Chapter 1

How the theory of relativity came into being (a brief historical sketch.)

1.1 Special vs. general relativity.

The name "relativity" covers two physical theories. The older one, called special relativity, published in 1905, is a theory of electromagnetic and mechanical phenomena taking place in reference systems that move with large velocities relative to an observer, but are not influenced by gravitation. It is at present considered to be a closed theory. Its parts had entered, as permanent components, the basic courses of classical mechanics, quantum mechanics and electrodynamics. Students of physics study these subjects before they begin to learn general relativity. Therefore, we shall not deal with special relativity here. Familiarity with it is, however, necessary for understanding the general theory. The latter was published in 1915. It describes the properties of time and space, and mechanical & electromagnetic phenomena in the presence of a gravitational field.

1.2 Space and inertia in Newtonian physics.

In Newtonian mechanics and gravitation theory the space was just a room to be filled with matter. It was assumed Euclidean, and, for a long time, no-one questioned this. The masses of matter particles were considered their internal properties, independent of any interactions with the remaining matter. However, from time to time suggestions appeared that not all the phenomena in the Universe can be explained using such an approach.

The best known among them was Mach's principle [1]. Mach started with the following observation: in Newtonian mechanics the assumption is tacitly made, that all the space points can be labelled, for example by assigning Cartesian coordinates to them. One then can observe the motion of matter by finding, in which point of space a given particle is

[1] Ernst Mach, The science of mechanics. Open Court Publishing 1974. (First edition: Ernst Mach, Die Mechanik in ihrer Entwickelung historisch-kritisch Dargestellt, 1883). This approach was originated by the English philosopher, bishop George Berkeley, in 1710, while Newton had still been alive.

located at a given instant. However, this is not actually possible. If we accept another basic assumption of Newton, that the space is Euclidean, then its points do not differ in any way. They can be labelled only by matter being present in the space. In truth, we thus can only observe the motion of one portion of matter relative to another portion of matter. Hence, a correctly formulated theory should only speak about relative motion (matter relative to matter), and not about absolute motion (matter relative to space). If so, then the motion of a single particle in a totally empty Universe would not be detectable. Without any other matter we could not establish whether the lone particle is at rest, or whether it is moving or experiencing acceleration. But the reaction of matter to acceleration is the only way to measure its inertia. Hence, that lone particle would have zero inertia. It follows then that inertia is, likewise, not an absolute property of matter, but is relative, and is induced by the remaining matter in the Universe, supposedly via the gravitational interaction.

This "Mach's principle" has never been formulated as a precise physical theory. It is just a collection of critical remarks and suggestions, partly based on calculations. Nevertheless, it inspired Einstein at the starting point of his work.

1.3 Newton's theory and the orbits of planets.

In addition to the theoretical problem pointed out by Mach, Newton's theory had an empirical problem. It was known already in the first half of 19th century that the planets revolve around the Sun on orbits that are not exactly elliptic. The real orbits are rosettes – curves that arise when a point goes around an ellipse, and at the same time the ellipse rotates slowly around its focus in the same direction (see Fig. 1.1). The reason of this phenomenon was easy to guess: An orbit of a planet is an exact ellipse only if we assume that the Sun has just one planet (plus that the Sun is exactly spherical, and that the space around the Sun is exactly empty). Since the Sun has several planets, they interact gravitationally between them and mutually perturb their orbits. When these perturbations are taken into account, the effect is *qualitatively* the same as observed.

However, in 1859, the French astronomer Urbain J. LeVerrier (the same who, a few years earlier, predicted the existence of Neptune on the basis of similar calculations) verified whether the calculated and observed motions of Mercury's perihelion agree [2]. It turned out that they do not – and that the discrepancy is much larger than the observational error. The calculated velocity of rotation of the perihelion was smaller than the one observed by 43" (arc seconds) per century (the modern result is 43.11 ± 0.45"/century [4]). Astronomers and physicists tried to explain this effect in various simple ways, e.g. by assuming that yet another planet, called Vulcan, revolves around the Sun inside Mercury's orbit and perturbs it; by allowing for gravitational interaction of Mercury with the interplanetary dust; or by assuming that the Sun is flattened in consequence of its rotation. In the last case, the gravitational field of the Sun would not be spherical, and a sufficiently large flattening would explain the additional rotation of Mercury's perihelion. All these

[2] Urbain J. Le Verrier, *Ann. de l'Obs. de Paris* **5**, 104 (1859); cited after

[3] Robert Henry Dicke, *The theoretical significance of experimental relativity.* Gordon and Breach, New York 1964.

[4] C. M. Will, *Theory and experiment in gravitational physics.* Cambridge University Press 1981.

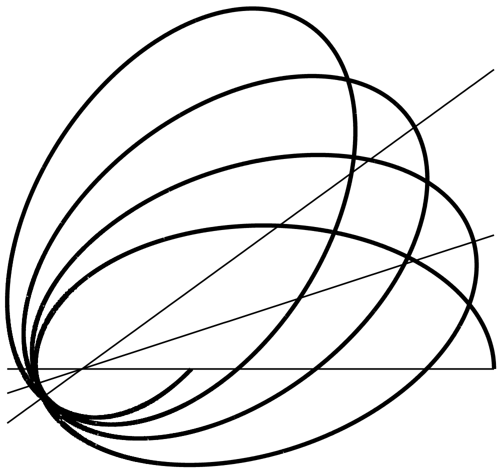

Figure 1.1: Real planetary orbits, in consequence of various perturbations, are not ellipses, but non-closed curves. The angle of revolution of the perihelion shown in the figure is greatly exaggerated. In reality, the greatest angle of perihelion motion observed in the solar system, for Mercury, equals approx. $1.5°$ per 100 years.

hypotheses did not pass the observational tests. The hypothetical planet Vulcan would have to be so massive that it would be clearly visible in telescopes, but wasn't. There was not enough of interplanetary dust to cause the observed effect. If the Sun were sufficiently flattened to explain Mercury's motion, it would cause yet another effect: the planes of the planetary orbits would swing periodically around their mean positions with an amplitude of $\propto 43''$ per century. That motion should have been detectable, but wasn't observed [3].

But nobody doubted in the correctness of Newton's theory. The general opinion was that Mach's critique would be answered by formal corrections in the theory, and the anomalous perihelion motion of Mercury would be explained by new observational discoveries. Nobody expected that any other theory could replace Newton's that had been going from one success to another for over 200 years. General relativity was not created in response to experimental or observational needs. It resulted from pure speculation, it preceded all but one experiments and observations that later confirmed it, and became broadly testable only about 50 years after it was created, in the 1960s. So much time did technology need to catch up and go beyond opportunities provided by astronomical phenomena.

We will now follow, in brief, the reasoning that was the starting point for relativity.

1.4 The basic assumptions of general relativity.

It would be interesting to follow the development of relativistic ideas in the same order in which they had actually appeared in literature. However, this was not a straight and smooth road. Einstein made a few mistakes and put forward a few hypotheses that he had to revoke later. He had been constructing the theory gradually, while at the same time learning the Riemannian geometry – the mathematical basis of relativity. If we followed that gradual progress, we would not only have to take into account some blind paths, but also competitors of Einstein, some of whom questioned the need of the (then) new theory, while some others tried to get ahead of Einstein, but without success. Learning relativity in this way would not be efficient, so we will take a shortcut. We shall begin by justifying the need for the relativity theory, then we shall present the basic elements of Riemann's geometry, and then we will present Einstein's theory in its final shape. The history of relativity's taking shape is presented in Mehra's book [5], and its original presentation can be found in the collection of classic papers [6].

The starting point for Einstein was a critique of Newton's theory based on Mach's ideas. The Newtonian physics said: in a space free from any interactions, material bodies would either remain at rest or would move by uniform rectilinear motion. Since, however, the real Universe is permeated by gravitational fields that cannot be shielded, all bodies in the Universe move on curved trajectories in consequence of gravitational interactions.

There is a problem here. When we say that a trajectory is curved, we assume that we can define a straight line. But how to do it when no real body follows a straight line? The terrestrial standards of straight lines are useful only because no distances on the Earth are truly great, and at short distances the deformation of "rigid" bodies due to gravitation is unmeasurably small. Could the path of a light ray be a good model of a straight line?

To verify, consider two Cartesian reference systems K and K' with axes (x, y, z) and (x', y', z') being parallel. Let K be inertial, and let K' move with respect to K along the z-axis with acceleration $g(t, x, y, z)$. Let the origins of K and K' coincide at $t = 0$. Then

$$x' = x, \qquad y' = y,$$

$$z' = z - \int_0^t \mathrm{d}\tau \int_0^\tau \mathrm{d}s\; g(s, x, y, z).$$

Hence, the equations of motion of a free particle, that in K are

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = 0,$$

in K' assume the form

$$\frac{\mathrm{d}^2 x'}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 y'}{\mathrm{d}t^2} = 0,$$

$$\frac{\mathrm{d}^2 z'}{\mathrm{d}t^2} = -g(t, x, y, z).$$

The quantity that we interpreted in K as acceleration, in K' we would interpret as the intensity of a gravitational field (with opposite sign). Gravitation can thus be simulated by accelerated motion. If so, then light in a gravitational field should behave as if it were observed in an accelerated reference system. How would we see a light ray in such a system? Imagine a space vehicle that flies (with a great velocity) across a light ray. Let the ray enter the vehicle through a window and fall on a screen on the other side of the vehicle

[5] J. Mehra, *Einstein, Hilbert and the theory of gravitation.* D. Reidel, Dordrecht 1974.

[6] A. Einstein, H. A. Lorentz, H. Weyl and H. Minkowski, *The principle of relativity. A collection of original papers on the special and general theory of relativity.* Dover Publications 1923.

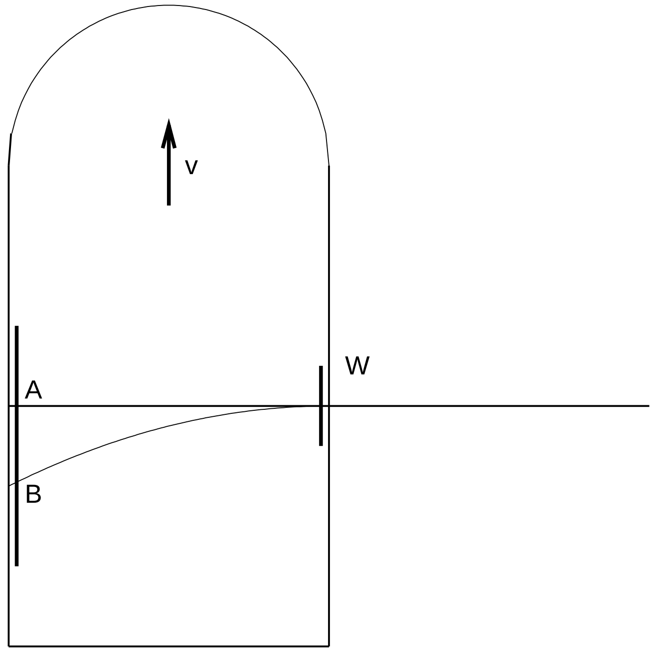

Figure 1.2: A space vehicle flying across a light ray. See explanation in text.

