Mechanics in Special Relativity

Let us now explain how classical mechanics can be reformulated in the context of Special Relativity. The trajectory that any point-like object follows in the course of time forms a curve in the spacetime M, which is called a *world line*. We begin this section by considering curves in M.

Let $I = (a,b) \subset \mathbb{R}$ be an open interval and consider a (parameterised) curve $\xi : I \to M$. We choose some inertial coordinate system x^μ to make an identification $X : M \to \mathbb{R}^4$, so that $X(\xi(s)) = \sum_{\mu=0}^{3} \xi^\mu(s) e_\mu$ with coordinates $\xi^\mu := x^\mu \circ \xi$. We also assume that the ξ^μ are C^1 (continuously differentiable) and we denote the derivatives w.r.t. the parameter $s \in I$ by a dot, e.g. $\dot{\xi}^\mu(s)$. We will also write $\dot{\xi}(s) := \sum_{\mu=0}^{3} \dot{\xi}^\mu(s) e_\mu$. Curves are a natural way to model the trajectories of particles or other physical objects.

We may distinguish several kinds of curves, which have a definite causal character and time orientation:

Definition 6.1 *A C^1 curve $\xi : I \to M$ is called* space-like, *resp.* light-like, *resp.* time-like, *resp.* causal, *when for all $s \in I$, $\eta(\dot{\xi}(s), \dot{\xi}(s)) > 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) = 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) < 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) \leq 0$.*

A causal C^1 curve ξ is future, *resp.* past directed *when for all $s \in I$, $\dot{\xi}^0(s) \geq 0$, resp. $\dot{\xi}^0(s) \leq 0$.*

Now consider a C^1 bijection $s : I' \to I$ with a C^1 inverse, where $I' = (a', b')$. We call the curve $\xi' : I' \to M$ defined by $\xi'(s') := \xi(s(s'))$ a reparametrisation of ξ. The image of ξ' coincides with the image of ξ. The speed with which this image is traced out may differ, but the velocity vectors are always parallel:

$$\dot{\xi}'^\mu(s') = \dot{\xi}^\mu(s(s')) \partial_{s'} s(s').$$

We say that ξ' has the same direction as ξ when $s : I' \to I$ preserves the orientation, i.e. when $s(a') < s(b')$. Otherwise we say that ξ' has the opposite direction as ξ.

Exercise 6.2 *Show that the causal character (space-like, light-like, time-like or causal) of a C^1 curve is independent of the choice of the parametrisation. Show that the time-orientation of a causal curve is preserved when the reparametrisation preserves the direction of the curves.*

Exercise 6.3 *All curves in this exercise are assumed to have C^1 coordinates:*

1. *Find a curve which is neither space-like nor causal.*
2. *Find a spacelike curve $\xi : (-2, 2)$ such that $\xi(1)$ lies to the future of $\xi(-1)$.*
3. *Find a causal curve which is neither light-like not time-like.*
4. *Find a causal curve which is neither future nor past directed.*
5. *Show that every time-like curve is either future or past directed.*

In analogy to the length of a curve in Euclidean geometry, we can define the *arc length* of a space-like C^1 curve by:

$$l(\xi) := \int_I \sqrt{\eta(\dot{\xi}(s), \dot{\xi}(s))} \mathrm{d}\, s$$

For a causal C^1 curve we similarly define the *proper time* to be

$$\tau(\xi) := \frac{1}{c} \int_I \sqrt{-\eta(\dot{\xi}(s), \dot{\xi}(s))} \mathrm{d}\, s.$$

Using the arc length and proper time we may find preferred parametrisations for all causal and space-like C^1 curves:

Definition 6.4 *A C^1 space-like curve is* parameterised by arc-length *when $\eta(\dot{\xi}, \dot{\xi}) = 1$.*

A C^1 time-like curve is parameterised by proper time *when $\eta(\dot{\xi}, \dot{\xi}) = -c^2$.*

Theorem 6.5 *A space-like curve ξ can always be parameterised by arc-length, without changing its direction. A time-like curve ξ can always be parameterised by proper time, without changing its direction (or time-orientation). These parametrisations are independent of the choice of inertial coordinates (up to a choice of units).*

Proof: In the space-like case we may choose $s'(s)$ such that $\partial_s s'(s) = \sqrt{\eta(\dot{\xi}(s), \dot{\xi}(s))}$ and in the time-like case such that $\partial_s s'(s) = \frac{1}{c}\sqrt{-\eta(\dot{\xi}(s), \dot{\xi}(s))}$. In this way we find orientation preserving changes of parameter and it is straight-forward to check that they implement the desired conditions. Independence of the choice of inertial coordinates follows from the independence of the spacetime interval (up to a choice of units). □

Exercise 6.6 *Show that the arc length and proper time are independent of the choice of coordinates x^μ and of the parametrisation of ξ. This shows that the parametrisations by arc length or proper time are independent of the choice of inertial coordinates.*

When a future directed time-like curve $\xi(\tau)$ is parameterised by proper time, we may define its velocity and acceleration as:

$$\begin{aligned}\dot{\xi}(\tau) &:= \sum_{\mu=0}^{3} \dot{\xi}^\mu(\tau) e_\mu, \\ \ddot{\xi}(\tau) &:= \sum_{\mu=0}^{3} \ddot{\xi}^\mu(\tau) e_\mu.\end{aligned}$$

