INTRODUCTION TO SPACETIME

A FIRST COURSE ON RELATIVITY

BERTEL LAURENT

World Scientific

INTRODUCTION TO SPACETIME

A FIRST COURSE ON RELATIVITY

INTRODUCTION TO SPACETIME

A FIRST COURSE ON RELATIVITY

BERTEL LAURENT

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 9128

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 73 Lynton Mead, Totteridge, London N20 8DH

Library of Congress Cataloging-in-Publication Data

Laurent, Bertel.
Introduction to spacetime : a first course on relativity / Bertel Laurent.
p. cm.
Includes index.
ISBN 9810219296
1. Relativity (Physics) 2. Space and time. 3. Electrodynamics.
I. Title.
QC173.55.L365 1994
530.1\1--dc20 94-34064
CIP

Printed in Singapore.

Contents

I PRINCIPLES. BASIC APPLICATIONS 7

1 Clocks and Acceleration 9

1.1 Measuring Time 9
1.2 Measuring Acceleration 11
1.3 The Principle of the Maximal Proper Time 13
1.4 Events. Space-Time 15
1.5 Parallel World Lines 17

2 Vector Algebra 21

2.1 Basic Properties 21
2.2 Scalar Product 25

3 Vector Characteristics 29

3.1 Timelike. Spacelike. Null-like 29
3.2 Comparison with Euclidean Space 33

4 Simultaneity and Space Distance 37

4.1 Simultaneity 37
4.2 Space Distance 39
4.3 The Orthogonal Space 41

5 Linear Independence 45

6 Relative Velocity and Four-Velocity 49

6.1 The Standard Velocity Split 50
6.2 Light Signals 50
6.3 Split of Null-Like Vectors 53

6.4 The Future and the Past 54

7 Two-Dimensional Spacetime **57**
7.1 Lorentz Transformation 57
7.2 Addition of Velocities 60
7.3 Lorentz Contraction 61

8 Plane Waves **63**
8.1 The Wave Four-Vector 63
8.2 Modulations 64
8.3 Doppler Shift and Aberration 67

9 Particle Reactions **71**
9.1 Four-Momentum 71
9.2 Particle Kinematics 75

10 Curved World Lines **81**
10.1 Four-Acceleration 81
10.2 Examples 84
10.2.1 Constant Acceleration 84
10.2.2 Fitting a Car into a Garage 86
10.2.3 Rotating Wheel 89

II TENSORS **91**

11 Definition and Examples **93**
11.1 Definition 93
11.2 Examples 95

12 Algebraic Properties **101**
12.1 First Rank Tensors 101
12.2 Generalized Tensors 104
12.2.1 Symmetries 104
12.2.2 Abstract Indices 105
12.3 Tensor Algebra 107
12.4 Expansion 109
12.5 Contraction 111

12.6 Tensors and the Scalar Product 114
12.6.1 Relation between Vectors and Dual Vectors . . . 114
12.6.2 The Metric Tensor 116

13 Tensor Fields 119
13.1 Scalar Fields. Gradients 119
13.2 Tensor Fields 124

14 Spacetime Volumes 129
14.1 Volume Without Metric 129
14.1.1 Basic Concepts. Parallelepipeds 129
14.1.2 Calculation of Volumes 132
14.1.3 Sums of Volumes 134
14.1.4 Volumes and Coordinates 135
14.2 Volume and Metric. Orientation 138
14.3 Hypersurface Volumes 141

15 Currents 143
15.1 Particle Flow. Four-Current Density 143
15.2 Gauss's Theorem 147
15.3 Different Kinds of Currents 150

III ELECTRODYNAMICS 153

16 Sourcefree Electromagnetism 157
16.1 The Wave Equation 157
16.2 The Field Tensor 159
16.3 The Field Equations 160
16.3.1 General Considerations 160
16.3.2 The Four-Potential Equations 162
16.3.3 The Field Tensor Equations 164
16.3.4 Gauge 167
16.4 Maxwell's Equations 168
16.4.1 The Electric and Magnetic Fields 168
16.4.2 The Field Equations 171

17 Electro-Magnetism with Sources **175**
17.1 The Field Equations 175
17.1.1 Sources 175
17.1.2 Maxwell's Equations 177
17.1.3 The Potential Equations 177
17.2 Energy-Momentum 178
17.2.1 Dust 178
17.2.2 Electromagnetism 182

18 Solution of the Wave Equation **185**
18.1 The Green's Function 185
18.2 The Lienard-Wiechert Potential 190

Foreword

This book contains a course on relativity for those who already believe in it. Most relativity courses spend a lot of time and energy on detailed arguments for the insufficiency of the Newtonian theory. Nowadays the theory of relativity is as well established as the Newtonian theory and the time may be ripe to present it without feeling forced to argue with sceptics at every corner. This does *not* imply that the experimental verification of the theory is unimportant. What it means is that we shall refrain from glancing furtively at the doubters when we present the theory.

We can therefore already from the start take advantage of the geometrical nature of the theory and there is no need for much of the standard procedures in the presentation of the theory. This includes the construction of inertial systems, the detailed discussion of the Michelson–Morley experiment, and the derivation of the Lorentz transformation from the constancy of the velocity of light.

We assume that the reader is familiar with vector calculus in ordinary three-dimensional Euclidean space. The calculations in this book are in fact very similar to ordinary vector calculations and this is actually the way that most relativistic calculations are made in scientific papers.

This book is based on a course which has been given at the Stockholm University through five years and I would like to thank all those students who have expressed their views on the course and thereby helped to shape it.

Bertel Laurent

Bertel Laurent in memoriam

Professor Bertel Laurent was a distinguished specialist in the theory of relativity and an excellent and inspiring teacher. He finished this book shortly before his untimely death on August 4, 1993. No changes have been made in his manuscript, except for corrections of obvious minor mistakes.

Stockholm, February 1994

Stig Flodmark

Part I

PRINCIPLES AND BASIC APPLICATIONS

Chapter 1

Clocks and Acceleration

1.1 Measuring Time

The most fundamental difference between Newtonian theory and the theory of relativity lies in the concept of *time*. It is therefore useful to start with a general discussion on the notion of time.

As all physical concepts time must be defined *operationally*, i.e., through a prescription of how it is to be measured. Time is measured with *clocks*, but there are so many different kinds of clocks and they are constructed after so many different blue prints. How come that they all measure the same quantity?

To discuss this point we must rely on the constancy of the laws of nature (which we shall not call in question). The typical clock makes use of some periodic process and 'measuring of time' consists in the counting of number of periods. Let us compare two clocks, each of them built according to given but not necessarily the same prescriptions. When we start to compare them each clock is in a definite state, its original state. Let us further say that when we compare the two clocks next time one of them has experienced n whole periods and the other m whole periods (where n and m are integers). Both clocks are then back in their original states and with reference to the constancy of the laws of nature we conclude that the following n periods of the first clock must again correspond to m periods of the second clock, etc. This implies that there exists a definite scale factor m/n which transforms the readings of the first clock to the readings of the second clock and this factor depends exclusively on how the clocks are constructed. This in turn implies that we can arbitrarily choose one clock as standard and calculate the readings of all other clocks from the readings of the standard clock. Choosing standard clock is tantamount to choosing the unit of time.

Obviously all clocks suffer from errors of measurement. One can easily establish this by comparing clocks of the same construction. The errors vary considerably between different types of clocks and for some of the experimental checks of relativity one has to use the very finest clocks which have been built.

So far we have assumed that a clock consists of a periodic process so that the clock returns over and over again to the same state and the typical clock is certainly of this kind. Using periodic clocks we may, however, 'calibrate' certain aperiodic processes so that they also can serve as clocks. A well-known example of this is the use of the decay of the radioactive carbon isotope C^{14} in archaeology. Also the processes of life can be used as clocks although such clocks are quite unreliable in comparison to the clocks we normally use in physics.

The definition 'time is that which is measured by clocks' is quite typical for definitions in physics. Through this definition we avoid a number of questions concerning the nature of time, which are important for philosophy and psychology. For us it suffices in principle as a definition of time to point at Big Ben and say: 'That is a clock; it measures time'.

So far everything we have said applies to Newtonian physics as well as to relativity, but now comes the first important difference. In Newtonian physics *the size* of a clock is of no consequence. In contrast to this we shall now require that *all clocks be small.* To be more precise they have to be small in comparison to the typical dimensions of the physical system we are considering. If we for instance study the motion of the earth and the other planets a sun dial is entirely unsuitable as clock, because it is based on the motion of the earth and the sun. Hence this clock in a sense covers a considerable part of the solar system. An atomic clock on the earth could, however, serve well and so could an atomic clock on board an artificial satellite circling the sun. I have met people who are quite upset by 'the prohibition against sun dials'. In fact there are no reasons to stop using sun dials except that the very powerful laws which we shall soon formulate don't apply directly to them.

1.2 Measuring Acceleration

The theory of relativity plays a very important part in physics today and its importance will become evident to everyone when fast space travel is commonplace. Relativity will then be taught with reference to experiences from space travel which are well-known to the students. For the sake of concreteness we shall pretend that we are already there and refer to such experiences as if they were facts. The simple rules which we shall formulate are of course not derived from space travel. They are nevertheless very well founded in experiences from many different experiments. With these simple rules as a basis we shall develop a theory which can be applied to much more than space travel. It is therefore possible to test it today and the positive results of an extensive testing has given it a very high degree of credibility.

Let us consider a space ship which is just floating through outer space (to use a science fiction term). It doesn't run its engines and the nearest star is far away so that there is no external influence like a solar wind or an electromagnetic field. Without any reference to the world outside the ship, the crew can find out that the ship is freely floating. The method is to measure the inertial forces in the ship. This can be done with equipment for inertial navigation. The crudest method is for the crew just to feel if they are pressed against some of the walls of the ship. It is most practical to turn this around to a definition: A space ship is freely floating if no inertial forces are felt on board. If on the other hand there are inertial forces on board we say that the ship is *accelerated.* Please, forget that you have ever been taught that acceleration is the time derivative of velocity. Such a definition of acceleration is unsuitable here because there is, as our future experience will tell us, no measurement on board which can tell you the velocity of the ship. The velocity of a ship must always be referred to something else like interstellar dust. This is used as a basic principle, the *principle of relativity* which has given its name to the entire theory of relativity.

The principle of relativity implies among other things that different kinds of clocks run with the same relative rates in all unaccelerated ships. Acceleration does, however, affect the running of clocks — no one can doubt that, who has dropped a sensitive clock on a hard floor — and it affects different kinds of clocks differently. How then can we measure the time unambiguously in an accelerating space ship?

One may approach this question from a few different angles. We choose to point at the observation that one can arrange an acceleration of any collection of clocks so that they all agree that the acceleration is of *short duration*, and this is true also if the acceleration is so great that it results in a considerable change of the motion of the clocks. A condition for this is of course that the clocks do not become seriously damaged. We shall idealize the situation and consider *sudden accelerations.* In a space ship which undergoes a sudden acceleration all its clocks show (nearly) the same time immediately after the acceleration as they did immediately before the acceleration, i.e., *the acceleration has not affected the time on board the ship.*

Let us as an example consider a special kind of space ship journey to which we will return repeatedly in the following: A space ship is constantly unaccelerated. For most purposes we may assume that it is the earth or the solar system. Another space ship is suddenly ejected from the first one, shot out by a catapult or some similar contraption. It has no motor on board and therefore flies unaccelerated far away. On its journey it collides with some heavy object and bounces in such a fortunate manner that it returns to the first space ship where it is landed with a crash. At three occasions the second ship is violently accelerated but we shall assume that its clocks (and its crew) survive undamaged. The total travel time which the crew in the second ship measures is the sum of the times for the unaccelerated journey out and the unaccelerated journey home. Negligible time elapses during the sudden accelerations.

Any space ship journey may be regarded as a limiting case of a series of sudden accelerations interspaced with unaccelerated journeys. The sum of the time intervals measured during free flight is called the ships *own time* or *eigentime* ('eigenzeit' means 'own time' in German) or *proper time* or sometimes *travel time.* It should be mentioned that there are practical limits for how close to one another the sudden accelerations should be executed. If a clock has been affected by an acceleration it may require some checking and adjusting to fulfill all specifications in the next period of free flight.

As we just mentioned the earth can for most purposes be regarded as a space ship with a common proper time. This is the reason why it has been believed for so long that there is a common time for all observers in the universe. As we shall soon discuss further this is not at all true. Actually clocks in different space ships cannot even be directly compared except when the ships are together. In general a comparison requires signalling between the ships and such a communication is not trivial.

1.3 The Principle of the Maximal Proper Time

In this section we shall formulate a rule which may be considered as one of the most important laws of nature all categories. It will be well-known to future students of physics through their experiences of space travel. The rule is the following:

> If two space ships part and meet and one of them is freely floating throughout its journey, then this ship will measure a longer travel time than the other one.

As it stands this rule has of course not been tested. 'Space ship' can, however, stand for almost anything carrying a clock. To test the rule Hafele and Keating put advanced clocks on board commercial aeroplanes which flew in opposite directions round the earth.[1] We cannot discuss this experiment in great detail at this stage but it is easy to understand the main feature of it through an idealization of the arrangement. Let us neglect gravitation and picture the earth as a merry-go-round all alone in empty space. The rim of the merry-go-round corresponds to the equator (see Figure 1.1). Let us further assume that two space ships are flying along the rim in opposite directions with such

[1] J.C. Hafele has written about the experiment in American Journal of Physics volume 40 (1972) page 81.

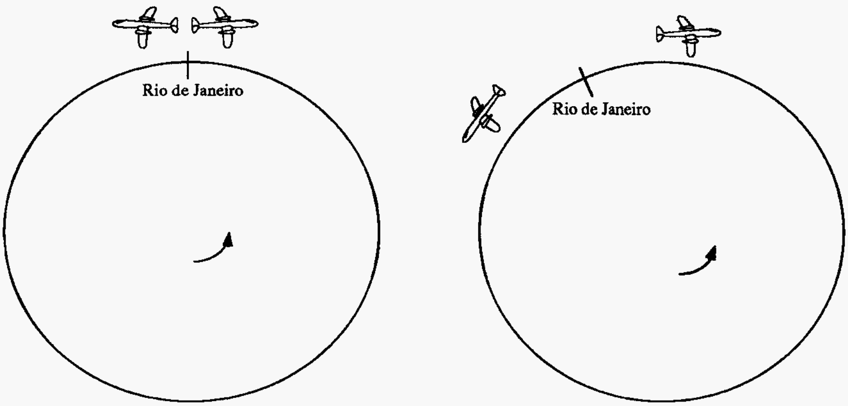

Figure 1.1: Idealization of the Hafele-Keating experiment. Two space ships (pictured as aeroplanes) are flying along the rim of a merry-go-round. One of them is feeling a centrifugal force. The other one, which does not run its engine, does not feel any inertial force. The latter one experiences a longer travel time during a round trip.

a speed that they circle once relative to the merry-go-round while it rotates one turn. This has the effect that centrifugal forces are felt on board one of the ships but not on board the other one. The clocks on board the latter one should therefore show a longer travel time when the clocks are compared after the arrival from a round trip. Hafele and Keating found such an effect in their experiment.