(see Fig. 1.2). If the vehicle were at rest, the light ray entering the vehicle at point W would hit the screen at point A. Since the vehicle keeps flying, it will move a bit before the light ray hits the screen, and the bright spot will appear in point B. Now assume that the ray indeed moves on a straight line when observed by an observer who is at rest. Then the path WB will be straight when the vehicle moves with a constant velocity, and it will be curved when the vehicle moves with acceleration. Hence, if gravitational field behaves analogously to the field of inertial forces, then light rays should be deflected also by gravitation. Consequently, they cannot be the standards of a straight line.

If we are unable to provide a physical model of a fundamental notion of Newtonian physics, let us try to do without this notion. Let us assume that no such thing exists as "gravitational forces" that curve the trajectories of celestial bodies, but that the geometry of space is modified by gravitation in such a way that the observed trajectories are paths of free motion. Such a theory might be more complicated in practical instances than the Newtonian one, but it will use only such notions that relate to actual observations, without using an abstract, unobservable background of the Euclidean space.

A modified geometry means non-Euclidean. A theory created for dealing with broad classes of non-Euclidean geometries is differential geometry. It is the mathematical basis of general relativity, and we will begin by studying it.

Part I

ELEMENTS OF DIFFERENTIAL GEOMETRY

Chapter 2

A short sketch of two-dimensional differential geometry.

2.1 Constructing parallel straight lines in a flat space.

The classical Greek methods of geometrical constructions, with the help of a ruler and compasses, fail when we consider large distances. For example, if we wish to construct a straight line parallel to the momentary velocity of the Earth that passes through a given point on the Moon, compasses and rulers do not help. Let us think then, what method to construct parallel straight lines might work in such a situation. For the beginning, let us assume that great distance is our only problem – that we live in a space without gravitation, so that we can use a light ray or the trajectory of a stone shot from a sling as a model of a straight line.

Assume that the observer is at point A (see Fig. 2.1) on the straight line p, and wants to construct a straight line through the point B that would be parallel to p. The following program is "technically realistic": we first determine the straight line passing through both A and B (for example, by directing a telescope toward B), then we measure the angle α between the lines p and AB, then, from B, we construct a straight line q that is inclined to AB at the same angle α *and lies in the same plane as* p *and* AB. The second condition requires that we can control points of q other than B, and it can pose some problems. However, assuming that our observer is able to construct parallel straight lines that are not too distant from the given one, he/she can carry out the following operation: the observer moves from A to A_1, constructs a straight line $p_1 \parallel p$, then moves on to A_2, constructs a straight line $p_2 \parallel p_1$, etc, until, in the n-th step, he/she reaches B and constructs $q = p_n \parallel p_{n-1}$ there.

This construction can be generalised. The observer does not have to move from A to B on a straight line. He can start from A in an arbitrary direction and, at an A_1, construct a straight line parallel to p; it has to lie in the plane pAA_1 and be inclined to AA_1 at the same angle as p. Then, from A_1 the observer can continue in still another arbitrary direction and at an A_2 repeats the construction: a straight line p_2 has to lie in the plane $p_1A_1A_2$ and be inclined to A_1A_2 at the same angle as p_1 was. When the broken line he/she is following reaches B, the last straight line will be the one we wanted to construct.

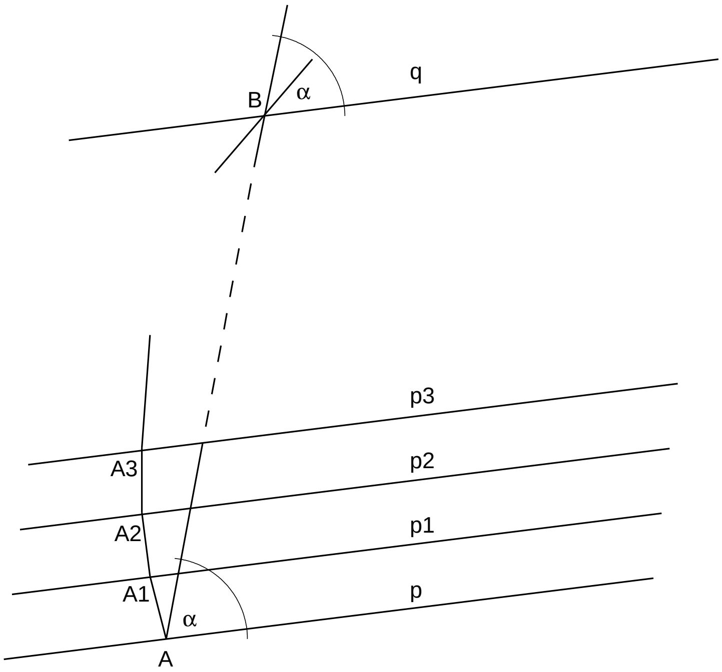

Figure 2.1: Constructing distant parallel straight lines in a flat space. See explanation in text.

We can imagine broken lines whose straight segments are becoming progressively shorter. In the limit, we conclude that we would be able to carry out this construction along an arbitrary differentiable curve. The plane needed in the construction will be in each step determined by the tangent vector of the curve and the last straight line we had constructed.

In this way, we arrived at the idea of constructing parallel straight lines by parallely transporting directions. Note that a straight line is privileged in this construction: this is the only line to which the parallely transported direction is inclined always at the same angle. In particular, a vector tangent to a straight line, when transported parallely along this line, remains tangent to it at every point. A straight line can be defined by this property, provided we are able to define what it means to be parallel without first invoking the notion of a straight line. One possible definition is this: a vector field $\mathbf{v}(x)$ defined along a curve $C \subset \mathbb{R}^n$ consists of parallel vectors (or, in other words, is parallely transported along C) when there exists a coordinate system such that $\partial v^i/\partial x^j \equiv 0$.

2.2 Generalizing the notion of parallelism to curved surfaces.

On a curved surface, the analogue of a straight line is a geodesic line. This is a curve whose arc PQ (see Fig. 2.2) is the shortest among all arcs connecting P and Q. Note that, unlike

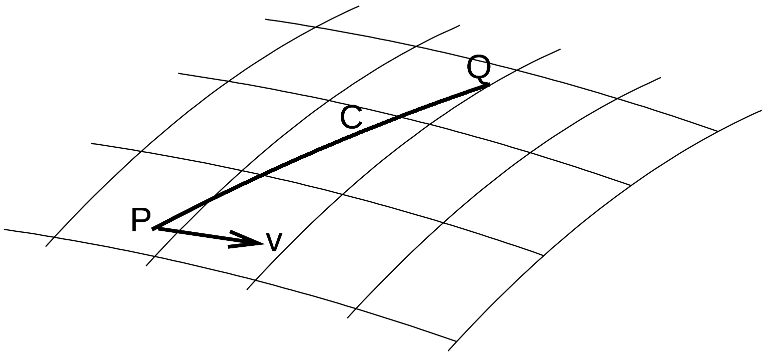

Figure 2.2: Parallel transport of vectors on a curved surface. See explanation in text.

on a plane or in a flat space, the vector tangent to a curve on a curved surface S is not a subset of this surface. The collection of all vectors tangent to the surface S at a point $p \in S$ spans a plane tangent to S at P.

On a curved surface S, parallel transport is defined as follows. Suppose we are given the pair of points P and Q, an arc of a curve C connecting P and Q and a vector tangent to S at P that we plan to parallely transport to Q. If C is a geodesic, then we transport the vector v along it in such a way that it is everywhere inclined to the tangent vector of C at the same angle. If C is not a geodesic, then:

1. We divide the arc PQ into n segments;
2. We connect the ends of each arc by a geodesic;
3. We transport $\mathbf{v}$ parallely along each geodesic arc.
4. We imagine (that is, calculate) the result of this operation when $n \to \infty$.

The result of parallel transport thus defined depends on the path of transporting. For example, consider a sphere, its pole C and two points A and B lying on the equator, 90° away from each other (Fig. 2.3). Let $\mathbf{v}$ be the vector tangent to the equator at A. Transport $\mathbf{v}$ parallely to C along the arc AC, and then again along the broken line consisting of the arcs AB and BC. All three arcs are parts of great circles, which are geodesic lines, so at every point $\mathbf{v}$ is inclined at the same angle to the tangent vectors of the arcs. The first transport will yield a vector at C that is tangent to BC, while the second one will yield a vector at C perpendicular to BC. In consequence, if we transport (in differential geometry one says "drag") a vector along a closed loop, we will not get the same vector we started with. The curvature of the surface is responsible for this. The connection between the initial vector, the final vector and the curvature will be discussed further on.

We discussed two-dimensional surfaces in this chapter to visualise things more easily. However, this gave us an unfair advantage: on a two-dimensional surface, the direction inclined to a given tangent vector at a given angle is uniquely determined. In spaces of higher dimension we will need a definition of "parallelism at a distance" that will be analogous to $\partial v^i/\partial x^j = 0$ that we used in a flat space.

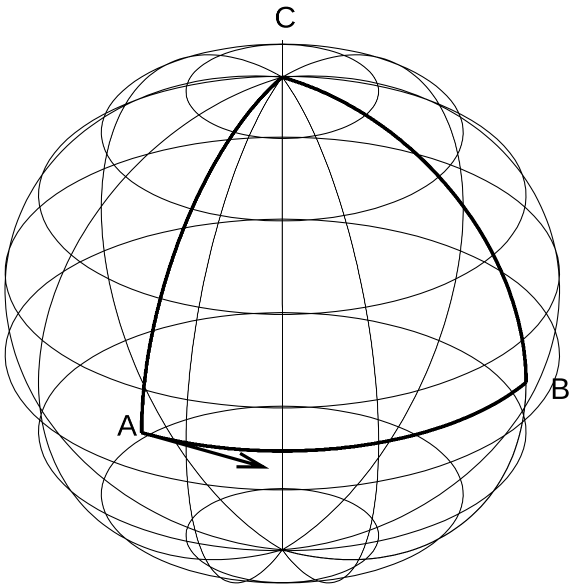

Figure 2.3: Parallel transport of vectors on a sphere. See explanation in text.

Chapter 3

Tensors, tensor densities.