The expressions on the left are independent of the choice of inertial coordinates, if we let a change of coordinates also affect the basis e_μ. By the *(rest) mass* m_0 of a particle we mean the mass that it has in an inertial coordinate system where it is at rest, i.e. a frame where it maintains a fixed position in space, so that it follows the trajectory $\tau \mapsto (c\tau, \mathbf{x}_0)$ for some fixed $\mathbf{x}_0$. When such a particle traverses a general, future directed, time-like curve ξ, we define its *energy-momentum* vector to be

$$P^\mu(\tau) := m_0 \dot{\xi}^\mu(\tau), \tag{8}$$

where τ is the proper time of the curve. In Special Relativity, Newton's first law then takes the form

$$F^\mu(\tau) = \partial_\tau P^\mu(\tau) \tag{9}$$

Let us now fix any choice of inertial coordinates and express the formulae above in terms of the coordinate time x^0, rather than the proper time τ. If a curve is time-like, then $\dot{\xi}^0(s) \neq 0$ for all $s \in I$ and (after reversing the orientation of the parameter, if necessary) we may assume that $\dot{\xi}^0(s) > 0$. We may then introduce a new parameter t, defined by $t(s) := c^{-1} \int_{s_0}^{s} \dot{\xi}^0$ for some $s_0 \in I$. (The map $s \mapsto t(s)$ will be an orientation preserving diffeomorphism on I, because $\dot{\xi}^0 > 0$.) The new parametrisation leads to

$$\dot{\xi}^0(s) = \partial_s \xi'^0(t(s)) = \dot{\xi}'^0(t)\dot{t}(s)$$

and hence $\dot{\xi}'^0(t) = c$. This means that, after changing parametrisation, we may assume that $\xi^0(t) = ct$ and $\xi^\mu(t) = (ct, \mathbf{x}(t))$.

The parameter derivative is now the same as the time derivative w.r.t. the inertial time coordinate $t = c^{-1}x^0$:

$$\dot{\xi}^\mu(t) = (c, \dot{\mathbf{x}}(t)) =: (c, \mathbf{v}(t)).$$

The formula above then shows that (in an arbitrary choice of inertial coordinates) the speed of the time-like curve satisfies $\|\mathbf{v}(t)\| < c$. This shows that time-like, future pointing curves model the trajectories of massive particles (also in the presence of forces). When ξ is a light-like curve (but $\dot{\xi} \neq 0$) we find in a similar way that $\|\mathbf{v}(t)\| = c$, which models the trajectory of a massless particle or light ray. Similarly, space-like curves may have speeds $\|\mathbf{v}(t)\| > c$, or even infinite speeds. Such curves do not model any known matter in the universe. Hypothetical particles that move faster than light are called *tachyons*.

A comparison of the parametrisation $\xi^\mu(t)$ with the parametrisation by proper time yields:

$$\begin{aligned}
\dot{\xi}^\mu(\tau) &= \partial_\tau t(\tau)\dot{\xi}^\mu(t(\tau)) = \gamma(\tau)\ \left(c, \mathbf{v}(t(\tau))\right), \\
\partial_\tau t(\tau) &= \gamma(\tau) := (1 - c^{-2}\|\mathbf{v}(t(\tau))\|^2)^{-\frac{1}{2}},
\end{aligned}$$

where the second line follows from the normalisation $c^2 = -\eta(\dot{\xi}^\mu(\tau), \dot{\xi}^\mu(\tau))$. It follows that

$$P^\mu(\tau) = m_0\gamma(\tau)\dot{\xi}^\mu(t(\tau)) = m_0\gamma(\tau)\ \left(c, \mathbf{v}(t(\tau))\right) =: (c^{-1}E(t(\tau)), \mathbf{P}(t(\tau))),$$

so that spatial components of the energy-momentum vector $P^\mu(\tau)$ in the given inertial frame take the form $\mathbf{P}(t) = m(t)\mathbf{v}(t)$, where the apparent (or inertial) mass in this coordinate frame is $m(t) = \gamma(\tau(t))m_0$. (Note that a high velocity implies a high apparent mass.) The energy in the given inertial frame satisfies

$$E(t) = \gamma(\tau(t))m_0c^2 = m_0c^2 + \frac{m_0\|\mathbf{v}(t)\|^2}{2} + \ldots,$$

where we made a Taylor expansion in $c^{-1}\|\mathbf{v}\|$. From this we see that the total energy in the given coordinate frame consists not only of the kinetic energy, but also a contribution from the rest mass m_0c^2 and higher order corrections. Note that the total energy is frame dependent: in the rest frame of the particle, only the first term remains.

Taking a further derivative we find

$$\begin{aligned} F^\mu(\tau) &= \partial_\tau P^\mu(\tau) = \gamma(\tau)\ \left(c^{-1}\partial_t E(t(\tau)), \partial_t \mathbf{P}(t(\tau))\right) \\ &= \gamma(\tau)\ \left(c^{-1}\partial_t E(t(\tau)), \mathbf{F}(t(\tau))\right), \end{aligned}$$

where $\mathbf{F}(t) = \partial_t \mathbf{P}(t)$ corresponds to the force appearing in Newton's second law (with a time-dependent mass $m(t)$) in the given inertial frame.

Observer Dependence and Paradoxes

To conclude we present some examples of effects in Special Relativity which are counter-intuitive and whose resolution relies on the fact that some of the quantities used are defined in a coordinate dependent way.

Throughout this section we will consider two sets of inertial coordinates, x^μ and x'^μ, related by a boost in the x^1-direction with relative speed v.

Length Contraction In the coordinate system x^μ we may use x^0 to measure time and $\mathbf{x}$ to measure distances and we consider a plank of length $l > 0$, whose endpoints at $x^0 = s$ are located at $(s, 0, 0, 0)$ and $(s, l, 0, 0)$, respectively, so the plank is at rest. In the coordinates x'^μ, the left and right endpoints are located at $(s\cosh(\theta), -s\sinh(\theta), 0, 0)$ and $(s\cosh(\theta) - l\sinh(\theta), l\cosh(\theta) - s\sinh(\theta), 0, 0)$, respectively, where θ is the rapidity of the boost.

