CHAPTER 4

Multivariable calculus

4.1. Smooth functions and changes of coordinates

4.1.1. Smooth functions. Let Ω be an open set of $\mathbb{R}^n$. 'Open' means that Ω is a possibly infinite union of *open balls*. The open balls of $\mathbb{R}^n$ are all of the form

$$B(a,r) = \{x \in \mathbb{R}^n \text{ such that } |x-a| < r\} \tag{4.1.1}$$

where a is any point of $\mathbb{R}^n$ and $r > 0$. It is the strict inequality in (4.1.1) that makes $B(a,r)$ an *open* ball.

Let $f : \Omega \to \mathbb{R}$ be a real-valued function in Ω. I assume you know what partial derivatives are.

DEFINITION 4.1.1. A function f is smooth, also written C^∞, if all partial derivatives, of any order, exist. That is, for any non-negative integers, $\alpha_1, \ldots, \alpha_n$,

$$\left(\frac{\partial}{\partial x^1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x^n}\right)^{\alpha_n} f(p) \tag{4.1.2}$$

exists for every point p of Ω. The set of all functions which are smooth in Ω is denoted by $C^\infty(\Omega)$. Then $C^\infty(\Omega)$ is an infinite-dimensional vector space.

REMARK 4.1.2. The *order* of the partial derivative in (4.1.2) is $\alpha = \alpha_1 + \cdots + \alpha_n$.

Recall that for smooth functions, partial differentiation with respect to different variables 'commutes' in the sense that

$$\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j} f(x) = \frac{\partial}{\partial x^j}\frac{\partial}{\partial x^i} f(x) = \tag{4.1.3}$$

for all $1 \leqslant i, j \leqslant n$.

REMARK 4.1.3. Smooth functions model 'scalar' physical quantities such as density, pressure, charge-density, temperature,...

4.1.2. Changes of coordinates. We have tacitly taken the $(x^1, \ldots, x^n)$ to be standard linear coordinates in $\mathbb{R}^n$ (i.e. associated by the standard basis of $\mathbb{R}^n$). It is fairly clear that the idea of 'smoothness' of functions should be independent of the choice of such linear coordinates, but here we are going to take things further by considering more general coordinate systems. As far as GR goes, this is a requirement of trying to make a theory that is 'generally covariant' (i.e. transforms predictably under general changes of coordinates).

So, what are changes of coordinates?

DEFINITION 4.1.4. A change of coordinates written in compact form $x = x(y)$ is a collection of smooth functions

$$\begin{array}{ccc}
x^1 & = & x^1(y^1, \ldots, y^n) \\
x^2 & = & x^2(y^1, \ldots, y^n) \\
\cdots & \cdots & \cdots \\
x^n & = & x^n(y^1, \ldots, y^n)
\end{array}$$

with the further properties:

- $x = x(y)$ gives a 1:1 correspondence between points $x \in \Omega$ and points y in some other open set Ω';

- The corresponding inverse map $y = y(x)$ is also smooth, i.e.
$$y^j = y^j(x^1, \ldots, x^n)$$
is smooth for each $j = 1, \ldots, n$.

EXAMPLE 4.1.5. Any affine linear transformation,
$$y^j = \sum L^j_k x^k + A^j \tag{4.1.4}$$
is invertible if the matrix L is invertible. (In this case, the Jacobian of the transformation is L.)

EXAMPLE 4.1.6. Plane polar coordinates.
$$x = r\cos\theta, y = r\sin\theta.$$
(To make this look like the above, write it as
$$x^1 = y^1 \cos y^2,\ x^2 = y^1 \sin y^2,$$
i.e. $(x, y) = (x^1, x^2)$, $(r, \theta) = (y^1, y^2)$.) The Jacobian is
$$\begin{bmatrix} x_r & x_\theta \\ y_r & y_\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$
This gives a change of coordinates between
$$\Omega = \{(x^1, x^2) \text{ such that} (x^1, x^2) \neq (t, 0) \text{ with } t \geqslant 0\}$$
and
$$\Omega' = \{(y^1, y^2) \text{ such that } y^1 > 0 \text{ and } 0 < y^2 < 2\pi\}$$

REMARK 4.1.7. Make sure you understand why we have to restrict the values of (y^1, y^2) to get a change of coordinates.

We can think of a change of coordinates more actively as follows. Given a function f in $C^\infty(\Omega)$, we get a new function $\widetilde{f} \in C^\infty(\Omega')$ by the formula
$$\widetilde{f}(y) = f(x(y)) \tag{4.1.5}$$
Similarly, given $\widetilde{g} \in C^\infty(\Omega')$ we get a new function $g \in C^\infty(\Omega)$ by the formula
$$g(x) = \widetilde{g}(y(x)). \tag{4.1.6}$$

REMARK 4.1.8. The fancy terminology for this is 'pull-back': $\widetilde{f}$ is obtained from f by pulling back by the change-of-coordinates map $\Omega' \to \Omega$. If you find this helpful (because you've seen it elswhere) fine. If you haven't, don't worry.