3.1 What are tensors good for?

In Newtonian physics, a preferred class of reference systems is used. They are the inertial systems – those in which the 3 Newtonian principles of dynamics hold true. However, it may be difficult in practice to identify them. As we have seen in chapter 1, the inertial force imitates the gravitational force, so it may not be easy to make sure whether a given object moves with acceleration or remains at rest in a gravitational field. Hence, the laws of physics should be formulated in such a way that no reference system is privileged. The choice of a reference system is an act of human will, while the laws of physics should not depend on our decisions. Tensors are objects that allow us to solve this problem.

For the beginning, we will settle for a vague definition that we will make more precise later. A tensor is a collection of functions defined on a certain n-dimensional space in such a way that when we change the coordinates $\{x^\alpha\}$, $\alpha = 1, 2, \ldots, n$ to new coordinates $\{x^{\alpha'}\}$, $\alpha' = 1, 2, \ldots, n$, the functions transform by the law that we will define later on.

3.2 Differentiable manifolds.

As already stated, in relativity we will be using non-Euclidean spaces. The most general class of spaces that we will consider are **differentiable manifolds**. This is a generalisation of the notion of a curved surface for which a tangent plane at its every point exists. An n-dimensional differentiable manifold of class p is a space M_n whose every point x has a neighbourhood $\mathcal{O}_x$ such that

1. There exists a one-to-one mapping κ_x of the neighbourhood $\mathcal{O}_x$ onto a subset of $\mathbb{R}^n$. The mapping κ_x is called a **map** of the neighbourhood $\mathcal{O}_x$. The coordinates of the image $\kappa_x(x)$ are called the coordinates of $x \in M_n$.

2. If the neighbourhoods $\mathcal{O}_x$ and $\mathcal{O}_y$ of $x, y \in M_n$ have a non-empty intersection ($\mathcal{O}_x \cap \mathcal{O}_y \neq \emptyset$), κ_x is a map of $\mathcal{O}_x$ and κ_y is a map of $\mathcal{O}_y$, then the mappings $(\kappa_y \circ {\kappa_x}^{-1})$ and $(\kappa_y \circ {\kappa_x}^{-1})^{-1}$ are mappings of class p of $\mathbb{R}^n$ into itself.

A **tangent space** to the manifold M_n at the point x is a vector space spanned by vectors tangent at x to curves in M_n that pass through x.

Note that if $\kappa_x(x) = \{x^1, \ldots, x^n\}$ are the coordinates of the point x, then the equation $x^i = \text{const}$, where $i \in \{1, \ldots, n\}$ is a fixed index, defines a hypersurface in $\mathbb{R}^n$, and thus also a hypersurface H in M_n. A coordinate system is thus a set of n one-parameter families of hypersurfaces; the parameter in the i-th family is the constant in $x^i = \text{const}$.

Now imagine for a moment that the manifold we consider is $\mathbb{R}^n$. Each hypersurface defines then a family of vectors: if $\Phi(x) = C$ (where C is an arbitrary constant) is the equation of a hypersurface H, then $\partial\Phi/\partial x^\alpha$, where $\{x^\alpha\}$ are the coordinates of the point x, is an equation of a family of vectors attached to H and orthogonal to it. Taking all the hypersurfaces of the $\Phi = C$ family, we get in M_n a family of curves tangent to the vectors $\partial\Phi/\partial x^\alpha$. Thus, each coordinate system in $\mathbb{R}^n$ defines a family of curves.

The converse is not true: not every n-parameter family of curves C_x in $\mathbb{R}^n$ defines a family of hypersurfaces orthogonal to C_x. It happens only when the vectors tangent to C_x have zero rotation (to be defined later). Hence, the set of all curves in M_n (and thus of all vector fields on M_n) is larger than the set of all coordinate systems in M_n.

The reason why, for this example, we had to take the special case of $M_n = \mathbb{R}^n$ is that, as we shall see later, in a general vector space vectors like the gradient of a function (called *covariant vectors*) and vectors like a tangent vector to a curve (called *contravariant vectors*) are unrelated objects of different kind. A relation between them exists only in spaces tangent to such manifolds in which a *metric* is defined, see Chapter 7. $\mathbb{R}^n$ is one of them. Without a metric, a covariant vector cannot be converted into a contravariant one, and a curve tangent to a field of covariant vectors cannot be constructed.

Let $U \subset M_n$ be an open subset. Suppose we are given a collection of n n-parameter families of curves such that n curves pass through each point $x \in U$. Suppose that the tangent vectors to these curves are linearly independent at every $x \in U$. Then the tangent vectors to these curves at the point x, $e_a(x)|_{a=1,\ldots,n}$ are a basis of the space tangent to M_n at x. Let $v(x)$ be an arbitrary vector tangent to M_n at x. Then

$$v = \sum_{a=1}^{n} v^a(x) e_a(x).$$

The coefficients $\{v^a\}$ are called the components of the vector v in the basis $e_a(x)$. The mapping $x \to v(x)$ assigns to each point $x \in U \subset M_n$ a vector tangent to M_n at x. This mapping is called a **vector field** on U.

Note: The vectors of a vector field are defined on tangent spaces to the manifold, while the components of vector fields are functions on the manifold. The vectors $e_a(x)$ can be identified with directional derivatives, and are then defined by coordinates as $e_a(x)\Phi(x) = (\partial/\partial x^a)\Phi(x)$ (that is, the a-th vector in the basis is the directional derivative down the a-th family of hypersurfaces in the coordinate system.) Then the v^a are components of the vector v in the coordinate system $\{x\}$. Again, they are functions on the manifold.

We adopt three conventions:

1. If an upper index and a lower index have the same symbol, a sum over all of their

values is implied. Hence, if α changes from 1 to n, while $\{V_\alpha\}$ and $\{U^\alpha\}$, $\alpha = 1, \ldots, n$ are collections of functions labelled by α, then

$$U^\alpha V_\alpha \equiv \sum_{\alpha=1}^{n} U^\alpha V_\alpha.$$

2. The collection of all coordinates $\{x^\alpha\}$, $\alpha = 1, \ldots, n$ will be denoted $\{x\}$, and a function $f(\{x\})$ will be denoted $f(x)$.

3. The derivative of f with respect to x^α will be denoted as

$$f_{,\alpha} \equiv \frac{\partial f}{\partial x^\alpha}.$$

3.3 Scalars.

A scalar is a function on a manifold whose value, when coordinates are transformed, changes by substituting the transformation in its argument:

$$\varphi'(x'(x)) = \varphi(x). \tag{3.1}$$

Examples of scalars are physical constants, rest masses of elementary particles, their electric charges, a density distribution in a continuous medium. The coordinates of a point in a manifold are scalars, too, since their values transform by (3.1) when coordinates are changed from $\{x'\}$ to $\{x''\}$:

$$x^\alpha(x''(x')) = x^\alpha(x').$$

3.4 Contravariant vectors.

The functions $v^\alpha(x)$, $\alpha = 1, \ldots, n$ are coordinates of a contravariant vector field when, by a change of coordinates on M_n, they transform by the law

$$v^{\alpha'}(x'(x)) = x^{\alpha'}{}_{,\alpha}\, v^\alpha(x). \tag{3.2}$$

Examples are vectors tangent to curves. Suppose that a curve C is given by the parametric equations $t \to x^\alpha(t)$, $\alpha = 1, \ldots, n$, $t \in [a, b] \subset \mathbb{R}^1$, where x^α is the value of a coordinate of a point on the curve. An arbitrary field of vectors tangent to C is then given by

$$t \to v^\alpha(t) = f(x)g(t)\frac{\mathrm{d}x^\alpha}{\mathrm{d}t},$$

where $f(x)$ is an arbitrary function of the coordinates and $g(t)$ is an arbitrary function of the parameter. When the coordinate system is changed from (x) to (x'), we have

$$v^{\alpha'}(t) = f(x'(x))g(t)\frac{\mathrm{d}x^{\alpha'}}{\mathrm{d}t} = f(x)g(t)x^{\alpha'}{}_{,\alpha}\frac{\mathrm{d}x^\alpha}{dt},$$

in agreement with (3.2).

3.5 Covariant vectors.

The functions $u_\alpha(x)$, $\alpha = 1, \ldots, n$ are coordinates of a covariant vector field when, by a change of coordinates on M_n, they transform by the law

$$u_{\alpha'}(x'(x)) = x^{\alpha}{}_{,\alpha'}\, u_\alpha(x). \tag{3.3}$$

Contravariant vectors are made graphically different from the covariant ones by placing their index as a superscript or a subscript, respectively. This notation makes it easier to remember the transformation law. An example of a covariant vector is the gradient of a scalar function $\varphi(x)$. For such a function we have

$$\varphi_{,\alpha'}\,(x'(x)) = x^{\alpha}{}_{,\alpha'}\, \varphi_{,\alpha}\,(x),$$

in agreement with (3.3).

Note that $v^\alpha u_\alpha$ is a scalar, since we have

$$v^{\alpha'} u_{\alpha'} = x^{\alpha'}{}_{,\alpha}\, v^\alpha x^{\beta}{}_{,\alpha'}\, u_\beta = \left(x^{\beta}{}_{,\alpha'}\, x^{\alpha'}{}_{,\alpha}\right) v^\alpha u_\beta$$

$$= x^{\beta}{}_{,\alpha}\, v^\alpha u_\beta = \delta^\beta{}_\alpha v^\alpha u_\beta = v^\alpha u_\alpha.$$

Another example of a scalar field is the directional derivative of a scalar field along a contravariant vector field, $v(\varphi) \overset{\text{def}}{=} v^\alpha \varphi_{,\alpha}$, as can be easily verified.

3.6 Tensors of second rank.

Scalars are called tensors of rank zero, to emphasise that they have no indices. The contra- and covariant vectors are collectively called tensors of rank 1. Tensors of rank 2 are objects whose components are labelled by two indices. There are three kinds of them:

1. Doubly contavariant tensors. Their components $T^{\alpha\beta}(x)$ transform in the following way under a coordinate transformation $x \to x'$ on M_n:

$$T^{\alpha'\beta'}(x'(x)) = x^{\alpha'}{}_{,\alpha}\, x^{\beta'}{}_{,\beta}\, T^{\alpha\beta}(x). \tag{3.4}$$

2. Doubly covariant tensors. Their components $T_{\alpha\beta}(x)$ transform by the rule

$$T_{\alpha'\beta'}(x'(x)) = x^{\alpha}{}_{,\alpha'}\, x^{\beta}{}_{,\beta'}\, T_{\alpha\beta}(x). \tag{3.5}$$

3. Mixed tensors. Their components transform by the rule:

$$T_{\alpha'}{}^{\beta'}(x'(x)) = x^{\alpha}{}_{,\alpha'}\, x^{\beta'}{}_{,\beta}\, T_\alpha{}^\beta(x). \tag{3.6}$$

As can be seen, the collection of components of an arbitrary second-rank tensor is a square matrix that transforms in a prescribed way when coordinates are changed.