If we now use x'^0 to measure time and $\mathbf{x}'$ to measure distances, then we can find the endpoints of the plank at a fixed time $x'^0 = s'$ as follows. For the left endpoint we set $s' = s\cosh(\theta)$, which leads to

$$(s\cosh(\theta), -s\sinh(\theta), 0, 0) = (s', -s'\tanh(\theta), 0, 0).$$

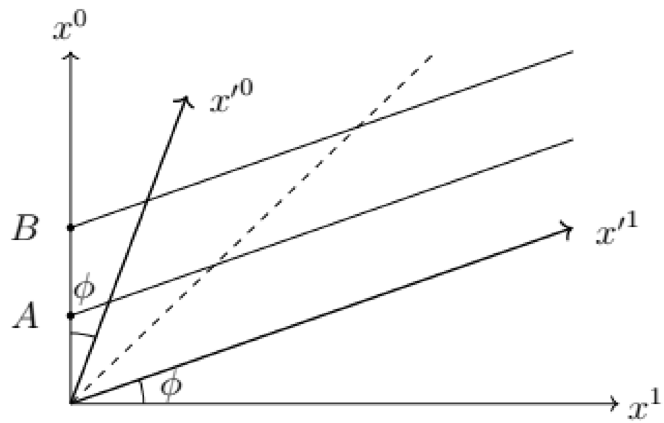
For the right endpoint we set $s' = s\cosh(\theta) - l\sinh(\theta)$, which leads to $s = s'\cosh(\theta)^{-1} + l\tanh(\theta)$ and hence

$$\begin{aligned} &(s\cosh(\theta) - l\sinh(\theta), l\cosh(\theta) - s\sinh(\theta), 0, 0) \\ = \ &(s', l\cosh(\theta)^{-1} - s'\tanh(\theta), 0, 0). \end{aligned}$$

In the primed coordinate system, the plank moves with a constant speed $v = c\tanh(\theta)$ in the negative x'^1-direction, as expected, but the length of the plank is

$$l' = l\cosh(\theta)^{-1} = l\gamma^{-1} < l. \tag{10}$$

This illustrates that length is a coordinate dependent notion. The plank is longest in the inertial coordinates in which it is at rest. The effect that in a boosted inertial frame all lengths in the direction of the boost are reduced, is called *length contraction*.[4] (Note that this effect is mutual: a plank which is at rest in the frame x'^{μ} will also appear contracted in the frame x^{μ}.)



Remark 7.1 *In many ways length contractions by boosts are very similar to the following situation: when the plank of length l, which lies along the x^1-axis, is rotated along the x^3-axis, say, then its projection onto the x^1-axis will have a contracted length.*

Time Dilation Consider an observer O who is at rest in the x^{μ} coordinate system and located at $\mathbf{x} = 0$, so his world line is $t \mapsto (ct, 0, 0, 0)$ and who carries a clock that measure the (proper) time $t = c^{-1}x^0$. We will use the fact that O can determine the coordinates x^{μ} of any event using operational

[4]Before Einstein proposed Special Relativity, Lorentz and Fitzgerald had already proposed that objects undergo a length contraction in their direction of motion. However, their interpretation was rather different: They assumed the existence of absolute time and space and, in addition, that a particular classical inertial frame can be singled out by the existence of a substance called ether, which pervades all space and is static in this frame. Length contractions were argued to be a physical process, caused by the motion w.r.t. the ether, and the underlying mechanisms were sought in electromagnetism. A consistent treatment of this idea leads to a theory which makes exactly the same predictions as Special Relativity, but which has a number of superfluous concepts that have no operational meaning: ether, absolute time and absolute space.

procedures (involving sending light rays), so he may use x^0 as a global time coordinate.

Now we consider a similar observer O' in the x'^μ coordinate system, so that each observer sees the other one moving away with a speed v. Let $A = (ct_1, 0, 0, 0)$ and $B = (ct_2, 0, 0, 0)$ be two events in the coordinates x^μ, which differ by a time interval $(t_2 - t_1)$ according to O. We may express these events in the coordinate frame x'^μ as

$$A = (ct_1 \cosh(\theta), -ct_1 \sinh(\theta), 0, 0), \qquad B = (ct_2 \cosh(\theta), -ct_2 \sinh(\theta), 0, 0)$$

in the primed coordinate system. According to O', the two events are therefore separated by a time interval

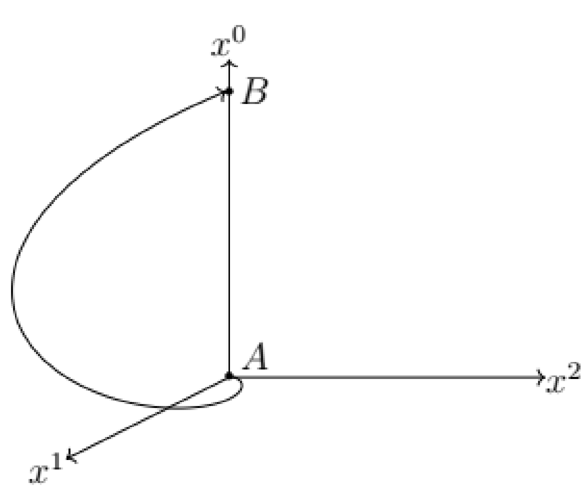
$$T' = \cosh(\theta)(t_2 - t_1) = \gamma T, \tag{11}$$

where $T = t_2 - t_1$. According to O', more time has elapsed between the two events, so O's clock is slow. In a similar way, O will find that the clock that O' uses is slow!

This effect is called *time dilation*. The fact that both observers find the other observer's clock to be slow violates the intuition that one clock must be faster than the other. However, this intuition is based on the false assumption that there exists an absolute time (and that both clocks run at a fixed rate compared with absolute time).

Exercise 7.2 *Consider an observer O', who starts running at event $A = (0, 0, 0, 0)$ along a curve of the form*

$$\chi(t) := (ct, 1 - \cos(\omega t), \sin(\omega t), 0).$$



For what values of ω is this curve time-like? For what values of ω is O' also present at the event $B = (ct_1, 0, 0, 0)$? What is the proper length of this curve between A and B? For what values of ω is this proper time less than half of the proper time t_1 along the straight line $t \mapsto (ct, 0, 0, 0)$? How fast must O' run to age only half as fast?