The fact that $\widetilde{f}$ is smooth if f is smooth follows from the chain rule:
$$\frac{\partial}{\partial y^j}\widetilde{f}(y) = \sum_{i=1}^n \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial x^i} f(x) \text{ where } x = x(y), \tag{4.1.7}$$
which I assume you've seen before and are happy with. The matrix whose components are $(\partial x^j/\partial y^i)$ entering in (4.1.7) is called the *Jacobian* of the coordinate transformation $x = x(y)$.

The Jacobian matrix $(\partial y^j/\partial x^i)$ of the inverse transformation $y = y(x)$ is the inverse of the Jacobian matrix $(\partial x^j/\partial y^i)$. As an exercise, you can verify this by using the chain rule to differentiate the equation
$$x^j(y(x)) = x^j \text{ for } j = 1, \ldots, n \tag{4.1.8}$$
which is true by definition of the transformations being inverse to each other. In particular for a change of coordinates, the Jacobian matrix must be invertible everywhere.

The explicit forms of these inverse relationships are
$$\sum_{j=1}^n \frac{\partial x^k}{\partial y^j}\frac{\partial y^j}{\partial x^i} = \delta^k_i,\ \ \sum_{j=1}^n \frac{\partial y^k}{\partial x^j}\frac{\partial x^j}{\partial y^i} = \delta^k_i. \tag{4.1.9}$$

REMARK 4.1.9. The *inverse function theorem* says that if $x = x(y)$ is just smooth for $y \in \Omega'$ and $x \in \Omega$ and if the Jacobian matrix if *invertible* at a point q, say, in Ω', then in fact $x = x(y)$ is invertible at least if you restrict the transformation to a small ball B' containing q inside Ω' and its image $W \subset \Omega$. That is, after restricting in this way, there is smooth $y = y(x)$, for $x \in W$ such that $y(x) \in B'$ inverting $x = x(y)$.

Thus the inverse function is a (partial) converse to the fact that Jacobians of coordinate transformations must be invertible.

EXAMPLE 4.1.10. In the case of polar coordinates, the determinant of the Jacobian is just r. This is invertible if and only if $r \neq 0$. This ties in with the fact that polar coordinates go wrong at $r = 0$.

4.2. Two types of vector

Many physical quantities are 'vectorial', as we know. We now consider vectorial quantities in the context of general coordinate transformations. A major subtlety is that there are *two different kinds* of vectorial quantities and we need to be clear on the difference between them.

4.2.1. Vector fields. A vector field in Ω is smooth first-order differential operator of the form

$$V = \sum_{j=1}^{n} V^j(x) \frac{\partial}{\partial x^j} \tag{4.2.1}$$

where the $V^j(x)$ are smooth functions in Ω. If $f \in C^\infty(\Omega)$ we obtain a new function $Vf \in C^\infty(\Omega)$ called the *derivative of* f along V,

$$Vf(x) = \sum_{j=1}^{n} V^j(x) \frac{\partial f}{\partial x^j}(x) \tag{4.2.2}$$

A vector field is also known as a tensor (field) of type $(1, 0)$.

4.2.2. Covector fields. A covector field in Ω is a quantity

$$\omega = \sum_{j=1}^{n} \omega_j(x)\, \mathrm{d}x^j \tag{4.2.3}$$

where the $\omega_j(x)$ are smooth functions in Ω.

If $f \in C^\infty(\Omega)$ we obtain a *covector field*

$$\mathrm{d}f = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, \mathrm{d}x^j \tag{4.2.4}$$

also known as the exterior derivative or differential of f. To match up the notation between (4.2.3) and (4.2.4), $\mathrm{d}f$ is the covector whose components are $\omega_j = \partial f/\partial x^j$.

REMARK 4.2.1. If ω is a covector field in Ω it is not generally true that $\omega = \mathrm{d}f$ for some function f. (Indeed the condition

$$\frac{\partial \omega_i}{\partial x^j} = \frac{\partial \omega_j}{\partial x^i}$$

is necessary for $\omega = \mathrm{d}f$, by (4.1.3).)

A covector field is also known as a tensor (field) of type $(0, 1)$.

4.2.3. Dual pairing (or contraction). Given a vector field V and a covector field ω, the *contraction* $\langle V, \omega \rangle$ is a scalar function defined in terms of components by

$$\langle V, \omega \rangle = \sum_{j=1}^{n} V^j(x) \omega_j(x). \tag{4.2.5}$$

The directional derivative is an example of this contraction: from the above formulae,

$$\langle V, \mathrm{d}f \rangle = Vf \tag{4.2.6}$$

4.2.4. Transformation laws for vector fields and covector fields. Let V be a vector field in Ω and let $x = x(y)$ (for $y \in \Omega'$) be a change of coordinates with inverse $y = y(x)$. We get a vector field $\widetilde{V}$ in Ω' as follows: $\widetilde{V}$ is supposed to differentiate functions in Ω'. We know how to differentiate functions in Ω. But we can transfer a function in Ω' to Ω by change of variables. So we *define*

$$\widetilde{V}\widetilde{g}\,(y) = (Vg)(x(y)) \tag{4.2.7}$$

where g and $\widetilde{g}$ are related as in (4.1.6).

