have a simplified expression for the contracted Christoffel symbol as ${\Gamma^\mu}_{\mu\lambda} = \frac{1}{2} g^{\nu\mu} \left(\partial_\lambda g_{\nu\mu} \right)$. We will now show that this expression can be reduced to the expression:

$${\Gamma^\mu}_{\mu\lambda} = \frac{1}{\sqrt{|g|}} \partial_\lambda \left(\sqrt{|g|} \right)$$

For that we prove some basic identities:

Identity 1:
For any (diagonalisable) matrix M, we have:

$$\mathrm{Tr}\Big(M^{-1} \partial_\lambda M \Big) = \partial_\lambda \ln |M|$$

Proof. Let us calculate the variation of the quantity in the right hand side due to some variation δx^λ in x^λ:

$$\begin{aligned} \delta \ln |M| &= \ln |M + \delta M| - \ln |M| \\ &= \ln \left(\frac{|M + \delta M|}{|M|} \right) \\ &= \ln \Big(|M^{-1}| |M + \delta M| \Big) \\ &= \ln \Big(|M^{-1} M + M^{-1} \delta M| \Big) \\ &= \ln \Big(|\mathbb{1} + M^{-1} \delta M| \Big) \end{aligned}$$

We now prove another small identity:

$$\ln |M| = \operatorname{Tr} \ln(M)$$

Since M is diagonalisable, we can write the following and then proceed:

$$\begin{aligned}
M &= PDP^{-1} \\
\Rightarrow \det M &= \det\left(PDP^{-1}\right) = \det(D) = \prod_i \lambda_i \\
\Rightarrow \ln \det M &= \ln\left(\prod_i \lambda_i\right) = \sum_i \ln \lambda_i \\
&= \operatorname{Tr}\left(\begin{bmatrix} \ln\lambda_1 & & \\ & \ln\lambda_2 & \\ & & \ddots \end{bmatrix}\right) \\
&= \operatorname{Tr}\left(P^{-1}P\begin{bmatrix} \ln\lambda_1 & & \\ & \ln\lambda_2 & \\ & & \ddots \end{bmatrix}\right) \\
&= \operatorname{Tr}\left(P\begin{bmatrix} \ln\lambda_1 & & \\ & \ln\lambda_2 & \\ & & \ddots \end{bmatrix}P^{-1}\right) \\
&= \operatorname{Tr}(\ln M)
\end{aligned}$$

So using this identity in the previous result, we have:

$$\begin{aligned}
\ln\left(|\mathbb{1} + M^{-1}\delta M|\right) &= \operatorname{Tr}\ln\left(\mathbb{1} + M^{-1}\delta M\right) \\
&= \operatorname{Tr}\left(M^{-1}\delta M - \frac{1}{2}(M^{-1})(\delta M)(M^{-1})(\delta M) + \dots\right) \\
&= \operatorname{Tr} M^{-1} \times \delta M + \dots
\end{aligned}$$

We somewhat proved this identity[1]. Now, we take the case when $M = g$, then we get:

$$\partial_\lambda \ln |g| = \operatorname{Tr}\left(g^{-1}\partial_\lambda g\right) = \left(g^{-1}\partial_\lambda g\right)^\rho{}_\rho = g^{\rho\sigma}\partial_\lambda g_{\sigma\rho} = 2\Gamma^\mu{}_{\mu\lambda}$$

Thus we get:

$$\Gamma^\mu{}_{\mu\lambda} = \frac{1}{2}\partial_\lambda \ln|g| = \frac{1}{2|g|}\partial_\lambda |g| = \frac{1}{\sqrt{|g|}}\partial_\lambda\left(\sqrt{|g|}\right)$$

Then we have a cute expression for the covariant divergence:

$$A^\mu{}_{;\mu} = \partial_\mu A^\mu + \frac{1}{\sqrt{|g|}}\partial_\lambda\left(\sqrt{|g|}\right)A^\lambda$$

We are not done yet, we can make it even cuter...using the product rule, we have:

$$\partial_\lambda(\sqrt{|g|}A^\lambda) = \sqrt{|g|}\partial_\lambda(A^\lambda) + \partial_\lambda(\sqrt{|g|})$$

Substituting this in the above expression, we get:

$$A^\mu{}_{;\mu} = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}A^\mu\right)$$

[1] This derivation is given in Weinberg's book of Cosmology and Gravitation

Let us check for the spherical polar coordinates where we had seen $g \equiv \mathrm{diag}(1, r^2, r^2 \sin\theta) \implies \sqrt{|g|} = r^2 \sin\theta$. Then we will have:

$$\begin{aligned}
A^{\mu}{}_{;\mu} &= \frac{1}{r^2 \sin\theta}\left(\frac{\partial(r^2 \sin\theta A^1)}{\partial r} + \frac{\partial(r^2 \sin\theta A^2)}{\partial\theta} + \frac{\partial(r^2 \sin\theta A^3)}{\partial\phi}\right) \\
&= \frac{1}{r^2 \sin\theta}\left(\sin\theta\frac{\partial(A^1 r^2)}{\partial r} + r^2\frac{\partial(A^2 \sin\theta)}{\partial\theta} + r^2\sin\theta\frac{\partial(A^3)}{\partial\phi}\right) \\
&= \frac{1}{r^2}\frac{\partial(A^1 r^2)}{\partial r} + \frac{1}{\sin\theta}\frac{\partial(A^2\sin\theta)}{\partial\theta} + \frac{\partial(A^3)}{\partial\phi}
\end{aligned}$$

Now, note that we had earlier done said something about ordinary vectors and how they relate to contravariant vector: $A^\mu = \frac{\widetilde{A_\mu}}{h_\mu}$ and we had also seen the relation between them in spherical coordinates. Using that we have:

$$\begin{aligned}
A^{\mu}{}_{;\mu} &= \frac{1}{r^2}\frac{\partial(\widetilde{A}_r r^2)}{\partial r} + \frac{1}{\sin\theta}\frac{\partial(\frac{\widetilde{A}_\theta}{r}\sin\theta)}{\partial\theta} + \frac{\partial(\frac{\widetilde{A}_\phi}{r\sin\theta})}{\partial\phi} \\
&= \frac{1}{r^2}\frac{\partial(\widetilde{A}_r r^2)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(\widetilde{A}_\theta\sin\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\frac{\partial(\widetilde{A}_\phi)}{\partial\phi}
\end{aligned}$$

Lol, this is the formula for normal divergence in spherical polar coordinates that we had been studying and so this new 'cute' formula makes sense! Let us now check for the Laplacian and curl too.

6.5.2. Covariant Laplacian

We calculate the following [1] for a scalar function ϕ:

$$D_\mu D^\mu \phi = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}\partial^\mu\phi\right)$$

Well, note that $\partial^\mu\phi = g^{\mu\nu}\partial_\nu$ and since the metric tensor is diagonal in this case, only $g^{rr} = 1, g^{\theta\theta} = \frac{1}{r^2}, g^{\phi\phi} = \frac{1}{r^2\sin^2\theta}$ will contribute. We then substitute it in the above:

$$\begin{aligned}
\frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}\partial^\mu\phi\right) &= \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}g^{\mu\nu}\partial_\nu\phi\right) \\
&= \frac{1}{r^2\sin\theta}\left(\frac{\partial}{\partial r}\left(r^2\sin\theta\frac{\partial\phi}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(r^2\sin\theta\frac{1}{r^2}\cdot\frac{\partial\phi}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(r^2\frac{1}{r^2\sin^2\theta}\cdot\frac{\partial\phi}{\partial\phi}\right)\right) \\
&= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\phi}{\partial\phi^2}
\end{aligned}$$

Wasn't this nice and simple, to derive the form of the Laplacian which has scared us for so long because it just looked ghastly and seemingly popped out of nowhere?

6.5.3. A Thing or Two about Levi-Civita

Curls and cross-products cannot be done without the mention of Levi-Civita symbol. So, let us see a few things about that thing first. We all know the famous ϵ_{ijk}, let us generalise this to higher number of indices and write it with a tilde. So, what we have is:

$$\widetilde{\epsilon}_{\mu_1\mu_2\ldots\mu_n} = \begin{cases} +1, \text{If } \mu_1\mu_2\ldots\mu_n \text{is even permutation of } \{0,1,2,\ldots,(n-1)\} \\ -1, \text{If } \mu_1\mu_2\ldots\mu_n \text{is odd permutation of } \{0,1,2,\ldots,(n-1)\} \\ +0, \text{otherwise} \end{cases}$$

[1] Note that the components of the covariant derivative of a scalar function are just the partial derivatives

We will call this object the Levi-Civita 'symbol', specifically, since this is not a tensor. Now note one identity:

Identity 2:

$$\tilde{\epsilon}_{\mu'_1\ldots\mu'_n}\,\det(M) = \tilde{\epsilon}_{\mu_1\ldots\mu_n} M^{\mu_1}{}_{\mu'_1} M^{\mu_2}{}_{\mu'_2} \ldots M^{\mu_n}{}_{\mu'_n}$$

Now take the matrix $M \equiv \frac{\partial x}{\partial x'}$, then we have from the above formula:

$$\tilde{\epsilon}_{\mu'_1\ldots\mu'_n}\left|\frac{\partial x}{\partial x'}\right| = \tilde{\epsilon}_{\mu_1\ldots\mu_n}\frac{\partial x^{\mu_1}}{\partial x_{\mu'_1}}\frac{\partial x^{\mu_2}}{\partial x_{\mu'_2}}\ldots\frac{\partial x^{\mu_n}}{\partial x_{\mu'_n}} \implies \tilde{\epsilon}_{\mu'_1\ldots\mu'_n} = \left|\frac{\partial x'}{\partial x}\right|\tilde{\epsilon}_{\mu_1\ldots\mu_n}\frac{\partial x^{\mu_1}}{\partial x_{\mu'_1}}\frac{\partial x^{\mu_2}}{\partial x_{\mu'_2}}\ldots\frac{\partial x^{\mu_n}}{\partial x_{\mu'_n}}$$

We just took the determinant in the denominator in the right hand side and then took the inverse matrix, since inverse of the determinant of a matrix is the determinant of its inverse. [1]However, note that the matrix is just the Jacobian and hence the Levi-Civita symbol is a *tensor density* of weight $+1$. Now, also remember that we had previously seen:

$$\mathrm{J} = \sqrt{\frac{|g|}{|g'|}}$$

Then, we will have:

$$\tilde{\epsilon}_{\mu'_1\ldots\mu'_n} = \sqrt{\frac{|g|}{|g'|}}\tilde{\epsilon}_{\mu_1\ldots\mu_n}\frac{\partial x^{\mu_1}}{\partial x_{\mu'_1}}\frac{\partial x^{\mu_2}}{\partial x_{\mu'_2}}\ldots\frac{\partial x^{\mu_n}}{\partial x_{\mu'_n}} \implies \sqrt{|g'|}\tilde{\epsilon}_{\mu'_1\ldots\mu'_n} = \sqrt{|g|}\tilde{\epsilon}_{\mu_1\ldots\mu_n}\frac{\partial x^{\mu_1}}{\partial x_{\mu'_1}}\frac{\partial x^{\mu_2}}{\partial x_{\mu'_2}}\ldots\frac{\partial x^{\mu_n}}{\partial x_{\mu'_n}}$$

Thus the quantity $\epsilon_{\mu_1\ldots\mu_n} = \sqrt{|g|}\tilde{\epsilon}_{\mu_1\ldots\mu_n}$ transforms as a tensor and we can do all the up and down game with the indices. This we call as **Levi-Civita tensor**. We also sometimes define the Levi-Civita symbol as $\tilde{\epsilon}^{\mu'_1\cdots\mu'_n}$ which is of weight -1[2] and then we get $\epsilon^{\mu_1\cdots\mu_n} = \frac{1}{\sqrt{|g|}}\tilde{\epsilon}^{\mu_1\cdots\mu_n}$.

Most of the time, we contract some of the indices of the Levi-Civita tensor and we obtain some Kronecker deltas. We have an identity for contracting p such indices:

$$\epsilon^{\mu_1\mu_2\cdots\mu_p\alpha_1\cdots\alpha_{n-p}}\epsilon_{\mu_1\mu_2\cdots\mu_p\beta_1\cdots\beta_{n-p}} = (-1)^s\,p!(n-p)!\,\delta^{[\alpha_1\ldots\alpha_{n-p}]}_{\beta_1\ldots\beta_{n-p}}$$

Now, wtf does the right hand side denote? Here s is the number of negative eigenvalues of the metric tensor but what about the weird delta symbol? Carroll writes it as an anti-symmetrised product of Kronecker deltas. We elaborate it a bit further. For any tensor T, the notation $T^{[\mu_1\cdots\mu_n]}$ is equal to:

$$T^{[\mu_1\ldots\mu_n]} = \frac{1}{n!}\sum_{\pi\in S_n}\mathrm{sgn}(\pi)T^{\mu_{\pi(1)}\mu_{\pi(2)}\cdots\mu_{\pi(n)}}$$

Well, did I make it more complicated? Perhaps, but complications lead to clarifications. S_n is the permutation group of order n, $\mathrm{sgn}(\pi)$ denotes the sign of the permutation and is positive if it is obtained from even number of exchanges and negative otherwise. Let us see for S_3, the group is then:

$$S_3 \equiv \{\underbrace{\{1,2,3\}}_{\pi_1},\underbrace{\{1,3,2\}}_{\pi_2},\underbrace{\{2,1,3\}}_{\pi_3},\underbrace{\{2,3,1\}}_{\pi_4},\underbrace{\{3,1,2\}}_{\pi_5},\underbrace{\{3,2,1\}}_{\pi_6}\}$$

[1]By the way, I really like these kind of reciprocal sentences. It's a cool thing about mathematics! Another example may be, like sum of trace is the trace of sum.

[2]This is numerically equal to $\mathrm{sgn}(g)\tilde{\epsilon}_{\mu_1\ldots\mu_n}$

Then we can check that we have $\mathrm{sgn}(\pi_i) = +1$ for $i = 1, 4, 5$ and -1 for $i = 2, 3, 6$ So, similarly the Kronecker delta can be written as:

$$\delta^{[\alpha_1 \ldots \alpha_{n-p}]}{}_{\beta_1 \ldots \beta_{n-p}} = \frac{1}{(n-p)!} \sum_{\pi \in S_n} \mathrm{sgn}(\pi) \prod_i \delta^{\alpha_{\pi(i)}}{}_{\beta(i)}$$

And now using Leibniz's formula for determinant[1], we have:

$$\delta^{[\alpha_1 \ldots \alpha_{n-p}]}{}_{\beta_1 \ldots \beta_{n-p}} = \frac{1}{(n-p)!} \det(\delta)$$

where,

$$\delta \equiv \begin{pmatrix} \delta^{\alpha_1}_{\beta_1} & \cdots & \delta^{\alpha_{n-p}}_{\beta_1} \\ \vdots & \ddots & \vdots \\ \delta^{\alpha_1}_{\beta_{n-p}} & \cdots & \delta^{\alpha_{n-p}}_{\beta_{n-p}} \end{pmatrix}$$

Then the final expression that we have is:

$$\epsilon^{\mu_1 \mu_2 \cdots \mu_p \alpha_1 \cdots \alpha_{n-p}} \epsilon_{\mu_1 \mu_2 \cdots \mu_p \beta_1 \cdots \beta_{n-p}} = (-1)^s p! \det\{\delta\}$$

Well, let us see one example, for this that we know. For this, we take $p = 1$ and $n = 3$, then $n - p = 2$. So, we have:

$$\epsilon^{ijk} \epsilon_{ilm} = (-1)^s \times 1! \times \det \begin{pmatrix} \delta^j_l & \delta^k_l \\ \delta^j_m & \delta^k_m \end{pmatrix} = (-1)^s \left(\delta^j_l \delta^k_m - \delta^j_m \delta^k_l \right)$$

Isn't this fun to finally see how this expression comes? Well, for me, yes but it is just an amusement for me, 'coz ultimately, to use it, we have to kinda remember it. Deriving things from scratch is a painful patch, with time to lose and little to match.