An unstable particle can be said to carry a (nonperiodic) clock, as it has an expected lifetime. Using magnetic fields one can keep unstable muons running round in so called storage rings. It is found that the (centrifugally) accelerated particles in the ring live considerably longer on the average than unaccelerated muons outside the ring. This is a striking confirmation of the principle of maximal proper time. There are a number of other verifications which we won't mention here.

A serious warning should be issued already here. The principle of maximal proper time must under no circumstances be interpreted so that the time effect is greater the greater the acceleration is. The precise formulation is quite important. As it stands it forms the basis

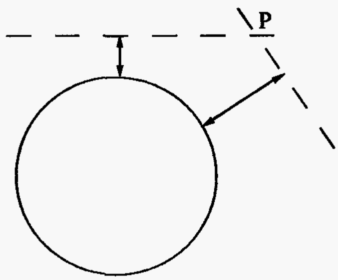

Figure 1.2: The height above *one specified point* on the surface of the earth can be defined as this figure shows. Obviously observers in different places on the earth generally judge the height to one and the same point P to be different.

for the special theory of relativity and in fact also for a part of the general theory of relativity.

What the principle certainly tells us is that time is personal. There is no universal or absolute time valid for all observers. The situation reminds us of the situation for the different observers on the surface of the earth. Each of them defines a personal 'height' (see Figure 1.2). Imagine a dramatic event on a space ship in outer space. It is just as meaningless to ask for the time at which this happened as to ask at what height it happened without further specification.

1.4 Events. Space-Time

When we are shown a graph we tend to ask as we were taught at school: "What's on the axes?" We are so indoctrinated by the Cartesian way of thinking that we tend to forget that we have been taught also another way of approaching graphs and figures: The Euclidean way. This course is going to use the Euclidean approach much more than the Cartesian one. The objects which we are going to discuss are very much like lines and surfaces; the quantities are very much like lengths and angles. Since Euclid's time we have obtained an excellent tool to handle such things: Vectors. We are going to make much use of them.

Describing the world the Cartesian way some statements depend

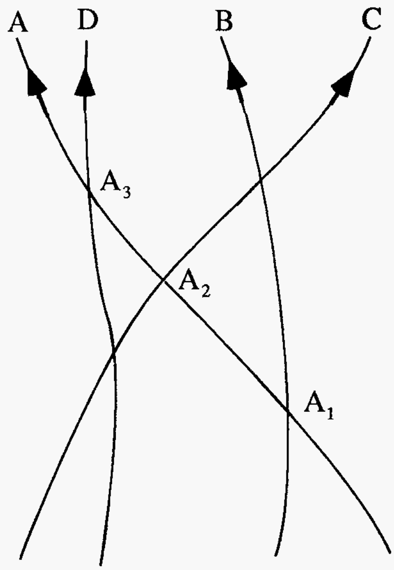

Figure 1.3: Survey of the history of four particles A, B, C, and D. To each one there is a corresponding world line. The events A1, A2, and A3 are the crossings of world line A with the world lines B, C, and D respectively. Corresponding statements hold for the other world lines.

on the choice of coordinate system and others don't. The latter kind of statements are said to be *invariant* under changes of coordinates. All physically meaningful statements must be invariant. Using the Euclidean approach the question of invariance does not arise. All statements are physically meaningful.

The history of a particle is called its *world line.* A brief encounter between two particles is a crossing of their world lines. Such a crossing is called an *event.* It is often useful to draw a picture of world lines and events (see Figure 1.3). Along each world line a number of events are marked and we can see in what order they appear, measured with a clock on board the particle.[2] The arrows show the direction from

[2] A 'particle' is any material body which is small in comparison to the typical extension of the system we are investigating. Remember that clocks are required to

earlier to later times. So far we are not interested in lengths and angles nor are we interested in if the lines are straight or bent.

An observer on board a (small) space ship follows the ship's world line. We shall see in Chapter 4 how such an observer can form a conception of the world around him through signals which he sends out and signals which he receives. The history of such signals are world lines crossing the observer world line. In this way the observer forms a personal opinion of his environment and he can interpret it in terms of his own time and a space of his own; a space with three dimensions. This reminds us again of the different observers on the surface of the earth discussed in Section 1.3. Each one defines his own height but also his own two-dimensional plane touching the surface of the earth.

From this we conclude that the world lines should be drawn in a four-dimensional space. Such a space cannot be visualized. In most applications we are, however, interested in a two- or three-dimensional subspace and consequently this is not such a severe problem as one would imagine at first.

All the possible world lines crossing each other form a net in a four-dimensional space and the points in this space are the events. The space is called *event space* or *spacetime.*

1.5 Parallel World Lines

We define a *straight world line* as the *history of an unaccelerated particle.* In spacetime diagrams such world lines will also be drawn as straight lines.

That the world line for a space ship is straight can be said to mean that the ship is on a 'steady course' in spacetime. Considering two different space ships we would like to give a meaning to the notion that they are on the same course, or that their world lines are *parallel.*

Figure 1.4a shows a construction for obtaining parallel lines in ordinary Euclidean geometry: If two rods are joined in their middle points, then the straight lines through their end points are parallel in pairs. The very similar Figure 1.4*b* illustrates a course of events in spacetime.

---

be small in this sense (see Section 1.1). Hence they can themselves be considered to be particles.

The space ships S_1 and S_2 are freely floating. Signals are sent between them. These signals consist of freely floating bullets B_1 and B_2. The first one is shot from S_1 and hits S_2. The second one goes the other way. These bullets contain clocks and recording devices. After it has been hit a ship communicates the records received to the other ship. If it then turns out that the two bullets *just barely escaped colliding when both were half way* (as judged from their clocks), then the two ships are by definition on the same course in spacetime — their world lines are parallel.

It is quite instructive to consider the corresponding situation in Newtonian theory: Two space ships are both unaccelerated. They shoot bullets at each other and both bullets hit. You easily realize that if these bullets almost collide and if this happens half way for both of them, then the two space ships are at rest relative to each other.

Returning to the relativistic description I would like to point at a few typical features:

- No general time, only proper times, occur in the description.
- Nowhere is 'space' (in contradistinction to 'spacetime') mentioned.
- Records must be kept of the different clocks. Only in retrospect can one obtain the full picture.
- If we actually would perform the experiment we could not hope to obtain just the right bullet world lines with only two bullets. It would be necessary to shoot a number of bullets and afterwards select from the records those which fit our requirements.

These features are valid not only for this particular experiment but for relativistic situations in general.

Look again at Figure 1.4*b*. The part AB of the world line S_1 represents a certain time interval, say 100 seconds measured with an atomic clock, on board the space ship S_1. Such a part of a space ship world line will be referred to as a space-ship *vector*. In pictures it will be drawn with an arrow pointing in the direction from earlier to later times according to the clocks on board the the space ship.

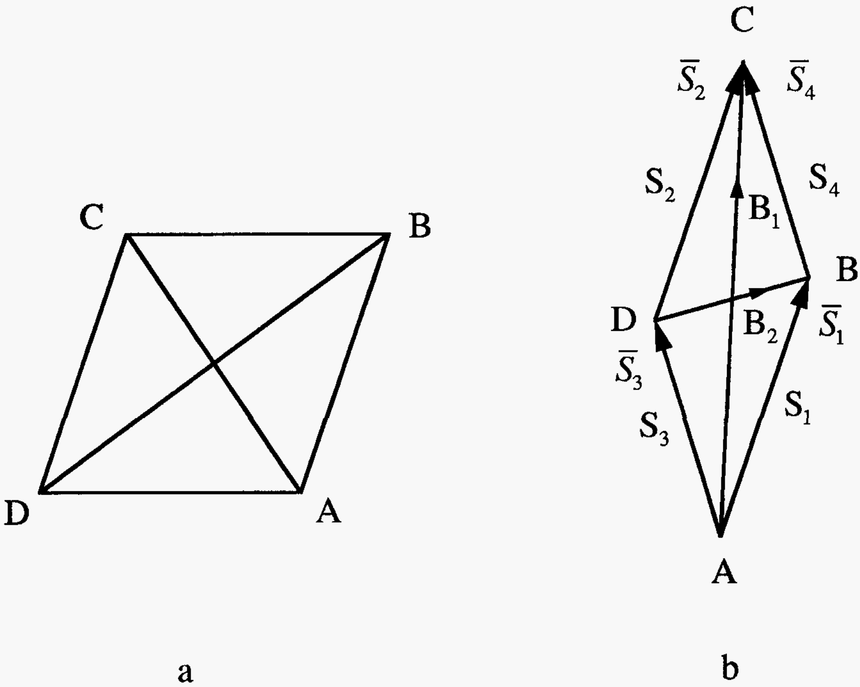

Figure 1.4: Figure a is an ordinary Euclidean figure. Assume that AC and BD cut each other in halves. Then $AB \parallel CD$ and $AD \parallel BC$. Also the lengths of these sides are equal in pairs: $|AB| = |CD|$ and $|AD| = |BC|$. This construction can hence be used to parallel transport vectors. Figure *b* illustrates a course of events in spacetime.

Starting from AB on S_1 the construction in Figure 1.4*b* marks out a time interval DC on board another space ship S_2. This construction is by definition an *elementary parallel displacement* of the vector AB.

A special case of elementary parallel displacement is obtained when we let the two world lines S_1 and S_2 coincide. Regarding Figure 1.4*b* in this limit we see that AB is in this case displaced in its own direction and the length of its corresponding time interval is unchanged by the displacement.

A *parallel displacement* is by definition composed of a series of elementary parallel displacements. Consider a parallel displacement on a certain space ship vector which moves the footpoint of the vector from some given event to another given event. Such a displacement can be composed in many (actually infinitely many) ways of elementary parallel displacements. *We shall assume that they all give the same vector as end result.* This is the *uniqueness rule for parallel displacements.* It is by no means trivial. The rule is in fact *not* true close to gravitating bodies. To treat this case we need *the theory of general relativity.* However, when gravitation is weak experience tells us that the uniqueness rule is true[3], something which we shall assume all through this book. We are then in the realm of *the theory of special relativity.*

There is one aspect of Figure 1.4*b* which we have so far neglected and that is the relation between the vectors AD and BC. They are connected through a slightly different type of experiment than the one we have considered so far: Two bullets are shot from the same ship and hit the other one. One of the bullets overtakes the other one and passes when both bullet clocks show half time. This kind of experiments constitutes an alternative way of defining elementary parallel displacement. We shall assume that the uniqueness rule is true also if we use this kind if elementary parallel displacement or even if we mix the two kinds in the definition of parallel displacements.

In short, the construction in Figure 1.4*b* constitutes the definition of elementary parallel displacements. More generally parallel displacement is obtained by a series of such constructions which can contain both experiments with two ways and with one way transmission (of bullets). Parallel displacement is assumed to be unique: Performing a parallel displacement of a given vector so that its foot point is moved to a given position (which we assume that we can choose arbitrarily) always gives the same resulting vector.

[3]More axiomatically we could *define* absence of gravity as the fulfilment of the uniqueness rule.

Chapter 2

Vector Algebra

2.1 Basic Properties

In Section 1.5 we introduced the term space-ship vector for a part of a space ship world line, i.e., a proper time interval on board a space ship, equipped with an arrow pointing in the positive time direction. You are certainly familiar with vectors from Euclidean geometry and we will use the same notation for vectors as there with a 'bar' on top of a letter, e.g., $\bar{A}$. Vectors in spacetime are often called *four-vectors* with reference to the four dimensions. You don't need to read the rest of this section (Section 2.1) if you are well acquainted with Euclidean geometry or if you are not so interested in the justification of the basic geometric assumptions. Most of the arguments and all the conclusions of this section could just as well belong to ordinary Euclidean space. All the essential conclusions are given in (2.2) – (2.4).

According to Section 1.5 a vector can be moved around in spacetime using parallel displacement. We shall adopt the convention that two vectors which are related by a parallel displacement are *one and the same vector*, and we shall denote them with the same symbol. Also this is something you should be used to from Euclidean geometry.

Choosing some arbitrary event as *origin* there is an event corresponding to each vector and vice versa. You find the event at the head of the vector if its foot is placed at the origin. We say that the vector is the *position vector* of the event.

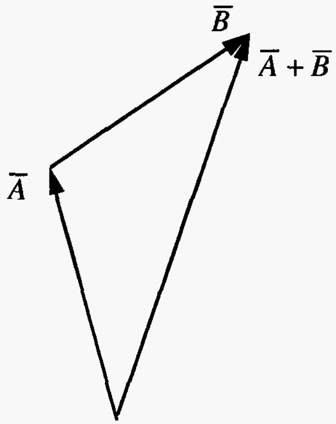

Figure 2.1: Addition of vectors.

Addition of two vectors $\bar{A}$ and $\bar{B}$ gives a vector $\bar{A}+\bar{B}$ defined exactly as in Euclidean space (see Figure 2.1). Figure 1.4*b* tells us immediately that vector addition is commutative. We conclude this from the equations $\bar{S}_1 + \bar{S}_4 = \bar{S}_3 + \bar{S}_2$ and $\bar{S}_1 = \bar{S}_2$ and $\bar{S}_3 = \bar{S}_4$ which give $\bar{S}_1 + \bar{S}_3 = \bar{S}_3 + \bar{S}_1$.

Studying Figure 2.2 we find that a parallel displacement of two vectors $\bar{A}$ and $\bar{B}$ automatically parallel displaces their sum $\bar{A}+\bar{B}$. This is obviously important. Without this property we would have had to discard either vector addition or the identification of parallel transported vectors.

Figure 2.3 shows that vector addition is associative, i.e., $(\bar{A}+\bar{B})+\bar{C} = \bar{A}+(\bar{B}+\bar{C})$.

Multiplying a vector (which is a part of a space ship world line) by a real number means simply multiplying the length of the corresponding proper time interval by the number. It is clear that $n(\bar{A}+\bar{B}) = n\bar{A}+n\bar{B}$, if n is a positive integer. Defining $\bar{C} \equiv n\bar{A}$ and $\bar{D} \equiv n\bar{B}$ we find $(1/n)\bar{C}+(1/n)\bar{D} = (1/n)(\bar{C}+\bar{D})$. Hence $x(\bar{C}+\bar{D}) = x\bar{C}+x\bar{D}$ for any positive rational number, $x = m/n$. With reference to continuity we generalize this to x being any positive real number.

Let $\bar{A}$ be some vector which is a part of a space ship world line and directed, according to our convention, in the positive proper time

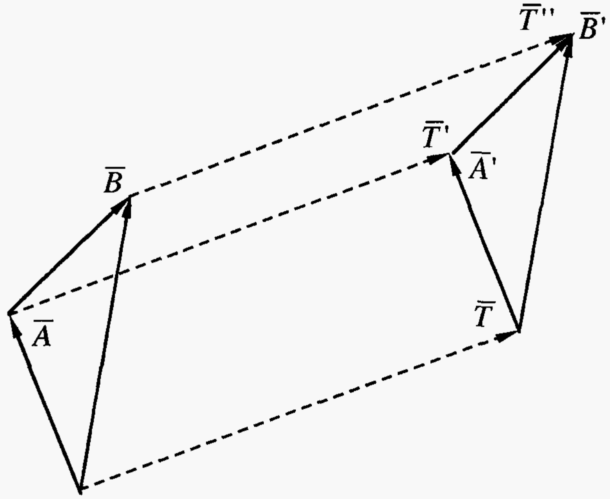

Figure 2.2: If $\bar{A}$ and $\bar{B}$ are parallel transported so is $\bar{A}+\bar{B}$: Assuming that $\bar{A}' = \bar{A}$ we know that $\bar{T}' = \bar{T}$ according to our construction of elementary parallel transport in Chapter 1. Similarly $\bar{B}' = \bar{B}$ gives $\bar{T}'' = \bar{T}'$. Hence $\bar{T}'' = \bar{T}$ from which follows $\bar{A}+\bar{B} = \bar{A}'+\bar{B}'$. Q.E.D.