PROPOSITION 4.2.2. *If $V = \sum V^j(\partial/\partial x^j)$ in terms of the x^j in Ω and $\widetilde{V} = \sum \widetilde{V}^j(\partial/\partial y^j)$ in terms of the y^j in Ω', then*

$$\widetilde{V}^i = \sum V^j \frac{\partial y^i}{\partial x^j} \tag{4.2.8}$$

PROOF. The chain rule tells us everything: from $g(x) = \widetilde{g}(y(x))$, we get

$$\frac{\partial g}{\partial x^j} = \sum_i \frac{\partial y^i}{\partial x^j}\frac{\partial \widetilde{g}}{\partial y^i}.$$

Multiplying by V^j and summing,

$$\sum_j V^j \frac{\partial g}{\partial x^j} = \sum_{i,j} V^j \frac{\partial y^i}{\partial x^j}\frac{\partial \widetilde{g}}{\partial y^i}.$$

But the RHS is supposed to be $\sum_i \widetilde{V}^i(\partial\widetilde{g}/\partial y^i)$, so the result follows by equating coefficients. $\square$

REMARK 4.2.3. This is an example of covariance (as opposed to invariance). The coefficients of a vector depend on a choice of coordinates, but they transform in a predictable and linear way. In particular *if the coefficients are all zero at a given point in one coordinate system then they are also zero in any other coordinate system.* This is as it should be: if there is no wind at a particular point (and time) in the atmosphere, then all observers should agree on this fact, regardless of how they choose their coordinates!

The classical formula

$$\mathrm{d}x^j = \sum_i \frac{\partial x^j}{\partial y^i}\,\mathrm{d}y^i \tag{4.2.9}$$

suggests that if $\sum_j \widetilde{\omega}_j \mathrm{d}y^j$ is to agree with $\sum_j \omega_j \mathrm{d}x^j$, then we should have

$$\sum_j \omega_j \mathrm{d}x^j = \sum_{i,j} \omega_j \frac{\partial x^j}{\partial y^i}\,\mathrm{d}y^i$$

and equating coefficients

$$\widetilde{\omega}_i = \sum_j \omega_j \frac{\partial x^j}{\partial y^i}. \tag{4.2.10}$$

Note that even allowing for the difference between upstairs and downstairs indices, (4.2.8) and (4.2.10) are *different* transformation laws.

The rule (4.2.10) gives a way of transferring a covector field ω in Ω to a new covector field $\widetilde{\omega}$ in Ω'. We already have a rule for transferring vector fields from Ω to Ω'. These are compatible in the sense that the contraction is *invariant*:

PROPOSITION 4.2.4. *Let ω and V be a covector field and a vector field on Ω and let $\widetilde{\omega}$ and $\widetilde{V}$ be the corresponding covector field and vector field on Ω'. Then we have*

$$\langle \widetilde{V}, \widetilde{\omega}\rangle = \langle V, \omega\rangle \tag{4.2.11}$$

where the LHS is calculated at y in Ω' and the RHS at $x = x(y)$ Ω.

PROOF. This follows at once from (4.2.8) and (4.2.10) for

$$\langle \widetilde{V}, \widetilde{\omega} \rangle = \sum_i \widetilde{V}^i \widetilde{\omega}_i = \sum_{i,j} V^j \frac{\partial y^i}{\partial x^j} \widetilde{\omega}_i. \tag{4.2.12}$$

Using the fact that the Jacobians $(\partial y^i/\partial x^j)$ and $(\partial x^i/\partial y^j)$ are inverse to each other, (4.2.10) is seen to be equivalent to

$$\omega_j = \sum_i \widetilde{\omega}_i \frac{\partial y^i}{\partial x^j}. \tag{4.2.13}$$

Hence the RHS of (4.2.12) can be written

$$\sum_{i,j} V^j \frac{\partial y^i}{\partial x^j} \widetilde{\omega}_i = \sum V^j \omega_j = \langle V, \omega \rangle.$$

This completes the proof. □

REMARK 4.2.5. It is possible to change the logic around: given the relation between V and $\widetilde{V}$ which is natural in terms of the way vector fields are supposed to differentiate functions, we could have *defined* $\widetilde{\omega}$ in terms of ω so that (4.2.11) holds. This would then have implied the transformation law (4.2.10) for the coefficients of the covector field ω.

4.2.5. Tangent space and cotangent space. We shall not make great use of the following, but they are really important in a more systematic development of these ideas.

If p is a point of Ω then $T_p\Omega$, the tangent space to Ω at p, is defined to be the space of all directional derivatives *acting at* p. A typical element of $T_p\Omega$ is thus written

$$V = \sum_{j=1}^n V^j \left. \frac{\partial}{\partial x^j} \right|_p \tag{4.2.14}$$

where the V^j are just numbers and by definition

$$\left. \frac{\partial}{\partial x^j} \right|_p f = \frac{\partial f}{\partial x^j}(p) \tag{4.2.15}$$

(i.e. differentiate the function then evaluate it at p).

Then $T_p\Omega$ is a vector space of dimension n (the dimension of the vector space in which Ω is sitting as an open set) and is independent of coordinates. One should think of $T_p\Omega$ as the set of arrows emanating from p, and pointing in every possible direction in Ω. The coordinate independence follows from equation (4.2.10).