6.5.4. Covariant Curl

In our normal vector calculus, we had often written the curl to be $(\nabla \times A)^i = \epsilon_{ijk} \partial_i A^j$ (Note that I am not caring about any index position here, just placing them randomly). The natural generalisation is to replace the normal derivative with covariant derivative and the epsilon with the Levi-Civita tensor and proceed. Let us compute the i^{th} component of the curl, that is, $C^i \equiv \epsilon^{ijk} D_j A_k = \frac{1}{\sqrt{|g|}} \widetilde{epsilon}^{ijk} D_j A_k$. Note that in the right side only two terms will contribute for which $j, k \neq i$. Consider $j \to k$ is symmetric while $k \to j$ is anti-symmetric.

$$\begin{aligned} C^i &= \frac{1}{\sqrt{|g|}} \left(\tilde{\epsilon}^{ijk} D_j A_k + \tilde{\epsilon}^{ikj} D_k A_j \right) \\ &= \frac{1}{\sqrt{|g|}} \left(D_j A_k - D_k A_j \right) \\ &= \frac{1}{\sqrt{|g|}} \left(\partial_j A_k - \cancel{\Gamma^\nu_{jk} A_\nu} - \partial_k A_j + \cancel{\Gamma^\nu_{kj} A_\nu} \right) \\ &= \frac{1}{\sqrt{|g|}} \left(\partial_j A_k - \partial_k A_j \right) \end{aligned}$$

[1] The formula says that for any matrix A, we have:

$$\det(A) = \sum_{\tau \in S_n} \mathrm{sgn}(\tau) \prod_{i=1}^n a^i_{\tau(i)} = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a^i_{\sigma(i)}$$

The Christoffel symbols cancel because we assumed no torsion. Thus we have a cool expression for the curl. Let us see an example, with our dear spherical coordinates:

$$\begin{aligned}\widetilde{C}_r = C_r &= \frac{1}{r^2 \sin\theta}\left(\partial_\theta A_\phi - \partial_\phi A_\theta\right) \\ &= \frac{1}{r^2 \sin\theta}\Big(\partial_\theta (r\sin\theta \widetilde{A}_\phi) - \partial_\phi (r\widetilde{A}_\theta)\Big) \\ &= \frac{1}{r \sin\theta}\Big(\partial_\theta (\sin\theta \widetilde{A}_\phi) - \partial_\phi (\widetilde{A}_\theta)\Big)\end{aligned}$$

$$\begin{aligned}\widetilde{C}_\theta = rC_\theta &= \frac{r}{r^2 \sin\theta}\left(\partial_\phi A_r - \partial_r A_\phi\right) \\ &= \frac{1}{r \sin\theta}\Big(\partial_\phi (\widetilde{A}_r) - \partial_r (r\sin\theta \widetilde{A}_\phi)\Big) \\ &= \frac{1}{r}\left(\frac{1}{\sin\theta}\partial_\phi (\widetilde{A}_r) - \partial_r (r\widetilde{A}_\phi)\right)\end{aligned}$$

$$\begin{aligned}\widetilde{C}_\phi = r\sin\theta C_\phi &= \frac{r\sin\theta}{r^2 \sin\theta}\left(\partial_r A_\theta - \partial_\theta A_r\right) \\ &= \frac{1}{r}\Big(\partial_r (r\widetilde{A}_\theta) - \partial_\theta (\widetilde{A}_r)\Big)\end{aligned}$$

We had nicely found the components of the curl in spherical coordinates just by using the above formula which is way easy to remember and then using the conversion between 'ordinary' and covariant vector components.

7. Curvature: coz being flat is boring!

Let us start with the normal meaning of curvature. For a one dimensional curve, to actually see it is 'curved', we need to visualise it in two dimension. Similarly, for two dimensional surface, to see its curvature, we need to embed it in three dimensional space. So, for an n dimensional space, the curvature properties which require it to be embedded in a $n+1$ dimensions is called *extrinsic curvature*.

Now, imagine a 2D creature (someone from *Flatland* perhaps) whose entirely life revolves in a 2D space, who doesn't have a taste for any other dimensions, how can that being say whether it is living in a flat space or curved space? One thing it can do is, it can just join three points with shortest distance lines and then measure the sum of the internal angles of the triangle. If they do not sum to $180°$, then it is not a flatland. This is called intrinsic curvature property, since we are not leaving the space.

7.1. Parallel Transport

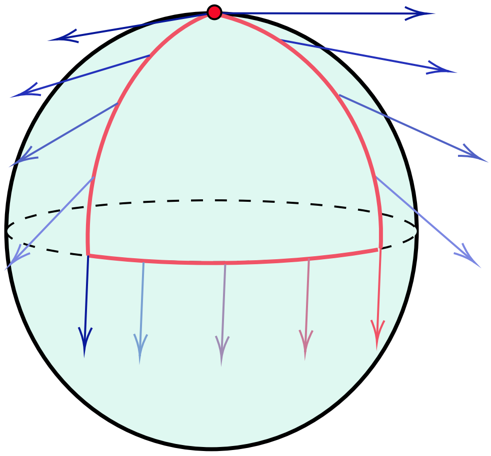


Figure 5: Parallel Transport of a vector on a sphere (pardon the distorted diagram tho!). We can observe that a vector starting from a point on the sphere, after moving along the closed loop, does not remain the same vector. This is an important indication of curvature of space.

7.2. Riemann Curvature Tensor

8. Bit of Differential Geometry: Dayuum!!

Let us look into a bit of differential geometry which is a formal way of treating this tensor thingy. We will try to be as intuitive and non-rigorous as possible (and thus increasing our chances of making a mathematician crazy!) but yeah, we will try to be rigorous enough so that I am satisfied.

8.1. Some prior things

Before touching manifolds, let us define what an abstract topological space is, since manifolds are special case of topological spaces.

Definition 1 (Topological Space):
A topological space is a set (X, τ) where $\tau \subset \mathcal{P}$ is a collection of subsets of X such that:

- $\emptyset, X \in \tau$.
- $U_\alpha \in \tau \implies \bigcup_{\alpha \in J} U_\alpha \in \tau$ (closed under arbitrary union).
- $U_i \in \tau \implies \bigcap_{i=1}^{n} U_i \in \tau$ (closed under finite intersection)[1].

2

Well well, this does not look anything like coffee cup and donut which most people associate topology with. That is a case of *homeomorphism* which will be discussed later (hopefully). However, for now let us proceed. The sets belonging to τ are called **open sets**. We define a **closed set** as a set whose complement is open. There are umpteen other definitions like **closure, boundary, interior, neighbourhood**, etc. Let define few of them 😭.

[2]Here α index is used when we want the indexing set J (indexing set means the set from where the incides to denote the elements of the set are taken from) to be arbitrary, meaning that the set $\{U_\alpha\}$ can be finite, countable or uncountable. On the other hand, index i is mostly used when the indexing set is finite.

- *Closure* of a set A is the smallest closed set containing A and is denoted by $\overline{A}$.
- *Interior* of a set A is the largest open set contained in A and is denoted by int(A).
- *Boundary* of a set A is the set of points which are neither in the interior nor in the exterior of A and is denoted by ∂A.
- If $p \in X$, then a *neighbourhood* of p is a set N such that there exists an open set $U \in \tau$ with $p \in U \subseteq N$.

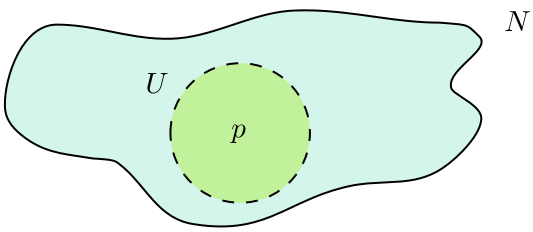


Figure 6: Neighbourhood of a point p in X

- *Hausdorff Space*: A topological space is called Hausdorff if for any two distinct points $x, y \in X$, there exist open sets $U, V \in \tau$ such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

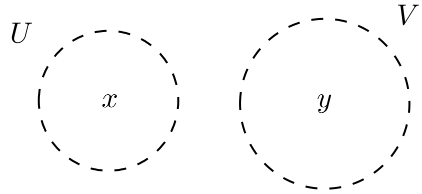
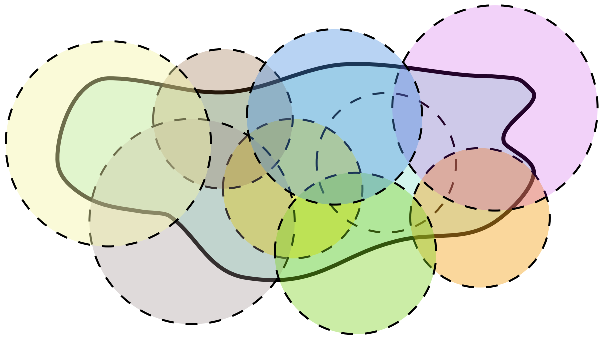


Figure 7: A **Hausdorff Space**, where the points x and y are separated by the open sets U and V.

- *Topological Continuity*: A function $f : X \to Y$ between two topological spaces is said to be continuous if for every open set $V \in \tau_Y$, the preimage $f^{-1}(V)$[1] $\in \tau_X$ is open in X.
- *Homeomorphism*: A homeomorphism is a bijective function $f : X \to Y$ between two topological spaces such that both f and its inverse f^{-1} are continuous. If such a function exists, we say that the two spaces are **homeomorphic** and we write $X \cong Y$.
- *Cover:* A cover of a topological space X is a collection of sets $\{U_\alpha\}$whose union is X that is, $\bigcup\limits_\alpha U_\alpha = X$. If each set is an open set, then it is called an **open cover**. If there exists a finite collection of subsets of the cover such that their union is X, that is, $\bigcup\limits_{i=1}^{k} U_i = X$ then it is called a **finite subcover**.

[1]The preimage of a set Y under the function f is defined as $f^{-1}(Y) = \{x | f(x) \in Y\}$. Note that this has nothing to do with inverse of a function (sadly we use the same notation)



$\{U_\alpha\}$

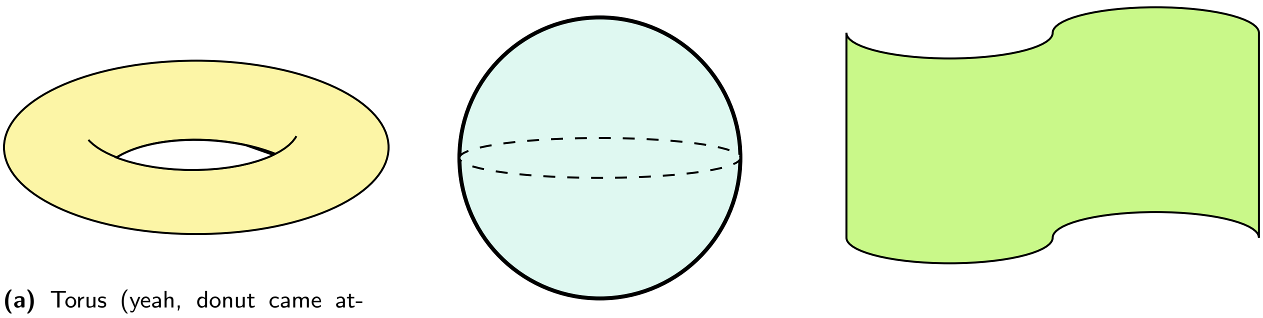
Figure 8: Open cover of a space

- *Subspace Topology* is the topology on a subset $Y \subseteq X$ induced by the topology of X. In this case, open sets of Y are basically the intersection of open sets of X with Y. So,

$$\tau_Y = \{U \cap Y | U \in \tau_X\}$$

8.2. Manifolds

Let us see some pictures.



(a) Torus (yeah, donut came at-last!)

(b) Sphere (like yo mama)

(c) A Waving Flag perhaps?

Figure 9: What's common in all these?

So, what is common in all these pictures? Note that they all look very different from each other but if we really ZOOM in we can see that each of them look alike, like a *flat plane*. Well, the road ahead of us looks flat but the road is on the freaking Earth which is, let's say to a physicist's satisfaction, a sphere. So, we can say that all of these things look 'locally' like the flat plane $\mathbb{R}^2$. This is essentially the idea behind a **manifold**, things which look locally Euclidean (like $\mathbb{R}^n$). Let us define manifolds formally:

Definition 2 (Differentiable Manifold):
$\mathcal{M}$ is a m-dimensional differentiable manifold if:

- $\mathcal{M}$ is a topological space.
- There exists an open cover $\{U_\alpha\}$ of $\mathcal{M}$ and for each α, there exists a homeomorphism $\phi_\alpha : U_\alpha \to V_\alpha$ where V_α is an open subset of $\mathbb{R}^m$.
- Two open sets U_i, U_j such that $U_i \cap U_j \neq \emptyset$, then the map $\psi_{ij} = \phi_i \circ \phi_j^{-1}$ is a smooth map [1], where $\psi_{ij} : \phi_j(U_i \cap U_j) \to \phi_i(U_i \cap U_j)$.

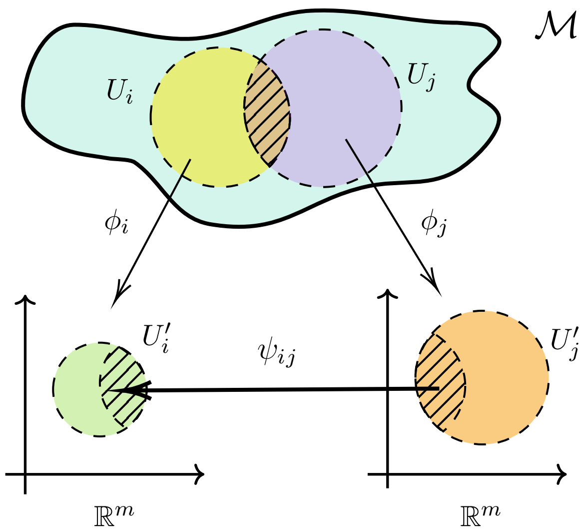


Figure 10: The figure shows the third point in the definition. So basically, we note where the homeomorphisms map the intersection of the open sets and then define the map ψ_{ij} between these two regions.

A point $p \in U \subset \mathcal{M}$ has the coordinates $(\phi^1(p), \ldots, \phi^m(p)) \in \mathbb{R}^m$ with respect to the chart (U, ϕ). The coordinate functions ϕ^i are defined in terms of the projection map $\text{proj}_i : \mathbb{R}^m \longrightarrow \mathbb{R}$ such that $\text{proj}_i(x) \mapsto x^i$[2] and $\phi^\mu : \mathcal{M} \to \mathbb{R}$ such that $p \mapsto \text{proj}_\mu(\phi(p))$. So basically ϕ sends p to some point in $\mathbb{R}^m$ and then to find the individual components, we take the projection of $\phi(p)$ along that coordinate. Henceforth we will use r^i to denote proj_i (cause brevity is the soul of wit!) and also, to be familiar with our common sense, we will denote the coordinate maps ϕ^i with x^i, which actually gives us the feel of some coordinates.

The terminologies used here are very much related to the geography of Earth. The pair (U_i, ϕ_i) is called a *chart* (maybe because they help us to locally "chart" the manifold, that is, understand it using some coordinates) and the collection of all charts is called an *atlas* (well, because it is collection of maps).

Now let us unfold this carefully. For the open set U_i, the map ϕ_i takes it to another open set in $\mathbb{R}^m$. So for all $x \in U_i$ we got a mapping to an Euclidean space. Same goes for U_j, that is we obtain

[1] A smooth map is a function which is infinitely differentiable, that is, all the derivatives exist and are continuous. Sometimes a smooth map f is said to belong to the class C^∞. In general, C^k is the class of functions which are k times continuously differentiable. We will consider mostly 'smooth' things here, so we will always in general take k to be ∞

[2] This just selects the i^{th} component of a m-dimensional vector x in $\mathbb{R}^m$

a mapping into another copy of $\mathbb{R}^m$. Now for points in the intersection of U_i and U_j, we have got two different mappings and we can go back and forth between the two copies of $\mathbb{R}^m$ since these mappings were homeomorphisms. This is what we had with the mapping ψ_{ij} (and thus, these are aptly called *transition functions*). It first maps with the inverse of ϕ_j and then applies ϕ_i. The net effect is that we are mapping between a point in one copy of $\mathbb{R}^m$ to another copy of $\mathbb{R}^m$.

Imagine the open sets as patches in the manifold. We can then patch together the whole manifold by taking the union of all the open sets, all of which can be viewed as a Euclidean space. There are also manifolds which do not have the smooth property on the transient function, only continuity is there. These are called in general **topological manifolds**.

] We also assume that our manifolds are **Hausdorff** and paracompact (which we will not define here).

Fun Fact:
For a differentiable manifold $\mathcal{M}$, the set of smooth real-valued functions on it $C^\infty(\mathcal{M})$ is a *ring*

8.2.1. Examples:

Now comes the good thing: examples! ☺ **The Space $\mathbb{R}^{\mathbf{n}}$**

Duh, it looks like $\mathbb{R}^n$ locally since it is $\mathbb{R}^n$ itself. A single chart is enough for the purpose and the homeomorphism is the identity map.

The Circle $\mathbb{S}^1$

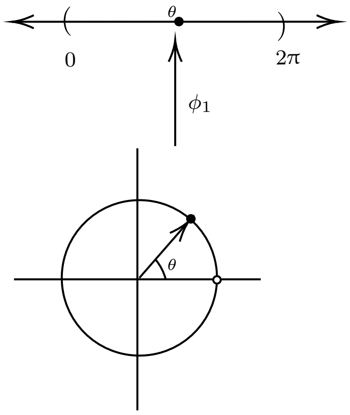
Circle is a curve in $\mathbb{R}^2$ with coordinates $(\cos\theta, \sin\theta)$. We mostly take $\theta \in [0, 2\pi)$ but we come across a problem. Note that the open sets on a circle are basically union of "open arcs". However, $[0, 2\pi)$ is not open. Thus we need atleast two charts to cover the circle.

We take two antipodal points on the circle and then define the charts as follows:

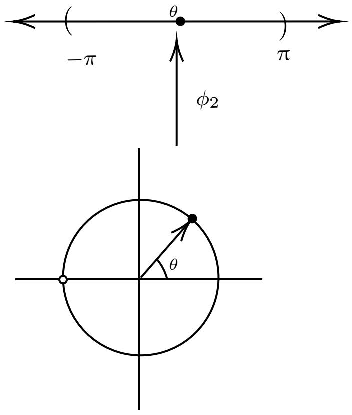
Let $U_1 = \mathbb{S}^1 \backslash \{(1,0)\}, U_2 = \mathbb{S}^1 \backslash \{(-1,0)\}$. Then define the homeomorphisms as:

$$\phi_1 : \mathbb{S}^1 \backslash \{(1,0)\} \to (0, 2\pi) \qquad \phi_2 : \mathbb{S}^1 \backslash \{(-1,0)\} \to (-\pi, \pi)$$

These functions basically take the value of the angle θ that a point on the circle makes with the x-axis.