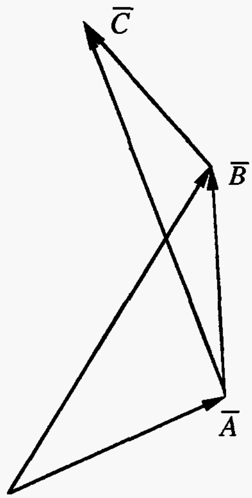

Figure 2.3: vector addition is associative: $(\bar{A}+\bar{B})+\bar{C} = \bar{A}+(\bar{B}+\bar{C})$.

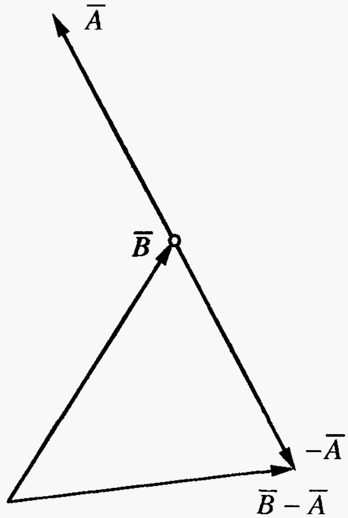

Figure 2.4: Addition of one positively directed space ship vector $\bar{B}$ and one negatively directed space ship vector $-\bar{A}$. The result is $\bar{B} - \bar{A}$.

direction, i.e., from earlier to later times. The vector $-\bar{A}$ is naturally defined as the same vector as $\bar{A}$ except for being directed *opposite* to the time direction. In Figure 2.4 we have illustrated the addition of one positively and one negatively directed space ship vector. As we shall see in Section 3.1 the result, $\bar{B} - \bar{A}$, may be a new type of vector which cannot be a part of a space ship world line. Such vectors must, however, be included to cover the entire spacetime with help of position vectors (and some chosen origin). An arbitrary vector $\bar{V}$ can be written as follows

$$\bar{V} = a\bar{A} + b\bar{B} + c\bar{C} + d\bar{D} \tag{2.1}$$

where $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$ are four given space ship vectors and a, b, c, and d are arbitrary real numbers.

Summing up, spacetime vectors have the following properties, well known also from Euclidean space

$$\bar{A} + \bar{B} = \bar{B} + \bar{A} \tag{2.2}$$

$$(\bar{A} + \bar{B}) + \bar{C} = \bar{A} + (\bar{B} + \bar{C}) \tag{2.3}$$

$$x(\bar{A} + \bar{B}) = x\bar{A} + x\bar{B}. \tag{2.4}$$

2.2 Scalar Product

Let me remind you of the properties of the scalar product in ordinary Euclidean space. It is a real number tied to each pair of vectors. One usually denotes it by the two vectors with a dot between them. Hence $\bar{a} \cdot \bar{b}$ is the number tied to the pair of vectors $\bar{a}$, $\bar{b}$ — it is the scalar product between the two vectors.

The scalar product has the following properties:

- It is *commutative*, i.e.,

$$\bar{a} \cdot \bar{b} = \bar{b} \cdot \bar{a}$$

- It is *linear*[1] in each vector, i.e.,

$$\bar{a} \cdot (\beta\bar{b} + \gamma\bar{c}) = \beta\bar{a} \cdot \bar{b} + \gamma\bar{a} \cdot \bar{c}$$

  where β and γ are arbitrary real numbers.

- The scalar product of a vector with itself is its length squared, i.e.,

$$\bar{a} \cdot \bar{a} = l_a^2$$

  where l_a is the length of the vector $\bar{a}$.

[1]Using a more exact terminology the property we refer to here should be named 'linearity and homogeneity'. See further in Chapter 5.

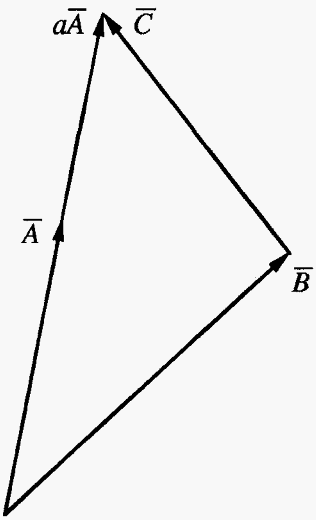

Figure 2.5: Thought experiment to obtain the scalar product $\bar{A} \cdot \bar{B}$.

We shall assume the existence of a scalar product with these properties also in spacetime, where proper time plays the rôle of length. These properties of the scalar product determines it uniquely in terms of measurable quantities as we shall now see. The only physical assumption involved is therefore the *existence of a scalar product.*

Consider two vectors $\bar{A}$ and $\bar{B}$, which both are parts of straight space-ship world-lines (see Figure 2.5). We may think of B as a capsule shot away from space ship A. At the tip of $\bar{B}$ the capsule turns sharply homeward through a powerful acceleration. It hits the space ship at the tip of the vector $a\bar{A}$. The proper times of all the vectors and the number a can be determined by clock readings. Obviously

$$a\bar{A} = \bar{B} + \bar{C} \tag{2.5}$$

or

$$\bar{C} = a\bar{A} - \bar{B}. \tag{2.6}$$

'Square' this equation and we obtain

$$\bar{C}\cdot\bar{C} = (a\bar{A}-\bar{B})\cdot(a\bar{A}-\bar{B}) = (a\bar{A}-\bar{B})\cdot a\bar{A} - (a\bar{A}-\bar{B})\cdot\bar{B} =$$
$$a\bar{A}\cdot a\bar{A} - \bar{B}\cdot a\bar{A} - a\bar{A}\cdot\bar{B} + \bar{B}\cdot\bar{B} =$$
$$a^2\bar{A}\cdot\bar{A} - 2a\bar{A}\cdot\bar{B} + \bar{B}\cdot\bar{B}. \tag{2.7}$$

That is

$$\bar{A}\cdot\bar{B} \equiv \frac{1}{2a}[a^2\bar{A}\cdot\bar{A} + \bar{B}\cdot\bar{B} - \bar{C}\cdot\bar{C}]. \tag{2.8}$$

Thus we see that the scalar product $\bar{A}\cdot\bar{B}$ at least in principle can be calculated from the readings of clocks.

It should be made clear that there is a physical content in the mere existence of a scalar product. The existence of a scalar product leads e.g. to the following identity

$$(\bar{A}+\bar{B})^2 + (\bar{A}-\bar{B})^2 \equiv 2\bar{A}^2 + 2\bar{B}^2 \tag{2.9}$$

which contains only squares of vectors and therefore only proper times. In principle we may arrange an experiment to test (2.9) (see Figure 2.6).

Nothing prevents us in principle from taking over the rules for scalar products unchanged from Euclidean geometry with lengths just replaced by proper times, and a number of authors do in fact do so. For reasons purely of convenience we shall, however, make a slight change. We shall define the square of a space-ship vector as *minus* the square of the corresponding proper time. This may appear very strange at the moment but it will soon turn out to be quite practical.

Thus our rules for scalar products are

$$\bar{A}\cdot\bar{B} = \bar{B}\cdot\bar{A} \tag{2.10}$$

$$\bar{A}\cdot(\beta\bar{B}+\gamma\bar{C}) = \beta\bar{A}\cdot\bar{B} + \gamma\bar{A}\cdot\bar{C} \tag{2.11}$$

$$\bar{A}\cdot\bar{A} = -\tau_A^2 \tag{2.12}$$

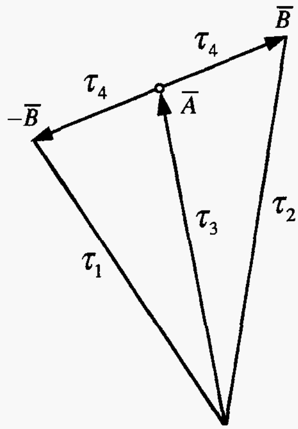

Figure 2.6: Four space ships are involved in this experiment to test the existence of a scalar product through equation (2.9), which can be written $\tau_1^2 + \tau_2^2 = 2\tau_3^2 + 2\tau_4^2$. The corresponding rule in ordinary Euclidean geometry is ascribed to Apollonios.

where β and γ are arbitrary real numbers. The last equation presupposes that $\bar{A}$ is a space ship vector and τ_A is the corresponding proper time.

This ends the account of the assumptions behind the special theory of relativity. In the following chapters we shall discuss the consequences of these assumptions.

Chapter 3

Vector Characteristics

3.1 Timelike. Spacelike. Null-like

Suppose that $\bar{U}$ and $\bar{V}$ are two vectors which both are along (possible) space-ship world-lines. Our convention from the end of Section 2.2 then tells us that

$$\bar{U}^2 = -\tau_U^2 < 0 \tag{3.1}$$

$$\bar{V}^2 = -\tau_V^2 < 0 \tag{3.2}$$

where τ_U and τ_V are the proper times of $\bar{U}$ and $\bar{V}$ respectively.

Let us now combine $\bar{U}$ and $\bar{V}$ linearly with coefficients u and v to form a new vector $\bar{A}$

$$\bar{A} = u\bar{U} + v\bar{V}. \tag{3.3}$$

Now scalar multiply both members of this equation by the vector $\bar{U}$ and we obtain

$$\bar{U} \cdot \bar{A} = u\bar{U}^2 + v\bar{U} \cdot \bar{V}. \tag{3.4}$$

Here u and v can be any real numbers. If we choose them so that they satisfy the following equation

$$u/v = -\bar{U} \cdot \bar{V}/\bar{U}^2 \tag{3.5}$$

equation (3.4) gives

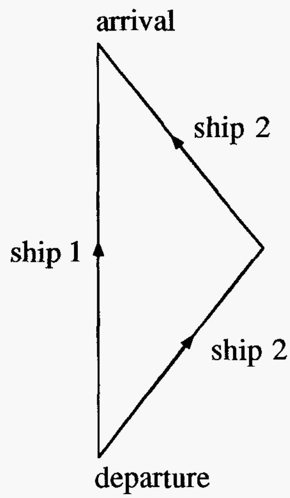

Figure 3.1: World lines for two space ships which part and meet again. One of them is unaccelerated all through its journey. The other one accelerates briefly and returns.

$$\bar{U} \cdot \bar{A} = 0. \tag{3.6}$$

We have thus constructed a vector $\bar{A}$ such that its scalar product with $\bar{U}$ vanishes. Two vectors whose scalar product vanishes are said to be *orthogonal* to one another.

Figure 3.1 illustrates a common situation in space travel, which we have actually already touched upon. An unaccelerated space ship shoots out another (smaller) space ship. The latter one is also unaccelerated for a while. Then it briefly runs its engine very hard so that it can return freely floating to the first ship.[1]

The vectors involved in a special space travel of this sort are drawn in Figure 3.2. The spacetime vector of the first ship from start to arrival is denoted $\bar{U}$. As we have just learnt there always exists a vector which we denote $\bar{A}$ orthogonal to $\bar{U}$. We can therefore construct the two

[1]The figure with its sharp corner presupposes an infinite acceleration. One can quite easily change the situation so that no infinite accelerations occur without changing the conclusions in this section.

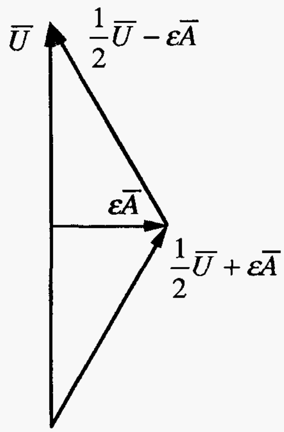

Figure 3.2: Vectors involved in the space travel illustrated in Figure 3.1.

vectors

$$\bar{V}_1 = (1/2)\bar{U} + \varepsilon\bar{A}; \quad \bar{V}_2 = (1/2)\bar{U} - \varepsilon\bar{A} \tag{3.7}$$

where ε is a small number. Obviously

$$\bar{V}_1 + \bar{V}_2 = (1/2)\bar{U} + \varepsilon\bar{A} + (1/2)\bar{U} - \varepsilon\bar{A} = \bar{U}. \tag{3.8}$$

Hence the ships will meet as required if we let $\bar{V}_1$ and $\bar{V}_2$ be the space-time vectors of the first and second parts of the journey of the second ship.

The proper time for the journey of the first ship is

$$\tau_1 = \sqrt{-\bar{U}^2}. \tag{3.9}$$

The total proper time for the second ship's journey is

$$\tau_2 = \sqrt{-(\frac{1}{2}\bar{U} + \varepsilon\bar{A})^2} + \sqrt{-(\frac{1}{2}\bar{U} - \varepsilon\bar{A})^2} =$$
$$\sqrt{-\frac{1}{4}\bar{U}^2 - \varepsilon\bar{U}\cdot\bar{A} - \varepsilon^2\bar{A}^2} + \sqrt{-\frac{1}{4}\bar{U}^2 + \varepsilon\bar{U}\cdot\bar{A} - \varepsilon^2\bar{A}^2}. \tag{3.10}$$

Using now our assumption that $\bar{A}$ and $\bar{U}$ are orthogonal, i.e., (3.6), we obtain from this equation

$$\tau_2 = \sqrt{-\bar{U}^2}\sqrt{1 + 4\varepsilon^2 \bar{A}^2/\bar{U}^2}. \tag{3.11}$$

The principle of the maximal proper time in Section 1.3 tells us that

$$\tau_1 > \tau_2 \quad \text{if} \quad \varepsilon \bar{A} \neq 0. \tag{3.12}$$

This in combination with (3.9) and (3.11) gives

$$\bar{A}^2 > 0 \quad \text{if} \quad \bar{A} \neq 0. \tag{3.13}$$

This has been obtained under the assumption that $\bar{U}^2$ is negative. If we had used the opposite convention, i.e., $\bar{U}^2 > 0$, we would now have been led to $\bar{A}^2 < 0$. Hence all vectors squared cannot be positive. The existence of negative vector squares is unfamiliar to Euclidean geometry and it is the impossibility to avoid them in spacetime geometry which constitutes the main difference between it and Euclidean geometry.

Obviously no space ship can travel along the vector $\bar{A}$. Its clocks would have shown an imaginary time which is absurd.

Let us now combine the vectors $\bar{U}$ and $\bar{A}$ which we have discussed above to a new vector $\bar{N}$

$$\bar{N} = u\bar{U} + a\bar{A} \tag{3.14}$$

where the coefficients satisfy

$$a^2/u^2 = -\bar{U}^2/\bar{A}^2. \tag{3.15}$$

The square of $\bar{N}$ is

$$\bar{N}^2 = u^2\bar{U}^2 + 2ua\bar{U}\cdot\bar{A} + a^2\bar{A}^2. \tag{3.16}$$

Using (3.15) and $\bar{U}\cdot\bar{A} = 0$ this gives

$$\bar{N}^2 = u^2\bar{U}^2 + a^2\bar{A}^2 = 0. \tag{3.17}$$

Hence there do in spacetime exist vectors which are not the null vector (i.e., $\bar{N} \neq 0$) but whose squares are nevertheless zero.