Similarly the *cotangent space* $T_p^*\Omega$ is the dual vector space to $T_p\Omega$; by definition this is the space of linear maps $T_p\Omega \to \mathbb{R}$ and a typical element has the form

$$\omega = \sum_{j=1}^n \omega_j \mathrm{d}x^j|_p. \tag{4.2.16}$$

This is also an n-dimensional vecto space, independent of choice of coordinates. The duality between $T_p\Omega$ and $T_p^*\Omega$ is given by

$$\langle V, \omega \rangle_p = \sum V^j \omega_j \tag{4.2.17}$$

as above but now producing a number rather than a function on the RHS.

4.3. The Einstein summation convention

In the above (and in the previous chapters), there are many expressions involving a summation over repeated indices, one upstairs and one downstairs. The Einstein summation convention is to omit the Σ symbol, so that whenever a repeated index appears in an expression it is to be understood that you sum over the range of that index (in this case from 1 to n).

In order for this to work, of course, it is essential that if an index is repeated then it must not occur anywhere else in the expression, so for example

$$A_iB^iC_i \text{ is not OK.}$$

Multiple sums (unfortunately) very often occur. For instance, if L and $\tilde{L}$ are matrices with components L^i_j and $\tilde{L}^i_j$ so that

$$LX \text{ has components } \sum_j L^i_j X^j = L^i_j X^j \text{ (summation convention)}$$

and

$$\tilde{L}X \text{ has components } \sum_j \tilde{L}^i_j X^j = \tilde{L}^i_j X^j$$

Then

$$L\tilde{L}X \text{ has components } L^i_p[\tilde{L}X]^p = L^i_p[\tilde{L}^p_q X^q] = L^i_p \tilde{L}^p_q X^q. \tag{4.3.1}$$

When the summation convention is in operation, repeated indices are *dummy indices* in the sense that

$$A_iB^i = A_pB^p = A_sB^s$$

as each of these is equal to

$$A_1B^1 + A_2B^2 + \cdots + A_nB^n.$$

The expression for the components of $L\tilde{L}X$ in (4.3.1) is unpacked as

$$\sum_{p=1}^n \sum_{q=1}^n L^i_p \tilde{L}^p_q X^q$$

and the summation over p corresponds to matrix multiplication of L and $\tilde{L}$ while the summation over q corresponds to the multiplication of $\tilde{L}$ by the column vector with components X^i.

Here is an example where you have to be careful to change the dummy indices to get an unambigous expression. Suppose α and β are covector fields and X and Y are vector fields. Consider

$$P = \langle X, \alpha\rangle\langle Y, \beta\rangle \tag{4.3.2}$$

We can write

$$\langle X, \alpha\rangle = X^i\alpha_i, \quad \langle Y, \beta\rangle = Y^i\beta_i.$$

Substitution of these into (4.3.2) give

$$P = X^i\alpha_iY^i\beta_i$$

However this is an ambiguous expression because the index i has been overworked, appearing 4 times. So before putting them together we should change one of the dummy indices, writing (say)

$$Y^i\beta_i = Y^j\beta_j.$$

Thus

$$P = X^i\alpha_iY^j\beta_j$$

is an unambigous way to write P using the summation convention.

Example 4.3.1. Write the expression

$$X^i\alpha_jY^j\beta_i$$

without indices, in terms of the pairing operation $\langle\cdot,\cdot\rangle$.

Definition 4.3.2. The Kronecker δ has components δ^j_k, equal to 1 if $j = k$ and 0 otherwise. This is the representation, in terms of indices, of the identity matrix.

Remark 4.3.3. If, later on in this chapter or the course, you find the expressions with multiple repeated indices confusing, it can help to put the Σ signs back in. The summation convention does take some getting used to.

4.4. Differentiation along a curve

Let $\Gamma(\tau)$ be a curve in Ω. That is, Γ is a smooth map from an interval $I = (\tau_1, \tau_2)$, say, into Ω.

If $f \in C^\infty(\Omega)$, then we get a function of τ

$$F(\tau) = f(\Gamma(\tau)). \tag{4.4.1}$$

This is 'the function f along the curve'. If Γ is the world-line of an observer and f is some physical quantity (like pressure) then $F(\tau)$ would be the pressure measured by the observer at different times along her worldline.

If the components of Γ are $(x^1(\tau), \ldots, x^n(\tau))$ as before, then we compute

$$\frac{\mathrm{d}F}{\mathrm{d}\tau} = \dot{x}^j(\tau)\frac{\partial f}{\partial x^j} \tag{4.4.2}$$

so that Γ defines the vector field

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}\tau} = \dot{x}^j\frac{\partial}{\partial x^j} \tag{4.4.3}$$

along Γ. Here the LHS will be used as short-hand for the RHS!

In contrast to the vector fields we've considered before, this one is only defined along the curve Γ and not in an open set Ω. Notice that the definition (4.4.2) is independent of any choice of coordinates and gives a suitably invariant definition of tangent vector to the curve Γ.

In so far as Γ is a mapping from a subset of $\mathbb{R}$ into a subset of V, its derivative $\dot{\Gamma}(\tau)$ is a mapping from I into V. Again, it is better to regard $\dot{\Gamma}(\tau)$ as being in the tangent space to $\Gamma(\tau)$,

$$\dot{\Gamma}(\tau) \in T_{\Gamma(\tau)}\Omega.$$

Then $\dot{\Gamma}(\tau)$ is called the tangent vector to Γ at the point $\Gamma(\tau)$

4.5. Tensor fields of rank 2

Higher order (or higher rank) tensors are, from the naive point of view, objects with more indices, upstairs or downstairs, or both.