(a) Here the point $(1, 0)$ is removed and ϕ_1 maps the rest of the circle to the interval $(0, 2\pi)$.



(b) Here the point $(-1, 0)$ is removed and ϕ_2 maps the rest of the circle to the interval $(-\pi, \pi)$.

Figure 11: ϕ_1 and ϕ_2 are invertiable and continuous (easily seen). The two charts intersect in the upper and lower semicircles (as the antipodal points are removed). The transition function is given by: $\phi_1 \circ \phi_2^{-1}(\theta) = \begin{cases} \theta & \text{if } \theta \in (0, \pi) \\ \theta + 2\pi & \text{if } \theta \in (-\pi, 0) \end{cases}$ which is *smooth* on each of the semicircles as required. Thus this is a valid chart for the circle.

The same circle can be described by another chart using the stereographic projection, resulting in the **Mercator Atlas**. For this let us consider $U_1 = \mathbb{S}\backslash(0, 1)$ and $U_2 = \mathbb{S}\backslash(0, -1)$. Then what we do is, we define maps $\phi_1 : U_1 \to \mathbb{R}$ ($\phi_2 : U_2 \to \mathbb{R}$) which maps a point $p \in U_1(U_2)$ to some point on the x-axis. This is called a stereographic projection. Let us find explicit forms for these maps:

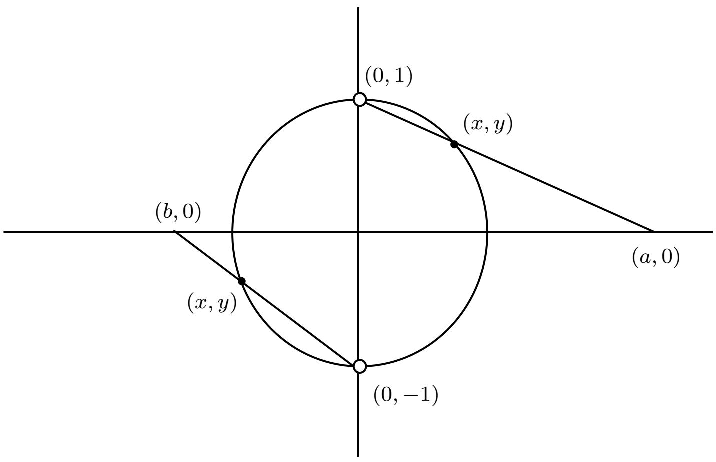


Figure 12: Stereographic projection of the circle

In the above picture, the equation of the straight line for ϕ_1 is given by:

$$\frac{y-1}{x-0} = \frac{0-y}{a-x}$$

from where we obtain $a = \frac{x}{1-y}$. Similarly, we obtain $b = \frac{x}{1+y}$. Then what we have is $(x, y) \mapsto \frac{x}{1-y}$ (we take only the x-component since y-component is always 0 and thus the codomain is a real number) for ϕ_1 and $(x, y) \mapsto \frac{x}{1+y}$ for ϕ_2. Suppose we are given a point $(t, 0) \in \mathbb{R} \times \{0\}$, and we want to compute the inverse image under the stereographic projection φ_1^{-1}.

We are given that $\varphi_1(x,y) = \dfrac{x}{1-y} = t$. Solving for x, we get:

$$x = t(1-y)$$

Substitute this into the unit circle equation $x^2+y^2=1$:

$$\begin{aligned} x^2+y^2 &= 1 \\ \implies t^2(1-y)^2+y^2 &= 1 \\ \implies t^2(1-2y+y^2)+y^2 &= 1 \\ \implies t^2-2t^2y+t^2y^2+y^2 &= 1 \\ \implies (1+t^2)y^2-2t^2y+(t^2-1) &= 0 \end{aligned}$$

This is a quadratic in y. Solving it gives:

$$y = \frac{2t^2 \pm \sqrt{4}}{2(1+t^2)} = \frac{t^2 \pm 1}{t^2+1}$$

So the two roots are:

$$y = 1, \quad y = \frac{t^2-1}{t^2+1}$$

Since $y=1$ corresponds to the north pole (which is excluded from the domain of φ_1), we discard it. Thus, we have:

$$y = \frac{t^2-1}{t^2+1}$$

Now substitute back to find x:

$$x = t(1-y) = t\left(1-\frac{t^2-1}{t^2+1}\right) = t\left(\frac{(t^2+1)-(t^2-1)}{t^2+1}\right) = t\left(\frac{2}{t^2+1}\right) = \frac{2t}{t^2+1}$$

Hence, the inverse map $\varphi_1^{-1} : \mathbb{R} \to S^1 \setminus \{(0,1)\}$ is given by:

$$\boxed{\varphi_1^{-1}(t) = \left(\frac{2t}{1+t^2}, \frac{t^2-1}{1+t^2}\right)}$$

Note that this map is continuous and hence this is a homeomorphism. Similarly, we can prove it for ϕ_2 also. Now, if we take a point from the intersection, $U_1 \bigcap U_2$ (that is, the circle without the poles), then we can never obtain the point $(0,0)$ since both maps give $(0,0)$ only when one of the input is along the y-xis. Thus, $\phi_1(U_1 \bigcap U_2) = \mathbb{R}\backslash 0, \phi_2(U_1 \bigcap U_2) = \mathbb{R}\backslash 0$. Now let's see the transition map:

$$\phi_2 \circ \phi_1^{-1} = \phi_2\left(\frac{2t}{1+t^2}, \frac{t^2-1}{t^2+1}\right) = \frac{\frac{2t}{t^2+1}}{\frac{t^2-1}{t^2+1}-1} = \frac{2t}{2t^2} = \frac{1}{t}$$

which is smooth on $\mathbb{R}\backslash 0$. Similarly for the other case also, we can show. Thus $\{(U_1,\varphi_1),(U_2,\varphi_2)\}$ for an atlas for the circle and is called the Mercator Atlas. Similarly we can prove that the n-dimensional sphere is a manifold.

Product Manifold:

Let $\mathcal{M}$ be an m-dimensional manifold with atlas $\{U_i,\varphi_i\}$ and $\mathcal{N}$ be an n-dimensional manifold with atlas $\{V_i,\psi_i\}$, then we define the product manifold as an $(m+n)-$ dimensional manifold with atlas $\{U_i \times V_j, (\phi_i,\psi_j)\}$. So basically we took the Cartesian product of the two coordinate neighbourhoods to define the new neighbourhood and then used the ordered pair of the two homeomorphisms to define the new coordinate function.

Example. The torus is a product manifold of two circles, that is, $T^2 = \mathbb{S}^1 \times \mathbb{S}^1$. Note that by our definition, it is a two $(1+1)$ dimensional manifold. We can generalise the notion of torus by taking multiple product of circles:

$$T^n = \underbrace{\mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{n}$$

8.2.2. Differentiable Maps

Definition 3 (Differentiable Map):
A map $f : \mathcal{M}_m \to \mathcal{N}_n$ between two manifolds is *differentiable* at p if for any charts (U, ϕ) on $\mathcal{M}$ and (V, ψ) on $\mathcal{N}$ (where $p \in U$ and $f(p) \in V$), the map

$$\psi \circ f \circ \phi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$$

is a smooth map, that is, the map belongs to C^∞, that is, infinitely continuously differentiable.

WTH...😟 what is this weird map that we want to be smooth? Let us see in more details. So, what does f do? It is a map from manifold $\mathcal{M}$ to manifold $\mathcal{N}$ and it maps $p \mapsto f(p)$. So far good...Now, we know that we can obtain coordinate representations using the homeomorphisms. Let them be:

$$\phi(p) \equiv \{x^\mu\} \qquad \psi(f(p)) \equiv \{y^\alpha\}$$

So what $\psi \circ f \circ \phi^{-1}$ does is, it takes a vector from $\mathbb{R}^m$, sends it back to the manifold $\mathcal{M}$, applies f to it to send it to manifold $\mathcal{N}$ and then sends it to $\mathbb{R}^n$, so finally we obtain a n element output from m element input, using f. So this is just the usual map like what we write $\mathbf{y} = f(\mathbf{x})$. Lol, this is nothing but the differentiability of a function, albeit in a more stylish (and appropriate) way.

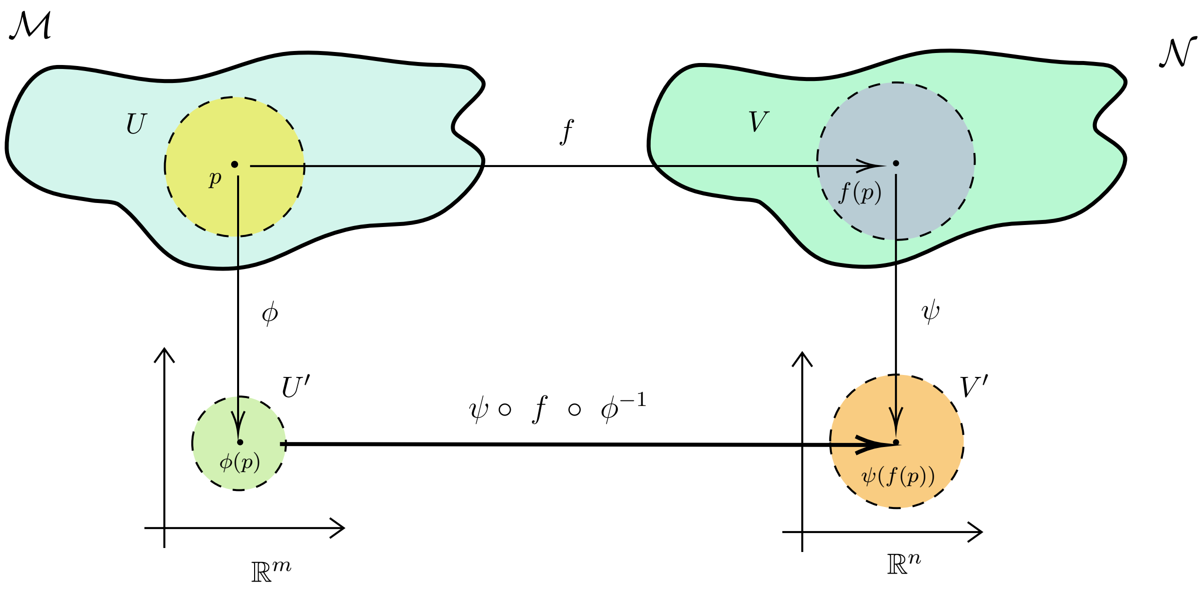


Figure 13: Representation of a differentiable map

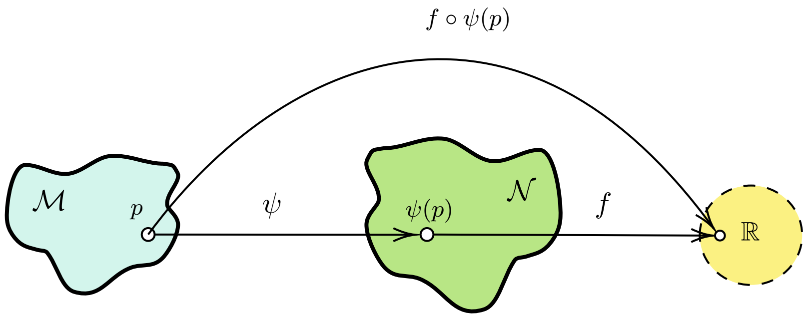
Next to come to another *-morphism* related to the differentiable manifolds.

Definition 4 (Diffeomorphism):
Let $f : \mathcal{M}_m \to \mathcal{N}_n$ be a homeomorphism and ψ and ϕ be the coordinate functions as defined before. If $\psi \circ f \circ \phi^{-1}$ is invertible and both $\psi \circ f \circ \phi^{-1}$ and its inverse $\phi \circ f^{-1} \circ \psi^{-1}$ are smooth maps, then f is called a *diffeomorphism*. The manifolds are then said to be diffeomorphic to each other.

Well, we had earlier seen that homeomorphisms characterise spaces which can be 'continuously' deformed into each other. Diffeomorphism does one thing extra, it characterises spaces which are transformed 'smoothly' into each other. Evidently, a diffeomorphism is also a homeomorphism. If two spaces a diffeomorphic, then their dimensions are same, that is $\dim \mathcal{M} = \dim \mathcal{N}$

The set of diffeomorphisms $f : \mathcal{M} \to \mathcal{M}$ is a group and is denoted by $\mathrm{Diff}(\mathcal{M})$.

8.2.3. Pullback of a function



Given a smooth map $\psi : \mathcal{M} \to \mathcal{N}$ and $f \in C^\infty(\mathcal{N})$, the pullback of f along ψ is defined to be:

$$\psi^* f \equiv f \circ \psi : \mathcal{M} \to \mathbb{R}$$

Since f and ψ both are smooth maps, their composition (which is a real-valued function) is too! Thus, ψ^* can be thought of as a map from $C^\infty(\mathcal{N}) \to C^\infty(\mathcal{M})$. Thus, it sort of 'pulls back' to the original manifold from which ψ was defined. [1] Also, note the following properties of the pullback for two functions $f, g \in C^\infty(\mathcal{N})$:

- $\psi^*(f+g)(p) \equiv (f+g) \circ \psi(p) = f(\psi(p)) + g(\psi(p)) = \psi^* f(p) + \psi^* g(p)$
- $\psi^*(fg)(p) = (fg) \circ \psi(p) = f(\psi(p))g(\psi(p)) = (\psi^* f(p))(\psi^* g(p))$

8.2.4. Curves

Definition 5 (Curve):
An open curve in an m-dimensional manifold $\mathcal{M}$ is a map $\sigma : (a, b) \to \mathcal{M}$ such that $a < 0 < b$ (a and b can be $\pm\infty$ also). A closed curve is a map $\sigma : \mathbb{S}^1 \to \mathcal{M}$

What this means is that suppose we take $t \in (a, b)$, then $\sigma(t)$ is a point on the manifold. As t is varied in the interval, the points $\sigma(t)$ kinda resembles a trajectory (which we intuitively had called a 'curve' so far). We had included zero in the interval for convenience.

[1] This has something to do dual space and one-forms also, apparently, which maybe seen later.

8.2.5. Vectors

In our usual sense, we imagine vectors as straight arrows drawn from the origin but in manifolds, which is 'curved', firstly, straight arrows cannot be drawn in general and secondly, there is no origin from which the arrow can be drawn. If the arrow is 'small' enough we can locally obtain something like a "vector space". Suppose the tail of the arrow is at a point $p \in \mathcal{M}$ and the tip is at a point p' which is 'close' to p, then the vector space can be approximated by the *tangent space* at the point p.

8.3. Tangent Space

Intuitively, if we use the word tangent space, we can think of some space which contains the tangent at a point of the manifold. Like in the following diagram:

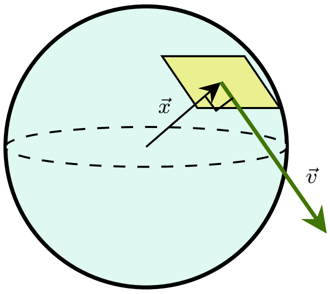


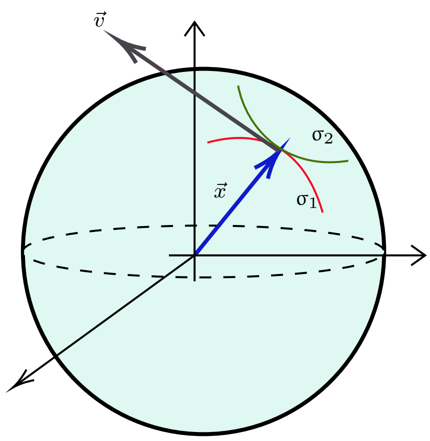
Figure 14: A very badly made representation of a tangent space. Here a sphere is considered and from the centre of the sphere, a vector $\mathbf{x}$ is drawn and then, a plane is drawn which just touches the sphere at point $\mathbf{x}$. This constitutes the tangent plane.

Note that since the tangent is always perpendicular to the vector $\mathbf{x}$, we can somewhat define the tangent space at the point $\mathbf{x} \in S^n \subset \mathbb{R}^{n+1}$ by:

$$\mathrm{T}_x S^n = \{\mathbf{v} \in \mathbb{R}^{n+1} | \mathbf{x} \cdot \mathbf{v} = 0\}$$

See an apparent problem with this: This approach depends on the sphere being embedded in a higher dimension vector space and the tangent space being a specific linear subspace of this larger space. It would be better to find some 'intrinsic' definition of tangent space which does not depend on this embedding of the manifold in a higher dimensional space.

For this, we can think of the tangent vector as a tangent to a curve on the manifold. The vector is tangent to a curve on the manifold at point $\mathbf{x}$. Note that this curve lies on the manifold and hence the notion of a higher-dimensional embedding is done away with.