To summarize the results of this section there exist three classes of vectors (in addition to the null vector) in spacetime. There are vectors with

- negative square called *timelike vectors*
- positive square called *spacelike vectors*
- zero square called *null-like vectors*

Another result is (3.13) which tells us that *if $\bar{U}$ is timelike and $\bar{A}$ is orthogonal to $\bar{U}$, then $\bar{A}$ is either the null vector (i.e., $\bar{A} = 0$) or it is spacelike.* You should memorize this rule as it is very important in many applications.

3.2 Comparison with Euclidean Space

Spacetime is four-dimensional while the ordinary Euclidean space, which you are used to, is three-dimensional. In many applications this is not so much noticed because one considers a three- or even two-dimensional subspace of spacetime.

A much more significant difference has its root in the principle of maximal proper time. If we compare proper time in spacetime and distance in Euclidean space this rule should be compared with the Euclidean rule that the straight line has the *shortest* distance between two given points. Thus in one case there is a maximum principle and in the other case there is a minimum principle.

As we have seen in Section 3.1 this difference leads to another important difference: There are three classes of vectors in spacetime while there is only one class in Euclidean space. This must always be kept in mind when discussing and calculating in spacetime. You will have to keep book, remembering to which class a certain vector belongs. If you don't know you may have to consider all the different possibilities as separate cases.

When you draw a spacetime figure on a piece of paper you draw it of course *in Euclidean space.* Hence you cannot make clear in the drawing to which class a certain vector belongs and if you don't have this in mind such a drawing can be very misleading. In fact you must not think that proper times can be found by measuring the lengths of vectors in a drawing. The same is true for angles. There have been a number of attempts to overcome this difficulty. In fact even books have been written on the subject but I think that it is fair to say that none is very helpful in visualizing spacetime.

One can ask what good there is in a figure if you cannot rely on lengths nor on angles. In fact purists tend to give the following piece of good advice: *Never trust a figure. All results must be obtained by vector algebra.*

If you follow this advice you are certainly safe. On the other hand, if you never use a drawing you do deprive yourself of the insights which in spite of everything are to be gained from figures. Here is a list of what properties of a spacetime situation you can illustrate by a figure in Euclidean space:

- You can represent the summation of vectors. You can also represent parallel vectors and the relative length of parallel vectors. (We did all this before the scalar product was introduced in Section 2.2.) As a result you can draw figures so that every pair of similar triangles in the figure corresponds to a pair of similar triangles in spacetime. This can be of help as work with similar triangles is an important part of ordinary Euclidean geometry.

- If two spacetime vectors are orthogonal we shall often draw them as perpendicular. The price we have to pay is that other vectors which are orthogonal will in general not appear as perpendicular to one another.

- As long as we don't give values to the scalar products the algebra in spacetime is the same as in Euclidean space. We can therefore often use the same equations as those obtained from a Euclidean figure.

- Symmetries in a spacetime situation can sometimes be very nicely represented in a Euclidean figure.

- There are certain insights to be gained from a qualitative comparison between a spacetime situation and a Euclidean one. An example is given below.

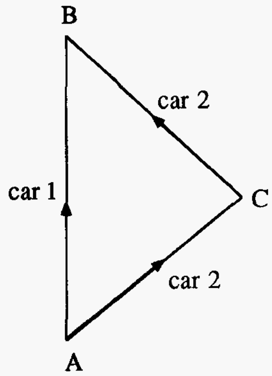

Figure 3.3: Euclidean figure. The distances AB and ACB are measured by cars.

You should not make a great effort to understand all these points at this stage. Hopefully they will become clear as we go along. But always remember: In the last instant it is the algebra that counts. You should use figures only as long as they assist you and not let them become a burden.

As an example of the last point on the list let us turn to a question which inevitably arises in discussions with laymen: What *causes* the difference in time on board two different space ships? It cannot be the velocity because none of the ships feels the velocity and it cannot be the acceleration because we have agreed that acceleration does not affect the time on board. So *logically*, the layman says, there cannot be a difference.

Behind this argument lies the conviction so deeply rooted in us that two identical clocks always must agree on elapsed time. In other words: if clocks don't agree on elapsed time they must have been physically influenced so that they actually are not identical. The relativistic view is entirely different. Ideal clocks are considered to be unaffected by acceleration as well as velocity. The reason for the time difference is thus not some physical influence on the pace of the clocks (whatever that would mean) but the fact that *the two clocks measure the travel times for two different spacetime objects.* To make this thought clear a reference to the corresponding situation in Euclidean geometry comes in handy. Figure 3.3 is meant to be just what it seems to be: An ordinary triangle in the plane. If such a triangle is of suitable size you can measure the distance from A to B simply by driving a car along AB and reading the distance meter, and you can measure the distance ACB in the same manner. Obviously you will find different distances in the two cases. It would not occur to you that this should be ascribed to some physical influence on the distance meter. Quite the contrary you say that if identical distance meters are used in both measurements they must show different distances for the two paths.

Chapter 4

Simultaneity and Space Distance

An observer is bound to a space ship. He may even be regarded as a space ship himself. Primarily he can observe only what takes place on his own world line, but he tries to obtain information about his surroundings by using signals of different types. These signals can also be regarded as space ships if they contain clocks. They may consist of capsules sent away and received by the observer.

4.1 Simultaneity

Let us consider an observer with a straight world line. Purely as a convention the observer may use the method illustrated in Figure 4.1 to *define when* an event P occurs which is not on his own world line. He sends away a capsule which turns around at P and returns to him. He arranges so that the proper time τ which the capsule measures is the same both ways. The observer measures the time interval $2t$ between the emission P_1 and the reception P_3 of the capsule. He defines the event P_2 in the middle of this interval to be *simultaneous* with P.

Let us use the notation $\bar{r}$ for the vector having its head at the event P and its foot at the event on the observer line which is simultaneous

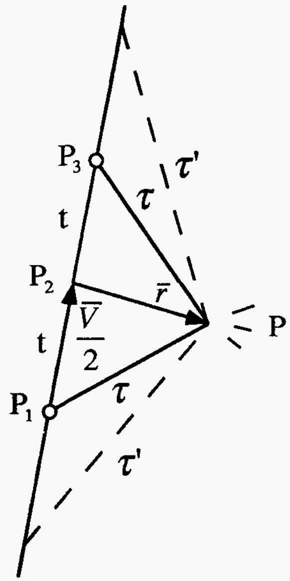

Figure 4.1: Method for defining simultaneity.

with P. Consulting Figure 4.1 we obtain

$$\tau^2 = -(\frac{1}{2}\bar{V} + \bar{r})^2 = -(\frac{1}{2}\bar{V} - \bar{r})^2 \tag{4.1}$$

which gives

$$\bar{V} \cdot \bar{r} = 0. \tag{4.2}$$

In Figure 4.1 the world lines for two different capsules are drawn, one of them as a broken line. Whichever you use you find the same point P_2 as simultaneous with P. Alternatively you could say that whichever capsule you use you find the same vector $\bar{r}$ and this follows from the fact that (4.2) determines the vector $\bar{r}$ uniquely. In case you don't believe this you may read the following formal proof.

The vector $\bar{r}$ has the following properties: 1. Its foot is on the observer world line. 2. Its head is at the event P. 3. It is orthogonal to the observer world line (see (4.2)). These characteristics suffice to define $\bar{r}$ uniquely as we shall now see:

Let us introduce two vectors $\bar{r}_1$ and $\bar{r}_2$ which both satisfy these conditions and show that the two vectors must be equal. Using (4.2) for each one of them we obtain two equations which we subtract obtaining $\bar{V} \cdot (\bar{r}_2 - \bar{r}_1) = 0$. The conditions 1. and 2. do however tell us that the difference between $\bar{r}_2$ and $\bar{r}_1$ must be in the direction of the observer world line, i.e., $\bar{r}_2 - \bar{r}_1 = \lambda\bar{V}$, where λ is a (real) number. Hence $\lambda\bar{V}^2 = 0$. The vector $\bar{V}$ is, however, timelike so that $\bar{V}^2 \neq 0$. Hence $\lambda = 0$ from which follows that $\bar{r}_2 = \bar{r}_1$. Q.E.D.

According to (4.2) and the final rule in Section 3.1 the vector $\bar{r}$ is spacelike, i.e., $\bar{r}^2 > 0$. The *length* of $\bar{r}$ is defined as $l \equiv \sqrt{\bar{r}^2}$.

4.2 Space Distance

Let us now consider a family of straight parallel world lines L_0, L_1, $L_2 \ldots$

$$\bar{R}_n = \lambda_n \bar{U} + n\bar{\rho} \quad \text{where} \quad n = 0, 1, 2 \ldots \tag{4.3}$$

Here both $\bar{U}$ and $\bar{\rho}$ are timelike and $\lambda_0, \lambda_1, \lambda_2 \ldots$ are parameters along the lines. See Figure 4.2.

This could be a spacetime picture of a fleet of a number of identical space ships which fly unaccelerated in formation head to tail in contact with each other without moving relative to each other. The history of the first ship is between the lines L_0 and L_1 the history of the second one between L_1 and L_2 etc. Across the formation from one ship to the next an observer is moving along the world line L'. He counts the number of ships that he passes (in Figure 4.2 the number is six) and returns to the ship where he started (along the world line L'' to its intersection with L_0). There he reports what he has found. This procedure constitutes a method for the observer on L_0 to find out how far it is from one end to the other of the fleet. The distance will be expressed in the unit 'standard space ship'.

There is an alternative way to measure this distance. There is one and only one vector which bridges L_0 and L_n and is orthogonal to $\bar{U}$. As one easily verifies this vector is

$$\bar{r}_n = n\bar{\rho} - (n\bar{\rho} \cdot \bar{U})\bar{U}/\bar{U}^2. \tag{4.4}$$

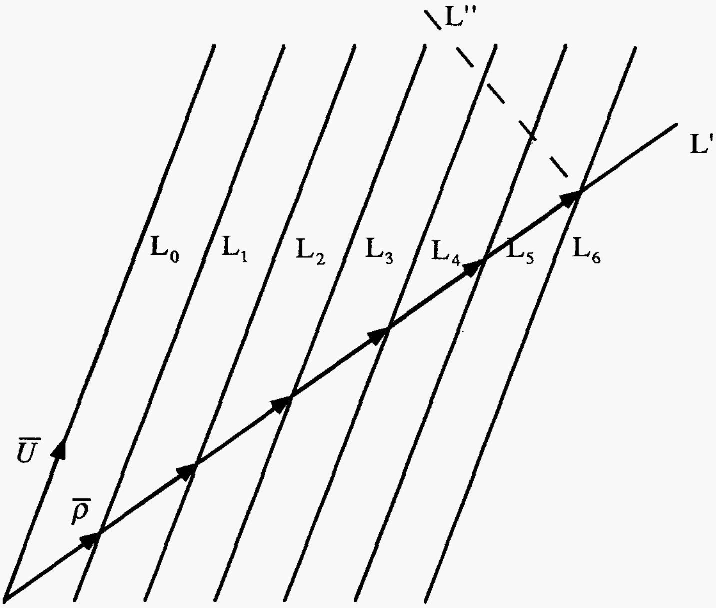

Figure 4.2: World lines of the head and tails in a fleet of six space ships in immediate contact with one another. A messenger ship is sent from one end of the fleet to the other and back.

We notice that this vector and consequently also its length $\sqrt{\bar{r}^2}$ is proportional to n. Thus if we can measure the length of $\bar{r}$ (here and in the following we suppress the index n on $\bar{r}$) we may use it to determine distance. We must only once and for all determine the scale factor between the length of $\bar{r}$ and the number of space ships.[1]

How the length of $\bar{r}$ can be measured can be seen from Figure 4.1. Introducing the notation $\bar{U}$ for $(1/2)\bar{V}$ (4.1) gives

$$-(\bar{U}+\bar{r})^2 = \tau^2 \tag{4.5}$$

[1]There is an assumption involved here that this scale factor is independent of where and in what direction we measure. This is an aspect on the homogeneity and isotropy of spacetime.

or

$$t^2 - \bar{r}^2 = \tau^2 \tag{4.6}$$

or

$$l^2 \equiv \bar{r}^2 = t^2 - \tau^2. \tag{4.7}$$

The great advantage with this way of determining distance is obvious. We are now independent of the space ship fleet except for fixing the unit of distance. Later on we shall find a still more practical method for measuring distance.

The unit of distance has so far been entirely arbitrary. We are free to choose the standard space ship as we like. Actually nothing prevents us from choosing the unit of distance so that $\sqrt{\bar{r}^2}$ itself gives the distance, which means that the scaling factor equals unity. This is actually what we will do from now on. Thus as is seen from (4.7) we give time and distance the same dimension of unit. This is in the theory of relativity as natural as giving the same unit of dimension to height and breadth. As a matter of fact the most precise measurements of distance in the laboratory are actually time measurements and in astronomy one uses lightyears[2] as a unit of distance. We shall soon find that this is the unit of distance corresponding to the time unit 'year'. From the point of view of the theory of relativity we could just as well say that a certain star is ten years distant as saying that it is ten light years distant, and this is in principle what we will do.

4.3 The Orthogonal Space

To every unaccelerated observer corresponds a straight world line and such a line is characterized by the timelike unit vector along the line. As you are probably used to we will denote unit vectors by letters topped by hats, e.g., $\hat{U}$. If $\hat{U}$ is a timelike vector it satisfies $\hat{U}^2 = -1$. Given an observer with a corresponding unit vector $\hat{U}$ there is a set of vectors satisfying (4.2), i.e.,

$$\hat{U} \cdot \bar{r} = 0. \tag{4.8}$$

[2]Professional astronomers use 'parsec' which from a theoretical point of view must be regarded as a more accidental unit.

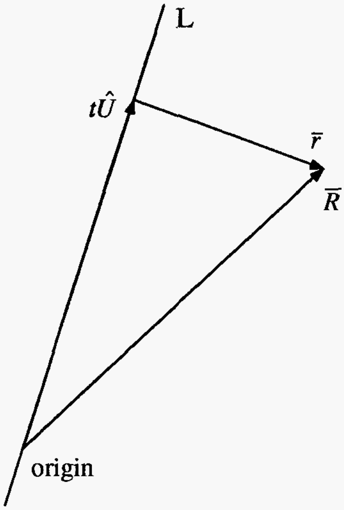

Figure 4.3: Standard split of the vector $\bar{R}$ with respect to the observer with world line in the direction $\hat{U}$.

Together they form the *orthogonal space belonging to the observer.* It is obviously a vector space in the usual sense that if two vectors belong to it then also every linear combination of them belongs to it. It is actually a Euclidean vector-space because all the vectors $\bar{r}$ are spacelike, i.e., they have positive squares (see Section 4.1). The condition (4.8) on the vectors $\bar{r}$ makes the orthogonal space three-dimensional. Remember also that $\sqrt{\bar{r}^2}$ is to be interpreted as the space distance from the observer (see Section 4.2) and you realize that the orthogonal space is the space which is used in Newtonian physics. The difference from Newtonian physics is that in the theory of relativity each observer constructs his own space.