An example of a doubly covariant tensor is a matrix of a quadratic form

$$\Phi(A) = \Phi_{\alpha\beta} A^{\alpha} A^{\beta},$$

where A^{α} are components of a contravariant vector, and the value of $\Phi(A)$ is a scalar.

An example of a mixed tensor of rank 2 is a matrix of mapping of one vector space into another

$$V^{\alpha} = B^{\alpha}{}_{\beta} W^{\beta},$$

where V^{α} and W^{β} are contravariant vectors in different vector spaces.

We will meet examples of doubly contravariant tensors later in this course. The simplest example is the inverse matrix to a matrix of a quadratic form, but in order to be able to prove this we have to learn about some other objects.

The scalar $T^{\alpha}{}_{\alpha}$ (i.e. the sum of diagonal components of a mixed tensor) is called the trace of T. Quantities like $\sum_{\alpha} T^{\alpha\alpha}$ and $\sum_{\alpha} T_{\alpha\alpha}$ do not transform like tensors. Summations over indices standing on the same level occur only exceptionally in differential geometry – for example, when a calculation is done in a chosen coordinate system.

3.7 Tensor densities.

A tensor density differs from the corresponding tensor in that, when transformed from one coordinate system to another, it gets multiplied by a certain power of the Jacobian of the transformation. The exponent of the Jacobian is called the **weight** of the density. For example, a scalar density of weight w transforms as follows

$$\Phi'(x') = \left[\frac{\partial(x')}{\partial(x)}\right]^{w} \Phi(x), \tag{3.7}$$

a contravariant vector density of weight w transforms by the rule

$$v^{\alpha'}(x') = \left[\frac{\partial(x')}{\partial(x)}\right]^{w} x^{\alpha'}{}_{,\alpha}\, v^{\alpha}(x), \tag{3.8}$$

and so on. An example of a scalar density is the element of volume in a multi-dimensional integral, which transforms by the law

$$\mathrm{d}_n x = \frac{\partial(x)}{\partial(x')}\, \mathrm{d}_n x'. \tag{3.9}$$

Its weight is seen to be $+1$.

An arbitrary tensor is by definition a tensor density of weight zero.

3.8 Tensor densities of arbitrary rank.

The components of a tensor density of weight w, k times contravariant and l-times covariant, transform by the law

$$T^{\alpha'_1\alpha'_2\ldots\alpha'_k}_{\beta'_1\beta'_2\ldots\beta'_l}(x'(x)) = \left[\frac{\partial(x')}{\partial(x)}\right]^w x^{\alpha'_1}{}_{,\alpha_1} x^{\alpha'_2}{}_{,\alpha_2} \ldots x^{\alpha'_k}{}_{,\alpha_k} x^{\beta_1}{}_{,\beta'_1} x^{\beta_2}{}_{,\beta'_2} \ldots x^{\beta_l}{}_{,\beta'_l} T^{\alpha_1\alpha_2\ldots\alpha_k}_{\beta_1\beta_2\ldots\beta_l}(x). \tag{3.10}$$

Such an object is called a tensor density of type $[w, k, l]$.

For general tensor densities one can carry out an operation that is analogous to finding the trace of a mixed tensor of rank 2. This operation is called **contraction**. It consists in making an upper index equal to a lower index, and summing over all its allowed values. The resulting density is of type $[w, k-1, l-1]$, thus

$$[\text{contracted}T]^{\alpha_1\ldots\alpha_{i-1}\alpha_{i+1}\ldots\alpha_k}_{\beta_1\ldots\beta_{j-1}\beta_{j+1}\ldots\beta_l}(x) = T^{\alpha_1\ldots\alpha_{i-1}\rho\alpha_{i+1}\ldots\alpha_k}_{\beta_1\ldots\beta_{j-1}\rho\beta_{j+1}\ldots\beta_l}(x) \tag{3.11}$$

(Note: sum over all values of ρ is implied above.) The indices over which the summation is carried out are called "dummy" as they do not show up in the transformation law of the contracted density.

The contraction may be done over several pairs of indices at the same time. One must take care to give different names to each pair of dummy indices, to avoid confusion.

3.9 Algebraic properties of tensor densities.

Here is a list of the most basic properties of tensor densities:

1. If $T^{\cdots\cdots\cdots}_{\cdots\cdots\cdots} \equiv 0$ in one coordinate sysem, then $T^{\cdots\cdots\cdots}_{\cdots\cdots\cdots} \equiv 0$ in all coordinate systems (follows from the transformation law).

2. A linear combination of two tensor densities of type $[w, k, l]$ is a tensor density of the same type. (A sum of tensor densities of different types is not a tensor density.)

3. The collection of quantities obtained when each component of one tensor density is multiplied by each component of another tensor density is called a **tensor product** of the two densities. For example, out of the vectors u_α and v^α one can form such tensor products as $v^\alpha v^\beta$, $v^\alpha u_\beta$, $u_\alpha u_\beta$, $v^\alpha v^\beta v^\gamma$, $v^\alpha u_\beta v^\gamma$, $v^\alpha u_\alpha v^\beta u_\gamma v^\delta$, and so on. The tensor product is denoted by $\otimes$, thus for example $v^\alpha u_\beta = (v \otimes u)^\alpha{}_\beta$. The tensor product of tensor densities of types $[w, k, l]$ and $[w', k', l']$ is a tensor density of type $[w + w', k + k', l + l']$.

4. If a tensor density does not change its value when two indices (either both upper or both lower) are interchanged, then it is called **symmetric** with respect to this pair of indices. If it only changes sign, then it is called **antisymmetric** in this pair of indices. The property of being symmetric or antisymmetric in a given pair of indices is preserved under transformations of coordinates, i.e. it is independent of the coordinate system chosen.

5. This last property allows us to define **symmetrisation** and **antisymmetrisation**

of a tensor density. The **completely symmetric part** of $T_{\alpha_1 \ldots \alpha_k}$ is the quantity

$$T_{(\alpha_1 \ldots \alpha_k)} := \frac{1}{k!} \sum_{\substack{\text{over all permutations} \\ i_1, \ldots i_k}} T_{\alpha_1 \ldots \alpha_k}. \tag{3.12}$$

The **completely antisymmetric part** of $T_{\alpha_1 \ldots \alpha_k}$ is the quantity

$$T_{[\alpha_1 \ldots \alpha_k]} := \frac{1}{k!} \sum_{\substack{\text{over all permutations} \\ i_1, \ldots i_k}} (\text{sign of the permutation}) T_{\alpha_1 \ldots \alpha_k}. \tag{3.13}$$

One can carry out the symmetrisation or antisymmetrisation (i.e., respectively, *symmetrise* or *antisymmetrise*) also with respect to the upper indices, and, in each case, over only a subset of all the indices that a tensor density has. For example, a tensor of rank 4 $T_{\alpha\beta\gamma\delta}$ can be symmetrised over only two of its indices

$$T_{(\alpha\beta)\gamma\delta} = \frac{1}{2} \left(T_{\alpha\beta\gamma\delta} + T_{\beta\alpha\gamma\delta} \right),$$

or over three indices

$$T_{(\alpha\beta\gamma)\delta} = \frac{1}{6} \left(T_{\alpha\beta\gamma\delta} + T_{\alpha\gamma\beta\delta} + T_{\beta\alpha\gamma\delta} + T_{\beta\gamma\alpha\delta} + T_{\gamma\alpha\beta\delta} + T_{\gamma\beta\alpha\delta} \right).$$

If we want to emphasise that a certain index is excluded from symmetrisation or antisymmetrisation, then we put it between vertical strokes. For example, the antisymmetrisation of $T_{\alpha\beta\gamma\delta}$ with respect to α and γ only would be denoted $T_{[\alpha|\beta|\gamma]\delta}$.

The symmetrisation and the antisymmetrisation with respect to just two indices are complementary operations, thus

$$T_{(\alpha\beta)\gamma\delta\ldots} + T_{[\alpha\beta]\gamma\delta\ldots} \equiv T_{\alpha\beta\gamma\delta\ldots}.$$

This is no longer true for symmetrisations/antisymmetrisations over larger numbers of indices. The other parts of $T_{\alpha\beta\gamma\delta\ldots}$ that enter neither $T_{(\alpha\beta\gamma)\delta\ldots}$ nor $T_{[\alpha\beta\gamma]\delta\ldots}$ are obtained when different rules of assigning signs to the different terms in (3.12) and (3.13) are chosen. We will not encounter those operations in this course.

If antisymmetrisation is done with respect to a larger number of indices than the dimension of the manifold, then the result is identically zero. This is because, with n possible different values of each index, there must be at least one value that is repeated in a set of $m > n$ indices. Then, each term in (3.13) has its counterpart that is identical, but enters the sum with an opposite sign.

3.10 The completely antisymmetric symbol.

Let $\epsilon_{\alpha_1 \ldots \alpha_n}$ be a covariant tensor density (whose weight is as yet unknown) defined on a differentiable manifold M_n, and let it be antisymmetric with respect to every pair of its indices. Because of antisymmetry, each component that has a pair of identical indices will be equal to zero. Only those components can be different from zero in which all the α_i-s are different, i.e. for which the set $\{\alpha_1, \ldots, \alpha_n\}$ is a permutation of the set $\{1, \ldots, n\}$. For even permutations $\epsilon_{\alpha_1 \ldots \alpha_n} = \epsilon_{1 \ldots n}$, for odd permutations $\epsilon_{\alpha_1 \ldots \alpha_n} = -\epsilon_{1 \ldots n}$. Hence, it is enough to define the component $\epsilon_{1 \ldots n}$ to have the whole tensor density $\epsilon_{\alpha_1 \ldots \alpha_n}$ defined. We thus define

$$\epsilon_{1 \ldots n} = +1. \tag{3.14}$$

The quantity $\epsilon_{\alpha_1 \ldots \alpha_n}$ is called the *completely antisymmetric Levi-Civita symbol.*

Let $A^\alpha{}_\beta$ be an arbitrary matrix. Let us investigate the quantity

$$[D(A)]_{\beta_1 \ldots \beta_n} := \epsilon_{\alpha_1 \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_n}{}_{\beta_n}. \tag{3.15}$$