However, note that, as in the above picture, many different curves can be drawn to satisfy this condition. Hence, the tangent vector can be rather thought as an equivalence class of all the curves satisfying this relation. 🤯

Definition 6 (Tangent):
Let $\mathcal{M}$ be a manifold and $p \in \mathcal{M}$. Then two curves σ_1 and σ_2 are tangent to the manifold at p if:

- $\sigma_1(0) = \sigma_2(0) = p$
- Given a chart (U, ϕ), the two curves are tangent in the usual way, that is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\phi \circ \sigma_1(t)\right)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\phi \circ \sigma_2(t)\right)\Big|_{t=0} \implies \frac{\mathrm{d}}{\mathrm{d}t}\left(\phi^i \circ \sigma_1(t)\right)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\phi^i \circ \sigma_2(t)\right)\Big|_{t=0} \quad \forall i = 1, \ldots, m$$

If σ_1 and σ_2 are 'tangent' in one system, then they are tanget in any other coordinate system local around p, and hence this definition is independent of any coordinates. We define the tangent vector at $p \in \mathcal{M}$ as the euivalence class of curves, that is:

$$[\sigma_i] = \{\sigma | \sigma \sim_t \sigma_i\}$$

where $\sim_t$ is the equivalence relation[1] satisfying that two curves are tangent to the manifold. Given σ_1, σ_2, either $[\sigma_1] = [\sigma_2]$ or $[\sigma_1] \bigcap [\sigma_2] = \emptyset$ (either same or disjoint). Thus, the equivalence classes form a 'partition' of the set of all curves. The tangent space $T_p\mathcal{M}$ to manifold $\mathcal{M}$ at point p, is the set of all tangent vectors at the point p.

8.3.1. Vector Space Structure on the Tangent Space

The intuitive idea of tangent space as a plane implies that vectors can be added and multiplied with some scalar, leading to a vector space structure. We now prove that the tangent space $T_p\mathcal{M}$ also carries a structure of a real vector space.

Theorem 2:
The tangent space $T_p\mathcal{M}$ carries a structure of a real vector space.

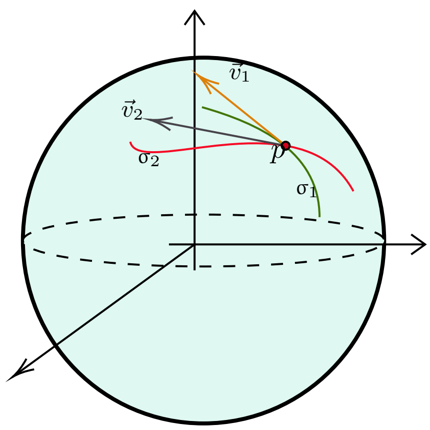


Figure 15: Two tangent vectors and their representative curves are shown on the manifold at point p.

[1]It can easily be shown to be reflexive, symmetric and transitive

In the above figure, we have two tangent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ at p with σ_1 and σ_2 as their representative curves (that is, $\mathbf{v}_1 = [\sigma_1]$ and $\mathbf{v}_2 = [\sigma_2]$). We use a local chart (U, ϕ) around p such that $\phi(p) = \mathbf{0} \in \mathbb{R}^m$.

Consider the maps $\phi \circ \sigma_1$ and $\phi \circ \sigma_2$. Since a curve σ maps an open interval to the manifold and the homeomorphism maps an open set in the manifold to an open set in $\mathbb{R}^m$, the above composite functions maps an open interval to an open set in $\mathbb{R}^m$.

Now, the addition $\phi \circ \sigma_1(t) + \phi \circ \sigma_2(t)$ is valid, since these are elements of $\mathbb{R}^m$ (actually, these are also curves in $\mathbb{R}^m$). Then we consider the map:

$$\phi^{-1} \circ (\phi \circ \sigma_1(t) + \phi \circ \sigma_2(t))$$

which is a curve back on the manifold and passes through p when $t = 0$. We then define:

1. $v_1 + v_2 := [\phi^{-1} \circ (\phi \circ \sigma_1 + \phi \circ \sigma_2)]$
2. $rv := [\phi^{-1} \circ (r\phi \circ \sigma)] \; \forall \; r \in \mathbb{R}$

It can then be shown that these definitions are independent of the charts and the representative curves, and using these, the tangent space can be shown to be a vector space.

8.3.2. Tangents as derivatives

We saw a geometric way of defining tangents using the tangent space approach. Now, we will define it algebraically. The directional derivative of a function f along a tangent vector v is defined as:

$$v(f) := \frac{\mathrm{d}}{\mathrm{d}t} f(\sigma(t))\bigg|_{t=0}$$

where $v = [\sigma]$. This definition helps us to view v as a differential operator on the space $C^\infty(\mathcal{M})$ of real-valued functions on the manifold.

Definition 7 (Derivations):
A derivation at a point $p \in \mathcal{M}$ is a map $v_p : C^\infty(\mathcal{M}) \to \mathbb{R}$ such that:

- $$\begin{aligned} v_p(f+g) &= v_p(f) + v_p(g) \; \forall \; f, g \in C^\infty(\mathcal{M}) \\ v_p(rf) &= r v_p(f) \; \forall \; f \in C^\infty(\mathcal{M}), r \in \mathbb{R} \end{aligned}$$
- $v_p(fg) = f(p)v_p(g) + g(p)v_p(f) \; \forall \; f, g \in C^\infty(\mathcal{M})$

The set of all derivations at p is denoted as $D_p\mathcal{M}$

The first condition tells us that v should be a linear map and the second condition (kind of like the product rule) is the actual thing for which we call it *derivation*. We can define:

$$\begin{aligned} (v_1 + v_2)(f) &:= v_1(f) + v_2(f) \; \forall v_1, v_2 \in D_p\mathcal{M} \\ (rv)(f) &:= rv(f) \end{aligned}$$

Also, note that:

$$\begin{aligned}(v_1+v_2)(fg) &= v_1(fg)+v_2(fg)\\ &= fv_1(g)+gv_1(f)+fv_2(g)+gv_2(f)\\ &= f(v_1+v_2)(g)+g(v_1+v_2)(f)\end{aligned}$$

This gives a natural vector space structure to the space of derivations $D_p\mathcal{M}$. Note that if we try to do the same analysis for $v_p \in D_p\mathcal{M}$ and $w_q \in D_q\mathcal{M}$, these $v_p + w_q$ will not satisfy the product rule and then will not be a valid vector. So, vectors at different point on a manifold cannot be added 🙄

We can also show that $T_p\mathcal{M}$ and $D_p\mathcal{M}$ are actually *isomorphic* and thus, both approaches can be taken to describe the tangent-space structure (However, the algebraic approach is a bit problematic for infinite-dimensional space). We shall henceforth use $T_p\mathcal{M}$ and $D_p\mathcal{M}$ as the same. We now define a set of derivations of a function f at p, with the help of a local chart (U,ϕ) where $f \in C^\infty$ and ϕ^μ are the coordinate components.

$$\left(\frac{\partial}{\partial x^\mu}\right)_p f := \frac{\partial}{\partial r^\mu} f\circ\phi^{-1}\bigg|_{\phi(p)} \qquad \mu = 1,2,\ldots,m$$

Note that in the right hand side, the partial derivative with respect to the coordinate is well defined, but in the left hand side, even though the notation is of partial derivative, the action is not well defined since the function is defined on the manifold which doesn't have the vector space structure required for definition of partial derivatives. Sometimes we use the notation $\frac{\partial f}{\partial x^\mu}$ directly. Then, using this notation and also the fact that $\phi^{-1}\circ\phi(p) = p$, we can write:

$$\left(\frac{\partial f}{\partial x^\mu}\right)(p) = \left(\frac{\partial f}{\partial x^\mu}\right)\left(\phi^{-1}\circ\phi(p)\right) = \frac{\partial}{\partial r^\mu}\left(f\circ\phi^{-1}\right)\phi(p)$$

Then we can write

$$\left(\frac{\partial f}{\partial x^\mu}\right)\circ\phi^{-1} \equiv \frac{\partial}{\partial r^\mu}\left(f\circ\phi^{-1}\right)$$

Fun Fact:
Tangent vector acting on a constant function is zero, that is, $v_p(c) = 0$

Proof. Suppose we have a constant real valued function $f \equiv c \in \mathbb{R}$. First, note that $v_p(1) = v_p(1\cdot 1) = 1v_p(1) + 1v_p(1) = 2v_p(1) \implies v_p(1) = 0$. Then we will have $v_p(c) = v_p(c\cdot 1) = cv_p(1) = 0$

Proposition 1:
Suppose (U,ϕ) is a chart with coordinates x^i. Then $\frac{\partial x^i}{\partial x^j} = \delta^i{}_j$

Proof. From the previous result, we have:

$$\frac{\partial x^i}{\partial x^j}(\phi(p)) = \frac{\partial x^i\circ x^{-1}}{\partial r^j}(\phi(p)) = \frac{\partial r^i\circ\phi\circ x^{-1}}{\partial r^j}(\phi(p)) = \frac{\partial r^i}{\partial r^j}(\phi(p)) = \delta^i{}_j$$

We just used the coordinate definition using the projection maps.

These quantitites form a basis set for the vector space $D_p\mathcal{M}$. For that we have to check for linear independence and span of these quantitites.

Proposition 2:
The set $\left\{\left(\frac{\partial}{\partial x^i}\right)_p\right\}$ is linearly independent.

Proof. $a^i\left(\frac{\partial}{\partial x^i}\right)_p = 0 \implies a^i\left(\frac{\partial x^j}{\partial x^i}\right)_p \implies a^i\delta^j{}_i = 0 \implies a^j = 0$

Proposition 3:
Every $v_p \in T_p\mathcal{M}$ can be written as a linear combination of $\left(\frac{\partial}{\partial x^i}\right)_p$

Proof. For the proof, we need to use the Mean-Value Theorem. For a continuously differentiable function $f : [a, b] \to \mathbb{R}$, $\exists\, c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a} \implies f(b) = f(a) + f'(c)(b - a)$.

For a multi-variable function, let $U \subset \mathbb{R}^n$ be an open set and $H : U \to \mathbb{R}$ be a continuously differentiable function. Let $\mathbf{a}, \mathbf{b} \in U$ and $l(\mathbf{a}, \mathbf{b}) \equiv \{(1 - t)\mathbf{a} + t\mathbf{b} | t \in [0, 1]\} \subset U$.

Then define $h : [0, 1] \to \mathbb{R}$ such that $t \mapsto H((1 - t)\mathbf{a} + t\mathbf{b}) = H(\mathbf{a} + t(\mathbf{b} - a))$.

Note that $h(0) = H(\mathbf{a}), h(1) = H(\mathbf{b})$ and h is continuously differentiable. Also,

$$h'(t) = \frac{\mathrm{d}}{\mathrm{d}t}H(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) = (b^i - a^i)\left.\frac{\partial H}{\partial r^i}\right|_{a+t(b-a)}$$

Now, we can apply the MVT for $\mathbb{R}$ on h and we have for some $t_0 \in (0, 1)$:

$$\begin{aligned} h(1) &= h(0) + (1 - 0)h'(t_0) \\ \implies H(b) &= H(a) + (b^i - a^i)\left.\frac{\partial H}{\partial r^i}\right|_{a+t_0(b-a)} \end{aligned}$$

Now, let us begin the actual proof. So let us take a chart $(U, \phi) on \mathcal{M}$ and $f \in C^\infty(\mathcal{M})$. Then $f \circ \phi^{-1} : \phi(U) \to \mathbb{R}$ is a smooth map.

Let $p \in U \implies \phi(p) \in \phi(U)$. As $\phi(U)$ is open subset of $\mathbb{R}^m$, there exists a $r > 0$ such that $\mathcal{B}(\phi(p), r) \subset \phi(U)$, from the basis definition of an open set. Then, let $q \in \phi^{-1}(\mathcal{B}(\phi(p), r))$. Then $\phi(p)$ and $\phi(q)$ satisfy the condition $\{(1 - t)\phi(p) + t\phi(q) | t \in [0, 1]\} \subset \phi(U)$. Then we can apply the mean value theorem as seen before and we have:

$$\begin{aligned} f(q) - f(p) &= f \circ \phi^{-1} \circ \phi(q) - f \circ \phi^{-1} \circ \phi(p) \\ &= f \circ \phi^{-1}(x^1(q), x^2(q), \ldots, x^m(q)) - f \circ \phi^{-1}(x^1(p), x^2(p), \ldots, x^m(p)) \\ &= (x^i(q) - x^i(p))\left.\frac{\partial f \circ \phi^{-1}}{\partial r^i}\right|_{\phi(p)+t(\phi(q)-\phi(p))} \end{aligned}$$

Now, if we fix p, then

$$\left.\frac{\partial f \circ \phi^{-1}}{\partial r^i}\right|_{\phi(p)+t(\phi(q)-\phi(p))}$$

depends only on q and we denote it by $g_i(q)$. This implies that $g_i(p) = \left.\frac{\partial f \circ \phi^{-1}}{\partial r^i}\right|_{\phi(p)} \equiv \left(\frac{\partial}{\partial x^i}\right)_p f$.

We finally have:

$$f(q) = f(p) + (x^i(q) - x^i(p))g_i(q) \implies f \equiv f(p) + (x^i - x^i(p))g_i$$

Now let us apply some tangent vector v_p to f. We have:

$$\begin{aligned}
v_p(f) &= v_p(f(p) + (x^i - x^i(p))g_i) \\
&= \cancelto{0}{v_p(f(p))} + v_p((x^i - x^i(p))g_i) \quad (\because p \text{ fixed, } f(p) \text{ constant}) \\
&= v_p(g_i)\cancelto{0}{(x^i - x^i(p))_p} + v_p(x^i - x^i(p))g_i(p) \\
&= (v_p(x^i) - v_p(x^i(p)))g_i(p) \\
&= v_p(x^i)g_i(p) \\
&= v_p(x^i)\left(\frac{\partial}{\partial x^i}\right)_p f
\end{aligned}$$

Thus, we see that every tangent vector can be written as a combination of the set of derivations.

So from the above two propositions, we proved that these quantities are indeed a basis for the tangent space. So any vector v_p can be written as:

$$v_p = c^i\left(\frac{\partial}{\partial x^i}\right)_p \implies v_p(x^j) = c^i\left(\frac{\partial x^j}{\partial x^i}\right)_p = c^i\delta^j{}_i = c^j \implies v_p = v_p(x^i)\left(\frac{\partial}{\partial x^i}\right)_p$$

We now define something which we had already seen earlier, the Jacobian. For that, let $\mathcal{M}$ and $\mathcal{N}$ be two manifolds and (U,ϕ) and (V,ψ) be two charts on these manifolds respectively and $F:\mathcal{M}\to N$ be a smooth map between the manifolds such that $F(U)\subset V$. Let F^i denote the i^{th} component of F in chart (V,ψ). Then we have:

$$F^i := \psi^i \circ F = r^i \circ \psi \circ F$$

We call the matrix $[\frac{\partial F^i}{\partial x^j}]$ to be the Jacobian matrix of F relative to these charts and if the two manifolds have same dimension, we can also calculate the determinant of this matrix. If the manifolds are same and we have two overlapping charts, then the transition map $\gamma = \psi\circ\phi^{-1}$ is from $\mathbb{R}^m$ to $\mathbb{R}^m$. Previously, in the definition, we had $F:\mathcal{M}\to\mathcal{N}$. If we take $F=\gamma$, then $\mathcal{N},\mathcal{M}=\mathbb{R}^m$ whose coordinates are r^μ. Then

$$\frac{\partial\gamma^i}{\partial r^j}(\phi(p)) = \frac{\partial\left(\psi\circ\phi^{-1}\right)^i}{\partial r^j}(\phi(p)) = \frac{\partial r^i\circ\left(\psi\circ\phi^{-1}\right)}{\partial r^j}(\phi(p)) = \frac{\partial\psi^i\circ\phi^{-1}}{\partial r^j}(\phi(p)) = \frac{\partial\psi^i}{\partial\phi^j}(p)$$

What this tells us is that the Jacobian matrix of the transition map is the matrix of the partial derivatives at p. This is very similar to what we had earlier seen, the Jacobian written as $\frac{\partial x'}{\partial x}$.

Theorem 3 (Inverse Function Theorem):
For $\mathbb{R}^n$: Let $F:W\to\mathbb{R}^n$ be a smooth map where W is an open subset of $\mathbb{R}^n$. For $p\in W$, F is locally invertible at p iff determinant of the Jacobian matrix is not zero.

For manifolds: Let $F:\mathcal{M}\to\mathcal{N}$ be a smooth map between two manifolds and $p\in\mathcal{M}$. Let (U,ϕ) and (V,ψ) be two charts about p and $F(p)$ respectively and $F(U)\subset V$. Let $F^i=\psi^i\circ F$. Then F is locally invertible at p iff Jacobian determinant is non-zero.

This is a local result and hence, directly translated to manifolds too from $\mathbb{R}^n$.

8.3.3. Push-forward (differential)

Remember the pull-back? Huh, now this is its counterpart.