We have already seen examples of such objects with two indices at least in the context of vector spaces. First of all, a bilinear form on $\mathbb{R}^n$ is an object with two lower indices. We have seen that if we choose a basis of $\mathbb{R}^n$, such that its elements are identified with column vectors with components X^j, then a bilinear form B has components B_{ij}, such that

$$B(X,Y) = B_{ij}X^iY^j \text{ (summation convention).}$$

4.5.1. Tensor fields of type $(0,2)$. A tensor field of type $(0,2)$ is an object of the form

$$B = B_{ij}\mathrm{d}x^i\,\mathrm{d}x^j \text{ (summation convention).} \tag{4.5.1}$$

(In the mathematical literature, you will often see the notation $\mathrm{d}x^i \otimes \mathrm{d}x^j$ on the RHS. I think this can be fairly safely ignored in this course. $\otimes$ is pronounced 'tensor', by the way.)

The transformation law for the components B_{ij} are deduced as for covector fields: from (4.2.9),

$$B = B_{ij}\mathrm{d}x^i\mathrm{d}x^j = B_{ij}\left(\frac{\partial x^i}{\partial y^p}\mathrm{d}y^p\right)\left(\frac{\partial x^j}{\partial y^q}\mathrm{d}y^q\right) \tag{4.5.2}$$

so if $\widetilde{B}_{pq}$ are the components of B in the y coordinates,

$$\widetilde{B}_{pq} = B_{ij}\frac{\partial x^i}{\partial y^p}\frac{\partial x^j}{\partial y^q} \tag{4.5.3}$$

Note that if X and Y are vector fields and B is a tensor field of type $(0,2)$ we can form

$$\omega = \omega_i\mathrm{d}x^i, \quad \omega_i = B_{ij}Y^j. \tag{4.5.4}$$

As a differential geometer, I might write this as $\omega = B(\cdot, Y)$ if I wanted to avoid using indices and components.

PROPOSITION 4.5.1. *ω defined from B and Y as above is a covector field.*

The proof is left as an exercise: you have to write down the transformation laws and check that the components of ω transform correctly.

We can further form the scalar

$$B(X,Y) = B_{ij}X^iY^j. \tag{4.5.5}$$

As the notation suggests, this is a well-defined scalar function; its value at a point does not depend on the coordinates used to write out the components of B, X and Y.

REMARK 4.5.2. The formula (4.5.5) gives another way to think about tensors of type $(0,2)$. Namely, you can reverse the logic and *define* such a B to be a smoothly varying bilinear form B_p on the tangent space $T_p\Omega$, for each $p \in \Omega$. Smoothness can be defined by saying that the coefficients B_{ij} are smooth functions in Ω for any choice of coordinates x. If we do this for the tangent space, and require $B(X,Y)$ to be invariant (i.e. independent of choice of coordinates), then we say that B is a tensor of type $(0,2)$.

Tensors of type $(0,2)$ are important because the metric tensor, which is the fundamental object in GR, is an example.

REMARK 4.5.3. It is very important not to switch the order of the $\mathrm{d}x^i$ symbols in computations of this kind. In other words, $\mathrm{d}x^i\mathrm{d}x^j \neq \mathrm{d}x^j\mathrm{d}x^i$. Indeed the first one is the bilinear form B such that $B(X,Y) = 0$ unless $X = \partial_i$, $Y = \partial_j$, whereas the second represents the bilinear form C such that $C(X,Y) = 0$ unless $X = \partial_j$, $Y = \partial_i$.

4.5.2. Tensor fields of type $(1,1)$. Suppose that for each p in Ω, we have linear map $A(p)$ from T_pV to T_pV which varies smoothly with p. Such a thing is called a smooth tensor field of type $(1,1)$. In coordinates, A has an expression of the form

$$A = A^k_j \mathrm{d}x^j \frac{\partial}{\partial x^k} \tag{4.5.6}$$

where the A^k_j are a collection of n^2 functions of x. A can be pictured as an $n \times n$ matrix whose entries are smooth functions of x.

We obtain the transformation law under a change of coordinates by substituting

$$\frac{\partial}{\partial x^k} = \frac{\partial y^q}{\partial x^k}\frac{\partial}{\partial y^q},\ \mathrm{d}x^j = \frac{\partial x^j}{\partial y^p}\mathrm{d}y^p \tag{4.5.7}$$

into (4.5.6), getting

$$A = A^k_j\left(\frac{\partial y^q}{\partial x^k}\frac{\partial}{\partial y^q}\right)\left(\frac{\partial x^j}{\partial y^p}\mathrm{d}y^p\right) = \left(A^k_j\frac{\partial x^j}{\partial y^p}\frac{\partial y^q}{\partial x^k}\right)\mathrm{d}y^p\frac{\partial}{\partial y^q}. \tag{4.5.8}$$

Hence the transformation law is

$$\widetilde{A}^q_p = A^k_j\frac{\partial x^j}{\partial y^p}\frac{\partial y^q}{\partial x^k} \tag{4.5.9}$$

EXAMPLE 4.5.4. The identity matrix is an example of a $(1,1)$ tensor. Its components in any coordinate system are the Kronecker δ, δ^j_k.