The Twin Paradox Consider the same Observer O in the coordinate system x^μ and an observer O' traveling through $A = (ct_1, 0, 0, 0)$ to $B = (ct_2, 0, 0, 0)$ via any C^1 time-like and future pointing curve $\xi : I \to M$. Then the following holds:

Theorem 7.3 *Let $\tau_\xi(A, B)$ be the proper time interval between A and B along ξ. Then $\tau_\xi(A, B) \leq (t_2 - t_1)$, with equality if and only if the (proper) velocity satisfies $\dot{\xi}^\mu \equiv (c, 0, 0, 0)$ between A and B.*

Proof: Because ξ is time-like and future pointing we have $\dot{\xi}^0(s) > 0$ for all $s \in I$, so we may change the parametrisation such that $t = c^{-1}x^0$ becomes the new parameter and $\xi^\mu(t) = (ct, \mathbf{x}(t))$. The proper time interval between A and B is then

$$\begin{aligned} \tau_\xi(A, B) &= c^{-1} \int_{t_1}^{t_2} \sqrt{c^2 - \|\dot{\mathbf{x}}(t)\|^2} \mathrm{d}\, t \\ &\leq c^{-1} \int_{t_1}^{t_2} c = (t_2 - t_1), \end{aligned}$$

which proves the estimate. Note that we have equality if and only if $\dot{\mathbf{x}} \equiv 0$, which means that $\dot{\xi}^\mu = (c, 0, 0, 0)$. In that case t is the proper time coordinate along ξ and we have $\xi^\mu(t) = (ct, 0, 0, 0)$, so ξ does indeed go from A to B. □

Theorem 7.3 is completely analogous to the familiar result in Euclidean space that two points A and B can be connected by curves of different lengths, and there is a unique shortest curve. However, in the present context, the interpretation of the theorem is that the observer O, who followed the linear curve, has aged more than the observer O', who followed the curve ξ. This exhibits a violation of the idea of absolute time. The effect is most striking when O' first travels away from O at a constant speed and then turns around to return. ξ is then also a linear curve most of the way, so in view of time dilation, which is symmetric, it is then paradoxical that both observers agree on the fact that O has aged more than O'.

The resolution of the paradox lies in the fact that we cannot (in general) choose an inertial frame in which O' is at rest all the way, so there is no symmetry between O and O'. We will see in Section 9.5 that the essential aspect of Theorem 7.3 is that the linear curve that O follows is a geodesic, whereas ξ is not, in general. (Using this idea one may reformulate Theorem 7.3 in a way that is independent of the choice of inertial coordinates: all that matters is that O follows the linear curve between A and B in any inertial coordinate frame.)

In the case where ξ is linear most of the time, at least two inertial coordinate frames are relevant. In this setting it is tempting to use the time coordinates of those inertial frames globally, to find out "when" the aging of O with respect to O' takes place. Some interpretations ascribe this aging to the acceleration of O' at the turning point. However, the main issue is that this question is ill-posed, because the term "when" makes no global sense in Special Relativity, especially not when several inertial frames are involved. Indeed, it is a non-trivial issue for O' to compute inertial (time) coordinates for events taking place elsewhere (i.e. not on his world line) and the coordinates that he computes (using light signals) depend on his state of motion. The change of inertial frame by O' thus forces a change in the notion of simultaneity, which causes a jump in the age of O, as computed by O'.

Part II
General Relativity

Introduction

Special Relativity was already a wonderful scientific revelation, which unified Newtonian mechanics and electromagnetism by weakening the assumed intrinsic structure of space and time. However, Newton's theory of gravitation does not fit into this new framework, because it does not behave well under Poincaré transformations. Indeed, Newtonian gravity involves an instantaneous action at a distance, but when the notion of simultaneity at non-zero distances is no longer available, such an action makes no sense anymore. Newton's theory had been criticised before for its action at a distance, notably by Descartes. Descartes had proposed an alternative theory of gravity, based on vortices, which, however, contradicted empirical evidence.

It was not until Einstein formulated his General Theory of Relativity that Newton's theory was replaced by a field theory of gravity. Moreover, with General Relativity Einstein achieved a variety of other deep and subtle goals. It drops e.g. the assumption that the set M of all events should admit a bijection onto $\mathbb{R}^4$. It also explains why the (heavy) mass appearing in Newton's law of gravitation is the same as the (inertial) mass that appears in his first law of mechanics (the *weak equivalence* principle). This equivalence means that the trajectory of a freely falling body is completely determined by its initial position and velocity and it is independent of the object's mass or shape. In General Relativity this is explained by the fact that these preferred free-fall trajectories are a part of the structure of the spacetime M.

Note that all bodies are influenced by gravity, so it is now inconceivable that one may produce the idealised clocks and measuring rods which make up the inertial coordinate systems of Special Relativity. Instead, General Relativity will treat all coordinate systems on an equal footing. Only on very small scales, where variations in the gravitational field can be neglected, can we consider inertial coordinates, which are associated to freely falling clocks and rods. As before we require that all such local inertial frames are equivalent, i.e. the outcome of any local experiment in a freely falling laboratory is independent of the initial position and velocity of the laboratory. (Together with the weak equivalence principle, this forms the *strong equivalence principle.*)

The important conceptual step that makes all this possible is to turn the background structure of spacetime, which determines the spacetime intervals, into a dynamical structure, like a physical field, which must satisfy Einstein's equation of motion. This conceptual change also resolves Newton's "rotating bucket" paradox: the bucket is not rotating with respect to empty space, but with respect to the gravitational field, which is a physical quantity in its own right.

Mathematical Preliminaries

In this section we will present the mathematical tools needed to formulate General Relativity. These mathematical tools had been developed before General Relativity, in particular by Riemann (building on earlier work by

Gauss). Most of these tools deal with the idea that the set of all events M may no longer be identifiable with $\mathbb{R}^4$, which requires the mathematical theory of manifolds (i.e. differential geometry). The idea that there should be a field which tells us how to measure time intervals and distances locally, requires the theory of (pseudo)-Riemannian manifolds. First, however, we will describe some notations for tensors, that will make the subsequent developments run more smoothly.

Calculus of Tensors

Let V be a finite dimensional real vector space of dimension $n \in \mathbb{N}$ and let $\{e_1, \ldots, e_n\}$ be a basis for V. Then each vector $v \in V$ can be written in a unique way as $v = \sum_{\mu=1}^{n} v^\mu e_\mu$ and we call the real numbers v^μ the components of v in the basis $\{e_\mu\}$. To simplify our notations we introduce the following convention:

Convention 9.1 (Einstein's Summation Convention) *Whenever an expression contains an index that appears once as a superscript and once as a subscript, then a summation over the range of this index (e.g. $1, \ldots, n$) is implied, unless explicitly stated otherwise.*

This means that we may write $v = v^\mu e_\mu$, dropping the summation symbol.