Let $\mathcal{M}$ and $\mathcal{N}$ be two differentiable manifolds and $\psi : \mathcal{M} \to \mathcal{N}$ be a smooth map. The map ψ induces a map between $T_p\mathcal{M}$ and $T_{\psi(p)}\mathcal{N}$ denoted by ψ_{*p} or $d\psi_p$ which is called the differential of ψ (basically, $p \mapsto \psi(p)$ under ψ and the tangent plane at p is mapped to tangent plane at $\psi(p)$) using this *differential*. If $v_p \in T_p\mathcal{M}$, then $\psi_{*p}(v_p) \in T_{\psi(p)}\mathcal{N}$ and is defined by:

$$\psi_{*p}(v_p)(f) \equiv v_p(\psi^* f) \;\forall\; f \in C^\infty(\mathcal{N})$$

Let us now check that $\psi_{*p}(v_p)$ is indeed a member of the tangent plane at $\psi(p)$ in N.

$$\begin{aligned}
\psi_{*p}(v_p)(c_1 f_1 + c_2 f_2) &= v_p(\psi^*(c_1 f_1 + c_2 f_2)) \\
&= v_p(c_1(\psi^* f_1) + c_2(\psi^* f_2)) && (\psi^* \text{ is a linear map}) \\
&= c_1 v_p(\psi^* f_1) + c_2 v_p(\psi^* f_2) && (v_p \text{ is a linear map}) \\
&= c_1 \psi_{*p}(v_p)(f_1) + c_2 \psi_{*p}(v_p)(f_2)
\end{aligned}$$

$$\begin{aligned}
\psi_{*p}(v_p)(fg) &= v_p(\psi^*(fg)) \\
&= v_p((\psi^* f)(\psi^* g)) \\
&= (\psi^* f) v_p(\psi^* g) + (\psi^* g) v_p(\psi^* f) && \text{(Liebniz rule)} \\
&= (\psi^* f)\psi_{*p}(v_p)(g) + (\psi^* g)\psi_{*p}(v_p)(f)
\end{aligned}$$

Now, let us check that ψ_{*p} is a linear map from $T_p\mathcal{M} \to T_{\psi(p)}\mathcal{N}$. For that, consider a linear combination of vectors from the tangent space of $\mathcal{M}$.

$$\begin{aligned}
\psi_{*p}(c_1 v_p + c_2 w_p)(f) &= (c_1 v_p + c_2 w_p)(\psi^* f) \\
&= c_1 v_p(\psi^* f) + c_2 w_p(\psi^* f) \\
&= c_1 \psi_{*p}(v_p)(f) + c_2 \psi_{*p}(w_p)(f)
\end{aligned}$$

Thus, we proved the linearity of the push-forward. Now using this differential, we will obtain an expression which is analogous to the Jacobian seen earlier. For that, take charts (U, ϕ) and (V, χ) on $\mathcal{M}$ and $\mathcal{N}$ respectively, with coordinate functions $(x^1, x^2, \ldots, x^m)$ and $(y^1, y^2, \ldots, y^n)$ and $\psi : \mathcal{M} \to \mathcal{N}$ and consider $\psi_{*p}\left(\frac{\partial}{\partial x^i}\right)_p \in T_{\psi(p)}\mathcal{N}$.

$$\begin{aligned}
\psi_{*p}\left(\frac{\partial}{\partial x^i}\right)_p &= \psi_{*p}\left(\frac{\partial}{\partial x^i}\right)_p \left(y^j\right)\left(\frac{\partial}{\partial y^j}\right)_{\psi(p)} && \text{(expanding with basis elements)} \\
&= \left(\frac{\partial}{\partial x^i}\right)_p (\psi^* y^j)\left(\frac{\partial}{\partial y^j}\right)_{\psi(p)} \\
&= \left(\frac{\partial}{\partial x^i}\right)_p (y^j \circ \psi)\left(\frac{\partial}{\partial y^j}\right)_{\psi(p)} \\
&= \left(\frac{\partial (y^j \circ \psi)}{\partial x^i}\right)_p \left(\frac{\partial}{\partial y^j}\right)_{\psi(p)}
\end{aligned}$$

This looks very similar to the Jacobian thingy that we saw earlier and thus, push-forward is also kind of like the Jacobian matrix.

Pushing tangents to curves:

Let $\psi : \mathcal{M} \to \mathcal{N}$ and $\sigma : (a, b) \to \mathcal{M}$ be a curve on $\mathcal{M}$. Then $\psi \circ \sigma : (a, b) \to \mathcal{N}$ is a curve on $\mathcal{N}$. Then tangent vector v to this curve at $(\psi \circ \sigma)(0)$ satisfies for any $f \in C^\infty(\mathcal{N})$:

$$v(f) = \frac{\mathrm{d}}{\mathrm{d}t} f \circ (\psi \circ \sigma)\Bigg|_{t=0}$$

Now, these can be alternatively written as:

$$v(f) = \frac{\mathrm{d}}{\mathrm{d}t}(\psi \circ \sigma)^* f\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\sigma^*(f \circ \psi)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \psi) \circ \sigma\bigg|_{t=0} = w(f \circ \psi)$$

where w is the tangent vector to the curve at $\sigma(0)$, acting on the function $f \circ \psi$ which can also be written as $w(\psi^* f)$. Using the definition of push-forward, this is nothing but $\psi_{*\sigma(0)}(w)(f)$
Now, since this is true for any smooth function f, we can write $v \equiv \psi_{*\sigma(0)}(w)$. In other words, the Jacobian of ψ pushes the tangent vectors of σ to tangent vectors of the image of σ under ψ.

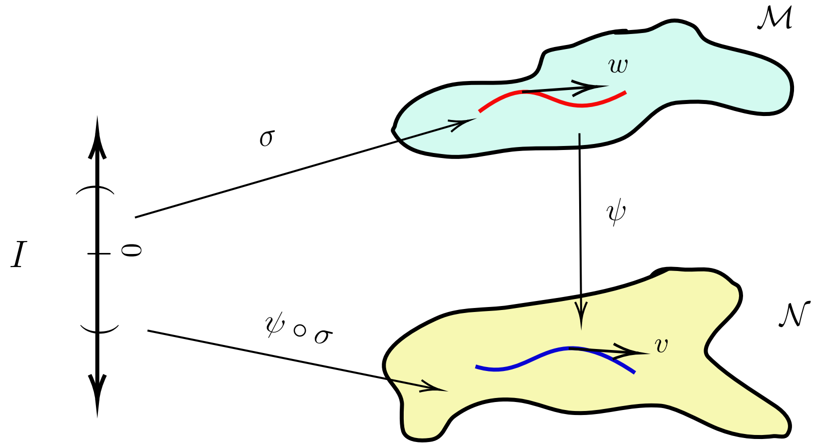


Figure 16: Diagram showing the pushing-forward of tangent to curves in different manifolds.

8.3.4. Tangent Bundles

The tangent bundle of a manifold $\mathcal{M}$ is the union of all the tangent spaces, hence:

$$T\mathcal{M} = \bigcup_{p \in \mathcal{M}} T_p\mathcal{M}$$

We can define a 'natural' (coordinate or chart independent) map π from the tangent bundle to the manifold given by $\pi(v) = p, \quad v \in T_p\mathcal{M}$ (since tangent bundle is the union of tangent spaces, it indeed consists of elements from $\{T_p\mathcal{M}\}$). The map π is sometimes called the *canonical projection*. Notice that the tangent bundle is just a set, with no structure or anything else. Let us modify this a bit. Suppose we have a chart (U, ϕ), then we will have:

$$TU = \bigcup_{p \in U} T_pU = \bigcup_{p \in \mathcal{U}} T_p\mathcal{M}$$

The last two things can be shown to be equal. Since the set of derivations form a basis for the tangent space at p, we can write the tangent vector uniquely as a linear combination:

$$T_p\mathcal{M} \ni v_p = \sum_i^n \dot{q}^i(v_p) \frac{\partial}{\partial x^i}\bigg|_p$$

Here, $\dot{q}^i : \pi^{-1}(U) \to \mathbb{R}$ are n real-valued functions [1]. Let us also define $q^i : \pi^{-1}(U) \to \mathbb{R}, q^i \equiv x^i \circ \pi = \pi^* x^i$ and $p\bar{h}i : \phi^{-1}(U) \to \mathbb{R}^{2n}$

$$\bar{\phi}(v_p) \equiv \Big(q^1(v_p), q^2(v_p), \ldots, q^n(v_p), \dot{q}^1(v_p), \dot{q}^2(v_p), \ldots, \dot{q}^n(v_p)\Big)$$

[1] The notation is borrowed from Lagrangian formulation of classical mechanics, where the manifold becomes the configuration space of the mechanical system

Note that $\bar{\phi}$ has an inverse such that $(q^1(v_p), q^2(v_p), \ldots, q^n(v_p), \dot{q}^1(v_p), \dot{q}^2(v_p), \ldots, \dot{q}^n(v_p)) \mapsto \sum_i^n \dot{q}^i(v_p) \frac{\partial}{\partial x^i}\Big|_p$ and is thus a bijection. Also,

$$q^i(v_p) = x^i \circ \pi(v_p) = x^i(p)$$
$$v_p = \dot{q}^i(v_p) \frac{\partial}{\partial x^i}\Big|_p \implies \dot{q}^i(v_p) = v_p[x^i]$$

Thus, the function $\bar{\phi}$ becomes:

$$\bar{\phi}(v_p) \equiv \left(x^1(p), x^2(p), \ldots, x^n(p), v_p[x^1], v_p[x^2], \ldots, v_p[x^n]\right)$$

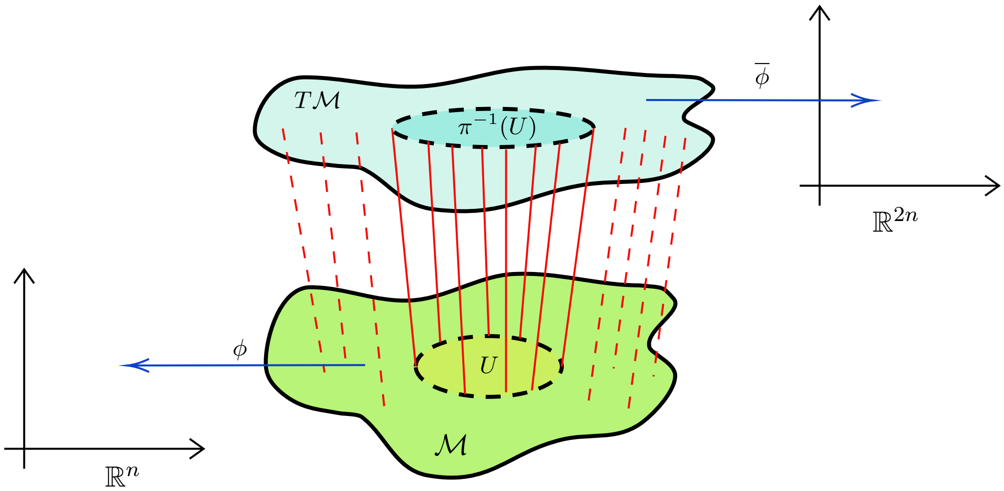


Figure 17: Diagram showing the chart $\left(\pi^{-1}(U), \bar{\phi}\right)$. The red lines represent the tangent space at each point of the manifold. Solid lines represent the tangent spaces for points belonging in U

> **Lemma 1:**
> The pair $\left(\pi^{-1}(U), \bar{\phi}\right)$ is a chart on $T\mathcal{M}$, often called the *bundle chart*.

The tangent bundle can also be shown to be *Hausdorff* and *second countable*. Well, tangent bundles are a special case of something called a fibre bundle.

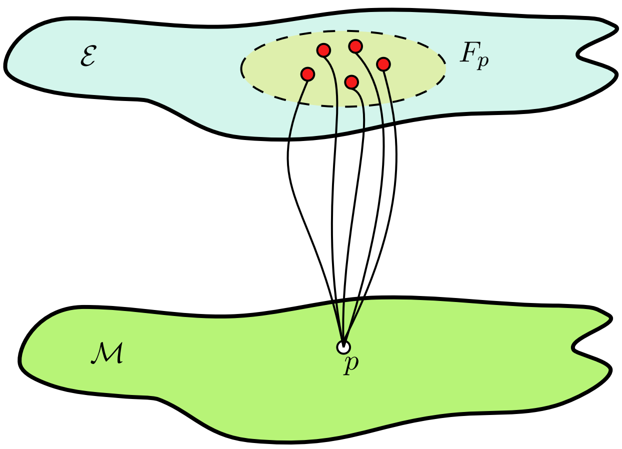


Figure 18: Diagram showing the base space, total space and the fibre of a point p

Definition 8 (Bundle):
Given a map $\pi : \mathcal{E} \to M$, the inverse image $pi^{-1}(p)$ to be the *fibre* at $p \in \mathcal{M}$, denoted as E_p. For two maps $\pi : \mathcal{E} \to M$ and $\pi : \mathcal{E}' \to M$, a map ϕ is said to be fibre-preserving if $\phi(E_p) \subset E'_p$.

A bundle is a triple $(\mathcal{E}, \mathcal{M}, \pi)$ consisting of manifolds $\mathcal{M}$ which is called the *base space* and the manifold $\mathcal{E}$, called the *total space* and π is a continuous, surjective map from the total space to the base space.

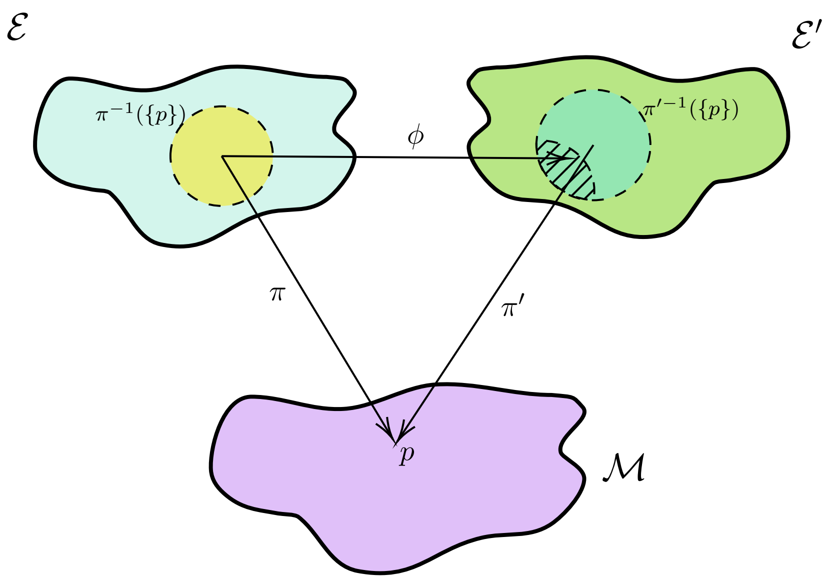


Figure 19: Commutative diagram for a bundle, showing the fibre preserving map ϕ

Fibres can be thought of as points in $\mathcal{E}$ which 'strings' together with p in $\mathcal{M}$ (hence the name bundle, like a bundle of strings perhaps...) In a bundle, different points of the base manifold may have (topologically) different fibres. If all points have a fibre, which is togologically equivalent to a single space, say F, then we call that bundle a *fibre bundle*. In other words, fibre bundles have a single fibre.

We now define a section of a bundle. Given a bundle $(\mathcal{E}, \mathcal{M}, \pi)$, the section is a map $\sigma : \mathcal{M} \to \mathcal{E}$ such that $\pi \circ \sigma = \mathbb{1}_{\mathcal{M}}$, the identity map of the base space. What it means that, this map σ takes p to some point in the fibre F_p such that when π is acted on $\sigma(p)$, we obtain the same point p.

Okay, so what was this thing really needed for? Well, we had earlier seen the tangent bundle, which we had written as $T\mathcal{M}$. More appropriately, we should have written it as $(T\mathcal{M}, \mathcal{M}, \pi)$. Then a vector field is nothing but a section of the tangent bundle which assigns a tangent vector (belonging to the tangent space) to each point p in the manifold.

8.4. Vector Fields

Definition 9 (Vector Field):
A vector field X on a smooth manifold is a map from $\mathcal{M} \to T\mathcal{M}$ with $\pi \circ X = \mathrm{id}_{\mathcal{M}}$, which assigns a tangent vector $X_p \in T_p\mathcal{M}$ at each point $p \in \mathcal{M}$.

X takes a point p from $\mathcal{M}$ and assigns a tangent vector X_p to the point. Now, if we apply the canonical projection π, then we obtain p back.

As specified earlier, a vector field is a section of the tangent bundle. A vector field is smooth if the map $X : \mathcal{M} \to T\mathcal{M}$ is a smooth section of the tangent bundle. We will focus mostly on smooth vector fields only. Since a vector field assigns a tangent vector to each point of its domain and the tangent vector (from algebraic approach) maps differentiable functions to real numbers, we can define a map $\mathbf{X}f$ from $\mathcal{M}$ to $\mathbb{R}$ as:

$$(\mathbf{X}f)(p) \equiv \mathbf{X}_p[f]$$

From this, we can define the map $X : C^\infty(\mathcal{M}) \to C^\infty(\mathcal{M})$ such that $f \mapsto Xf$. This gives us another interpretation of a vector field as a *"map from smooth functions to smooth functions"*. The image Xf is often called the *Lie derivative* of f along the vector field X, denoted by $\mathcal{L}$. Then, from the definition of derivations, we have the following:

$$\mathbf{X}(af + bg) = a\mathbf{X}f + b\mathbf{X}g \qquad \mathbf{X}(fg) = f\mathbf{X}g + g\mathbf{X}f \quad \forall\ f, g \in C^\infty(\mathcal{M}) \ \text{ and } \ a, b \in \mathbb{R}$$

We also have stated that $\left(\frac{\partial}{\partial x^i}\right)$ form a basis for the tangent space (or space of derivation which is isomorphic to the tangent space, so we treat them as the 'same')

The set of all vector fields $\mathfrak{X}(\mathcal{M})$ on a manifold carries a structure of a real vector space, that is:

$$(aX + bY)f := aXf + bYf\ \forall\ f \in C^\infty(\mathcal{M})$$

This definition can be extended to give $\mathfrak{X}(\mathcal{M})$ a *module* structure over the ring $C^\infty(\mathcal{M})$. Thus, if $g, h \in C^\infty(\mathcal{M})$ and $X, Y \in \mathfrak{X}(\mathcal{M})$, then:

$$(gX + hY)_p f := g(p)X_p f + h(p)Y_p f\ \forall p \in \mathcal{M}, f \in C^\infty(\mathcal{M})$$

8.4.1. Coordinate Basis Vector Field

If (U, ϕ) is a chart on the manifold $\mathcal{M}$ with coordinate functions $(x^1, ., x^n)$, then we can naturally define n vector fields $\left(\frac{\partial}{\partial x^i}\right)$ defined by:

$$\left(\frac{\partial}{\partial x^i}\right)(p) := \left(\frac{\partial}{\partial x^i}\right)_p$$

The RHS is just the tangent vector to the i^{th} coordinate cuve passing through p. Now, using the previous definitions, for any function $f \in C^\infty(\mathcal{M})$,

$$\left(\frac{\partial}{\partial x^i}f\right)(p) = \frac{\partial}{\partial x^i}_p[f] = \frac{\partial}{\partial r^i}\Big(f \circ \phi^{-1}\Big)(\phi(p)) = \left(\frac{\partial}{\partial r^i}\Big(f \circ \phi^{-1}\Big) \circ \phi\right)(p)$$

Thus from here we obtain:

$$\frac{\partial}{\partial x^i}f = \frac{\partial}{\partial r^i}\Big(f \circ \phi^{-1}\Big) \circ \phi$$

This is something which is good tbh, this tell us that the "partial derivative" of a function defined on the manifold is just the partial derivative of the "proxy function" $f \circ \phi^{-1}$ composed with ϕ.