The transformation law (4.5.9) means that if X is a vector field on Ω then $Y = AX$, with components $A^j_kX^k$ is again a vector field on Ω. This means that under coordinate transformations we have $\widetilde{Y} = \widetilde{A}\widetilde{X}$ where the relation between $\widetilde{A}$ and A is given by (4.5.9) and the transformation law (4.2.8) is used for the components of the vector fields X and Y.

4.5.3. Tensor fields of type $(2,0)$. We've seen tensors with two downstairs indices and one up and one down. The zoo of two-index tensors is completed by the ones with two upstairs indices.

We give the transformation law first:

DEFINITION 4.5.5. A tensor field H of type $(2,0)$ is an object whose components H^{ij} after a choice of coordinates transform according to the rule

$$\widetilde{H}^{pq} = H^{ij}\frac{\partial y^p}{\partial x^i}\frac{\partial y^q}{\partial x^j}. \tag{4.5.10}$$

4.5.4. Algebraic operations on tensors. We mention various interrelations between these types of tensors.

EXAMPLE 4.5.6. Outer product. If X and Y are two vector fields, then there is a tensor $X \otimes Y$ whose components are $X^i Y^j$. This is an example of a tensor field of type $(2,0)$. The verification of this is straightforward.

EXAMPLE 4.5.7. Similarly if α is a covector field then $\alpha \otimes X$ is a tensor whose components are $\alpha_i X^j$ and this is a tensor field of type $(1,1)$.

We can also decrease the number of indices.

EXAMPLE 4.5.8. If α is a covector field and H is of type $(2,0)$, then the contraction of $H\alpha$ is a $(1,0)$ vector field with components

$$H^{ij}\alpha_j.$$

In fact there are two such contractions this one, and the one with components

$$H^{ij}\alpha_i$$

EXAMPLE 4.5.9. If H is a $(2,0)$ tensor and B is a $(0,2)$ tensor, then there is a tensor of type $(1,1)$ with components

$$H^{ik}B_{kj}$$

In fact there are four generally different tensors of this kind:

$$H^{ik}B_{jk}, \quad H^{ki}B_{kj}, \quad H^{ki}B_{jk}$$

along with the one above.

The verification that the components of these objects transform correctly is straightforward, as long as we remember that

$$\frac{\partial \widetilde{x}^i}{\partial x^j}\frac{\partial x^j}{\partial \widetilde{x}^k} = \delta^i_k$$

EXAMPLE 4.5.10. The trace of a $(1,1)$ tensor with components A^i_j is the scalar A^i_i. We can write this as the contraction

$$A^i_i = \delta^i_j A^j_i$$

with the Kronecker δ.

4.6. General tensors

DEFINITION 4.6.1. A tensor field of type (r,s) is an object whose components in any basis are of the form

$$T^{j_1\ldots j_r}_{i_1\ldots i_s} \tag{4.6.1}$$

and which transform under change of coordinates according to the rule

$$\widetilde{T}^{q_1\ldots q_r}_{p_1\ldots p_s} = T^{j_1\ldots j_r}_{i_1\ldots i_s}\frac{\partial x^{i_1}}{\partial y^{p_1}}\cdots\frac{\partial x^{i_s}}{\partial y^{p_s}}\frac{\partial y^{q_1}}{\partial x^{j_1}}\cdots\frac{\partial y^{q_r}}{\partial x^{j_r}} \tag{4.6.2}$$

It is possible to give the components the interpretation of a more geometric object, as we have done for vectors, covectors, and tensors of type $(0,2)$ and $(1,1)$. We shall not do this now.

By way of motivation, note that when we differentiate a function we get a covector. In other words, differentiation seems to increas the number of indices. Thus if we differentiate a function twice, we might expect to get a tensor of type $(0,2)$ and if we differentiate a vector field twise we might expect to get a tensor of type $(1,2)$.

This is true if we are working in a subset of a vector space and restrict ourselved only to linear or affine transformations. It is not, however, true for more general types of transformations. We can see this in the simplest case: starting from

$$\frac{\partial f}{\partial x^j} = \frac{\partial f}{\partial y^q}\frac{\partial y_q}{\partial x^j} \tag{4.6.3}$$

we differentiate again, getting

$$\frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial y^p \partial y^q} \frac{\partial y^p}{\partial x^i} \frac{\partial y^q}{\partial x^j} + \frac{\partial f}{\partial y^q} \frac{\partial^2 y^q}{\partial x^i \partial x^j} \tag{4.6.4}$$

This is *not* the correct transformation law for a $(0, 2)$ tensor (compare with (4.5.3)) unless the Jacobian $\partial y^q / \partial x^j$ is constant. And the Jacobian is only constant if the original change of coordinates is affine-linear.

Although this seems like a serious problem, we shall see in the next chapters a work-around: there is a way to change the way we differentiate vectors (and covectors) by a term which also transforms in such a way as to compensate for the 'bad term' on the RHS of (4.6.4).