Figure 4.3 shows a split which is very common. Let the timelike unit vector $\hat{U}$ correspond to a definite observer. An arbitrary vector $\bar{R}$ can be split in two vectors one of which is parallel to $\hat{U}$ and one which is orthogonal to $\hat{U}$

$$\bar{R} = t\hat{U} + \bar{r} \quad \text{where} \quad \hat{U} \cdot \bar{r} = 0. \tag{4.9}$$

According to Figure 4.3 the observer 'measures' the 'time' t and the 'space position' $\bar{r}$ for the event corresponding to the position vector $\bar{R}$. The space distance to the event is $\sqrt{\bar{r}^2}$.

Equation (4.9) can easily be solved for t and $\bar{r}$: Multiply it by $\hat{U}$ and you obtain $t = -\hat{U} \cdot \bar{R}$. Insert this in (4.9) and you obtain $\bar{r} = \bar{R} + (\hat{U} \cdot \bar{R})\hat{U}$. Obviously t and $\bar{r}$ are uniquely given by $\bar{R}$ and $\hat{U}$.

Chapter 5

Linear Independence

This chapter constitutes a short mathematical interlude.

Look at the special case of (2.11) in which $\beta\bar{B} = -\gamma\bar{C}$ and you obtain

$$\bar{A} \cdot 0 = 0. \tag{5.1}$$

This property of the function $F(\bar{V}) = \bar{A} \cdot \bar{V}$ of $\bar{V}$ is expressed by saying that F is *homogeneous* in $\bar{V}$. Thus (2.11) implies homogeneity. A function F of $\bar{V}$ which is *linear* but not necessarily homogeneous is given by $F(\bar{V}) = \bar{A} \cdot \bar{V} + const.$

Consider a number of vectors, e.g., four-vectors $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$. If some of them can be written as a linear combination of the others then we say that the vectors are *linearly dependent.* If none of them can be written as a linear combination of the others the vectors are said to be *linearly independent.* A symmetrical way of defining linear independence is the following: The vectors $\bar{A}, \bar{B}, \bar{C} \ldots$ are linearly independent if and only if the following equation

$$a\bar{A} + b\bar{B} + c\bar{C} + \ldots = 0 \tag{5.2}$$

has $a = b = c = \ldots = 0$ as its only solution.

That spacetime is four-dimensional means that there exist sets of four spacetime vectors which are linearly independent but that five or more vectors cannot be linearly independent. Assume that $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$ are linearly independent and write down the equation

$$a\bar{A} + b\bar{B} + c\bar{C} + d\bar{D} + e\bar{E} = 0 \tag{5.3}$$

where $\bar{E}$ is an arbitrary vector. According to what we have just stated there must exist a solution to this equation apart from $a = b = c = d = e = 0$. In this solution $e \neq 0$ because otherwise we must have $a = b = c = d = 0$ due to the linear independence of the first four vectors. Hence the arbitrary vector $\bar{E}$ can be expressed as a linear combination of any four linearly independent vectors.

The coefficients in such an expansion are unique which is quite easy to see: Assume that the vector $\bar{E}$ can be expanded in two different ways as linear combinations of the same set of linearly independent vectors

$$\bar{E} = a\bar{A} + b\bar{B} + c\bar{C} + d\bar{D} \tag{5.4}$$

and

$$\bar{E} = a'\bar{A} + b'\bar{B} + c'\bar{C} + d'\bar{D}. \tag{5.5}$$

Subtracting these equations we obtain

$$(a' - a)\bar{A} + (b' - b)\bar{B} + (c' - c)\bar{C} + (d' - d)\bar{D} = 0. \tag{5.6}$$

Since $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$ are linearly independent this leads to

$$a' = a,\ b' = b,\ c' = c,\ d' = d. \tag{5.7}$$

Q.E.D.

We are used to from Euclidean geometry that two orthogonal vectors always are linearly independent. This is *not* generally true in spacetime. Let me only remind you that any null-like vector is orthogonal to itself.

There is, however, a similar but weaker theorem which is valid in spacetime and plays a great rôle there: If a vector is orthogonal to *all* vectors, then it is the null vector. Otherwise expressed:

$$\text{If } \ \bar{V} \cdot \bar{A} = 0 \ \text{ for all } \bar{A} \text{ then } \ \bar{V} = 0. \tag{5.8}$$

We can easily prove this by first choosing $\bar{A}$ as a timelike vector, say $\bar{U}$. From $\bar{U} \cdot \bar{V} = 0$ follows that $\bar{V}$ is in the orthogonal space to $\bar{U}$.

Choosing then $\bar{A}$ as vectors in this orthogonal space we have brought the situation back to the Euclidean case and we conclude that $\bar{V} = 0$.

We can use this theorem to construct a test on linear independence of four spacetime vectors $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$: Construct the following determinant

$$\begin{vmatrix} \bar{A}\cdot\bar{A} & \bar{A}\cdot\bar{B} & \bar{A}\cdot\bar{C} & \bar{A}\cdot\bar{D} \\ \bar{B}\cdot\bar{A} & \bar{B}\cdot\bar{B} & \bar{B}\cdot\bar{C} & \bar{B}\cdot\bar{D} \\ \bar{C}\cdot\bar{A} & \bar{C}\cdot\bar{B} & \bar{C}\cdot\bar{C} & \bar{C}\cdot\bar{D} \\ \bar{D}\cdot\bar{A} & \bar{D}\cdot\bar{B} & \bar{D}\cdot\bar{C} & \bar{D}\cdot\bar{D} \end{vmatrix}$$

We propose that *a necessary and sufficient condition for the vectors to be linearly dependent is that this determinant vanishes.*

To prove this let us first assume that the vectors are linearly dependent which means that (5.2) is valid with at least one of the coefficients different from zero. From this follows that we can perform the corresponding linear combination of the rows in the determinant and obtain a row of zero elements. Thus the determinant vanishes as a result of (5.1).

Let us now turn to the converse part of the theorem and assume that the determinant is zero. In that case there exists a linear combination of its rows which produces a row with four zeros. Thus there exists a combination

$$\bar{V} = a\bar{A} + b\bar{B} + c\bar{C} + d\bar{D} \tag{5.9}$$

where a, b, c, and d are not all zero such that $\bar{V}$ is orthogonal to all the vectors $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$. Now assuming that these vectors are linearly independent leads us into a contradiction because if this is true each vector can be written as a linear combination of them and hence $\bar{V}$ is orthogonal to every vector so that $\bar{V} = 0$, which according to (5.9) means that our vectors are linearly dependent contrary to our assumption. We have thus proven that $\bar{A}, \bar{B}, \bar{C}$, and $\bar{D}$ are linearly dependent. This concludes the proof.

It should be stressed that the theorem is valid for four spacetime vectors but that there is no such theorem for arbitrary subspaces, i.e., for two or three vectors.

Chapter 6

Relative Velocity and Four-Velocity

Consider an unaccelerated observer measuring on an unaccelerated particle. The particle passes close to the observer and we choose this event as the origin. Figure 4.3 is a spacetime picture of the situation with L and L' as the world lines of the observer and the particle respectively. Referring to equation (4.9) and the interpretations of t and $\bar{r}$ the quantity $\bar{r}/t$ is obviously the *velocity* of the particle as measured by the observer:

$$\bar{v} \equiv \bar{r}/t. \tag{6.1}$$

The unit vector along the particle's world line is denoted $\hat{V}$. The velocity $\bar{v}$ tells us how $\hat{V}$ is directed *in relation to* $\hat{U}$. Thus the velocity always requires an observer for its definition. To make this clear we shall often refer to it as the *relative* velocity. To make a distinction we shall refer to the unit vector $\hat{V}$ itself as the *four-velocity* of the particle. In general every timelike unit vector pointing in the positive time direction is said to be a four-velocity because it can be the direction of some particle's world line. Don't mix up the two notions 'four-vector' and 'four-velocity'. ' Four-velocity' is a special case of 'four-vector'.

In the present chapter we will discuss developments which are connected with the notions 'relative velocity' and 'four-velocity'.

6.1 The Standard Velocity Split

Introducing $\hat{V}$ as the four-velocity of the particle in Figure 4.3, i.e., $\bar{R} \equiv \tau \hat{V}$, (4.9) can be written as

$$\hat{V} = \gamma(\hat{U} + \bar{v}) \quad \text{with} \quad \bar{v} \cdot \hat{U} = 0 \tag{6.2}$$

where

$$\gamma \equiv t/\tau. \tag{6.3}$$

You should memorize the standard velocity split (6.2) which has a great number of applications. Squaring (6.2) you obtain

$$1 = \gamma^2(1 - \bar{v}^2) \tag{6.4}$$

which together with (6.3) gives

$$t/\tau = \gamma = 1/\sqrt{1 - \bar{v}^2}. \tag{6.5}$$

We have here arrived at the famous equation for *time dilation*. It can be used to calculate quantitatively the kind of effects which we touched upon in Section 1.3.

Now multiply (6.2) by $\hat{U}$ to obtain

$$\gamma = -\hat{U} \cdot \hat{V} \tag{6.6}$$

which also is often used in applications. Together with (6.5) it gives

$$\hat{U} \cdot \hat{V} = -1/\sqrt{1 - \bar{v}^2}. \tag{6.7}$$

Hence we can conclude that the absolute value $v \equiv \sqrt{\bar{v}^2}$ of $\bar{v}$ is symmetric in $\hat{U}$ and $\hat{V}$ so that the 'particle' judges the 'observer' to have the same v as the 'observer' judges the 'particle' to have.

6.2 Light Signals

So far we have not discussed null-like lines which form a limiting case between timelike and spacelike lines. Let us now, however, assume that L in figure 4.3 is a null-like line. In this case

$$\tau^2 = -\bar{R}^2 = 0. \tag{6.8}$$

Squaring (4.9) we thus obtain

$$\bar{r}^2 = t^2 \tag{6.9}$$

and hence

$$v^2 = \bar{v}^2 = 1. \tag{6.10}$$

Notice here the striking fact that this is true for an arbitrary observer. *The absolute value of the velocity of a particle with a null-like world line is the same for all observers.* It does not follow from the theory of relativity as such that particles following null-like lines do exist. It would obviously be meaningless to contemplate putting a clock on board such a particle as its proper time does not develop. Nonetheless experience tells us that certain *signals* can travel along null-like world lines. Light signals is an example of this. For this reason the unit velocity, $v = 1$, is usually called the *'velocity of light'*.

In Section 4.1 in connection with Figure 4.1 we discussed the measurement of simultaneity and space distances. We performed a gedanken experiment with a capsule whose velocity from and towards the observer was chosen so that its clock measured the same proper time both ways. Employing light signals simplifies such experiments considerably and this is almost always utilized in actual measurements. Replacing the capsule with a light signal in Section 4.1 the experimentalist is *automatically* guaranteed that the proper time is the same both ways, because $\tau = 0$ both ways. Otherwise expressed the observer knows the velocity of the light signal ($v = 1$) and can therefore directly calculate the distance from his own measurement of the time interval from emission to reception.

The existence of signals with unit velocity is thus of practical importance. It is not, however, a necessary prerequisite for the theory of relativity.

At the origin of Figure 6.1 a lamp twinkles once and emits a pulse of light which is spreading in all directions. The center line is the world line for some observer who passes close to the lamp at the moment when it twinkles. He can use the methods of Section 4.1 to judge

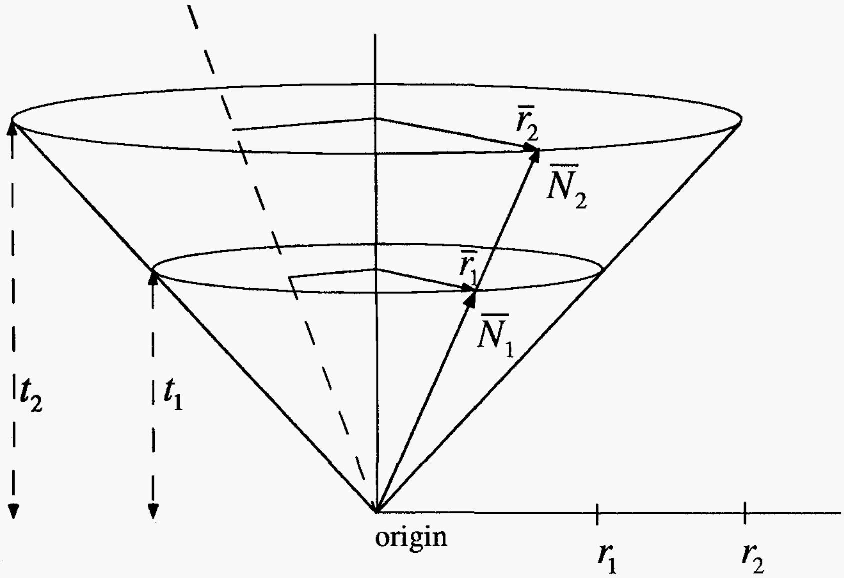

Figure 6.1: The future light cone. $\sqrt{\bar{r}^2} = t$. The vectors $\bar{N}_1$ and $\bar{N}_2$ are null-like.

(retrospectively) how the light front has been moving and he will find that it has been spreading with unit velocity outwards — it has followed the equation (6.9), where $\bar{r}$ and t are referred to our observer. In three dimensions this is the equation for a cone as we have drawn it in Figure 6.1. At each moment of time t the figure shows a circle which in four-dimensional spacetime corresponds to a sphere. Equation (6.9) is referred to as *'the light cone'*.

Actually Figure 6.1 illustrates only a part of the light cone — the *future light cone* $\sqrt{\bar{r}^2} = t$. There is also the *past light cone* with equation $\sqrt{\bar{r}^2} = -t$. This is the equation for a spherical light front *contracting* with unit velocity to the origin. In Figure 6.1 it could have been drawn as a cone with its apex directed upwards.

You must keep in mind that the light cone is a physical phenomenon which exists independently of any observer and that the derivation of (6.9) did not involve a particular observer. Hence the same equation is valid for any observer. If you choose another observer as indicated by the dotted line in Figure 6.1 and let time and space be related to this observer the light cone appears exactly the same as for the original observer and not oblique as expected from the figure. This is a point where a figure drawn in Euclidean space is misleading and you must be aware of it.