Let α_i label the rows, and β_j the columns of the matrix. The expression above is a sum of n-tuple products of elements of the matrix $A^\alpha{}_\beta$. In each product, each factor comes from a different row (because, if two indices α_i and α_j were equal, then $\epsilon_{\alpha_1 \ldots \alpha_n} = 0$). The i-th factor in each product is always from the same (i-th) column, but each time from a different row, and, in the whole sum, runs over all rows. For even permutations $\alpha_1 \ldots \alpha_n$, the product enters the sum with a $+$-sign, for odd permutations – with a $-$-sign. The quantity (3.15) is antisymmetric in all the β_i indices, as can be seen from the following. (The additional indices under the α-s refer to the position of a given α in the set.)

$$\epsilon_{\alpha_1 \ldots \underset{i}{\alpha_i} \ldots \underset{j}{\alpha_j} \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_i}{}_{\beta_i} \ldots A^{\alpha_j}{}_{\beta_j} \ldots A^{\alpha_n}{}_{\beta_n} =$$

$$\epsilon_{\alpha_1 \ldots \underset{i}{\alpha_j} \ldots \underset{j}{\alpha_i} \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_j}{}_{\beta_i} \ldots A^{\alpha_i}{}_{\beta_j} \ldots A^{\alpha_n}{}_{\beta_n},$$

by interchanging the names of the dummy indices α_i and α_j, which does not change the value of the expression, and further

$$= -\epsilon_{\alpha_1 \ldots \underset{i}{\alpha_i} \ldots \underset{j}{\alpha_j} \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_i}{}_{\beta_j} \ldots A^{\alpha_j}{}_{\beta_i} \ldots A^{\alpha_n}{}_{\beta_n},$$

since $\epsilon_{\ldots}$ is antisymmetric with respect to every pair of indices, while the factors $A^{\alpha_i}{}_{\beta_i}$ can be interchanged at will (this is an ordinary commutative product). Hence, the quantity $[D(A)]_{\beta_1 \ldots \beta_n}$ from (3.15) vanishes if any two indices β_i, β_j are equal (i.e. if the same column of the matrix A appears in the positions (i, j)).

It is thus seen that $[D(A)]_{\beta_1 \ldots \beta_n}$ has all the properties of the determinant of A. It will be equal $+\det(A)$ when the permutation $\{1, \ldots, n\} \to \{\beta_1, \ldots, \beta_n\}$ is even, and it will equal $[-\det(A)]$ when the permutation is odd. Hence

$$\epsilon_{\alpha_1 \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_n}{}_{\beta_n} = \epsilon_{\beta_1 \ldots \beta_n} \det(A). \tag{3.16}$$

Now let us apply this formula to the matrix of derivatives (the Jacobi matrix) $x^\alpha{}_{,\alpha'}$ of the coordinate transformation $\{x\} \to \{x'\}$. We have

$$\epsilon_{\alpha_1 \ldots \alpha_n} x^{\alpha_1}{}_{,\alpha'_1} \ldots x^{\alpha_n}{}_{,\alpha'_n} = \epsilon_{\alpha'_1 \ldots \alpha'_n} \frac{\partial(x)}{\partial(x')}, \tag{3.17}$$

or, equivalently

$$\epsilon_{\alpha'_1 \ldots \alpha'_n} = \frac{\partial(x')}{\partial(x)} x^{\alpha_1}{}_{,\alpha'_1} \ldots x^{\alpha_n}{}_{,\alpha'_n} \epsilon_{\alpha_1 \ldots \alpha_n}. \tag{3.18}$$

This shows that $\epsilon_{\alpha_1 \ldots \alpha_n}$ is a tensor density of type $[1, 0, n]$.

By a similar method one can verify that $\epsilon^{\alpha_1 \cdots \alpha_n}$ is a tensor density of type $[-1, n, 0]$.

3.11 Multidimensional Kronecker deltas.

Let us recall the definition of the ordinary Kronecker delta:

$$\delta^\alpha{}_\beta = \begin{cases} 1 & \text{when} \quad \alpha = \beta \\ 0 & \text{when} \quad \alpha \neq \beta \end{cases} \tag{3.19}$$

(this is the unit matrix in n dimensions). This is a tensor, since, on the one hand

$$\delta'^\alpha{}_\beta = \delta^\alpha{}_\beta \tag{3.20}$$

by definition, and on the other hand

$$\delta'^{\alpha'}{}_{\beta'} = x^{\alpha'}{}_{,\beta'} = x^{\alpha'}{}_{,\rho}\, x^\rho{}_{,\beta'} = x^{\alpha'}{}_{,\rho}\, x^\sigma{}_{,\beta'}\, \delta^\rho{}_\sigma, \tag{3.21}$$

which is the transformation law of a mixed tensor of rank 2.

A multidimensional Kronecker delta is defined as follows

$$\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} = \begin{vmatrix} \delta^{\alpha_1}{}_{\beta_1} & \ldots & \ldots & \delta^{\alpha_1}{}_{\beta_k} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \delta^{\alpha_k}{}_{\beta_1} & \ldots & \ldots & \delta^{\alpha_k}{}_{\beta_k} \end{vmatrix}. \tag{3.22}$$

From the definition we have at once

$$\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} = \delta^{[\alpha_1 \ldots \alpha_k]}_{\beta_1 \ldots \beta_k} = \delta^{\alpha_1 \ldots \alpha_k}_{[\beta_1 \ldots \beta_k]}, \tag{3.23}$$

and from here it follows that k must not be greater than the dimension of the manifold, n, or else $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} \equiv 0$. Other than this, k is in general unrelated to n.

Let us consider the case $k = n$. Since an object that is antisymmetric in all n indices has just one independent component, it must be proportional to the Levi-Civita symbol with the same indices. For the upper indices we thus have

$$\delta^{\alpha_1 \ldots \alpha_n}_{\beta_1 \ldots \beta_n} = T_{\beta_1 \ldots \beta_n} \epsilon^{\alpha_1 \ldots \alpha_n}. \tag{3.24}$$

But the quantity $T_{\beta_1 \ldots \beta_n}$ defined above is antisymmetric in $\{\beta_1, \ldots, \beta_n\}$, so it must be proportional to $\epsilon_{\beta_1 \ldots \beta_n}$, thus

$$\delta^{\alpha_1 \ldots \alpha_n}_{\beta_1 \ldots \beta_n} = \lambda \epsilon_{\beta_1 \ldots \beta_n} \epsilon^{\alpha_1 \ldots \alpha_n}. \tag{3.25}$$

To calculate λ it is now enough to substitute in the above formula any sets of indices for which both sides are nonzero. We substitute $\{\alpha_1, \ldots, \alpha_n\} = \{\beta_1, \ldots, \beta_n\} = \{1, \ldots, n\}$, and we see that $\lambda = 1$. Hence, finally

$$\delta^{\alpha_1 \ldots \alpha_n}_{\beta_1 \ldots \beta_n} = \epsilon_{\beta_1 \ldots \beta_n} \epsilon^{\alpha_1 \ldots \alpha_n}. \tag{3.26}$$

This, in consequence of the properties of both ϵ-s, is a tensor (i.e. has the weight zero).

From (3.22) one sees that $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} \neq 0$ only when the values of $\{\alpha_1, \ldots, \alpha_k\}$ are all different and are a permutation of $\{\beta_1 \ldots \beta_k\}$. If any pair of upper indices has equal values, or if any pair of lower indices has equal values, then the determinant in (3.22) has two identical rows or two identical columns and is zero. If any index β_i of the set $\{\beta_1 \ldots \beta_k\}$ is different from all the α-s in $\{\alpha_1, \ldots, \alpha_k\}$, the determinant in (3.22) has only zeros in the i-th column and is zero again. From the antisymmetry property it is also seen that if $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} \neq 0$, then $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} = +1$ when the lower indices are an even permutation of the upper ones, and $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} = -1$ when they are an odd permutation.

Using these properties one can verify that

$$\delta^{\alpha_1 \ldots \alpha_k \rho}_{\beta_1 \ldots \beta_k \rho} = (n-k) \delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k}. \tag{3.27}$$

This is seen as follows. If $\{\beta_1, \ldots, \beta_k\}$ are not a permutation of $\{\alpha_1, \ldots, \alpha_k\}$, or if any index in any of the two sets is repeated, then both sides of (3.27) are zero and the equation holds. Then, when all $\{\alpha_1, \ldots, \alpha_k\}$ are different, while $\{\beta_1, \ldots, \beta_k\}$ are their permutation, then each term in the sum on the left-hand side is equal to $\delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k}$ when $\rho \notin \{\alpha_1, \ldots, \alpha_k\}$, and is zero when $\rho \in \{\alpha_1, \ldots, \alpha_k\}$. Consequently, there are $(n-k)$ values of ρ with which there are nonzero contributions on the left-hand side, and each contribution is equal to the delta on the right. □

The following equations are simple consequences of (3.27):

$$\delta^{\alpha_1 \ldots \alpha_{n-1} \rho}_{\beta_1 \ldots \beta_{n-1} \rho} = \delta^{\alpha_1 \ldots \alpha_{n-1}}_{\beta_1 \ldots \beta_{n-1}}, \tag{3.28}$$

$$\delta^{\rho_1 \ldots \rho_s}_{\rho_1 \ldots \rho_s} = (n-s+1)(n-s+2) \ldots n = \frac{n!}{(n-s)!}, \tag{3.29}$$

$$\delta^{\rho_1 \ldots \rho_n}_{\rho_1 \ldots \rho_n} = n!, \tag{3.30}$$

$$\delta^{\alpha_1 \ldots \alpha_k \rho_{k+1} \ldots \rho_n}_{\beta_1 \ldots \beta_k \rho_{k+1} \ldots \rho_n} = (n-k)! \; \delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k}. \tag{3.31}$$

3.12 Examples of applications of the Levi-Civita symbol and of the multidimensional Kronecker delta.

The Levi-Civita symbols and the multidimensional deltas are useful in calculations in which determinants or antisymmetrisations appear because they allow to replace tricky reasonings with simple rules of calculation.

With the help of the equations (3.16), (3.26) and (3.30) we can easily verify that

$$\det(A^{\cdot}{}_{\cdot}) = \frac{1}{n!} \delta^{\beta_1 \ldots \beta_n}_{\alpha_1 \ldots \alpha_n} A^{\alpha_1}{}_{\beta_1} \ldots A^{\alpha_n}{}_{\beta_n}, \tag{3.32}$$

and so the determinant of a mixed tensor is a scalar;

$$\det(B_{..}) = \frac{1}{n!}\epsilon^{\alpha_1\ldots\alpha_n}\epsilon^{\beta_1\ldots\beta_n}B_{\alpha_1\beta_1}\ldots B_{\alpha_n\beta_n}, \tag{3.33}$$

and so the determinant of a doubly covariant tensor is a scalar density of weight (-2);

$$\det(C^{..}) = \frac{1}{n!}\epsilon_{\alpha_1\ldots\alpha_n}\epsilon_{\beta_1\ldots\beta_n}C^{\alpha_1\beta_1}\ldots C^{\alpha_n\beta_n}, \tag{3.34}$$

and so the determinant of a doubly contravariant tensor is a scalar density of weight $(+2)$.