Now, since $\left(\frac{\partial}{\partial x^i}\right)_p$ form a basis of the tangent space, for any vector field X, we can write:

$$X(p) \equiv X_p = X^i(p)\left(\frac{\partial}{\partial x^i}\right)(p) = \left(X^i\frac{\partial}{\partial x^i}\right)(p)$$

Since p is any arbitrary point on the manifold, we have:

$$X = \left(X^i\frac{\partial}{\partial x^i}\right)$$

Note that the vector field X has an expansion which is very similar to a tangent vector but there, the coefficients were real numbers while here, the coefficients are functions and the basis vectors here are vector fields and not tangent vectors. To find the expression of the function X^i note that,

$$(Xx^i)(p) = X_p[x^i] = X^j(p)\left(\frac{\partial}{\partial x^j}\right)(p)[x^i] = X^j(p)\delta^i{}_j = X^i(p)$$

Now, from here we have the nice expression that $X^i = Xx^i$. Let us see the nicest thing to ever come, which provides some kind of redemption to all these "mind-hurting" mathematical juggleries:

If (U, ϕ) and (V, χ) be two chart of manifold $\mathcal{M}$ and if the coordinate functions are $(x^1, x^2, \ldots, x^n)$ and $(x'^1, x'^2, \ldots, x'^n)$ respectively, then for $X \in \mathfrak{X}$:

$$X'^i = Xx'^i = X^j\left(\frac{\partial}{\partial x^j}\right)x'^i = X^j\left(\frac{\partial x'^i}{\partial x^j}\right)$$

Damn, the sanctimonious transformation rule is obtained as a result of these intricate definitons!! Idk, but this really made my day...

Example: Let $\mathcal{M} = \mathbb{R}^2$ and consider two charts:

- $(U, \phi) = (\mathcal{M}, \mathrm{id})$ with coordinate function x, y
- (V, χ) where $V = \mathbb{R}^2 \backslash \{(x, y) | x \leq 0, y = 0\}$ and χ is such that $p \mapsto (r, \theta)$.The origin had to be taken out anyways since at zero, the value of θ is not well-defined. The transformations are:

$$x = r\cos\theta \qquad y = r\sin\theta$$
$$r = \sqrt{x^2 + y^2} \qquad \theta(p) = \int_{\mathcal{C}} \frac{xdy - ydx}{x^2 + y^2}$$

  Here, $\mathcal{C}$ is a curve joining $(1, 0)$ to point p. We took this form of θ instead of the usual $\tan^{-1}\left(\frac{y}{x}\right)$ since it is not really well-defined at $x = 0$.

Now, if X has components $2xy$ and $(1 - x^2 + y^2)$ in the basis $\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right\}$, then what are components in the basis $\left\{\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right\}$

Note that $\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right\}$ is not the partial derivative. These are vector fields. We can write

$$X = 2xy\frac{\partial}{\partial x} + (1 - x^2 + y^2)\frac{\partial}{\partial y}$$

. Now, we have:

$$Xr = 2xy\frac{\partial r}{\partial x} + (1 - x^2 + y^2)\frac{\partial r}{\partial y}$$

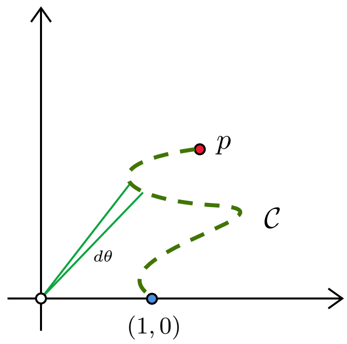


Figure 20: How we defined θ. Note that a small change in the integrand is exactly equal to $d\theta$ and then when we integrate, we get our value of θ_p.

Calculating the above expression, we will obtain $Xr = (1+r^2)\sin\theta$. Similarly, we have

$$X\theta = 2xy\frac{\partial\theta}{\partial x} + (1 - x^2 + y^2)\frac{\partial\theta}{\partial y}$$

Now, from the integral definition of θ, we have $\frac{\partial\theta}{\partial x} = -\frac{y}{x^2+y^2}, \frac{\partial\theta}{\partial y} = \frac{x}{x^2+y^2}$. Then, evaluating the expression would give $X\theta = \frac{1-r^2}{r}\cos\theta$. Then,

$$X = X^r\frac{\partial}{\partial r} + X^\theta\frac{\partial}{\partial\theta} = Xr\frac{\partial}{\partial r} + X\theta\frac{\partial}{\partial\theta} = (1+r^2)\sin\theta\frac{\partial}{\partial r} + \left(\frac{1-r^2}{r}\right)\cos\theta\frac{\partial}{\partial\theta}$$

Fun Fact:

$$\hat{r} = \frac{\partial}{\partial r} \qquad \hat{\theta} = \frac{1}{r}\frac{\partial}{\partial\theta}$$

8.4.2. Lie stuffs (it's pronounced lee)

We will now see when 'multiplying' two vector fields will yield another field (The multiplication of two fields is intuitively defined as their composition, that is, $(X\circ Y)f = X(Yf)$). For this, note that the linearity property of vector field is easily satisfied by $X\circ Y$ but the main problem is the *product rule*:

$$\begin{aligned} X\circ Y(fg) &= X(fYg + gYf) \\ &= X(fYg) + X(gYf) \\ &= XgYf + gXYf + XfYg + fXYg \end{aligned} \qquad \begin{aligned} Y\circ X(fg) &= Y(fXg + gXf) \\ &= Y(fXg) + Y(gXf) \\ &= YgXf + gYXf + YfXg + fYXg \end{aligned}$$

Note that on its own these two quantities are not vector fields, since these do not satisfy the 'product rule', however, if we subtract these quantities, we get:

$$\begin{aligned} (X\circ Y - Y\circ X)(fg) &= \cancel{XgYf} + gXYf + \cancel{XfYg} + fXYg - (\cancel{YgXf} + gYXf + \cancel{YfXg} + fYXg) \\ &= f(X\circ Y - Y\circ X)g + g(X\circ Y - Y\circ X)f \end{aligned}$$

This difference satisfies the product rule and is a vector field. So we define a new operation on $\mathfrak{X}(\mathcal{M})$.

Definition 10 (Lie Bracket):
The Lie Bracket of two vector fields X, Y on $\mathcal{M}$ is defined by:

$$[X,Y]f = X(Yf) - Y(Xf) \quad \forall f \in C^\infty(\mathcal{M})$$

The Lie Bracket trivially satisifes lineartiy, that is,

$$[X,Y](af+bg) = a[X,Y]f + b[X,Y]g$$

And as seen above, it also satisfies the product rule, that is,

$$[X,Y](fg) = f[X,Y]g + g[X,Y]f$$

Thus, using this definition of Lie Bracket, we can say that if $X, Y \in \mathfrak{X}(\mathcal{M})$ and $f \in C^\infty(\mathcal{M})$ then $[X,Y] \in \mathfrak{X}(\mathcal{M})$. Note that, then $[\cdot\,,\cdot] : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ is a binary operation. Hence, using the Lie Bracket, we can define an *algebra* [1] satisfying the bilinear property:

$$\begin{aligned}
[aX+bY,Z]f &= (aX+bY)(Zf) - Z(aX+bY)f \\
&= aX(Zf) + bY(Zf) - aZ(Xf) - bZ(Yf) \\
&= a(XZ-ZX)(f) + b(YZ-ZY)(f) \\
&= (a[X,Z,+]b[YZ,)](f)
\end{aligned}$$

As f is arbitrary, we get $[aX+bY,Z] = a[X,Z] + b[Y,Z]$. The linearity of the second factor can be similarly shown. Thus it is a valid algebra, called the *Lie Algebra*. We can also check that $[X,Y]f = -[Y,X]f \implies [X,Y] = -[Y,X]$, hence it is an *anti-commutative* algebra.

While associativity of the Lie Bracket does not hold, we have another property, where given three vectors fields X, Y, Z, if we perform the Lie Bracket using a cyclic permutation, we get:

$$\begin{aligned}
&([X,[Y,Z]] + [Z,[X,Y]] + [Y,[Z,X]])(f) \\
&= X[Y,Z]f - [Y,Z]Xf + Z[X,Y]f - [X,Y]Zf + Y[Z,X]f - [Z,X]Yf \\
&= (XYZ)f - (XZY)f - (YZX)f + (ZYX)f + (ZXY)f - (ZYX)f - (XYZ)f \\
&\quad + (YXZ)f + (YZX)f - (YXZ)f - (ZXY)f + (XZY)f \\
&= 0
\end{aligned}$$

Thus, we have an important property, $([X,[Y,Z]] + [Z,[X,Y]] + [Y,[Z,X]]) = 0$, which is often termed as the *Jacobi property*.

Lemma 2:
Let $X, Y \in \mathcal{M}$ be two vector fields and $f, g \in C^\infty(\mathcal{M})$ be two smooth functions. Then we have,

$$[fX, gY] = fg[X,Y] + f(Xg)Y - g(Yf)X$$

Proof. Let us take $h \in C^\infty(\mathcal{M})$. Then,

$$\begin{aligned}
[fX,gY]h &= fX(gYh) - gY(fXh) \\
&= f((Xg)(Yh) + gX(Yh)) - g((Yf)(Xh) + fY(Xh)) \\
&= fg(XYh - YXh) + f(Xg)Yh - g(Yf)Xh \\
&= (fg[X,Y] + f(Xg)Y - g(Yf)X)h
\end{aligned}$$

Since h is any arbitrary smooth function, we proved the lemma.

We now that any vector field can be expanded in terms of $\frac{\partial}{\partial x^i}$ and thus, it becomes kinda imperative to calculated the Lie Bracket of these quantitites. Note:

$$\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] f = \frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial x^j}\right) f - \frac{\partial}{\partial x^j}\left(\frac{\partial}{\partial x^i}\right) f$$

[1] It is a non-associative algebra

At the first glance, this seems to be zero (and it actually is). However, we have to be careful that these quantities are not the partial derivatives of f since f is a function on the manifold where partial derivatives are not defined. Instead we had earlier seen:

$$\frac{\partial f}{\partial x^i} = \frac{\partial}{\partial r^i}\left(f \circ \phi^{-1}\right) \circ \phi$$

Using this notation, we have:

$$\frac{\partial}{\partial x^i}\frac{\partial f}{\partial x^j} = \frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) \circ \phi\right)$$

Now, let us take $T_j = \left(\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) \circ \phi\right)$, then we have:

$$\begin{aligned}
\frac{\partial T_j}{\partial x^i} &= \frac{\partial}{\partial r^i}\left(T_j \circ \phi^{-1}\right) \circ \phi \\
&= \frac{\partial}{\partial r^i}\left(\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) \circ \phi \circ \phi^{-1}\right) \circ \phi \\
&= \frac{\partial}{\partial r^i}\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) \circ \phi
\end{aligned}$$

Note that in the above, it is indeed the partial derivative since $f \circ \phi^{-1}$ is a function on $\mathbb{R}^n$.

$$\begin{aligned}
\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] f &= \frac{\partial}{\partial x^i}\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) \circ \phi - \frac{\partial}{\partial x^j}\frac{\partial}{\partial r^i}\left(f \circ \phi^{-1} \circ \phi\right) \\
&= \left(\frac{\partial}{\partial x^i}\frac{\partial}{\partial r^j}\left(f \circ \phi^{-1}\right) - \frac{\partial}{\partial x^j}\frac{\partial}{\partial r^i}\left(f \circ \phi^{-1}\right)\right) \circ \phi
\end{aligned}$$

Now, since partial derivatives commute, we can safely say that the Lie Bracket is zero 😄. Now, using Lemma 2, taking $f = X^i$ and $g = Y^j$, we can write:

$$\begin{aligned}
[X, Y] &= \left[X^i \frac{\partial}{\partial x^i}, Y^j \frac{\partial}{\partial x^j}\right] \\
&= X^i Y^j \cancelto{0}{\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right]} + X^i \frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\frac{\partial}{\partial x^i} \\
&= \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\right)\frac{\partial}{\partial x^i} \\
&= \left(XY^i - YX^i\right)\frac{\partial}{\partial x^i}
\end{aligned}$$

Here comes an example, let us take the following vector fields:

$$\begin{aligned}
X &= (1 + r^2)\sin\theta \frac{\partial}{\partial r} + \left(\frac{1 - r^2}{r}\right)\cos\theta \frac{\partial}{\partial \theta} \\
Y &= -(1 + r^2)\cos\theta \frac{\partial}{\partial r} + \left(\frac{1 - r^2}{r}\right)\sin\theta \frac{\partial}{\partial \theta} \\
Z &= \frac{\partial}{\partial \theta}
\end{aligned}$$

While we can calculate the components of the Lie Bracket using the above equation, but that would be too much to handle, since we have to calculate the coefficient and then act the vector field on it for each component. Instead, we will use the fact that the component of any vector field is simply given by the vector field acting on the coordinate. Thus, we will have,

r component: $$[X,Y]\,r = X(Yr) - Y(Xr) = X\Big(-(1+r^2)\cos\theta - 0\Big) - Y\Big((1+r^2)\sin\theta + 0\Big)$$
$$= -\left((1+r^2)\sin\theta\frac{\partial}{\partial r} + \left(\frac{1-r^2}{r}\right)\cos\theta\frac{\partial}{\partial\theta}\right)(1+r^2)\cos\theta$$
$$- \left(-(1+r^2)\cos\theta\frac{\partial}{\partial r} + \left(\frac{1-r^2}{r}\right)\sin\theta\frac{\partial}{\partial\theta}\right)(1+r^2)\sin\theta$$
$$= -2r(1+r^2)\sin\theta\cos\theta + \frac{1-r^4}{r}\cos\theta\sin\theta$$
$$+ 2r(1+r^2)\sin\theta\cos\theta - \frac{1-r^4}{r}\sin\theta\cos\theta$$
$$= 0$$

θ component: $$[X,Y]\,\theta = X(Y\theta) - Y(X\theta) = X\left(\frac{1-r^2}{r}\right)\sin\theta - Y\left(\frac{1-r^2}{r}\right)\cos\theta$$
$$= \left((1+r^2)\sin\theta\frac{\partial}{\partial r} + \left(\frac{1-r^2}{r}\right)\cos\theta\frac{\partial}{\partial\theta}\right)\left(\frac{1-r^2}{r}\right)\sin\theta$$
$$- \left(-(1+r^2)\cos\theta\frac{\partial}{\partial r} + \left(\frac{1-r^2}{r}\right)\sin\theta\frac{\partial}{\partial\theta}\right)\left(\frac{1-r^2}{r}\right)\cos\theta$$
$$= 2(\sin^2\theta + \cos^2\theta)\left[(1+r^2)\frac{\partial}{\partial r}\left(\frac{1-r^2}{r}\right) - \left(\frac{1-r^2}{r}\right)^2\right]$$
$$= 2(1+r^2)\left(-\frac{1}{r^2} - 1\right) - \left(\frac{1-r^2}{r}\right)^2$$
$$= -2\left(\frac{1-r^4-(1-r^2)^2}{r^2}\right)$$
$$= -4$$

Since we obtained the components, we can finally write:

$$[X,Y] = 0\times\frac{\partial}{\partial r} - 4\times\frac{\partial}{\partial\theta} = -4\times\frac{\partial}{\partial\theta} = -4Z$$

Similarly if calculated, we will see:

$$[Y,Z] = -X \qquad\qquad [Z,Y] = -Y$$

8.4.3. Giving vector fields a push!

We had earlier seen the push-forward of a tangent vector. Now, the question remains is whether vector fields can be pushed too in a similar way, that is, if we start with a vector field X on $\mathcal{M}$, do we get a vector field on $\mathcal{N}$ using $\{\psi_{*p}Y_p\}_{p\in\mathcal{M}}$. There lies a problem here.😢

Remember that the push-forward basically mapped the tangent vector at one point to the tangent vector at another point of the other manifold. So, suppose our function $\psi : \mathcal{M} \to \mathcal{N}$ is not injective, hence $\exists\ p \neq q \in \mathcal{M}$ such that $\psi(p) = \psi(q)$ (both mapped to same point in $\mathcal{N}$).