6.3 Split of Null-Like Vectors

Any vector can be split in a vector parallel to a given four-velocity $\hat{U}$ and a vector orthogonal to $\hat{U}$. Perform the general split of a *null-like vector* $\bar{N}$ (cf. (6.2))

$$\bar{N} = \kappa(\hat{U} + \bar{n}) \tag{6.11}$$

where κ is a number and $\bar{n}$ is orthogonal to $\hat{U}$. Using $\bar{N}^2 = 0$ you find immediately that $\bar{n}^2 = 1$ so that $\bar{n}$ is a spacelike unit vector. Let us rename it to $\hat{n}$ so that

$$\bar{N} = \kappa(\hat{U} + \hat{n}). \tag{6.12}$$

Let us assume that a vector $\bar{K}$ is orthogonal to $\bar{N}$ and investigate what this implies for $\bar{K}$. Splitting $\bar{K}$ we write

$$\bar{K} = \kappa'(\hat{U} + \bar{k}) \tag{6.13}$$

with $\bar{k} \cdot \hat{U} = 0$. The assumed orthogonality between $\bar{K}$ and $\bar{N}$ gives

$$0 = \bar{K} \cdot \bar{N} = \kappa\kappa'(-1 + \bar{k} \cdot \hat{n}). \tag{6.14}$$

There are two possibilities here. One of them is that $\kappa' = 0$, the other one is that $\bar{k} \cdot \hat{n} = 1$. Remembering that $\bar{k}$ and $\hat{n}$ are vectors in the same three-dimensional Euclidean space — the orthogonal space to $\hat{U}$ — this latter condition can be written

$$|k| = 1/(\hat{k} \cdot \hat{n}) = 1/\cos\theta. \tag{6.15}$$

where $|k|$ is the norm of $\bar{k}$ and θ is the angle between $\bar{k}$ and $\hat{n}$. Now $\cos\theta \leq 1$ so that $|k| \geq 1$. From (6.13) then follows that $\bar{K}^2 \geq 0$. Thus

$\bar{K}$ is spacelike or null-like. The latter case corresponds to the equality sign, i.e., $\cos\theta = 1$, $|k| = 1$, which implies that $\bar{k} = \hat{n}$. From (6.12) and (6.13) follows that in this case $\bar{K}$ and $\bar{N}$ are parallel. Summing up

If a non-vanishing vector $\bar{K}$ is orthogonal to a null-like vector $\bar{N}$ there are two possibilities

- *$\bar{K}$ is spacelike*
- *$\bar{K}$ is parallel to $\bar{N}$*

A corollary to this rule is that *two orthogonal null-like vectors are parallel.* Both the main rule and the corollary are of great practical importance and well worth to memorize.

6.4 The Future and the Past

Let us place a timelike vector $\bar{V}$ with its foot at the origin of Figure 6.1. Split the vector with respect to the four-velocity $\hat{U}$ of some observer as in (4.9)

$$\bar{V} = t\hat{U} + \bar{r} \quad \text{with} \quad \hat{U}\cdot\bar{r} = 0. \tag{6.16}$$

Square this and you obtain

$$\bar{r}^2 - t^2 = \bar{V}^2 < 0. \tag{6.17}$$

We say that *every timelike vector is pointing inside the light cone* and vice versa. It may point either to the future or to the past. Replacing $\bar{V}$ with a spacelike vector we find that *every spacelike vector is pointing outside the light cone* and vice versa. Finally, *every null-like vector is* (of course) *pointing along the light cone.*

Consider now two timelike vectors $\bar{U}$ and $\bar{V}$ with a negative scalar product

$$\bar{U}^2 < 0, \quad \bar{V}^2 < 0, \quad \bar{U}\cdot\bar{V} < 0. \tag{6.18}$$

From these vectors we construct

$$\bar{W} = u\bar{U} + v\bar{V} \quad \text{with} \quad u \geq 0, \quad v \geq 0, \quad \text{and} \quad u + v \neq 0, \tag{6.19}$$

which gives

$$\bar{W}^2 = u^2\bar{U}^2 + v^2\bar{V}^2 + 2uv\bar{U}\cdot\bar{V} < 0. \tag{6.20}$$

By varying u and v in (6.19) we can thus continuously transfer the vector $\bar{U}$ to the vector $\bar{V}$ through solely timelike vectors. This could not have been true if one of the vectors were pointing in the future light cone and the other one in the past light cone. Hence *two timelike vectors with a negative scalar product are both pointing in the same light cone.* It is interesting to note that the proof and the result goes through unchanged if one vector is timelike and the other one null-like.

If two timelike vectors $\bar{U}$ and $\bar{V}$ have a positive scalar product the vectors $\bar{U}$ and $-\bar{V}$ are timelike and have a negative scalar product. They are therefore pointing in the same light cone, from which follows that $\bar{U}$ and $\bar{V}$ are pointing in different light cones. From this follows that two timelike vectors pointing in the same light cone have a scalar product which is either negative or zero. It cannot, however, be zero due to the final rule in Section 3.1. Thus we have found that the above theorem has a converse: *Two timelike vectors pointing in the same light cone have a negative scalar product.* Also this proof and result goes through unchanged if one vector is timelike and the other one null-like.

From this theorem we can draw the conclusion that the sum of two space ship vectors is another (possible) space ship vector: Suppose that $\bar{U}$ and $\bar{V}$ are both timelike and future directed. Then

$$(\bar{U}+\bar{V})^2 = \bar{U}^2 + \bar{V}^2 + 2\bar{U}\cdot\bar{V} < 0 \tag{6.21}$$

and

$$(\bar{U}+\bar{V})\cdot\bar{V} = \bar{U}\cdot\bar{V} + \bar{V}^2 < 0. \tag{6.22}$$

So obviously $\bar{U}+\bar{V}$ is timelike and future directed. Q.E.D.

Any arbitrary sum of space ship vectors must therefore be timelike and future directed. No ship can therefore travel into the past and arrive before it has started (see Figure 6.2). If two space ships part and meet they always agree that they part before they meet. No matter how a space ship tries it can never break through the light cone.[1] The

[1] It is obvious also from another argument that no observer can reach the velocity of light: A null-like signal cannot be at rest with respect to any observer because such a signal has the unit velocity with respect to every observer.

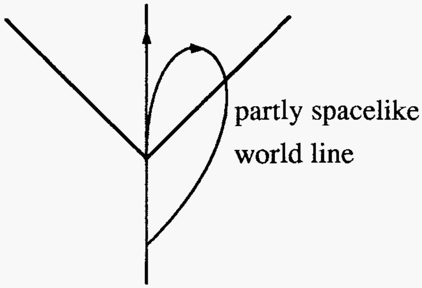

Figure 6.2: It is forbidden for a space ship to break through the light cone. It can therefore not arrive before it has departed.

science fiction dream of a time machine travelling backwards in time can therefore not be realized within the frame work of special relativity.

Chapter 7

Two-Dimensional Spacetime

The present chapter is devoted to situations in which all interesting spacetime vectors lie in a two-dimensional subspace of spacetime. This means that all interesting vectors can be written as linear combinations of two linearly independent vectors. Many text book problems fall in this category and a two-dimensional problem is often a part of a more complicated problem as in Euclidean geometry.

Our interest is now concentrated on cases in which the subspace we consider contains at least one timelike future directed unit vector. You can then regard this vector as the four velocity of an observer. Every vector can be split with respect to this four velocity and described by the observer in terms of a time t and a distance r from the observer. Hence the situation corresponds to a one-dimensional situation in Newtonian physics.

7.1 Lorentz Transformation

Consider the situation in Euclidean space which is illustrated in Figure 7.1 with two pairs of orthogonal unit vectors rotated with respect to one another. The corresponding situation occurs frequently in spacetime.

It is characterized by the following equations for the four spacetime vectors $\hat{U}, \hat{r}$ and $\hat{V}, \hat{s}$

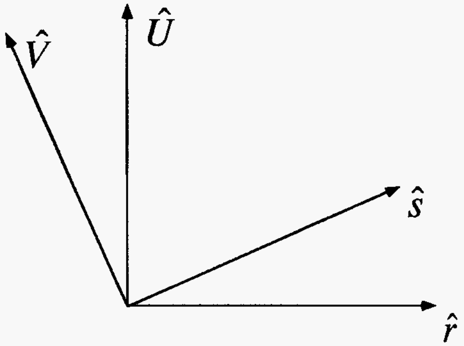

Figure 7.1: Rotation of a pair of orthogonal vectors in Euclidean space.

$$\hat{U}^2 = \hat{V}^2 = -1 \tag{7.1}$$

$$\hat{r}^2 = \hat{s}^2 = 1 \tag{7.2}$$

$$\hat{U} \cdot \hat{r} = \hat{V} \cdot \hat{s} = 0. \tag{7.3}$$

Obviously $\hat{U}$ and $\hat{r}$ are linearly independent because one of them is timelike and the other one spacelike. Hence $\hat{V}$ can be written as a linear combination of them so that $\hat{V}$ is proportional to say $\hat{U} + \nu\hat{r}$. The vector $\hat{s}$ must then be proportional to $\hat{r} + \nu\hat{U}$ to be orthogonal to $\hat{V}$. Now

$$-\left(\hat{U} + \nu\hat{r}\right)^2 = \left(\hat{r} + \nu\hat{U}\right)^2 = 1 - \nu^2 \tag{7.4}$$

Therefore to satisfy $\hat{V}^2 = -1$ and $\hat{s}^2 = 1$ the following relations must be valid

$$\hat{V} = \pm\frac{1}{\sqrt{1-\nu^2}}(\hat{U} + \nu\hat{r}) \tag{7.5}$$

$$\hat{s} = \pm\frac{1}{\sqrt{1-\nu^2}}(\hat{r} + \nu\hat{U}). \tag{7.6}$$

Let us now further demand that $\hat{V} \to \hat{U}$ and $\hat{s} \to \hat{r}$ when $\nu \to 0$ (cf. the Euclidean case in Figure 7.1). We must then choose the plus signs in (7.5) and (7.6). Comparing (7.5) with (6.2) we see that $|\nu|$ is the absolute value v of the velocity between $\hat{U}$ and $\hat{V}$, i.e., $|\nu| = v$.

Furthermore $\nu\hat{r}$ is the velocity of $\hat{V}$ with respect to $\hat{U}$, i.e., $\bar{v} = \nu\hat{r} = \pm v\hat{r}$. Thus

$$\hat{V} = \gamma(\hat{U} + \nu\hat{r}) \tag{7.7}$$

$$\hat{s} = \gamma(\hat{r} + \nu\hat{U}) \tag{7.8}$$

where γ is given by (6.5). The equations (7.7) and (7.8) form together the famous *Lorentz transformation.* In the present case when the vector $\hat{r}$ (and $\hat{s}$) is a part of the situation it is practical to use the velocity ν which can be both positive and negative while the absolute value v by definition cannot be negative. When ν is positive the velocity of $\hat{V}$ in relation to $\hat{U}$ is in the direction $\hat{r}$. When ν is negative the velocity is in the opposite direction. You should compare with the Euclidean situation where the rotation can either be such that $\hat{V}$ is leaning towards $\hat{r}$ or such that $\hat{V}$ is leaning away from $\hat{r}$. In Figure 7.1 the latter is the case.

Sometimes you are interested in splitting *one and the same vector* $\bar{A}$ with respect to two different observers

$$\bar{A} = t\hat{U} + x\hat{r} \tag{7.9}$$

and

$$\bar{A} = t'\hat{V} + x'\hat{s}. \tag{7.10}$$

Introducing (7.7) and (7.8) into the second equation and equating the two expressions for $\bar{A}$ you obtain

$$[t - \gamma(t' + \nu x')]\hat{U} + [x - \gamma(x' + \nu t')]\hat{r} = 0. \tag{7.11}$$

As $\hat{U}$ and $\hat{r}$ are linearly independent this implies

$$t = \gamma(t' + \nu x') \tag{7.12}$$

$$x = \gamma(x' + \nu t') \tag{7.13}$$

This relation between the measurements of two different observers is also referred to as a *Lorentz transformation.* To distinguish between the two types of Lorentz transformation (7.7), (7.8) is named the *active* and (7.12), (7.13) the *passive* Lorentz transformation. The latter one is not going to be used very much in this book and hence there is no real need to make the distinction in the following.

7.2 Addition of Velocities

If you, in Euclidean space, perform first one rotation and then another one the result is a third rotation. Similarly you can perform two consecutive Lorentz transformations with parameters ν_1 and ν_2. It should be obvious from Section 7.1 that the result is a Lorentz transformation and the only remaining question is what its parameter ν is expressed in ν_1 and ν_2. The following is a short argument giving the answer.

Using the Lorentz transformation (7.7) and (7.8) twice starting from $\hat{U}, \hat{r}$ we find that the final timelike unit vector is *proportional to*

$$(\hat{U} + \nu_1 \hat{r}) + \nu_2(\hat{r} + \nu_1 \hat{U}) = (1 + \nu_1 \nu_2)\hat{U} + (\nu_1 + \nu_2)\hat{r}. \tag{7.14}$$

We can here apply a simple reasoning which often is helpful: Looking at (7.7) we realize that *the velocity of a Lorentz transformation can be found as the ratio between the second and first coefficients in the expression for the resulting timelike unit vector.* Returning to our above expression we thus find that the resulting velocity is

$$\nu = (\nu_1 + \nu_2)/(1 + \nu_1 \nu_2). \tag{7.15}$$

This is the relativistic formula for *addition of velocities*. It has been tested in quite a direct way through experiments with light propagation in running liquids.

Choosing $\nu_2 = -\nu_1$ in (7.15) we find $\nu = 0$. Thus we obtain the *inverse* Lorentz transformation by changing the sign of the velocity parameter.

If you like you could quite easily combine two Lorentz transformations in all detail. It is in fact a good idea to do so to confirm that it all works as we have stated here. In case you do this keep the relation (6.5) in mind. Something else you could do while you are at it is to solve (7.7) and (7.8) for $\hat{U}$ and $\hat{r}$ obtaining the inverse of (7.7), (7.8). You will find the expected result that the inverse transformation has the same form as the original one with only a change of sign of the velocity ν.

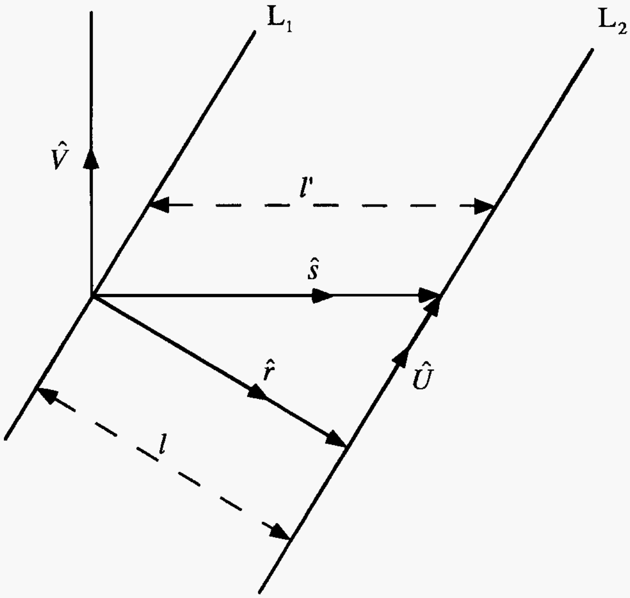

Figure 7.2: The history of a space ship lies between the world lines L_1 and L_2. The observers with four-velocities $\hat{U}$ and $\hat{V}$ each measures the length of the ship orthogonally to his own four-velocity.