Now let V^* be the *dual vector space* of V, i.e. the vector space of all linear maps $\omega : V \to \mathbb{R}$, where the vector space structure is given by pointwise addition and multiplication:

$$(\lambda_1\omega_1 + \lambda_2\omega_2)(v) := \lambda_1\omega_1(v) + \lambda_2\omega_2(v).$$

There is a natural basis of V^* which is dual to e_μ, namely $e^{*1}, \ldots, e^{*n}$ such that

$$e^{*\mu}(e_\nu) = \delta^\mu{}_\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases} . \tag{12}$$

Because any $\omega \in V^*$ is uniquely determined by its values on the basis vectors e_μ, we see that $\{e^{*\mu}\}$ is indeed a basis and that V^* also has the dimension n. We may write $\omega = \omega_\mu e^{*\mu}$ with unique components ω_μ in the basis $\{e^{*\mu}\}$. Note that

$$\omega(v) = \omega_\mu v^\nu e^{*\mu}(e_\nu) = \omega_\mu v^\mu,$$

because the bases of V and V^* are dual to each other.

Exercise 9.1 *Let V^{**} be the double dual vector space, i.e. the dual vector space of Vv. Show that there is a linear map $\iota : V \to V^{**}$ defined by $\iota(v)(\omega) = \omega(v)$ for all $\omega \in V^*$. Moreover, show that ι is an isomorphism of vector spaces, which is independent of any choice of basis for V. (For this reason it suffices to consider V and V^* and no further duals are needed.)*

By a *tensor T of type (k, l)* we will mean a multi-linear map[5]

$$T : \underbrace{V^* \times \ldots \times V^*}_{k \text{ times}} \times \underbrace{V \times \ldots \times V}_{l \text{ times}} \to \mathbb{R},$$

i.e. a map $T(\omega_1, \ldots, \omega_k, v_1 \ldots, v_l)$ which is linear in each ω_i and v_j when all other arguments are fixed. The space of all such tensors is denoted by

$$\underbrace{V \otimes \ldots \otimes V}_{k \text{ times}} \otimes \underbrace{V^* \otimes \ldots \otimes V^*}_{l \text{ times}}$$

and it carries a natural vector space structure (by pointwise addition and scalar multiplication).

A tensor of type $(0, 1)$ is an element of V^*, whereas a tensor of type $(1, 0)$ is an element of $V^{**} \simeq V$. In this way tensors generalise vectors and dual vectors. Note that the tensor product $V \otimes V$ has a natural basis determined by the e_μ, namely $e_\mu \otimes e_\nu$ with $\mu, \nu = 1, \ldots, n$. (The number of such vectors is n^2, which is indeed the dimension of $V \otimes V$.) Any element of $w \in V \otimes V$ can therefore be written in terms of components as $w = w^{\mu\nu} e_\mu \otimes e_\nu$. Similarly, for $L \in V \otimes V^*$ we may write $L = L^\mu{}_\nu e_\mu \otimes e^{*\nu}$ and an analogous expansion into components works for arbitrary tensors.

Of course the components of tensors depend heavily on the choice of basis e_μ. Nevertheless, it is possible to write down expressions which are independent of this choice of basis. To see how this goes we will now consider how a change of basis acts on the various components.

Let $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$ be another basis of V. Then there is a unique, invertible linear map $L : V \to V$ such that $\tilde{e}_\mu = L \cdot e_\mu$ for all $\mu = 1, \ldots, n$. We can view L as an element of $V \otimes V^*$, using

$$L : V^* \times V \to \mathbb{R} : (\omega, v) \mapsto \omega(L(v)),$$

[5] The symbol $\times$ denotes the Cartesian product of sets, i.e. the set whose elements are of the form $(\omega_1, \ldots, \omega_k, v_1 \ldots, v_l)$.

which is bilinear. The components of L in the basis $e_\mu \otimes e^{*\nu}$ are

$$L^\mu{}_\nu = e^{*\mu}(\tilde{e}_\nu) = e^{*\mu}(L(e_\nu)).$$

Note in particular that

$$\tilde{e}_\nu = L^\mu{}_\nu e_\mu, \qquad e^{*\mu} = L^\mu{}_\nu \tilde{e}^{*\nu},$$

where the second equality follows from the fact that both sides of the equation take the same values on the basis $\tilde{e}_\rho$.

For any element $v \in V$ we may now compare the components v^μ and $\tilde{v}^\mu$ in the two bases:

$$v^\mu e_\mu = v = \tilde{v}^\nu \tilde{e}_\nu = L^\mu{}_\nu \tilde{v}^\nu e_\mu \qquad \Leftrightarrow \qquad v^\mu = L^\mu{}_\nu \tilde{v}^\nu.$$

Similarly, for any $\omega \in V^*$ we have

$$\tilde{\omega}_\mu \tilde{e}^{*\mu} = \omega = \omega_\nu e^{*\nu} = \omega_\mu L^\mu{}_\nu \tilde{e}^{*\nu} \qquad \Leftrightarrow \qquad \tilde{\omega}_\nu = \omega_\mu L^\mu{}_\nu.$$

After interchanging the roles of e_μ and $\tilde{e}_\mu$ and using L^{-1} instead of L this leads to:

$$\omega_\nu = (L^{-1})^\mu{}_\nu \tilde{\omega}_\mu.$$

A similar relation can be obtained for arbitrary tensors:

$$T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} = L^{\mu_1}{}_{\rho_1} \cdots L^{\mu_k}{}_{\rho_k} (L^{-1})^{\sigma_1}{}_{\nu_1} \cdots (L^{-1})^{\sigma_l}{}_{\nu_l} \tilde{T}^{\rho_1\cdots\rho_k}{}_{\sigma_1\cdots\sigma_l}. \tag{13}$$

This is called the *tensor transformation law.*

Given two tensors T and S of the same type (k, l), the equality $T = S$ of multi-linear maps is equivalent to the equality of the components

$$T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} = S^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} \tag{14}$$

using a basis e_μ of V. This is because the tensors are uniquely determined by their values on the basis elements, and these values are exactly the components. Note that this equality holds in any basis e_μ. Sometimes, however, it is convenient to write equality (14) for the components of a tensor T in some given basis e_μ and a set of numbers $S^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l}$ which may not be related to a tensor, but which emerge in some other way from a particular physical problem. Such an equality does depend on the choice of basis, because the numbers in S do not transform as a tensor. To distinguish true tensor equations from other equalities using indices we introduce an additional convention:

Convention 9.2 (Abstract Index Notation) *The components of a tensor in a certain basis will be denoted by Greek indices. When the choice of basis is arbitrary, we will use a symbol with latin indices, e.g.* $T^{a_1\cdots a_k}{}_{b_1\cdots b_l}$*, to denote the type of the tensor. (These are not numbers or components of the tensor in some basis.) In particular, an equality between tensors, which holds in any basis, will be written using latin indices: it is an equality between multi-linear maps, rather than between real numbers.*

For general developments it is often nicer to use abstract indices, but in concrete examples it is often easier to choose a particular set of coordinates, which is adapted to the symmetries of the problem. This is why we will use abstract indices for now, but we will mostly revert to particular coordinates when considering applications.

There are two widely used operations on tensors, which we will now describe. Firstly, given a tensor S of type (k, l) and a tensor T of type (k', l') we may define the *outer product* as the tensor of type $(k + k', l + l')$ defined by

$$(S \otimes T)^{a_1\cdots a_{k+k'}}{}_{b_1\cdots b_{l+l'}} := S^{a_1\cdots a_k}{}_{b_1\cdots b_l} T^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}}.$$

(Note that this definition is independent of the choice of basis.) Furthermore, given any tensor T of type (k, l) with $k \geq 1$ and $l \geq 1$ we may define the contraction CT of the ith upper index and the jth lower index as the tensor

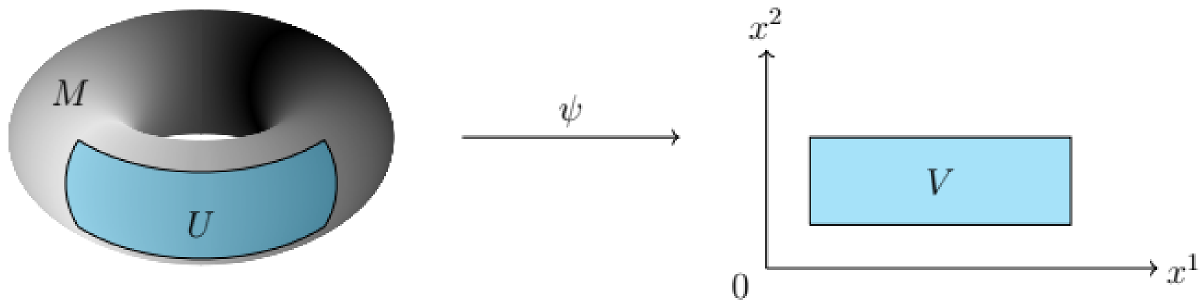
$$(CT)^{a_1\cdots \widehat{a_i}\cdots a_k}{}_{b_1\cdots \widehat{b_j}\cdots b_l} := T^{a_1\cdots c\cdots a_k}{}_{b_1\cdots c\cdots b_l},$$

where we recall that a sum is implied. As an example we consider a vector v^a and a dual vector ω_b, for which the outer product is $v^a\omega_b$ and the contraction of the outer product is $v^a\omega_a = \omega(v)$.

Exercise 9.2 *Verify that the equality* $v^a\omega_a = \omega(v)$ *holds in any basis* e_μ*.*

Manifolds

In order to describe the set M of all events with as few unphysical assumptions as possible, we introduce the mathematical concept of a manifold. This is ultimately based on ideas from cartography.



Definition 9.3 *A* C^∞ manifold of dimension $n \in \mathbb{N}$ *is a non-empty set* M *together with a collection of bijective maps* $\psi_i : U_i \to V_i$*, where* $U_i \subset M$*,* $V_i \subset \mathbb{R}^n$ *is an open set and* $i \in \mathcal{I}$ *some index set, such that:*

1. *the sets* U_i *cover* M*,* $\bigcup_{i\in\mathcal{I}} U_i = M$*, i.e. each* $x \in M$ *lies in some* U_i*,*
2. *if* $U_i \cap U_j \neq \emptyset$ *for some* $i, j \in \mathcal{I}$*, then* $\psi_i(U_i \cap U_j)$ *and* $\psi_j(U_i \cap U_j)$ *are open subsets of* $\mathbb{R}^n$ *and* $\psi_j \circ \psi_i^{-1}$ *is a* C^∞ *map with a* C^∞ *inverse between these sets.*

The maps $\psi_i : U_i \to V_i$ *are called* charts *(or* coordinate systems*) and the collection* $\{\psi_i\}_{i\in\mathcal{I}}$ *is called an* atlas *for* M*. The dimension* n *of* M *is written as* $\dim(M)$*.*

Let us consider some examples, to illustrate this notion:

Example 9.4

- $M = \mathbb{R}^n$ *is an* n*-dimensional manifold if we choose the atlas to consist of the single chart* $\psi : M \to \mathbb{R}^n$*, where* ψ *is the identity map.*
- *The unit circle* $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 |\ x^2 + y^2 = 1\}$ *is a manifold of dimension 1. We may cover the set by four charts*

$$\begin{aligned}
\psi_{y+} &: \{(x,y) \in \mathbb{S}^1 |\ y > 0\} \to (-1,1) \ :\ (x,y) \mapsto x \\
\psi_{y-} &: \{(x,y) \in \mathbb{S}^1 |\ y < 0\} \to (-1,1) \ :\ (x,y) \mapsto x \\
\psi_{x+} &: \{(x,y) \in \mathbb{S}^1 |\ x > 0\} \to (-1,1) \ :\ (x,y) \mapsto y \\
\psi_{x-} &: \{(x,y) \in \mathbb{S}^1 |\ x < 0\} \to (-1,1) \ :\ (x,y) \mapsto y.
\end{aligned}$$

The compatibility of these charts is straightforward to check. E.g., for ψ_{x+} *and* ψ_{y+} *the intersection of the domains of definition is* $\{(x, y) \in \mathbb{S}^1 |\ x > 0, y > 0\}$ *and we have* $\psi_{x+} \circ \psi_{y+}^{-1}(x) = \sqrt{1 - x^2}$*, which is* C^∞ *and has the* C^∞ *inverse* $\psi_{y+} \circ \psi_{x+}^{-1}(y) = \sqrt{1 - y^2}$*.*

- *The n-dimensional unit sphere* $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} | \sum_{k=1}^{n+1}(x^k)^2 = 1\}$ *is a manifold of dimension n. It can be covered by charts in a similar way as* $\mathbb{S}^1$*, but now* $2(n+1)$ *charts are needed.*

Notice that a general manifold cannot be covered by a single chart.