If X is a vector field on $\mathcal{M}$, then it is possible that the tangent vector X_p and X_q might not be mapped to the same tangent vector using the push-forward, that is, $\psi_{*p}X_p \neq \psi_{*q}X_q$. Since vector field at point $\psi(p)$ assigns only one tangent vector to the point, there is an ambiguity to which of the tangent vector be assigned to $\psi(p)$.

Definition 11:
Let $\psi : \mathcal{M} \to \mathcal{N}$ be a differentiable map and $X \in \mathfrak{X}(\mathcal{M}), Y \in \mathfrak{X}(\mathcal{N})$. We say that X and Y are *ψ-related* if,

$$Y_{\psi(p)} = \psi_{*p}X_p \qquad \forall\ p \in \mathcal{M}$$

So, the previous problem is circumvented since $\psi(p) = \psi(q)$ and hence Y assigns the same vector to the point.

Proposition 4:
X and Y are ψ-related iff $\psi^*(Y(f)) = X(\psi^* f)$ for all $f \in C^\infty(\mathcal{N})$

*Proof.*Let $p \in \mathcal{M}$. Then,

$$(Yf)(\psi(p)) = Y_{\psi(p)}[f] = \psi_{*p}X_p[f] = X_p[\psi^* f] = X(\psi^* f)(p)$$

Theorem 4:
If $\psi : \mathcal{M} \to \mathcal{N}$ is a diffeomorphism, then for any $X \in \mathfrak{X}(\mathcal{M})$, there exists a unique $Y \in \mathcal{N}$ such that X and Y are ψ-related. We denote this smooth vector field as $\psi_* X$ and call it the *push-forward* of X by ψ.

Proof. Let $q \in \mathcal{N}$. As ψ is a diffeomorphism, it is invertible and thus there exists a unique $p \in \mathcal{M}$ such that $\psi(p) = q$. Then define,

$$Y_q := \psi_{*p}X_p$$

Then Y is a well-defined vector field since ψ is both injective and surjective. To check smoothness of Y, consider a chart (U, ϕ) on $\mathcal{M}$ with coordinate functions $(x^1, \ldots, x^n)$. As ψ is a diffeomorphism, then $(F(U), \phi \circ \psi^{-1})$ is a valid chart on $\mathcal{N}$, with coordinate functions say, $(y^1, \ldots, y^n)$ (since it is diffeomorphism, the dimensions of both the manifold are same).

Then we can expand the tangent vector at p using the basis and then apply the push-forward:

$$X_p = X_p(x^i) \frac{\partial}{\partial x^i}\bigg|_p \implies \psi_{*p}X_p = X_p(x^i)\psi_{*p}\left(\frac{\partial}{\partial x^i}\bigg|_p\right) = Y_q$$

Now, the push-forward is a tangent vector at q and hence can be expanded using the basis in the following way:

$$\psi_{*p}\left(\frac{\partial}{\partial x^i}\bigg|_p\right) = b^j \frac{\partial}{\partial y^j}\bigg|_q$$

where the coeffients are $b^j := \psi_{*p}\left(\frac{\partial}{\partial x^i}\Big|_p\right)(y^j) = \frac{\partial}{\partial x^i}\Big|_p(\psi^* y^j) = \left(\frac{\partial}{\partial x^i}(\psi^* y^j)\right)(p)$ Then, using this we can write the final expression for Y_q to be:

$$Y_q = \left[X_p(x^i)\left(\frac{\partial}{\partial x^i}\left(\psi^* y^j\right)\right)(p)\right]\frac{\partial}{\partial y^j}\bigg|_q$$

The coefficient is a smooth map since $X_p(x^i)$ is smooth, as X is smooth and $\frac{\partial}{\partial x^i}(\psi^* y^j)$ is smooth since $\psi^* y^j = y^j \circ \psi$, that is, the 'proxy' function, is a smooth map and hence proved.

Example:

Let $p \in \mathcal{M}$ and v_p be the tangent vector at p. We will consider the bundle chart where the homeomorphism was defined by:

$$\bar{\phi}(v_p) \equiv \Big(x^1(p), x^2(p), \ldots, x^n(p), v_p[x^1], v_p[x^2], \ldots, v_p[x^n]\Big)$$

Then for $f \in C^\infty(\mathcal{M})$, we have:

$$\pi_{*v}\left(\frac{\partial}{\partial x^i}\right)_v [f] = \left(\frac{\partial}{\partial x^i}\right)_v [\pi^* f] = \left(\frac{\partial}{\partial x^i}\right)_v [f \circ \pi]$$

Now, $f \circ \pi$ is a function on the manifold and then we define the partial derivative:

$$\begin{aligned}\pi_{*v}\left(\frac{\partial}{\partial x^i}\right)_v [f] = \lim_{t\to 0}\frac{1}{t}\Big[&(f\circ\pi)\circ\bar{\phi}^{-1}\Big(x^1(p), \ldots, x^i(p)+t, \ldots, x^n(p), v_p[x^1], v_p[x^2], \ldots, v_p[x^n]\Big) \\ &- (f\circ\pi)\circ\bar{\phi}^{-1}\Big(x^1(p), \ldots, x^i(p), \ldots, x^n(p), v_p[x^1], v_p[x^2], \ldots, v_p[x^n]\Big)\Big]\end{aligned}$$

Now, note that,

$$(f\circ\pi)\circ\bar{\phi}^{-1} = (f\circ\phi^{-1})(\phi\circ\pi\circ\bar{\phi}^{-1})$$

Let us decode this step by step: as seen earlier, $\bar{\phi}^{-1}$ takes the $2n$-dimensional vector and returns the original tangent vector. Then π will act on the tangent vector and then map it back to the point on the manifold. Then ϕ will map it to a vector in $\mathbb{R}^n$ with coordinate functions as the component. So essentially, the above line means that $f\circ\phi^{-1}$ acts on the vector in $\mathbb{R}^n$. Thus,

$$\begin{aligned}\pi_{*v}\left(\frac{\partial}{\partial x^i}\right)_v [f] &= \lim_{t\to 0}\frac{1}{t}\Big[(f\circ\phi)\Big(x^1(p), \ldots, x^i(p)+t, \ldots, x^n(p)\Big) \\ &\qquad - (f\circ\phi)\Big(x^1(p), \ldots, x^i(p), \ldots, x^n(p)\Big)\Big] \\ &= \left(\frac{\partial}{\partial x^i}\right)_p [f]\end{aligned}$$

Thus, from here we obtain that $\frac{\partial}{\partial q^i}$ and $\frac{\partial}{\partial x^i}$ are $\pi-$related, that is,

$$\pi_{*v}\left(\frac{\partial}{\partial x^i}\right)_v = \left(\frac{\partial}{\partial x^i}\right)_p$$

8.4.4. Connection with Lie Brackets

Let $\psi : \mathcal{M} \to \mathcal{N}$ and let $X_1, X_2 \in \mathfrak{X}(\mathcal{M})$ and $Y_1, Y_2 \in \mathfrak{X}(\mathcal{N})$ be ψ-related respectively. Then using the previous proposition, we have, for any $f \in C^\infty(\mathcal{N})$

$$\begin{aligned}(Y_1 f)\circ\psi &= X_1(f\circ\psi) \\ (Y_2 f)\circ\psi &= X_2(f\circ\psi)\end{aligned}$$

Let us calculate what happens with the Lie Bracket:

$$\begin{aligned}([Y_1, Y_2]\, f)\circ\psi &= (Y_1(Y_2 f) - Y_2(Y_1 f))\circ\psi \\ &= (Y_1(Y_2 f))\circ\psi - (Y_2(Y_1 f))\circ\psi \\ &= X_1(Y_2 f\circ\psi) - X_2(Y_1 f\circ\psi) \quad \because Y_2 f, Y_1 f \in C^\infty(\mathcal{N}) \\ &= X_1(X_2(f\circ\psi)) - X_2(X_1(f\circ\psi)) \\ &= [X_1, X_2]\,(f\circ\psi)\end{aligned}$$

Thus we see that $[X_1, X_2]$ and $[Y_1, Y_2]$ are ψ related.

8.5. Covectors and one-forms

We will begin to learn some things about dual spaces, dual vectors, 1-forms and other high-sounding mathemaitcal names. Remember that the dual space was the set of all linear functionals, that is, a dual vector acts on the vector and produces a real number[1]. Now, we had also seen that a tangent vector acting on a function produced a real number, that is, $v_p[f] \in \mathbb{R}$. It would be nice to somehow connect these things which brings us to the following definition.

Definition 12:
Let $\mathcal{M}$ be a manifold and $f \in C^\infty(\mathcal{M})$. Then the differential of f at a point $p \in \mathcal{M}$, denoted by df_p, is a map that acts on $T_p\mathcal{M}$ to produce a real-number:

$$df_p : T_p\mathcal{M} \to \mathbb{R} \qquad df_p(v_p) \equiv v_p[f]$$

Also, note that df_p is a linear map (due to linearity of tangent vectors) and hence is a linear functional. Thus, $df_p \in T_p^*\mathcal{M}$. The elements of $T_p^*\mathcal{M}$ are called *co-vectors* [2] and $T_p^*\mathcal{M}$ is itself called, without any surprise, the *co-tangent plane* at p. Since dual spaces have same dimension as the original space and the tangent space was a vector space of dimension n (same as the manifold), the cotangent plane is also an n dimensional vector space.

We can similarly define the cotangent bundle as

$$T^*\mathcal{M} = \bigcup_p T_p^*\mathcal{M}$$

Fun Fact:
This is extremely useful in classical mechanics, where $\mathcal{M}$ represents the *configuration space*, that is, space of all possible configurations of the system and $T^*\mathcal{M}$ is, in most cases, what we know to be the *phase space*.

We can also define the covector field exactly in same way as the vector field, that is:

Definition 13 (Co-vector Field):
A co-vector field α on $\mathcal{M}$ is a map from $\mathcal{M} \to T^*\mathcal{M}$, which assigns at each point $p \in \mathcal{M}$, a covector, that is,

$$\alpha(p) = \alpha_p \in T_p^*\mathcal{M}$$

A co-vector field α acts on a vector field X, to produce a function $\alpha(X)$ which, acting on a point p on the manifold gives:

$$(\alpha(X))(p) \equiv \alpha_p(X_p) \in \mathbb{R}$$

The action $\alpha(X)$ is called *contraction* or *interior product* and is sometimes denoted by $X \lrcorner \alpha$ or $\iota_X \alpha$

A co-vector field α is smooth if $\forall\ X \in \mathfrak{X}(\mathcal{M}),\quad \alpha(X) \in C^\infty(\mathcal{M})$. These smooth co-vector fields are called *differential forms* or $1-forms$. The set of all 1-forms is denoted by $\Lambda^1(\mathcal{M})$ which is actually a module under the ring $C^\infty(\mathcal{M})$ such that:

$$(\alpha+\beta)_p \equiv \alpha_p + \beta_p \qquad (f\alpha)_p \equiv f(p)\alpha_p \alpha, \beta \in \Lambda^1(\mathcal{M})$$

[1] For now let's keep it real, without increasing the 'complex'ity.

[2] This is basically the covariant vectors from earlier. We will call contravariant vectors to be just 'vectors' and covariant vectors to be the 'covectors' to 'vectors'

From our previous discussion on the differential of a function at a point on the manifold, we can define a covector field df (differential of a function) such that $df(p) = df_p$. Now for $X \in \mathfrak{X}(\mathcal{M})$,

$$(df(X))(p) = df_p(X_p) = X_p[f] = (Xf)(p)$$

We obtain $df(X) \equiv Xf \in C^\infty(\mathcal{M})$ and hence df is a smooth vector field (1-form). We can say that the operator d changes the function f to a 1-form, that is,

$$d : C^\infty(\mathcal{M}) \to \Lambda^1(\mathcal{M}), \qquad f \mapsto df$$

Let $X \in \mathfrak{X}(\mathcal{M})$ and $f, g \in C^\infty(\mathcal{M}), a, b \in \mathbb{R}$. Then we have:

$$(d(af + bg))(X) = X(af+bg) = a\ Xf + b\ Xg = a\ df(X) + b\ dg(X) = (a\ df + b\ dg)(X)$$
$$\implies d((af + bg)) \equiv (a\ df + b\ dg)$$

Thus, d is a linear operator. Also, note that:

$$d(fg)(X) = X(fg) = f\ Xg+g\ Xf = f\ dg(X) + g\ df(X) = (f\ dg + g\ df)(X)$$
$$\implies d(fg) = f\ dg + g\ df$$

We had earlier mentioned of the dual basis $\{\phi^i\}$ such that, if v^i is a basis for V, then:

$$\phi^i(v_j) = \delta^i{}_j$$

If (U, ϕ) is a chart on $\mathcal{M}$ with coordinate functions x^i, then according to the differential of functions, taking $f = x^i$, we have:

$$\left(dx^i \left(\frac{\partial}{\partial x^j} \right) \right) \bigg|_p = \left(\frac{\partial}{\partial x^j} \right)_p [x^i] = \delta^i{}_j$$

We can see the similarity very clearly between the two facts. Thus, $\{dx^i_p\}$ provides a basis for the cotangent plane. Now, let us take $\alpha_p \in T^*_p\mathcal{M}$. Then we have:

$$\begin{aligned}
\alpha_p(v_p) &= \alpha_p\left(v_p[x^i] \left(\frac{\partial}{\partial x^i} \right)_p \right) \\
&= v_p[x^i]\alpha_p\left(\left(\frac{\partial}{\partial x^i} \right)_p \right) \qquad \because\ v_p[x^i] \text{ is a number} \\
&= \alpha_p\left(\left(\frac{\partial}{\partial x^i} \right)_p \right) dx^i_p(v_p) \qquad \because\ dx^i_p(v_p) = v_p[x^i] \\
&= \left(\alpha_p\left(\left(\frac{\partial}{\partial x^i} \right)_p \right) dx^i_p(v_p) \right)
\end{aligned}$$

From here we obtain an expansion of the covector:

$$\alpha_p = \alpha_p\left(\left(\frac{\partial}{\partial x^i} \right)_p \right) dx^i_p$$

Then in a similar way as expansion of vector field, we obtain:

$$\alpha = \alpha\left(\frac{\partial}{\partial x^i} \right) dx^i \equiv \alpha_i dx^i$$

Let us take the case of the differential of function. Then according to this,

$$df = df\left(\frac{\partial}{\partial x^i}\right)dx^i = \left(x^i\right)f = \frac{\partial f}{\partial x^i}dx^i$$

Damn, this is like EPIC!! 🔥 Looks exactly like the 'total derivative' thing on $\mathbb{R}^n$ that we are used to see. Well, half of the credit goes to the carefully selected notations which make this analogy successful!

Change of coordinate:

Let α be a covector field and let us denote a coordinate transformation $(x^1, \dots, x^n) \to (x'^1, \dots, x'^n)$. We shall now see how α'_j will be in terms of α_j

$$\alpha'_j = \alpha\left(\frac{\partial}{\partial x'^j}\right) = (\alpha_i dx^i)\left(\frac{\partial}{\partial x'^j}\right) = \alpha_i\left(\frac{\partial}{\partial x'^j}\right)(x^i) = \alpha_i\left(\frac{\partial x^i}{\partial x'^j}\right)$$

Another 🔥! We recover the transformation rule of covariant vectors, though, using a more sophisticated and elegant formalism.

8.5.1. Keep Integrating!!

Given a curve $\sigma : [a, b] \to \mathcal{M}$, the integral of a 1-form α over σ is defined by:

$$\int_\sigma \alpha = \int_a^b \alpha_{\sigma(t)}(\sigma'_t)dt$$

So, for a particular $t \in [a, b]$ mapped to a point $\sigma(t)$ on the manifold, α assigns the covector $\alpha_{\sigma(t)}$ which acts on the tangent vector at $\sigma(t)$, denoted by σ'_t, giving a number and then we integrate it. It so happens that $\int_\sigma \alpha$ depends only on the image of σ and the direction of traversal.

If $\alpha = df$ is a differential of a function f, then:

$$\int_\sigma df = \int_a^b df_{\sigma(t)}(\sigma'_t) = \int_a^b \sigma'_t[f]dt = \int_a^b \frac{\mathrm{d}(\sigma^* f)}{\mathrm{dt}}dt = (\sigma^* f)(b) - (\sigma^* f)(a) = f(\sigma(b)) - f(\sigma(a))$$

If σ is a closed curve, then $\sigma(b) = \sigma(a) \implies \int_\sigma df = 0$, often denoted by $\oint_\sigma df = 0$

Example:

Let $\mathcal{M} = \mathbb{R}\backslash\{(0,0)\}$ and $\alpha = \frac{xdy - ydx}{x^2+y^2}$. Let a curve $\mathcal{C} : [0, 2\pi] \to \mathcal{M}, \mathcal{C}(t) = (\cos t, \sin t)$. Then,

$$\int_{\mathcal{C}} \alpha = \int_0^{2\pi} \frac{\cos t \cdot \cos t dt - \sin t \cdot (-\sin t dt)}{\cos^2 t + \sin^2 t} = \int_0^{2\pi} dt = 2\pi \neq 0$$

From this, we can conclude that the given 1-form is not the differential of any function. However, note that, we had been studying the an expression of the form $M\ dx + N\ dy$ is exact if $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$. Here we have:

$$M = \frac{-y}{x^2 + y^2} \qquad N = \frac{x}{x^2 + y^2}$$

$$\frac{\partial M}{\partial y} = -\left(\frac{1}{x^2+y^2} - \frac{2y^2}{(x^2+y^2)^2}\right) = \frac{y^2-x^2}{(x^2+y^2)^2}$$
$$\frac{\partial N}{\partial x} = \left(\frac{1}{x^2+y^2} - \frac{2x^2}{(x^2+y^2)^2}\right) = \frac{y^2-x^2}{(x^2+y^2)^2}$$

We see that in this case, the partial derivatives are indeed equal but it is not an exact differential. In fact, the equality of the partial derivatives is a necessary condition but not sufficient for differential to be exact. If the manifold is in form of $\mathbb{R}^n$, then only it becomes exact, but for any other type of manifold, equality of partial derivatives is, in general, not sufficient.