One can also verify that the antisymmetrisation with respect to any set of indices can be written as follows:

$$T_{[\alpha_1\ldots\alpha_k]} = \frac{1}{k!}\delta^{\rho_1\ldots\rho_k}_{\alpha_1\ldots\alpha_k}T_{\rho_1\ldots\rho_k}; \tag{3.35}$$

and similarly for upper indices.

3.13 Exercises.

1. Prove that the cofactor of the element $A^\alpha{}_\beta$ in a mixed tensor $\{A^\alpha{}_\beta\}$ is

$$M^\beta{}_\alpha = \frac{1}{(n-1)!}\delta^{\beta\beta_1\ldots\beta_{n-1}}_{\alpha\alpha_1\ldots\alpha_{n-1}}A^{\alpha_1}{}_{\beta_1}\ldots A^{\alpha_{n-1}}{}_{\beta_{n-1}}. \tag{3.36}$$

2. Find the formulae for the cofactors of the elements $B_{\alpha\beta}$ and $C^{\alpha\beta}$ in a doubly covariant and a doubly contravariant tensor, respectively. Note that the cofactor has in each case its indices positioned opposite to its corresponding element.

3. Find the formula for the coefficient of λ^i in the characteristic equation for a matrix $M^\mu{}_\nu$:

$$\det\left(M^\mu{}_\nu - \lambda\delta^\mu{}_\nu\right) = 0.$$

Prove that all the coefficients of this polynomial are scalars. Note the coefficients of λ^0 and of λ^{n-1} – what functions of the matrix are they?

Chapter 4

Covariant derivatives.

4.1 Differentiation of tensors.

Let us calculate the derivative of a contravariant vector field v^α after it had been transformed from the $\{x\}$-coordinates to the $\{x'\}$-coordinates. We have

$$v^{\alpha'}{}_{,\beta'} = \left(x^{\alpha'}{}_{,\alpha}\, v^\alpha\right)_{,\beta'} = x^{\alpha'}{}_{,\alpha\beta'}\, v^\alpha + x^{\alpha'}{}_{,\alpha}\, v^\alpha{}_{,\beta'} = x^{\alpha'}{}_{,\alpha\beta}\, x^\beta{}_{,\beta'}\, v^\alpha + x^{\alpha'}{}_{,\alpha}\, x^\beta{}_{,\beta'}\, v^\alpha{}_{,\beta}\,. \quad (4.1)$$

This is not a tensor, in consequence of the term $x^{\alpha'}{}_{,\alpha\beta}\, x^\beta{}_{,\beta'}\, v^\alpha$. An analogous result would be obtained for most other tensors. The derivative of an *arbitrary* tensor field transforms like a tensor only under linear transformations, for which $x^{\alpha'}{}_{,\alpha\beta} = 0$. There are only a few special cases in which the derivatives of tensor fields are themselves tensors with respect to arbitrary coordinate transformations. One example we already know – it is the derivative of a scalar field, which is a covariant vector. The three other examples are:

1. The derivatives of the Levi-Civita symbols and of all the Kronecker deltas are identically equal to zero.

2. If $T_{\alpha_1\ldots\alpha_k}$ is a tensor (of weight 0), then $T_{[\alpha_1\ldots\alpha_k,\alpha_{k+1}]}$ is a tensor, too. This quantity is a generalisation of the rotation of a vector field.

3. If $T^{\alpha_1\ldots\alpha_k}$ is a tensor density of weight -1, and is completely antisymmetric in all the indices, then $T^{\alpha_1\ldots\alpha_k}{}_{,\alpha_k}$ is also a tensor density of weight -1. This is a generalisation of the divergence of a vector field.

The first example is trivial, while the second one is easy to verify (hint: we consider only functions of class C^2, for which second derivatives commute). We will verify the third example because it provides an application of multidimensional deltas in a calculation.

By assumption, when the coordinates are transformed from $\{x\}$ to $\{x'\}$, $T^{\alpha_1\ldots\alpha_k}$ transforms as follows

$$T^{\alpha'_1\ldots\alpha'_k} = \left(\frac{\partial(x')}{\partial(x)}\right)^{-1} x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_k}{}_{,\alpha_k}\, T^{\alpha_1\ldots\alpha_k}. \quad (4.2)$$

Let us differentiate this by $x^{\alpha'_k}$ and contract the result by α'_k. In the term where differentiation acts on $\left(\frac{\partial(x')}{\partial(x)}\right)^{-1} = \frac{\partial(x)}{\partial(x')}$, we use (3.32). We have

$$\begin{aligned}
T^{\alpha'_1\ldots\alpha'_k}{}_{,\alpha'_k} &= \left(\frac{1}{n!}\sum_{i=1}^{n} \delta^{\rho'_1\ldots\rho'_n}_{\sigma_1\ldots\sigma_n} x^{\sigma_1}{}_{,\rho'_1}\ldots x^{\sigma_i}{}_{,\rho'_i\alpha'_k}\ldots x^{\sigma_n}{}_{,\rho'_n}\right) x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_k}{}_{,\alpha_k} T^{\alpha_1\ldots\alpha_k} \\
&+ \frac{\partial(x)}{\partial(x')}\sum_{i=1}^{k-1} x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_i}{}_{,\alpha_i\alpha'_k}\ldots x^{\alpha'_k}{}_{,\alpha_k} T^{\alpha_1\ldots\alpha_k} \\
&+ \frac{\partial(x)}{\partial(x')} x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_{k-1}}{}_{,\alpha_{k-1}} x^{\alpha'_k}{}_{,\alpha_k\alpha'_k} T^{\alpha_1\ldots\alpha_k} \\
&+ \frac{\partial(x)}{\partial(x')} x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_k}{}_{,\alpha_k} T^{\alpha_1\ldots\alpha_k}{}_{,\rho}\, x^{\rho}{}_{,\alpha'_k}\,.
\end{aligned} \tag{4.3}$$

In the first term we use the property that δ is antisymmetric with respect to both upper and lower indices, so interchanging simultaneously a pair of upper indices and a pair of lower indices does not change its sign. Hence, the products in the first term can be ordered so that the factor with the second derivative is contracted with the last indices of the delta. After such an ordering we see that all the n components of the sum are identical, so

$$\text{first term} = \left[\frac{1}{(n-1)!}\delta^{\rho'_1\ldots\rho'_n}_{\sigma_1\ldots\sigma_n} x^{\sigma_1}{}_{,\rho'_1}\ldots x^{\sigma_{n-1}}{}_{,\rho'_{n-1}} x^{\sigma_n}{}_{,\rho'_n\alpha'_k}\right] x^{\alpha'_1}{}_{,\alpha_1}\ldots x^{\alpha'_k}{}_{,\alpha_k} T^{\alpha_1\ldots\alpha_k}. \tag{4.4}$$

In the second term of (4.3), each component of the sum contains the expression $x^{\alpha'_i}{}_{,\alpha_i\alpha'_k}\, x^{\alpha'_k}{}_{,\alpha_k} \equiv x^{\alpha'_i}{}_{,\alpha_i\alpha_k}$, which is symmetric in (α_i, α_k). This expression is contracted with respect to both (α_i, α_k) with $T^{\alpha_1\ldots\alpha_k}$, which is antisymmetric in these same two indices. Such a contraction is always identically zero, hence the second term is zero.

In the third term, we note that $x^{\alpha'_k}{}_{,\alpha_k}$ is an element of the inverse matrix to $[x^{\alpha}{}_{,\alpha'}]$. Consequently, $x^{\alpha'_k}{}_{,\alpha_k}$ is equal to the cofactor of the element transposed to $\left({}^{\alpha'_k}{}_{\alpha_k}\right)$ in the matrix $[x^{\alpha}{}_{,\alpha'}]$, divided by the determinant of $[x^{\alpha}{}_{,\alpha'}]$. The element transposed to $\left({}^{\alpha'_k}{}_{\alpha_k}\right)$ in $[x^{\alpha}{}_{,\alpha'}]$ is $x^{\alpha_k}{}_{,\alpha'_k}$, so its cofactor is

$$\frac{1}{(n-1)!}\delta^{\alpha'_k\mu'_1\ldots\mu'_{n-1}}_{\alpha_k\nu_1\ldots\nu_{n-1}} x^{\nu_1}{}_{,\mu'_1}\ldots x^{\nu_{n-1}}{}_{,\mu'_{n-1}}\,,$$

while $\det[x^{\alpha}{}_{,\alpha'}] = \frac{\partial(x)}{\partial(x')}$. Hence we have

$$\left(x^{\alpha'_k}{}_{,\alpha_k}\right)_{,\alpha'_k} = \left[\left(\frac{\partial(x)}{\partial(x')}\right)^{-1} \frac{1}{(n-1)!}\delta^{\alpha'_k\mu'_1\ldots\mu'_{n-1}}_{\alpha_k\nu_1\ldots\nu_{n-1}} x^{\nu_1}{}_{,\mu'_1}\ldots x^{\nu_{n-1}}{}_{,\mu'_{n-1}}\right]_{,\alpha'_k}. \tag{4.5}$$

The differentiation of $x^{\nu_i}{}_{,\mu'_i}$ by $x^{\alpha'_k}$ will give zero contributions because $x^{\nu_i}{}_{,\mu'_i\alpha'_k}$, are symmetric in (μ'_i, α'_k) and will be contracted with the delta which is antisymmetric in the same indices. The only nonzero contribution will be from the derivative of the determinant, so

$$\begin{aligned}
&\left(x^{\alpha'_k}{}_{,\alpha_k}\right)_{,\alpha'_k} \\
&= -\left(\frac{\partial(x)}{\partial(x')}\right)^{-2}\left[\frac{1}{n!}\delta^{\rho'_1\ldots\rho'_n}_{\sigma_1\ldots\sigma_n} x^{\sigma_1}{}_{,\rho'_1}\ldots x^{\sigma_n}{}_{,\rho'_n}\right]_{,\alpha'_k} \frac{1}{(n-1)!}\delta^{\alpha'_k\mu'_1\ldots\mu'_{n-1}}_{\alpha_k\nu_1\ldots\nu_{n-1}} x^{\nu_1}{}_{,\mu'_1}\ldots x^{\nu_{n-1}}{}_{,\mu'_{n-1}}
\end{aligned}$$

$$= -\left(\frac{\partial(x)}{\partial(x')}\right)^{-2} \cdot \frac{1}{(n-1)!} \cdot \delta^{\rho'_1 \ldots \rho'_n}_{\sigma_1 \ldots \sigma_n} x^{\sigma_1}{}_{,\rho'_1} \ldots x^{\sigma_{n-1}}{}_{,\rho'_{n-1}} x^{\sigma_n}{}_{,\rho'_n \alpha'_k} \times$$
$$\times \frac{1}{(n-1)!} \delta^{\alpha'_k \mu'_1 \ldots \mu'_{n-1}}_{\alpha_k \nu_1 \ldots \nu_{n-1}} x^{\nu_1}{}_{,\mu'_1} \ldots x^{\nu_{n-1}}{}_{,\mu'_{n-1}} . \tag{4.6}$$