Exercise 9.5 *Show that* $\mathbb{S}^2$ *is a manifold, give the coordinate charts in analogy to* $\mathbb{S}^1$ *and show that the changes of coordinate charts are smooth. (You may use symmetries to simplify the problem.)*

In addition to specific examples, there are general constructions which allow us to construct new manifolds from old ones. We mention the most important ones.

First, we call a set $O \subset M$ an *open subset* if and only if $\psi_i(O \cap U_i)$ is an open subset of $\mathbb{R}^n$ for all charts in the atlas. Any open subset $O \subset M$ is a manifold in its own right, where the charts are given by $\psi_i|_O : (O \cap U_i) \to \psi_i(O \cap U_i)$ with $i \in \mathcal{I}$.

Given two manifolds M and M', the product set $M \times M'$ is a manifold, where the charts are given by all maps $\psi_i \times \psi'_j : U_i \times U'_j \to V_i \times V_j$, where $\psi_i : U_i \to V_i$ is any chart on M and $\psi'_i : U'_i \to V'_i$ on M'. Note that $\dim(M \times M') = \dim(M) + \dim(M')$.

Example 9.6

The n-dimensional torus is $\mathbb{T}^n := \mathbb{S}^1 \times \ldots \times \mathbb{S}^1$ *(n factors). Note that* $\mathbb{T}^1$ *is just the circle and* $\mathbb{T}^2$ *has the shape of a doughnut.*

Exercise 9.7 *Sketch* $\mathbb{T}^2$ *as a subset of* $\mathbb{R}^3$ *and sketch a typical product chart obtained from the charts of* $\mathbb{S}^1$ *given in Example 9.4.*

A map $f : M \to M'$ between two manifolds is called k-times continuously differentiable, or C^k, when $\psi'_i \circ f \circ \psi_j^{-1}$ is k-times continuously differentiable for all charts ψ_j in the atlas for M and all charts ψ'_i in the atlas for M'. As an example we note that, in any chart $\psi_i : U_i \to V_i$, the coordinates $x^k : \mathbb{R}^n \to \mathbb{R}$ can be used to define smooth maps $x^k \circ \psi_i$ on U_i. Other than such local coordinates, we will mostly be concerned with curves, $\xi : I \to M$, where $I \subset \mathbb{R}$ is an open interval (which is also a manifold of dimension 1).

All of our manifolds will have some additional properties, which we mention here without a very detailed discussion, because they will not be needed explicitly:

1. The atlas of a manifold M is called *maximal* when it has the following property: Let $\psi : U \to V$ be a bijective map between some $U \subset M$ and an open set $V \subset \mathbb{R}^n$ and suppose that for all $i \in \mathcal{I}$, ψ is compatible with ψ_i in the sense of the second condition in the definition of a manifold. Then ψ is already contained in the atlas $\{\psi_i\}_{i\in\mathcal{I}}$, i.e. $\psi = \psi_j$ for some $j \in \mathcal{I}$. Any atlas of a manifold M can always be extended in a unique way to a *maximal* one. We will always assume that our manifolds are equipped with a maximal atlas.

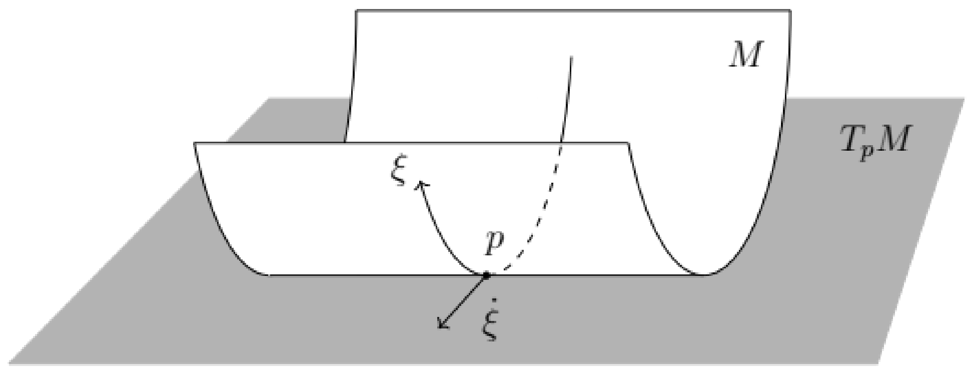
2. All our manifolds are path-connected. This means that for any two points $p_1, p_2 \in M$ there is a continuous curve $\xi : [0,1] \to M$ such that $\xi(0) = p_1$ and $\xi(1) = p_2$.

3. All our manifolds are *Hausdorff* topological spaces. This means that for any two points $p_1, p_2 \in M$ there are open sets $U_1, U_2 \subset M$ with $p_i \in U_i$, $i = 1, 2$, but $U_1 \cap U_2 = \emptyset$. (This is an additional assumption on M, which we will always make.)

4. All our manifolds are *second countable*. This means that there is a countable collection $\{O_n\}_{n\in\mathbb{N}}$ of open sets $O_n \subset M$ such that every open set $O \subset M$ contains some O_n.

All of the examples of manifolds M we have seen so far can be embedded into $\mathbb{R}^m$ (for some $m \geq \dim(M)$). However, the importance of manifolds is that we can investigate them in a framework which is independent of this embedding. For example, $\mathbb{S}^1$ is defined as a subset of $\mathbb{R}^2$ and by taking products $\mathbb{T}^2$ can be viewed as a subset of $\mathbb{R}^4$. However, $\mathbb{T}^2$ can also be viewed as a subset of $\mathbb{R}^3$. Moreover, for the set M that models all events in the universe, it is not clear a priori if it can be embedded into any $\mathbb{R}^m$ at all! Also the shape of M is not a priori clear − why should we assume that it be covered by a single chart like Minkowski space? For these reasons, any information about a manifold should really be formulated in terms of the intrinsic structure of the manifold itself, independent of any embedding. We have already done this for the set M and its topological and differential structure (i.e. we know what C^∞ maps on M are). We now turn to the notion of tangent vectors.