8.6. Tensors

We had earlier defined to be a multi-linear map, blah blah. We will now focus on tensors of type $(0,k)$ which is also called a covariant tensor of rank k. Thus,

$$t_p = \bigotimes_{i=1}^{k} T_p\mathcal{M} \to \mathbb{R}$$

So, it will take k tangent vectors and map it to a real number. It must also satisfy multilinearity, that is, linearity in each argument:

$$t_p(\dots, a\ v_p + b\ w_p, \dots) = a\ t_p(\dots, v_p, \dots) + b\ t_p(\dots, w_p, \dots)$$

The set of all $(0,k)$ tensors denoted by $T_p^{(0,k)}\mathcal{M}$ forms a real vector space. Let $a, b \in \mathbb{R}$ and $t_p, s_p \in T_p^{(0,k)}\mathcal{M}$, then for all $v_{ip} \in T_p\mathcal{M},\ i = 1, 2, \dots, k$, we have:

$$(a\ t_p + b\ s_p)(v_{1p}, \dots, v_{kp}) \equiv a\ t_p(v_{1p}, \dots, v_{kp}) + b\ s_p(v_{1p}, \dots, v_{kp})$$

To check that this is indeed another tensor, we need to check for multilinearity. For that,

- Let $v_p, w_p \in T_p\mathcal{M}$ be two vectors. Then,

$$\begin{aligned}(at_p + bs_p)(v_{1p}\dots, v_{ip} + w_{ip}\dots v_{kp}) &\equiv a\ t_p(v_{1p}\dots, v_{ip} + w_{ip}\dots v_{kp}) + b\ s_p(v_{1p}\dots v_{ip} + w_{ip}\dots v_{kp}) \\ &= a\ (t_p(v_{1p}\dots v_{ip}\dots v_{kp}) + t_p(v_{1p}\dots, w_{ip}\dots v_{kp})) \\ &\qquad + b\ (s_p(v_{1p}\dots v_{ip}\dots v_{kp}) + s_p(v_{1p}\dots w_{ip}\dots v_{kp})) \\ &= (at_p + bs_p)(v_{1p}\dots v_{ip}\dots v_{kp}) + (at_p + bs_p)(v_{1p}\dots, w_{ip}\dots v_{kp})\end{aligned}$$

- Let $c \in \mathbb{R}, v_{ip} \in T_p\mathcal{M}$. Then,

$$\begin{aligned}(at_p + bs_p)(v_{1p}\dots, cv_{ip}\dots v_{kp}) &\equiv a\ t_p(v_{1p}\dots, cv_{ip}\dots v_{kp}) + b\ s_p(v_{1p}\dots, cv_{ip}\dots v_{kp}) \\ &= c(a\ t_p(v_{1p}\dots, v_{ip}\dots v_{kp}) + b\ s_p(v_{1p}\dots, v_{ip}\dots v_{kp})) \\ &= c(at_p + bs_p)(v_{1p}\dots, v_{ip}\dots v_{kp})\end{aligned}$$

Thus, from these two conditions, we can conclude the condition for multi-linearity.

8.6.1. Tensor Product

We can form a product of two tensors, not neccessarily of the same rank. For that, let $t_p \in T_p^{(0,k)}\mathcal{M}, s_p \in T_p^{(0,l)}\mathcal{M}$. Then the tensor product is defined as:

$$(t_p \otimes s_p)(v_{1p}, \dots, v_{k+l,p}) \equiv t_p(v_{1p}, \dots, v_{kp})s_p(v_{k+1,p}, \dots, v_{k+l,p})$$

Using this, we can say that

$$t_p \otimes s_p : \bigotimes_{i=1}^{k+l} T_p\mathcal{M} \to \mathbb{R}$$

And checking for multilinearity will be easy (as well as boring). Hence we directly declare that it is a $(k+l)$ rank covariant tensor.

Properties of Tensor Product:

- Tensor product is distributive.

  Proof. Let us take three covariant tensors t_1, t_2 of rank k and s of rank l at point p and let $(v_1, \dots, v_{k+1})$ be a vector. Then,

$$\begin{aligned}
&((at_1 + bt_2) \otimes s)(v_1, \dots, v_{k+1}) \\
&= (at_1 + bt_2)(v_1, \dots, v_k)s(v_{k+1}, \dots, v_{k+l}) \\
&= (at_1(v_1, \dots, v_k) + bt_2(v_1, \dots, v_k))s(v_{k+1}, \dots, v_{k+l}) \\
&= at_1(v_1, \dots, v_k)s(v_{k+1}, \dots, v_{k+l}) + bt_2(v_1, \dots, v_k)s(v_{k+1}, \dots, v_{k+l}) \\
&= (a(t_1 \otimes s) + b(t_2 \otimes s))(v_1, \dots, v_{k+1})
\end{aligned}$$

  Similarly, we can prove that if it had been a linear combination in the right tensor $\sim s_1 + s_2$ and hence we conclude the distributive property.

- Tensor product is associative.

  Proof. Let $r \in T_p^{(0,k)}\mathcal{M}, s \in T_p^{(0,l)}\mathcal{M}, t \in T_p^{(0,m)}\mathcal{M}$. Then r⊗(s⊗t) and (r⊗s)⊗t are both $(k+l+m)$ rank tensor. Now, we have:

$$\begin{aligned}
(r \otimes (s \otimes t))(v_1, \dots, v_{k+l+m}) &= r(v_1, \dots, v_k)(s \otimes t)(v_{k+1}, \dots, v_{k+l+m}) \\
&= r(v_1, \dots, v_k)s(v_{k+1}, \dots, v_{k+l})t(v_{k+l+1}, \dots, v_{k+l+m}) \\
((r \otimes s) \otimes t)(v_1, \dots, v_{k+l+m}) &= (r \otimes s)(v_1, \dots, v_{k+l})t(v_{k+l+1}, \dots, v_{k+l+m}) \\
&= r(v_1, \dots, v_k)s(v_{k+1}, \dots, v_{k+l})t(v_{k+l+1}, \dots, v_{k+l+m})
\end{aligned}$$

  Since both these expressions are equal, we proved the associativity.

8.6.2. Sometimes all that we need is a coordinate...

Remember that any tangent vector can be written as:

$$v_p = v_p[x^i]\left(\frac{\partial}{\partial x^i}\right)_p = dx_p^i(v_p)\left(\frac{\partial}{\partial x^i}\right)_p$$

Then we can write the action of the tensor in the following way:

$$\begin{aligned}
t_p(v_1, \dots v_j \dots, v_k) &= t_p\left(dx_p^{i_1}(v_1)\left(\frac{\partial}{\partial x^{i_1}}\right)_p, \dots dx_p^{i_j}(v_j)\left(\frac{\partial}{\partial x^{i_j}}\right)_p\right) \dots dx_p^{i_k}(v_k)\left(\frac{\partial}{\partial x^{i_k}}\right)_p \\
&= t_p\left(\left(\frac{\partial}{\partial x^{i_1}}\right)_p, \dots \left(\frac{\partial}{\partial x^{i_j}}\right)_p \dots \left(\frac{\partial}{\partial x^{i_k}}\right)_p\right) \prod_{m=1}^{k} dx_p^{i_m}(v_m)
\end{aligned}$$

The second line follows from the multilinearity of tensors. Also, note that $\prod_{m=1}^{k} dx_p^{i_m}(v_m)$ can be written as $\left(\bigotimes_{m=1}^{k} dx_p^{i_m}\right)(v_1, \dots, v_k)$ (basically, each $dx_p^{i_m}$ acts on the corresponding v_m and we get the product thingy). Thus, we obtain:

$$t_p(v_1, \dots v_j \dots, v_k) = \left(t_{p,i_1 i_2 \dots i_k} \bigotimes_{m=1}^{k} dx_p^{i_m} \right)(v_1, \dots, v_k)$$

where,

$$t_{p,i_1 i_2 \dots i_k} = t_p\left(\left(\frac{\partial}{\partial x^{i_1}} \right)_p, \dots \left(\frac{\partial}{\partial x^{i_j}} \right)_p \dots \left(\frac{\partial}{\partial x^{i_k}} \right)_p \right)$$

Note that this is not a single quantity. Owing to Einstein's convention, the sum over all the indices has been nicely hidden from the above expression (and even then it looks so ghastly! 💀). However, there are indeed n^k sums and hence these are n^k numbers, which are the coordinates of the tensor t_p in the coordinate system. [1]

From this, we finally have:

$$t_p = t_{p,i_1 i_2 \dots i_k} \bigotimes_{m=1}^{k} dx_p^{i_m}$$

Now, suppose we have two set of coordinate functions (that is, we have two charts (U, ϕ) and (U', ϕ') and we consider $p \in U \bigcup U'$) given by $(x^1, \dots, x^n)$ and $(x'^1, \dots, x'^n)$. Then we can write:

$$t_p = t_{p,i_1 i_2 \dots i_k} \bigotimes_{m=1}^{k} dx_p^{i_m} = t'_{p,i_1 i_2 \dots i_k} \bigotimes_{m=1}^{k} dx_p'^{i_m}$$

Now,

$$\begin{aligned}
t'_{p,i_1 i_2 \dots i_k} &= t_p\left(\left(\frac{\partial}{\partial x'^{i_1}} \right)_p, \dots \left(\frac{\partial}{\partial x'^{i_j}} \right)_p \dots \left(\frac{\partial}{\partial x'^{i_k}} \right)_p \right) \\
&= \left(t_{p,j_1 j_2 \dots j_k} \bigotimes_{m=1}^{k} dx_p^{j_m} \right)\left(\left(\frac{\partial}{\partial x'^{i_1}} \right)_p, \dots \left(\frac{\partial}{\partial x'^{i_j}} \right)_p \dots \left(\frac{\partial}{\partial x'^{i_k}} \right)_p \right) \\
&= t_{p,j_1 j_2 \dots j_k} \prod_{m=1}^{k} dx_p^{j_m}\left(\left(\frac{\partial}{\partial x'^{i_m}} \right)_p \right)
\end{aligned}$$

Now, note the following notational jugglery: $dx_p^{j_q}\left(\left(\frac{\partial}{\partial x'^{i_q}} \right)_p \right) \equiv \left(\frac{\partial}{\partial x'^{i_q}} \right)_p [x^{j_q}] \equiv \left(\frac{\partial x^{j_q}}{\partial x'^{i_q}} \right)_p$ Using this we again have something nice:

$$t'_{p,i_1 i_2 \dots i_k} = t_{p,j_1 j_2 \dots j_k} \prod_{m=1}^{k} \left(\frac{\partial x^{j_m}}{\partial x'^{i_m}} \right)_p$$

We obtain the transformation rule for covariant tensors (yes, the rule that was apparently a definition of tensors, has now been derived explicitly, just from a simple definition of tensor as a multi-linear map!).

8.6.3. Tensor Field

A $\begin{pmatrix} 0 \\ k \end{pmatrix}$ tensor field t, like vector field, assigns a tensor to each point $p \in \mathcal{M}$. If t is a tensor field and $X_1, \dots, X_k$ are vector fields, then $t(X_1, \dots, X_k)$ is a function whose value at $p \in \mathcal{M}$ is:

$$t(X_1, \dots, X_k)(p) = t_p(X_{1p}, \dots, X_{kp})$$

[1] Physicists do all sort of weird stuffs with these components only. For them, tensor means these components.

If $X_i \in \mathfrak{X}(\mathcal{M}) \implies t(X_1, \ldots, X_k) \in C^\infty(M)$, then it is called a differentiable tensor field.

Now, we define the components of a tensor field as the n^k functions:

$$t_{i_1 \ldots i_k} \equiv t\Big(\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big)$$

Then using this definition we can write:

$$\begin{aligned} t(X_1, \ldots, X_k) &= t\Big(X_1^{i_1}\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, X_k^{i_k}\Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big) \\ &= t\Big(\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big) X_1^{i_1} \ldots X_k^{i_k} \\ &= t_{i_1 \ldots i_k}\, X_1^{i_1} \ldots X_k^{i_k} \end{aligned}$$

We can now say that the tensor field is differentiable iff the component functions are smooth.

Some definitions...

If s, t are $\begin{pmatrix} 0 \\ k \end{pmatrix}$ tensor fields, $a, b \in \mathbb{R}$ and f is a function on $\mathcal{M}$, we define,

$$(at + bs)_p := at_p + bs_p \qquad (ft)_p = f(p)t_p$$

If m is a $\begin{pmatrix} 0 \\ k \end{pmatrix}$ tensor field and r is a $\begin{pmatrix} 0 \\ l \end{pmatrix}$ tensor field, then $(m \otimes r)_p \equiv m_p \otimes r_p$ is a $\begin{pmatrix} 0 \\ k+l \end{pmatrix}$ tensor field.

$$\begin{aligned} t(p) = t_p &= t_p\left(\Big(\frac{\partial}{\partial x^{i_1}}\Big)_p, \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)_p\right)\, dx_p^{i_1} \otimes \ldots \otimes dx_p^{i_k} \\ &= t\Big(\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big)(p)\, dx^{i_1}(p) \otimes \ldots \otimes dx^{i_k}(p) \\ &= \Big(t\Big(\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big)\, dx^{i_1} \otimes \ldots \otimes dx^{i_k}\Big)(p) \end{aligned}$$

Thus, we obtain:

$$t \equiv t\Big(\Big(\frac{\partial}{\partial x^{i_1}}\Big), \ldots, \Big(\frac{\partial}{\partial x^{i_k}}\Big)\Big)\, dx^{i_1} \otimes \ldots \otimes dx^{i_k}$$

We can show that the tensor field is a $\mathcal{C}^\infty$ multi-linear map, that is, $f, g : \mathcal{M} \to \mathbb{R}$ and $\{X_i\}_{i=1}^k, Y_i$ are vector fields, then:

$$t(X_1, \ldots fX_i + gY_i, \ldots X_k) = ft(X_1, \ldots X_i, \ldots X_k) + gt(X_1, \ldots Y_i, \ldots X_k)$$

The converse of this statement is also true, that is, any map t following this property, is a $\begin{pmatrix} 0 \\ k \end{pmatrix}$ tensor field.

8.7. Riemannian Geometry

Okay, so whatever BS we had been seeing earlier (I admit, I got a bit carried away with diff geo part since it was so interesting), we will now use that to see our previous discussions on tensors in a more stronger form. We will re-discuss metrics, connections, covariant derivatives, etc. but now in a more sophisticated language (which goes against my initial intentions).

Appendices

A. Equivalence Relations and Equivalence Classes

B. Vector Space

C. Groups

A group is a

D. Ring

A structure $(R, +, \cdot)$ is a ring, with $R \neq \emptyset$ and $+$ and $\cdot$ two binary operations such that:

- **Addition**: $(R, +)$ is an abelian group that is associative, has a zero element, an inverse and is commutative.
- **Multiplication**: is associative.
- **Distribution:** $(a + b) \cdot c = a \cdot c + b \cdot c$ and $a \cdot (b + c) = a \cdot b + a \cdot c$

Well, since multiplication is not necessarily commutative in a ring, the distributive law is postulated as two laws.

E. Module

Let R be a ring. Then a *left R-module* is an abelian group $(M, +)$ together with a map $\cdot : R \times M \to M$ satisfying:

- $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y \; \forall \; \alpha \in R$ and $x, y \in M$
- $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot y \; \forall \; \alpha, \beta \in R$ and $x \in M$
- $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$
- $\mathbb{1} \cdot x = x$

The elements of the ring are sometimes called as *scalars*. Since the scalars always appear on the left, it is called a left module. We can similarly define the right module but by module, we will always mean left module.

F. Algebra

An algebra over a field K is a vector space A with a multiplication map $\mu : A \times A \to A$ denoted by $\mu(a, b) = a \cdot b$, such that for all $a, b, c \in A$ and $r \in K$ we have:

- Associativity: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- Right and Left Distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$
- Homogenity: $r(a \cdot b) = (ra) \cdot b = a \cdot (rb)$

The associativity property is not always required, though. If the binary operation is associative too, then we call it an *associative algebra*, otherwise *non-associative*.

The above conditions, apart from associativity, is equivalent of saying that the algebra satisfies the *bilinearity property*, that is, for $m, n \in K$ and $a, b \in A$:

Linearity in first factor:	$(m\ a + n\ b) \cdot c = m\ a \cdot c + n\ b \cdot c$
Linearity in second factor:	$c \cdot (m\ a + n\ b) = m\ c \cdot a + n\ c \cdot b$

If A, A' are algebras over field K, then an algebra homeomorphism is a linear map $\mathcal{L} : A \to A'$ such that multiplication is preserved, that is, $\mathcal{L}(ab) = \mathcal{L}(a)\mathcal{L}(b)\ \forall\ a, b \in A$.