7.3 Lorentz Contraction

Figure 7.2 illustrates a space ship history which fills the part of space-time between the world lines L_1 and L_2. These two lines are the histories of the stem and the stern, respectively. There is an observer travelling with the ship having the ship's four-velocity $\hat{U}$, but there is also another observer and he has the four-velocity $\hat{V}$. Each observer measures from his point of view the length of the ship from stem to stern. Both the vector $l\hat{r}$ and the vector $l'\hat{s}$ connect the stem with the stern. The first one is orthogonal to $\hat{U}$ and the latter one to $\hat{V}$. Hence the first observer measures the length l and the second one the length l'. The difference between the two vectors is obviously in the direction $\hat{U}$

$$l'\hat{s} - l\hat{r} \propto \hat{U}. \tag{7.16}$$

Multiplying this by $\hat{r}$ we obtain

$$l'\hat{r} \cdot \hat{s} - l = 0. \tag{7.17}$$

We recognize the arrangements of our vectors $\hat{U}, \hat{r}$ and $\hat{V}, \hat{s}$ from Section 7.1 and we may use (7.8). Multiplying this equation by $\hat{r}$ we obtain $\hat{r} \cdot \hat{s} = \gamma$ so that

$$l'/l = 1/\gamma = \sqrt{1 - v^2} \tag{7.18}$$

where v is the velocity between the two observers. This is the equation for *Lorentz contraction.* You may notice from Figure 7.2 that a very similar effect occurs when you slice a sausage: The size of the slice depends on the cutting angle. In that case the perpendicular cut gives the *shortest* slice while the observer at rest in the space ship measures the *greatest* length of the ship: $l' < l$.

Chapter 8

Plane Waves

8.1 The Wave Four-Vector

Sound waves are well known from Newtonian physics. In a sound wave the pressure P (e.g., the air pressure) depends on both space and time. It varies about the mean pressure P_0. In a *plane* sound wave the difference $p \equiv P - P_0$ has the following form

$$p = A \sin(-\omega t + \bar{k} \cdot \bar{r} + const) \tag{8.1}$$

where A and ω are constant numbers, $\bar{k}$ is a constant vector, t is the time and $\bar{r}$ is the position vector in space. Without restricting the physical situation we may assume *that ω is positive.* A wave with negative ω could obviously always be rewritten as a wave with positive ω.

Sound waves is just an example. Many other quantities than pressure can form plane wave patterns. Turning now to spacetime we consider *some* quantity which we denote ψ. On the basis of our example we say *by definition* that ψ forms a plane wave if

$$\psi = A \sin(\bar{K} \cdot \bar{R} + F_0) \tag{8.2}$$

where A and F_0 are constant numbers, $\bar{K}$ is a constant vector and $\bar{R}$ is the position vector in spacetime. The number A is named the *amplitude* and the vector $\bar{K}$ is named *the wave four-vector.* We shall see that this spacetime definition of plane waves leads to the 'old' expression (8.1) when we introduce an observer. Let the observer have the four-velocity $\hat{U}$ and split $\bar{R}$ and $\bar{K}$

$$\bar{R} = t\hat{U} + \bar{r} \quad \text{with} \quad \bar{r} \cdot \hat{U} = 0 \tag{8.3}$$

$$\bar{K} = \omega\hat{U} + \bar{k} \quad \text{with} \quad \bar{k} \cdot \hat{U} = 0. \tag{8.4}$$

Inserting this into (8.2) gives

$$\psi = A \sin(-\omega t + \bar{k} \cdot \bar{r} + F_0) \tag{8.5}$$

which obviously has the form (8.1). The number ω is the *angular frequency* and the vector $\bar{k}$ is the *wave vector*. It is important to notice that both of them *require an observer to be defined.* In that sense they are *not* inherent properties of the wave. We shall sometimes but not always stress the observer dependence of ω and $\bar{k}$ by putting an index on them writing ω_U and $\bar{k}_U$.

Multiplying (8.4) by $\hat{U}$ we obtain

$$\omega_U = -\bar{K} \cdot \hat{U} \tag{8.6}$$

which is an equation worth to remember.

Much of the usefulness of plane waves derives from *the principle of superposition* which states that the quantities A and F_0 can be chosen arbitrarily independently of $\bar{K}$ and that arbitrary sums of plane waves are legitimate waves. This will be used in the next section (Section 8.2).

8.2 Modulations

A plane wave constitutes quite a special situation. Generally a wave is a superposition of plane waves, so that Ψ is a sum of terms like (8.2) with different A, $\bar{K}$, and F_0 in different terms. A certain observer with four-velocity $\hat{V}$ does in general measure different angular frequencies ω_V for the different plane waves. A wave in which the occurring $\bar{K}$ vectors are all nearly the same is usually referred to as a *modulated* wave. Let us consider such a wave and ask ourselves if we can find an observer such that he measures *the same frequency for all the constituent plane* *waves.* With only one frequency present the time variation is just a harmonic oscillation and is thus of the form $\Psi = A(\bar{r})\cos[-\omega t + \phi(\bar{r})]$. The occurring $\bar{k}$-vectors are all nearly the same. Let us pick one of them, denote it $\bar{k}_0$, and write

$$\Psi = A(\bar{r})\cos[-\omega t + \bar{k}_0 \cdot \bar{r} + F(\bar{r})]. \tag{8.7}$$

The functions $A(\bar{r})$ and $F(\bar{r})$ represent *time independent* modulation patterns imprinted on a plane wave.

The condition that all frequencies be the same can be written $d\omega = 0$, where $d\omega$ stands for the difference in angular frequency of two arbitrary specimens among the plane waves. Assuming that the observer has the four-velocity $\hat{W}$ this takes the form

$$d(\hat{W} \cdot \bar{K}) = \hat{W} \cdot d\bar{K} = 0 \tag{8.8}$$

(see equation (8.6)). This must be true for all $d\bar{K}$ within the set of waves we consider. With respect to any four velocity other than $\hat{W}$ the modulation pattern is *moving* with a velocity known as the *group velocity*. The four-velocity $\hat{W}$ satisfying (8.8) is named the *the group four-velocity.*

Let us investigate the velocity of $\hat{W}$ with respect to some observer with the four-velocity $\hat{U}$. To that end we perform the splitting (6.2)

$$\hat{W} = \gamma(\hat{U} + \bar{w}) \quad \text{with} \quad \bar{w} \cdot \hat{U} = 0 \tag{8.9}$$

and the splitting (8.4) which gives

$$d\bar{K} = d\omega\hat{U} + d\bar{k} \quad \text{with} \quad d\bar{k} \cdot \hat{U} = 0 \tag{8.10}$$

where the angular frequency ω refers to the four-velocity $\hat{U}$, not to $\hat{W}$ as above. Taking the scalar product of (8.9) and (8.10) and using (8.8) you obtain

$$d\omega = \bar{w} \cdot d\bar{k}. \tag{8.11}$$

This is an equation in a Euclidean space and can be treated as such. Writing $\bar{k} = k\hat{k}$ (8.11) runs

$$d\omega = (\bar{w} \cdot \hat{k})dk + k\bar{w} \cdot d\hat{k}. \tag{8.12}$$

Assuming now a *dispersion equation*, i.e., an equation which gives ω as a function of k (but not involving $\hat{k}$) we can by choosing $dk = 0$ in (8.12) obtain $\bar{w} \cdot d\hat{k} = 0$, where $d\hat{k}$ can have an arbitrary direction perpendicular to $\hat{k}$. (Remember that $\hat{k}^2 = 1$ leads to $\hat{k} \cdot d\hat{k} = 0$.) Hence we can draw the conclusion that $\bar{w}$ is parallel to $\hat{k}$. Choosing $d\hat{k} = 0$ in (8.12) we find $d\omega = |\bar{w}|dk$. In summary

$$\bar{w} = (d\omega/dk)\hat{k}. \tag{8.13}$$

Here you recognize the elementary formula for the group velocity of the wave.

Modulations 'mark' the wave with moving patterns which provide a means of signalling. If a wave is used for signalling it should therefore posses a group four-velocity. According to (5.8) the defining equation (8.8) would, however, run into a contradiction if $d\bar{K}$ could have an arbitrary direction in spacetime (for a given modulated wave) because this would imply $\hat{W} = 0$, which of course is absurd. Hence there must be a restriction on $d\bar{K}$. Such a restriction arises from the differentiation of a condition on $\bar{K}$. For waves in matter such a condition is usually called a *dispersion equation* (cf. a few lines above). For waves in vacuum the freedom to construct conditions on $\bar{K}$ is much limited. Without the disposal of a matter four-velocity the only form in which $\bar{K}$ can enter the condition is the square $\bar{K}^2$. Due to the principle of superposition (see page 64) A and F_0 should not enter into the condition so that the only possibility is that $\bar{K}^2$ is a constant

$$\bar{K}^2 = -m^2 \tag{8.14}$$

where m is a given number. (It will soon be clear why we choose a negative constant.) Differentiate this and you obtain

$$\bar{K} \cdot d\bar{K} = 0 \tag{8.15}$$

which tells you that $d\bar{K}$ is restricted to be orthogonal to $\bar{K}$. In fact $d\bar{K}$ can have an arbitrary direction orthogonal to $\bar{K}$, so that (8.8) and (8.14) give $\hat{W}$ uniquely as

$$m\hat{W} = \bar{K}. \tag{8.16}$$

Notice that $\hat{W}$ is timelike (as it should) due to our choice of $\bar{K}$ as timelike in (8.14). Without any physical restriction we may assume that $\bar{K}$ is pointing towards the future (cf. the discussion of sound waves on page 63). With this choice m has to be positive because $\hat{W}$, being a four-velocity, must point towards the future.

The case in which the world lines of signals are null-like falls, strictly speaking, outside the conditions for the present discussion as $\hat{W}$ doesn't exist in this case. Introducing some arbitrary four-velocity $\hat{U}$ and requiring that $E = -\hat{U} \cdot \bar{K}$ be finite also when $\bar{K}$ is null-like we draw the conclusion from $-m\gamma = m\hat{W} \cdot \hat{U} = -E$ that $m \to 0$ in the limit for a null-like $\bar{K}$ because in this limit $\gamma \to \infty$. This limiting case includes light signals and is very important. We shall therefore not exclude the possibility that $\bar{K}$ is null-like so that m vanishes.

Signalling is actually not typically done with plane (modulated) waves. More typical is that the sender emits a spherical modulated wave which propagates in all (space) directions. If it is a light wave it propagates with light velocity. In fact each part of it which is small compared to the distance from the point of emission behaves very much like a plane wave. It is now obvious that independently of where the receiver is situated the position vector in spacetime between events of emission and reception of a certain light signal (a certain 'bump' on the light wave) is null-like.

So far we have not mentioned the *phase velocity* of waves. The reason is that there is no general definition of a well defined and observer independent phase velocity in relativity. In two-dimensional spacetime the observer independent equation $\hat{W}_{phase} \cdot \bar{K} = 0$ leads to the velocity ω/k if we introduce an observer but the same equation is insufficient in four dimensions as it defines the velocity only up to arbitrary vectors orthogonal to $\bar{K}$. Physically the reason for the failure is that if there is nothing to refer to except the wave you cannot find out how much you are moving along a wave plane. Another absurdity is that if $\bar{K}$ is timelike the four-velocity defined by this equation has to be spacelike.

8.3 Doppler Shift and Aberration

Let us now consider one wave and *two* observers with the four-velocities $\hat{U}$ and $\hat{V}$. Let us for simplicity assume that the wave has light velocity so that $\bar{K}^2 = 0$. We can now write down two splittings with respect to each of the observer four-velocities. Let us first choose $\hat{U}$, writing

$$\bar{K} = \omega_U(\hat{U} + \hat{k}_U) \quad \text{with} \quad \hat{k}_U \cdot \hat{U} = 0 \tag{8.17}$$

$$\hat{V} = \gamma(\hat{U} + \bar{v}_U) \quad \text{with} \quad \bar{v}_U \cdot \hat{U} = 0. \tag{8.18}$$

The particular form of the equation (8.17) is dictated by $\bar{K}^2 = 0$ which becomes obvious if you square it.

The two vectors $\bar{v}_U$ and $\hat{k}_U$ are both in the orthogonal space to $\hat{U}$ and this is just the Euclidean space constructed by the observer with the four-velocity $\hat{U}$. Thus he finds that the angle α_U between $\bar{v}_U$ and $\hat{k}_U$ is given by

$$v \cos \alpha_U = \bar{v}_U \cdot \hat{k}_U \tag{8.19}$$

where $v = |\bar{v}_U|$. Remembering (8.6) and taking the scalar product of (8.17) and (8.18) you obtain

$$-\omega_V = \bar{K} \cdot \hat{V} = \omega_U \gamma(-1 + \bar{v}_U \cdot \hat{k}_U) = -\omega_U \gamma(1 - v \cos \alpha_U). \tag{8.20}$$

Equation (8.20) accounts for the *Doppler effect.* Slightly rewritten it runs

$$\omega_V / \omega_U = (1 - v cos \alpha_U)/\sqrt{1 - v^2}. \tag{8.21}$$

This equation has many important applications. Let us just mention the broadening of spectral lines and the measuring of the velocity of stars and galaxies.

We could have chosen from the start to perform the splitting with respect to $\hat{V}$, exchanging the rôles of $\hat{U}$ and $\hat{V}$. With obvious notations this will result in the following exchanges: $\omega_U \leftrightarrow \omega_V$, $\bar{v}_U \leftrightarrow \bar{u}_V$, $\alpha_U \leftrightarrow \alpha_V$. The value of v will remain unchanged because the absolute value of the velocity of $\hat{V}$ with respect to $\hat{U}$ is the same as the absolute value of the velocity of $\hat{U}$ with respect to $\hat{V}$. Consider now the equation obtained from (8.21) by this exchange and multiply it with (8.21) thus obtaining the equation for *aberration*

$$(1 - v \cos \alpha_U)(1 - v \cos \alpha_V) = 1 - v^2. \tag{8.22}$$

We notice here that for values of v which are so small that v^2 terms may be neglected one obtains $\cos\alpha_V \approx -\cos\alpha_U$. The minus sign can be traced back to the fact that in our symmetrical treatment of the two observers they are for small velocities 'facing' each other, i.e., $\bar{u}_V \approx -\bar{v}_U$. Usually one prefers to use the angle $\tilde{\alpha}_V$ between $\hat{k}$ and $-\bar{v}_V$ rather than α_V. As $\cos\tilde{\alpha}_V = -\cos\alpha_V$ we obtain

$$(1 - v\cos\alpha_U)(1 + v\cos\tilde{\alpha}_V) = 1 - v^2. \tag{8.23}$$

Solving (8.23) for $\cos\alpha_V$ it is transformed into

$$\cos\tilde{\alpha}_V = (\cos\alpha_U - v)/(1 - v\cos\alpha_U) \tag{8.24}$$

which is the form in which aberration is presented in most text books. If you like to memorize an aberration formula I would, however, recommend that you rather try (8.22) or (8.23) which are so much more symmetric and therefore easier to remember.

Aberration has for a long time been a well known effect in astronomy, where it leads to a seasonal dependence of the angular separation between objects in the sky.

Chapter 9

Particle Reactions

9.1 Four-Momentum

In equation (6.2), i.e.,

$$\hat{V} = \gamma(\hat{U} + \bar{v}) \text{ with } \bar{v} \cdot \hat{U} = 0 \tag{9.1}$$

we have split a four-velocity $\hat{V}$ with respect to a four-velocity $\hat{U}$. Here $\hat{V}$ could be the four-velocity of some physical body and $\hat{U}$ could be the four-velocity of an observer. Then $\bar{v}$ is the velocity of the body relative to the observer.

Let us now consider the *Newtonian limit* by which we mean the case with $v \equiv |\bar{v}| \ll 1$. Equivalently the Newtonian limit can be defined as the case in which $\hat{U}$ and $\hat{V}$ are nearly parallel. Expanding γ in a Taylor series and keeping terms only up to second order in v we obtain

$$\hat{V} \approx (1 + \frac{1}{2}v^2)\hat{U} + \bar{v}, \tag{9.2}$$

where we have used the expression (6.5) for γ.