A C^1 curve $\xi : I \to \mathbb{R}^n$ with $0 \in I$ has a tangent vector at $p = \xi(0)$, which is given by $\dot{\xi}^\mu(0)$ (where the derivative is taken component-wise in some basis). The tangent vectors at $p \in \mathbb{R}^n$ form a vector space of dimension n, just by adding their components.



However, to formulate the idea of tangent vectors on a manifold it is helpful to take a different perspective. Let x^μ be Cartesian coordinates on $\mathbb{R}^n$ (with a corresponding basis e_μ). Any tangent vector v at p, with components v^μ, defines a directional derivative operator

$$v^\mu \partial_{x^\mu} : C^\infty(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R} : f \mapsto v^\mu \partial_{x^\mu} f(p).$$

This operator is linear and satisfies the Leibniz rule. Conversely, one may show that any such directional derivative operator corresponds to a unique tangent vector.

Exercise 9.8 *Show that the formula $v^\mu \partial_{x^\mu}$ is independent of the choice of basis e_μ.*

For a manifold M we now make the following

Definition 9.9 *A* tangent vector *v at $p \in M$ is an operator $v : C^\infty(M, \mathbb{R}) \to \mathbb{R}$ such that, for all $f_1, f_2 \in C^\infty(M, \mathbb{R})$ and $c_1, c_2 \in \mathbb{R}$,*

1. *$v(c_1 f_1 + c_2 f_2) = c_1 v(f_1) + c_2 v(f_2)$, and*

2. *$v(f_1 f_2) = f_1(p) v(f_2) + v(f_1) f_2(p)$ (Leibniz rule).*

The set of all tangent vectors at $p \in M$ is denoted by T_pM.

We note that T_pM forms a vector space, where

$$(c_1 v_1 + c_2 v_2)(f) := c_1 v_1(f) + c_2 v_2(f).$$

One way to obtain some examples $X_\mu \in T_pM$ is by fixing a chart $\psi : U \to V$ with $p \in U$ and setting

$$X_\mu(f) := \partial_{x^\mu}(f \circ \psi^{-1})(\psi(p)),$$

where the x^μ are Cartesian coordinates on $\mathbb{R}^n$. It is not hard to verify that the X_μ are indeed in T_pM. Moreover, one may show that they form a basis of T_pM, so that T_pM is a vector space of dimension $n = \dim(M)$. (To see this one first shows that v only depends on the values of f in any (small) neighbourhood of p. One may then use a chart near p to turn the question into a problem in ordinary calculus.)

In fact, we may use the chart $\psi : U \to V$ to define tangent vectors like X_μ at any point $p \in U$. In this way we obtain a basis for T_pM for any $p \in U$, which is called a *coordinate basis*. If we use a different chart $\psi' : U' \to V'$ near p, then the vectors X'_μ are related to the X_μ by the chain rule:

$$X'_\mu = \frac{\partial x^\nu(x')}{\partial x'^\mu} X_\nu,$$

where $x^\nu(x')$ is short-hand for $x^\nu(\psi \circ \psi'^{-1}(x'))$, which describes how the map $\psi \circ \psi'$ changes the coordinates x'^μ into coordinates x^ν. The quotient can also be written in the matrix notation $D^\nu{}_\mu(\psi \circ \psi'^{-1})$ (with bases corresponding to the Cartesian coordinates x^μ and x'^ν).

Similarly, any vector $v \in T_pM$ can be written as $v = v^\mu X_\mu = v'^\mu X'_\mu$ with

$$v'^\mu = \frac{\partial x'^\mu(x)}{\partial x^\nu} v^\nu. \tag{15}$$

This is just the tensor transformation law applied to a vector, except that the matrix involved may now depend on the point $p \in M$.

Another way to obtain tangent vectors in T_pM is to consider a C^1 curve $\xi : I \to M$ with $0 \in I$ and such that $\xi(0) = p$ and setting

$$D_0\xi(f) := \dot{\xi}_0(f) := \partial_s(f \circ \xi)(0).$$

In a chart $\psi : U \to V$ with $p \in U$, we can express ξ in terms of its components $\xi^\mu := x^\mu \circ \psi \circ \xi$ in Cartesian coordinates x^μ:

$$\begin{aligned} D_0\xi(f) &= \partial_s(f \circ \psi^{-1} \circ \psi \circ \xi)(0) \\ &= \partial_\mu(f \circ \psi^{-1})(\psi(p)) \cdot \partial_s\xi^\mu(0) \\ &= \dot{\xi}^\mu(0) X_\mu(f), \end{aligned}$$

i.e.

$$D_0\xi = \dot{\xi}^\mu(0)X_\mu.$$

Many curves can define the same tangent vector in T_pM, but every tangent vector in T_pM is of this form for some curve ξ.

More generally, let $\chi : M \to M'$ be a smooth map such that $\chi(p) = p'$. Any tangent vector $v \in T_pM$ gives rise to a tangent vector $D_p\chi(v) \in T_{p'}M'$ defined by

$$(D_p\chi(v))(f) := v(f \circ \chi). \tag{16}$$

Note that the map $D_p\chi : T_pM \to T_{p'}M'$ is linear in v. We recover $D_0\xi$ as a special case, where our notation suppresses the vector $v = e_1 \in T_0\mathbb{R}$, which is the unit vector which points in the positive direction.

As a general warning we emphasise that for a general manifold there is no natural way to identify tangent vectors at some point $p \in M$ with tangent vectors at some other point $q \in M$. We know that this is possible in $\mathbb{R}^n$, simply by applying a translation. (More precisely, there is a unique translation $\tau : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto x - (p - q)$ which maps p to q and $D_p\tau$ can be used to identify $T_p\mathbb{R}^n$ with $T_q\mathbb{R}^n$.) However, such translations are not defined for general manifolds and there is no natural analog.