The last line above is just $\left(\frac{\partial(x)}{\partial(x')} x^{\alpha'_k}{}_{,\alpha_k}\right)$. Using this in the third term of (4.3) we obtain

$$\text{third term} =$$
$$-\frac{1}{(n-1)!} \cdot \delta^{\rho'_1 \ldots \rho'_n}_{\sigma_1 \ldots \sigma_n} x^{\sigma_1}{}_{,\rho'_1} \ldots x^{\sigma_{n-1}}{}_{,\rho'_{n-1}} x^{\sigma_n}{}_{,\rho'_n \alpha'_k} x^{\alpha'_1}{}_{,\alpha_1} \ldots x^{\alpha'_k}{}_{,\alpha_k} T^{\alpha_1 \ldots \alpha_k}.$$

We see that the third term has the same absolute value as the first one, but is of opposite sign, so terms I and III cancel each other. The final result in (4.3) is the last term, thus

$$T^{\alpha'_1 \ldots \alpha'_k}{}_{,\alpha'_k} = \frac{\partial(x)}{\partial(x')} x^{\alpha'_1}{}_{,\alpha_1} \ldots x^{\alpha'_{k-1}}{}_{,\alpha_{k-1}} T^{\alpha_1 \ldots \alpha_k}{}_{,\alpha_k} , \tag{4.7}$$

where we used $x^{\alpha'_k}{}_{,\alpha_k} x^{\rho}{}_{,\alpha'_k} = x^{\rho}{}_{,\alpha_k} = \delta^{\rho}{}_{\alpha_k}$. Hence, $T^{\alpha_1 \ldots \alpha_k}{}_{,\alpha_k}$ is a tensor density of weight -1.

Note that the statement proved above is also correct for the case when T^{α} has just one index and is a contravariant vector density. In that case, terms I and III still cancel each other while term II in (4.3) simply does not exist.

The fact that the derivatives of tensor fields are not tensor fields themselves is unfortunate because the laws of physics are usually formulated as differential equations. Hence, those equations are not tensorial; they will change when coordinates are transformed. But we would like the laws of physics to have the form (a tensor) = 0, since such an equation would hold in all the coordinate systems. This suggests the following idea: let us define a "generalised differentiation", which will yield tensor fields when acting on tensor fields, and will coincide with ordinary differentiation when acting on scalars and the Kronecker deltas, for which the partial derivative does not destroy the tensor property. Then, we will replace the partial derivatives with the generalised derivatives in the laws of physics. We guess that this generalised differentiation, called **covariant differentiation**, will reduce to ordinary differentiation in certain privileged coordinate systems.

4.2 Axioms of the covariant derivative.

We want the covariant differentiation to have all the algebraic properties of an ordinary differentiation, but in addition we want it to yield tensor densities when acting on tensor densities. We will denote the covariant derivative by ∇_α, $_{|\alpha}$ or $D/\partial x^\alpha$. The symbols $T_i[w,k,l]$ will denote tensor densities whose indices we do not need to write out explicitly.

Specifically, we want the ∇_α to have the following properties:

1. To be distributive with respect to addition:

$$\nabla_\alpha \left(T_1[w,k,l] + T_2[w,k,l]\right) = \nabla_\alpha \left(T_1[w,k,l]\right) + \nabla_\alpha \left(T_1[w,k,l]\right). \tag{4.8}$$

2. To obey the Leibniz rule when acting on a tensor product:

$$\begin{aligned}
&\nabla_\alpha \left(T_1[w_1, k_1, l_1] \otimes T_2[w_2, k_2, l_2]\right) \\
&= \left(\nabla_\alpha T_1[w_1, k_1, l_1]\right) \otimes T_2[w_2, k_2, l_2] + \left(T_1[w_1, k_1, l_1]\right) \otimes \left(\nabla_\alpha T_2[w_2, k_2, l_2]\right). \qquad (4.9)
\end{aligned}$$

3. To reduce to the partial derivative when acting on a scalar:

$$\nabla_\alpha \Phi = \Phi_{,\alpha}\,. \qquad (4.10)$$

4. To yield zero when acting on the Levi-Civita symbols and Kronecker deltas:

$$\begin{aligned}
\nabla_\alpha \epsilon^{\alpha_1 \ldots \alpha_n} &= 0, \\
\nabla_\alpha \epsilon_{\alpha_1 \ldots \alpha_n} &= 0, \\
\nabla_\alpha \delta^\alpha{}_\beta &= 0. \qquad (4.11)
\end{aligned}$$

The last equation implies at once that

$$\nabla_\alpha \delta^{\alpha_1 \ldots \alpha_k}_{\beta_1 \ldots \beta_k} = 0 \qquad (4.12)$$

for any k. It also implies that ∇_α commutes with contraction.

5. When acting on a tensor density of type $[w, k, l]$, it produces a tensor density of type $[w, k, l+1]$, thus

$$\nabla_\alpha \left(T_1[w, k, l]\right) = T_2[w, k, l+1].$$

Only the last property is different for the covariant and for the partial derivative.

From these requirements we will now derive an operational formula for the covariant derivative.

4.3 A field of vector bases on a manifold and scalar components of tensors.

In every tangent space to an n-dimensional manifold M_n we can choose a set of n linearly independent contravariant vectors, $\{e_1{}^\alpha, \ldots, e_n{}^\alpha\}$. The indices a, b, c, ... will label vectors (as opposed to Greek indices that label coordinate components of tensors). After such a basis of the tangent space is chosen at every $x \in M_n$, let us consider the n vector *fields*:

$$x \to e_a{}^\alpha(x), \qquad a = 1, \ldots, n.$$

The collection of quantities $\{e_a{}^\alpha(x)\}$, $\alpha = 1, \ldots, n$, $a = 1, \ldots, n$ forms a matrix whose elements are functions on the manifold. Since all the vectors are linearly independent at every x, the matrix is nonsingular, so there exists an inverse matrix $e^a{}_\alpha$ that obeys

$$e^a{}_\alpha e_a{}^\beta = \delta^\beta{}_\alpha. \qquad (4.13)$$

Subsets of the matrix $||e^a{}_\alpha||$ defined by a fixed a are then covariant vectors that form a **dual basis** to $\{e_a{}^\alpha(x)\}$, $a = 1, \ldots, n$. One can verify that, in virtue of the $\{e_a{}^\alpha(x)\}$ being linearly independent, eq. (4.13) implies the following:

$$e^a{}_\alpha e_b{}^\alpha = \delta^a{}_b. \tag{4.14}$$

It follows that for any tensor field (i.e. of weight 0) $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$, the collection of quantities

$$T^{a_1\ldots a_k}_{b_1\ldots b_l} := e^{a_1}{}_{\alpha_1}\ldots e^{a_k}{}_{\alpha_k} e_{b_1}{}^{\beta_1}\ldots e_{b_l}{}^{\beta_l} T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}, \tag{4.15}$$

labelled by the indices $a_1, \ldots, a_k, b_1, \ldots, b_l = 1, \ldots, n$ is a set of n^{k+l} scalar fields that uniquely represents the set of n^{k+l} coordinate components of the tensor field $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$. This is because, in consequence of (4.13) – (4.14), an inverse formula to (4.15) exists that allows one to calculate $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$ when $T^{a_1\ldots a_k}_{b_1\ldots b_l}$ are given:

$$T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l} = e_{a_1}{}^{\alpha_1}\ldots e_{a_k}{}^{\alpha_k} e^{b_1}{}_{\beta_1}\ldots e^{b_l}{}_{\beta_l} T^{a_1\ldots a_k}_{b_1\ldots b_l}. \tag{4.16}$$

Let us denote

$$e := \det ||e_a{}^\alpha|| = \frac{1}{n!}\epsilon^{a_1\ldots a_n}\epsilon_{\alpha_1\ldots\alpha_n} e_{a_1}{}^{\alpha_1}\ldots e_{a_n}{}^{\alpha_n}. \tag{4.17}$$

Now, $\epsilon_{\alpha_1\ldots\alpha_n}$ is a tensor density of weight $+1$, while $\epsilon^{a_1\ldots a_n}$ is a set of scalars because it depends on the choice of basis in the vector space, and not on the coordinate system. Hence, e is a scalar density of weight $+1$.

The quantity e, together with the bases $\{e_a{}^\alpha\}$ and $\{e^a{}_\alpha\}$ can be used to represent arbitrary tensor densities by sets of scalars. Let $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$ now be a tensor density of type $[w, k, l]$; then each element of the set

$$T^{a_1\ldots a_k}_{b_1\ldots b_l} := e^{-w} e^{a_1}{}_{\alpha_1}\ldots e^{a_k}{}_{\alpha_k} e_{b_1}{}^{\beta_1}\ldots e_{b_l}{}^{\beta_l} T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l} \tag{4.18}$$

is a scalar. The set $T^{a_1\ldots a_k}_{b_1\ldots b_l}$ uniquely defines $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$ via (4.16) with the factor e^{+w} added. The weight w has to be given as an extra bit of information, since the set of scalars alone does not define the weight.

4.4 The affine connection.

We now define the set of quantities:

$$\Gamma^\alpha{}_{\beta\gamma} = -e_s{}^\alpha\left(\nabla_\gamma - \partial_\gamma\right) e^s{}_\beta. \tag{4.19}$$

The elements of this set are the coefficients of **affine connection**.When specified explicitly, they tell us how the covariant derivative acts on the basis vector fields. Later we will consider manifolds in which these coefficients can be calculated from more basic objects (see Chapter 7). For now, we consider manifolds in which the $\Gamma^\alpha{}_{\beta\gamma}$ are just given.