4.6.1. Algebra of tensors.

- For each (r, s), the set of all tensors (or tensor fields) forms a vector space. That is you can add any two tensors of the same type and multiply tensors by scalars.
- If T is of type (r, s) and S is of type (p, q), then the tensor product $T \otimes S$ is a tensor of type $(r + p, s + q)$. The components are just
$$T^{j_1 \dots j_r}_{i_1 \dots i_s} S^{m_1 \dots m_p}_{k_1 \dots k_q}$$
- If T is of type (r, s), then picking a pair of indices, one up and one down, we have a *contraction* of T, a tensor of type $(r - 1, s - 1)$. For example if we pick the first indices upstairs and downstairs, we get
$$T^{i j_2 \dots j_r}_{i i_2 \dots i_s}.$$
Recall that by the summation convention, this is actually a sum over the index i. Contraction of a different pair of indices will generally give a different tensor.

We note that the tensor product is distributive over addition.

4.7. Manifolds

A *manifold* is, roughly speaking, a topological space $\mathscr{M}$, with the additional structure necessary to be able to speak of smooth functions from $\mathscr{M}$ to $\mathbb{R}$. This additonal structure is called a smooth atlas and consists of systems of local coordinates satisfying certain compatibility conditions. A function $\mathscr{M} \to \mathbb{R}$ is then called smooth if it is smooth when written in terms of any of these local coordinate systems.

We are not going to get into the details of what a topological space is: it is a set of points with enough structure (open sets) to be able to define continuous functions.

As in the definition of curvilinear coordinates on an open subset of $\mathbb{R}^n$, suppose we have a set of n continuous functions $x(p) = (x^1(p), \dots, x^n(p))$ from $\mathscr{U} \subset \mathscr{M}$ with image some open set Ω of $\mathbb{R}^n$.

Definition 4.7.1. The functions $(x^1, \dots, x^n)$ from $\mathscr{U}$ to Ω form a *local coordinate system* on $\mathscr{M}$ if the map $\mathscr{U} \to \Omega$ is one-one and onto, and if the inverse is continuous.

Thus every point p of $\mathscr{U}$ gets labelled by an ordered set of n real numbers which we're calling the coordinates of p, and conversely if this set of labels is taken from Ω, then it is the label of one and only one point of $\mathscr{U}$.

Definition 4.7.2. If $p_0 \in \mathscr{U}$ is a given point, we say the coordinate system is *centred at* p_0 if $x^j(p_0) = 0$ for all j.

Definition 4.7.3. An *atlas* on $\mathscr{M}$ is a collection of local coordinate systems $x_\nu : \mathscr{U}_\nu \to \Omega'_\nu$, where the open sets $\mathscr{U}_\nu$ cover $\mathscr{M}$.

In this definition the subscript ν does not refer to the different components of the coordinate system, but rather to the different local coordinate systems needed to cover $\mathscr{M}$.

Remark 4.7.4. The individual local coordinate systems $x_\nu : \mathscr{U}_\nu \to \Omega_\nu$ are often called *charts*: thus an atlas is a set of a lot of charts (which made a lot more sense before everyone was using GPS to find their way around).

An atlas, without further conditions, is insufficient to define a consistent notion of when a function $\mathscr{M} \to \mathbb{R}$ should be smooth. To explain what the additional condition is, let's see how far we can get with our atlas. If F is a function from $\mathscr{M}$ to $\mathbb{R}$, then for each ν we get a function $F_\nu : \Omega_\nu \to \mathbb{R}$, defined by

$$F_\nu(x_\nu(p)) = F(p)$$

which makes sense for p in $\mathscr{U}$ and x_ν in Ω_ν.

Now since Ω_ν is an open set of $\mathbb{R}^n$, we know what it means for F_ν to be smooth: it is the classical condition that all partials of F_ν should exist. So we want to say that $F : \mathscr{M} \to \mathbb{R}$ is smooth if and only if all the F_ν are smooth. For this to be consistent, we need it not to matter which coordinate chart we use at p for those points which belong to two or more coordinate charts (which will definitely happen).

So suppose that $\mathscr{U}_\nu \cap \mathscr{U}_\mu \neq \emptyset$ and consider the functions F_ν and F_μ. We have, for

$$F_\nu(x_\nu(p)) = F_\mu(x_\mu(p)) \text{ for } p \in \mathscr{U}_\nu \cap \mathscr{U}_\mu.$$

Let Ω'_ν be the subset of Ω_ν consisting of $x_\nu(p$ with $p \in \mathscr{U}_\nu \cap \mathscr{U}_\mu$, and let Ω'_μ be the corresponding subset of Ω_ν. Then there is a 1:1 correspondence

$$\Omega'_\nu \longleftrightarrow \Omega'_\mu,\ x_\nu(p) = x_\mu(p). \tag{4.7.1}$$

For clarity, write this as $x = x(y)$, where $x \in \Omega'_\nu$ and $y \in \Omega'_\mu$. Then

$$F_\nu(x) = F_\mu(y(x))$$

and we want the LHS of this to be smooth whenever the RHS is smooth. This entails that the functions $y = y(x)$ with $x \in \Omega'_\nu$ and $y \in \Omega'_\mu$ should be a smooth change of coordinates in the sense of §4.1.2. Thus we build this into our definition of smooth atlas:

DEFINITION 4.7.5. An atlas as in Definition 4.7.3 is called *smooth* or *differentiable* if all the change-of-coordinates maps (4.7.1), for all possible pairs μ and ν, are smooth where they are defined[1]. A topological space, equipped with a smooth atlas, is called a smooth manifold of dimension n if the image sets Ω_ν are open subsets of $\mathbb{R}^n$.