Equation (4.19) can be rewritten in an equivalent form

$$\nabla_\gamma e^a{}_\beta = \partial_\gamma e^a{}_\beta - \Gamma^\alpha{}_{\beta\gamma} e^a{}_\alpha.$$

We will verify that the $\Gamma^\alpha{}_{\beta\gamma}$ do not depend on the choice of basis. Let us assume that $\{e_a{}^\alpha\}$ and $\{e_{a'}{}^\alpha\}$ are two different bases. The vectors of the second basis can then be decomposed in the first basis

$$e_{a'}{}^\alpha = A^b{}_{a'} e_b{}^\alpha, \tag{4.20}$$

and the elements of the transformation matrix

$$A^b{}_{a'} = e^b{}_\alpha e_{a'}{}^\alpha \tag{4.21}$$

are scalar fields. Hence, $A^b{}_{a'|\alpha} = A^b{}_{a',\alpha}$ and $\left(A^{-1}\right)^{c'}{}_{d|\alpha} = \left(A^{-1}\right)^{c'}{}_{d,\alpha}$. Then, calculating the $\Gamma^\alpha{}_{\beta\gamma}$ in the basis $\{e_{a'}{}^\alpha\}$ we have

$$\begin{aligned}\left(\Gamma^\alpha{}_{\beta\gamma}\right)_{e'} &= -e_{s'}{}^\alpha \left(\nabla_\gamma - \partial_\gamma\right) e^{s'}{}_\beta = -A^r{}_{s'} e_r{}^\alpha \left(\nabla_\gamma - \partial_\gamma\right) \left[\left(A^{-1}\right)^{s'}{}_s e^s{}_\beta\right] \\ &= -A^r{}_{s'} \left(A^{-1}\right)^{s'}{}_s e_r{}^\alpha \left(\nabla_\gamma - \partial_\gamma\right) e^s{}_\beta = -\delta^r{}_s e_r{}^\alpha \left(\nabla_\gamma - \partial_\gamma\right) e^s{}_\beta \\ &= -e_s{}^\alpha \left(\nabla_\gamma - \partial_\gamma\right) e^s{}_\beta = \left(\Gamma^\alpha{}_{\beta\gamma}\right)_e .\end{aligned} \tag{4.22}$$

Now let us note that the $\Gamma^\alpha{}_{\beta\gamma}$ *are not tensor fields.* When coordinates are transformed, these coefficients change as follows

$$\Gamma^{\alpha'}{}_{\beta'\gamma'} = -e_s{}^{\alpha'} \left(\nabla_{\gamma'} - \partial_{\gamma'}\right) e^s{}_{\beta'} = x^{\alpha'}{}_{,\alpha}\, x^\beta{}_{,\beta'}\, x^\gamma{}_{,\gamma'}\, \Gamma^\alpha{}_{\beta\gamma} + x^{\alpha'}{}_{,\rho}\, x^\rho{}_{,\beta'\gamma'} . \tag{4.23}$$

However, the antisymmetric part of $\Gamma^\alpha{}_{\beta\gamma}$

$$\Omega^\alpha{}_{\beta\gamma} \stackrel{\text{def}}{=} \Gamma^\alpha{}_{[\beta\gamma]}, \tag{4.24}$$

is a tensor, called the **torsion tensor**, since $x^\rho{}_{,[\beta'\gamma']} = 0$.

4.5 The explicit formula for the covariant derivative of tensor densities.

In order to obtain the explicit formula for the covariant derivative, we need to know two other properties of the connection coefficients:

$$(I) \quad \Gamma^\alpha{}_{\beta\gamma} = e^s{}_\beta \left(\nabla_\gamma - \partial_\gamma\right) e_s{}^\alpha. \tag{4.25}$$

The verification of this is an easy exercise.

$$(II) \quad \nabla_\alpha \left(e^w\right) = w e^{w-1} \nabla_\alpha e. \tag{4.26}$$

This can be verified in the following way. Let us consider the quantity

$$F_\alpha(w) := e^{-w} \nabla_\alpha \left(e^w\right). \tag{4.27}$$

Using the postulated properties of the covariant derivative we obtain:

$$F_\alpha \left(w_1 + w_2\right) = e^{-w_1} e^{-w_2} \left[\left(\nabla_\alpha e^{w_1}\right) e^{w_2} + e^{w_1} \left(\nabla_\alpha e^{w_2}\right)\right]$$

$$= e^{-w_1}\left(\nabla_\alpha e^{w_1}\right) + e^{-w_2}\left(\nabla_\alpha e^{w_2}\right) = F_\alpha(w_1) + F_\alpha(w_2). \tag{4.28}$$

Every continuous function that has the property $f(w_1 + w_2) = f(w_1) + f(w_2)$ for all real w_1 and w_2 also has the property $f(w) = f(1)w$. Hence

$$e^{-w}\nabla_\alpha\left(e^w\right) = we^{-1}\nabla_\alpha e,$$

which is equivalent to (4.26).

Equation (4.26) holds also for partial derivatives, so

$$e^{-w}\left(\nabla_\gamma - \partial_\gamma\right)\left(e^w\right) = we^{-1}\left(\nabla_\gamma - \partial_\gamma\right)e. \tag{4.29}$$

Now, using (3.32) and (3.36), we obtain

$$\begin{aligned}
\left(\nabla_\gamma - \partial_\gamma\right)e &= \frac{1}{(n-1)!}\delta^{\rho_1\ldots\rho_n}_{a_1\ldots a_n}e_{a_1}{}^{\rho_1}\ldots e_{a_{n-1}}{}^{\rho_{n-1}}\left(\nabla_\gamma - \partial_\gamma\right)e_{a_n}{}^{\rho_n} \\
&= ee^{a_n}{}_{\rho_n}\left(\nabla_\gamma - \partial_\gamma\right)e_{a_n}{}^{\rho_n} = e\Gamma^\rho{}_{\rho\gamma}.
\end{aligned} \tag{4.30}$$

At this point, we are prepared to deduce the general formula for the covariant derivative of an arbitrary tensor density. As an introductory exercise we do it first for contravariant and covariant vector densities.

Let us convert a contravariant vector density A^α of weight w to a scalar by (4.18). Since $A^a = e^{-w}A^\alpha$ is a scalar, axiom 3 implies

$$\left(\nabla_\gamma - \partial_\gamma\right)A^a = 0. \tag{4.31}$$

On the other hand, using now axiom 2 and (4.30), we apply $\left(\nabla_\gamma - \partial_\gamma\right)$ to the right-hand side of (4.18) and obtain

$$\begin{aligned}
&\left(\nabla_\gamma - \partial_\gamma\right)A^a = \\
&-we^{-w}\Gamma^\rho{}_{\rho\gamma}e^a{}_\alpha A^\alpha + e^{-w}\left[\left(\nabla_\gamma - \partial_\gamma\right)e^a{}_\alpha\right]A^\alpha + e^{-w}e^a{}_\alpha\left(\nabla_\gamma - \partial_\gamma\right)A^\alpha.
\end{aligned} \tag{4.32}$$

Now let us convert this equation back to coordinate components, by contracting it with $e^we_a{}^\alpha$ (first change the summation index α!). Using eqs. (4.19) and (4.25), we get

$$\left(\nabla_\gamma - \partial_\gamma\right)A^\alpha = w\Gamma^\rho{}_{\rho\gamma}A^\alpha + \Gamma^\alpha{}_{\rho\gamma}A^\rho. \tag{4.33}$$

From here, finally

$$\nabla_\gamma A^\alpha = \partial_\gamma A^\alpha + w\Gamma^\rho{}_{\rho\gamma}A^\alpha + \Gamma^\alpha{}_{\rho\gamma}A^\rho. \tag{4.34}$$

By similar calculations we obtain for a *covariant* vector density B_α of weight w:

$$\nabla_\gamma B_\alpha = \partial_\gamma B_\alpha + w\Gamma^\rho{}_{\rho\gamma}B_\alpha - \Gamma^\rho{}_{\alpha\gamma}B_\rho, \tag{4.35}$$

and for tensors of rank 2:

$$T^{\alpha\beta}{}_{|\gamma} = T^{\alpha\beta}{}_{,\gamma} + \Gamma^\alpha{}_{\rho\gamma}T^{\rho\beta} + \Gamma^\beta{}_{\rho\gamma}T^{\alpha\rho}; \tag{4.36}$$

$$T_{\alpha\beta|\gamma} = T_{\alpha\beta,\gamma} - \Gamma^\rho{}_{\alpha\gamma}T_{\rho\beta} - \Gamma^\rho{}_{\beta\gamma}T_{\alpha\rho}; \tag{4.37}$$

$$T^{\alpha}{}_{\beta|\gamma} = T^{\alpha}{}_{\beta,\gamma} + \Gamma^{\alpha}{}_{\rho\gamma}T^{\rho}{}_{\beta} - \Gamma^{\rho}{}_{\beta\gamma}T^{\alpha}{}_{\rho}. \tag{4.38}$$

For a general tensor density of weight w, $T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l}$, still by the same reasoning, we obtain

$$\begin{aligned}\nabla_\gamma T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l} &= \partial_\gamma T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l} + w\Gamma^{\rho}{}_{\rho\gamma}T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\beta_l} + \sum_{i=1}^{k}\Gamma^{\alpha_i}{}_{\rho_i\gamma}T^{\alpha_1\ldots\rho_i\ldots\alpha_k}_{\beta_1\ldots\beta_l} \\ &- \sum_{j=1}^{l}\Gamma^{\rho_j}{}_{\beta_j\gamma}T^{\alpha_1\ldots\alpha_k}_{\beta_1\ldots\rho_j\ldots\beta_l},\end{aligned} \tag{4.39}$$

where the sums run through all the positions of the respective indices.

Note that, unlike a partial derivative, the covariant derivative does not act on single components of tensor densities. It is an operator that acts on the whole tensor density and produces another tensor density.

4.6 Exercises.

1. What is the condition for the "covariant rotation" $T_{[\alpha|\beta]}$ of a covariant vector field T_α to coincide with the ordinary rotation $T_{[\alpha,\beta]}$?

2. Let $g_{\alpha\beta} = g_{(\alpha\beta)}$ be a doubly covariant tensor that is nonsingular, i.e. $\det||g_{\alpha\beta}|| \neq 0$. Let $g^{\alpha\beta}$ be its inverse matrix, i.e.

$$g^{\alpha\rho}g_{\rho\beta} = \delta^{\alpha}{}_{\beta}.$$

Show that the object defined as follows

$$\begin{Bmatrix}\alpha\\ \beta\gamma\end{Bmatrix} = \frac{1}{2}g^{\alpha\rho}\left(g_{\beta\rho,\gamma} + g_{\gamma\rho,\beta} - g_{\beta\gamma,\rho}\right)$$

transforms under coordinate transformations by the same law as the coefficients of affine connection. What is the torsion in this case?