REMARK 4.7.6. This is a daunting definition. In practice, some basic theorems are proved (using the definition) which give us our supply of manifolds.

EXAMPLE 4.7.7. The circle $x^2 + y^2 = 1$ is a smooth manifold, of dimension 1.

EXAMPLE 4.7.8. $\mathbb{R}^n$ is a smooth manifold, of dimension n. So is any open subset of $\mathbb{R}^n$.

EXAMPLE 4.7.9. The closed half-space $x \geqslant 0$ inside $\mathbb{R}^2$ is not a manifold. (Though it is a manifold with boundary.)

EXAMPLE 4.7.10. The closed quadrant $\{(x, y) : x \geqslant 0 \text{ and } y \geqslant 0\}$ is not a manifold. (Though it is a manifold with corners.)

EXAMPLE 4.7.11. The null cone $\{X \in M : \eta(X, X) = 0\}$ is not a manifold for the more serious reason that you can't introduce coordinates at the vertex $X = 0$.

EXAMPLE 4.7.12. Generalizing the example of the circle, if $f(x, y, z)$ is a smooth function of 3 variables then the level-set

$$\{(x, y, z) : f(x, y, z) = c\}$$

(any constant c) is a smooth manifold of dimension 2 if at least one of the partial derivatives

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z},$$

is non-zero for every point (x, y, z) on the level set.

[1]A lot of notation is involved in making this precise, and I'm sparing you the details, which you can find in any basic book on the subject

The previous example generalises to functions of any number of variables, but the condition on the non-vanishing of at least one of the partials at every point on the level-set $f = c$ continues to be essential.

It is interesting to note that the null cone, defined by

$$t^2 - x^2 - y^2 - z^2 = 0$$

exactly fails to satisfy this condition on the partials at the origin: all partials vanish there as well.

4.7.1. The tangent space. Let $\mathscr{M}$ be a smooth manifold. For each point p in $\mathscr{M}$, we can use the local coordinates defined near p to define the tangent space. We can either say that it is the abstract vector space spanned by the partials corresponding to any choice of local coordinates from the atlas or we can say that it is the space of directional derivatives at p (and then show that this space is an n-dimensional vector space). Either way $T_p\mathscr{M}$ is an n-dimensional vector space naturally associated with the point p.

We can now define a vector field X on $\mathscr{M}$ as a function which assigns to each p in $\mathscr{M}$, a vector X_p in $T_p\mathscr{M}$, which is required to vary smoothly with p. As in the case of open subset of $\mathbb{R}^n$ 'varying smoothly with p' means: when expanded as a linear combination of the $\partial/\partial x^j$, the coefficients are smooth (in the domain of the coordinate system).

4.7.2. The cotangent space. This is the dual to the tangent space. If f is a smooth function on $\mathscr{M}$, then $\mathrm{d}f$ is a smooth covector field on $\mathscr{M}$: at each point it is in the dual space $T^*_p\mathscr{M}$ and varies smoothly with p.

If X is a vector field and f is a function, then Xf is the directional derivative of f with respect to X. It is again a smooth function on $\mathscr{M}$. It can also be written $\langle X, \mathrm{d}f\rangle$, where $\langle\cdot,\cdot\rangle$ is the pairing between $T\mathscr{M}$ and $T^*\mathscr{M}$.

4.7.3. General tensors. Taking it further, we can extend the idea of tensor field of any type (r, s) to a manifold $\mathscr{M}$, using the low-tech definition above: in any chart a tensor is given by a collection of components, and these are required to transform according to (4.6.2) where $x = x(y)$ is any of the change-of-coordinates map arising from a smooth atlas on $\mathscr{M}$.

4.7.4. Other smooth gadgets. With the aid of a smooth atlas we can define more than just smooth function on $\mathscr{M}$. For example, a smooth curve Γ on $\mathscr{M}$ is defined as a continuous map $\Gamma : I \to \mathscr{M}$ (I is an interval) with the property that the corresponding maps $\Gamma_\nu : I \to \Omega_\nu$, are all smooth, where

$$\Gamma_\nu(\tau) = x_\nu(\Gamma(\tau)) \text{ if } \Gamma(\tau) \in \mathscr{U}_\nu.$$

Similarly (I'll omit the details) if $\mathscr{M}$ and $\mathscr{M}'$ are two manifolds, and $F : \mathscr{M} \to \mathscr{M}'$ is a mapping, we can define what it means for F to be a smooth map between manifolds. The idea is that we can look at F using the charts on $\mathscr{M}$ and $\mathscr{M}'$ and define F to be smooth if and only if all these functions are smooth. The interested reader is referred to any standard introductory book on differential geometry.

Remark 4.7.13. The formalism of general relativity works most naturally on the assumption that space-time is a smooth 4-dimensional manifold. This is particularly important when trying to understand black holes and the large-scale structure of the universe. For the purposes of this course we shall mostly work with space-times that are subsets of $\mathbb{R}^4$: but we shall need to work as if it were a manifold, in other words, without assigning any privileged role to the standard flat coordinates on $\mathbb{R}^4$.