For small velocities the classical Newtonian mechanics should be valid. Among other things each physical body has a certain *mass*, which tells us how inert the body is to forces acting on it. Multiplying (9.2) by the mass m of the body we obtain

$$m\hat{V} \approx (m + \frac{1}{2}mv^2)\hat{U} + m\bar{v}. \tag{9.3}$$

On the right hand side appears the kinetic energy $(1/2)mv^2$ and the momentum $m\bar{v}$ of the body.

True enough the mass of the body is defined in the Newtonian limit, but any body whatever its four-momentum can be regarded in the Newtonian limit. You just have to *choose an observer whose four-momentum is nearly parallel to the four-momentum of the body.* Thus every body has a well defined mass with a value which is characteristic for that body. Using the mass we define the *four-momentum* $\bar{P}$ of a body as

$$\bar{P} \equiv m\hat{V}, \tag{9.4}$$

which gives

$$\bar{P}^2 = -m^2. \tag{9.5}$$

The equations (9.4) and (9.1) give the following split of $\bar{P}$ with respect to $\hat{U}$

$$\bar{P} = E\hat{U} + \bar{p}, \tag{9.6}$$

where

$$E = m\gamma = m/\sqrt{1-v^2} \approx m + \frac{1}{2}mv^2 \tag{9.7}$$

and

$$\bar{p} = m\bar{v}\gamma = m\bar{v}/\sqrt{1-v^2} \approx m\bar{v}. \tag{9.8}$$

The quantity E is called the *energy* and the four-vector $\bar{p}$ is called the *momentum*. It is very important to keep in mind that unlike the mass, the energy and the momentum are not properties of the body itself. They are not given by the body's own four-momentum $\bar{P}$ alone. Their specification requires an extra four-velocity $\hat{U}$ often said to be the four-velocity of the observer. Thus different observers find different energy and momentum with one and the same body at one and the same occasion. This is actually not very surprising as the same thing happens in Newtonian mechanics.

The Newtonian mass is positive. Hence (9.4) tells us that $\bar{P}$ is time-like and future directed. A neat expression for the energy is obtained by taking the scalar product of (9.6) with $\hat{U}$

$$E = -\bar{P} \cdot \hat{U}. \tag{9.9}$$

This equation shows that the *energy is positive.*

The usefulness of momentum in Newtonian mechanics derives from the fact that it is *conserved.* By this is meant that the sum of momenta is the same before and after a collision between free bodies resulting in new free bodies. This is a part of a rule stating that *four-momentum is conserved,* which follows from basic physical principles and homogeneity of spacetime. The proof requires concepts and tools which are not within the framework of this course.

Conservation of four-momentum implies conservation of energy which is remarkable because in the Newtonian limit the energy is $E \approx \Sigma(m + \frac{1}{2}mv^2)$ where the sum is extended over all parts of the system considered (see (9.7)). The second term here is the kinetic energy and we know that the kinetic energy is *not* generally conserved. In the collision of macroscopic bodies the total kinetic energy usually decreases. The only way out of this situation is to admit that the first term changes, i.e., that *the sum of the masses can change* in a reaction contrary to what is assumed in Newtonian physics. In an inelastic collision the total kinetic energy decreases and the partaking bodies are heated. As the energy must remain the same we must draw the conclusion that if a body is heated its Newtonian mass increases.

In every day situations $v \ll 1$ so that the kinetic energy is much less than the 'mass energy': $1/2mv^2 \ll m$. Nevertheless the *changes* in these terms for a certain body are of the same order and only the changes are considered in Newtonian mechanics. This is the reason why we cannot be satisfied with keeping only the first term 'm' in the expansion (9.7). Another consequence is that m changes relatively speaking very little, which explains why it took so long before one could see a change in the sum of masses experimentally.

In a gas the molecules can with good approximation be regarded as free from one another. Therefore the total four-momentum $\bar{P}$ is simply the sum of all the four-momenta of the molecules. The 'Newtonian' limit as we have defined it here is obtained by choosing an observer with four-velocity parallel to this total four-momentum. This is true even if some of the molecules happen to have high velocities with respect to this observer. If he starts to 'push' the gas he will notice that its inertial mass m is given by $m^2 = -\bar{P}^2$. This mass is generally *not* the sum of the molecular masses but given as the *sum of the molecular energies* as measured by him, i.e., the *rest energy* of the gas.

What we have just discussed is the content of one of the most famous equations in physics, $E = mc^2$. A few comments are in order. First a reminder that we have chosen units so that the 'velocity of light' is unity, i.e., $c = 1$. Second that $E = m$ is true only for the observer who (momentarily) is at rest with respect to the body considered. This is obvious from (9.7). Actually we should therefore write $E_{\mathrm{rest}} = m$.

Some older textbooks introduce different kinds of masses as transverse and longitudinal mass and use the term 'rest mass' for the Newtonian mass used here. I don't find this a very good practice and we are not going to follow it in this course.

If you know the four-momenta of all but one of the particles occurring in some reaction you may use conservation of four-momentum to calculate the last one. Using this method one has found (with a certain experimental uncertainty) that particles exist for which the four-momentum square vanishes, $\bar{P}^2 = 0$. Defining the mass through $m^2 = -\bar{P}^2$ one has thus found *massless particles*. Examples are photons and neutrinos. Their four-momenta do not vanish. Otherwise the particles would not make themselves known to us in the four-momentum balance. This means that according to (9.9) and the rule at the end of Section 3.1 the energy is nonzero (and in fact positive). Choosing any observer we then find from (9.7) that $m = 0$ implies $v = 1$, which tells us that a massless particle follows a null-like line in spacetime.

9.2 Particle Kinematics

A particle reaction can either be a *decay* with one initial particle or a *collision* with two initial particles. In both cases there can be any number of final particles. Microscopic systems have discrete rest energies or masses. In the case of elementary particles, states with different masses have obtained different names. Therefore when we speak of an electron or a muon we refer to a state with fixed mass. Hence in a reaction involving given elementary particles the four-momenta satisfy

$$\bar{P}_r^2 = -m_r{}^2 \tag{9.10}$$

where index r goes from 1 to N where N is the number of particles involved in the reaction.

A further general restriction on the four-momenta is the conservation law from last section (Section 9.1)

$$\sum_r \bar{P}_{r,\text{in}} = \sum_s \bar{P}_{s,\text{out}}. \tag{9.11}$$

The equations (9.10) and (9.11) form the basis for *particle kinematics* as far as it will be treated in this course. In the end predictions and experimental results must be given as numbers. The numbers in particle kinematics are the scalar products between the four-momenta of the particles partaking in the reaction studied and the aim of particle kinematics is to find relations between these products caused by (9.10) and (9.11).

We are frequently interested in *interpreting* the scalar products with respect to some observer with say the four-velocity $\hat{U}$. What we do then is to split the four-momenta according to (9.6)

$$\bar{P}_r = E_r\hat{U} + \bar{p}_r. \tag{9.12}$$

This gives

$$\bar{P}_r \cdot \bar{P}_s = -E_r E_s + \bar{p}_r \cdot \bar{p}_s. \tag{9.13}$$

Thus we see that the scalar product on the left hand side can be written in terms of energies and a scalar product between momenta, all referring to the same observer. In a special case of (9.13) the indices r and s refer to the same particle

$$m_r{}^2 = -\bar{P}_r^2 = E_r{}^2 - \bar{p}_r^2. \tag{9.14}$$

A typical kinematical calculation involves the following two steps

- Either take the scalar product of (9.11) with particle four-momenta or rearrange it and square.
- Replace all squares of four-momenta with minus mass squares. Interpret the rest of the scalar products according to (9.13). If you like you can very well use different observers in the interpretation of the different scalar products. Only be careful to mark what observers the different energies and momenta refer to.

Let us illustrate this by considering the reaction in which two particles with four-momenta $\bar{P}$ and $\bar{Q}$ collide and form two particles with four-momenta $\bar{P}'$ and $\bar{Q}'$. — The decay of one particle into three particles can be treated in essentially the same way.— To be still more specific let us assume that the reaction is *elastic*. By this we mean that the particles do not change their masses in the collision, i.e.,

$$\bar{P}^2 = \bar{P}'^2 \equiv -M^2 \;\; ; \;\; \bar{Q}^2 = \bar{Q}'^2 \equiv -m^2. \tag{9.15}$$

The conservation of four-momentum runs

$$\bar{P} + \bar{Q} = \bar{P}' + \bar{Q}'. \tag{9.16}$$

Let us assume that this reaction takes place in a laboratory where the first particle is at rest and that we are interested in the angle measured in the laboratory between the momenta of the second and third particles (see Figure 9.1). The best strategy in this common situation is to rearrange (9.16) so that the four-momentum of the particle we have *not* mentioned (i.e., the fourth particle) stands alone on one side of the equation

$$\bar{P} + \bar{Q} - \bar{P}' = \bar{Q}'. \tag{9.17}$$

Squaring this we obtain

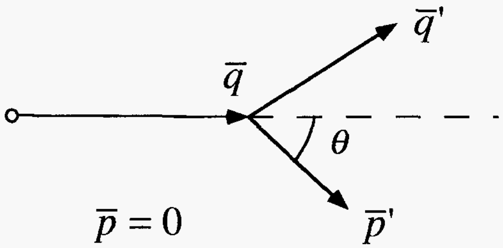

Figure 9.1: Figure illustrating the behaviour of particle momenta in a two-particles-in two-particles-out collision. The observer is assumed to be 'sitting on' particle 1 and all momenta refer to this observer. The figure is drawn in his orthogonal space and the momentum of particle 1 consequently vanishes. In the case that the collision is elastic the angle θ is the recoil angle.

$$\bar{P}^2 + \bar{Q}^2 + \bar{P}'^2 + 2\bar{P}\cdot\bar{Q} - 2\bar{P}\cdot\bar{P}' - 2\bar{Q}\cdot\bar{P}' = \bar{Q}'^2. \tag{9.18}$$

Here all four-momenta squared are to be replaced by minus the corresponding mass squares. The remaining scalar products are to be interpreted using (9.13) assuming an observer with four-velocity parallel to $\bar{P}$. We will, however, vary the procedure slightly and first simplify (9.18) using (9.15). A short calculation gives

$$(\bar{P} + \bar{Q})\cdot(\bar{P} - \bar{P}') = 0. \tag{9.19}$$

This is quite a useful equation valid in the case of two-particle elastic collisions.

Now interpret (9.19) assuming that the observer four-velocity is parallel to $\bar{P}$ so that $\bar{P} = M\hat{U}$, $\bar{Q} = E\hat{U} + \bar{q}$, and $\bar{P}' = E'\hat{U} + \bar{p}'$ where both $\bar{q}$ and $\bar{p}'$ are orthogonal to $\hat{U}$. Inserting this into (9.19) gives

$$-(M + E)(M - E') - \bar{q}\cdot\bar{p}' = 0. \tag{9.20}$$

The two momenta $\bar{q}$ and $\bar{p}'$ are both orthogonal to $\hat{U}$. Hence $\bar{q}$ and $\bar{p}'$ can be treated as 'ordinary' three-dimensional vectors in the orthogonal space to $\hat{U}$ so that we may write

$$\bar{q} \cdot \bar{p}' = qp' \cos\theta, \tag{9.21}$$

where θ is the angle we mentioned in the text below Equation (9.16) (see also Figure 9.1). The norms q and p' can be expressed as functions of E and E' respectively using (9.14). Thus (9.20) gives the angle θ as a function of the two energies E and E'. The result is

$$\cos\theta = \frac{E+M}{\sqrt{E^2-m^2}}\sqrt{\frac{E'-M}{E'+M}}. \tag{9.22}$$

In the treatment of more complicated processes it is often practical to regard all or a part of the out-particles as one system. The four-momentum of such a system is

$$\bar{P} = \sum_{r=1}^{N} \bar{P}_r \tag{9.23}$$

with one term for each constituent particle. The mass defined by $M^2 = -\bar{P}^2$ does not have a fixed value for such a system but it has a lower limit. Remember from Section 6.4 that the sum of timelike future directed vectors is itself such a vector. Hence there exists a four-velocity parallel to $\bar{P}$. The easiest way to find the limit is to introduce an observer with this four-velocity. Then

$$\bar{P}^2 = -E^2 = -(\sum_{r=1}^{N} E_r)^2 = -(\sum_{r=1}^{N} \sqrt{{m_r}^2 + \bar{p}_r^2}\,)^2. \tag{9.24}$$

From this it is obvious that

$$-(\sum_{r=1}^{N} \bar{P}_r)^2 \geq (\sum_{r=1}^{N} m_r)^2. \tag{9.25}$$

We realize that the equality here can actually occur. The only restriction on the momenta in (9.24) is that the sum of them must vanish and this does of course not prevent us from choosing all the momenta to be zero. In this case all the four-momenta are parallel to $\bar{P}$ so that all the particles are at rest relative to each other. The inequality (9.25) has important applications. Using it one can find conditions for certain reactions to take place, so called *threshold conditions*.

Notice a striking feature of (9.25): *It contains no reference to any observer.* What has happened to the observer we introduced in the proof? The simple answer is that we never introduced an independent observer. The use of an observer was in a way only a manner of speaking. We could have carried through the proof without mentioning an observer. We could have defined $\bar{P}$ as $\bar{P} = \sum \bar{P}_r$ and $\hat{U}$ as $\hat{U} = -\bar{P}/\sqrt{\bar{P}^2}$. You can even define $\bar{p}_r$ explicitly as

$$\bar{p}_r = \bar{P}_r + (\bar{P}_r \cdot \hat{U})\hat{U}.$$

In the literature you will find a number of such imagined observers with different epithets like 'centre of mass', 'brick wall', etc, which are used purely as calculational tools. They are rarely of any help when a situation is well understood but they can assist in giving a general view of a situation in the form of a momentum diagram like Figure 9.1. Using Lorentz transformations one can pass from one such observer to another one. This has been used very much as a calculational tool especially in older literature but for the most part it is awkward and unnecessary.

Chapter 10

Curved World Lines

10.1 Four-Acceleration

In Section 1.5 we introduced straight world lines defining them as world lines for unaccelerated space ships. Particles not influenced by forces travel along such lines. Space ships running their engines and particles influenced by forces are in consequence said to travel along *curved world lines*. There is a wide spread misconception that accelerated motion cannot be treated within special relativity. It has its roots in the fact that one for practical reasons usually avoids to consider accelerated *observers*. We shall follow this practice here and assume that the observers we employ are unaccelerated unless otherwise stated.

The most natural parameter along the world line of a particle is usually the proper time τ. The position vector $\bar{R}$ of the particle in spacetime is then given as a function of τ

$$\bar{R} = \bar{R}(\tau). \tag{10.1}$$

As we can add vectors and multiply them by real numbers we may differentiate (10.1) writing

$$\frac{d\bar{R}}{d\tau} \equiv \lim_{\Delta\tau \to 0} \frac{\bar{R}(\tau + \Delta\tau) - \bar{R}(\tau)}{\Delta\tau}. \tag{10.2}$$

As we have agreed that τ is the proper time along the world line we have

$$d\tau^2 = -(d\bar{R})^2 \tag{10.3}$$

which shows that

$$\hat{V} \equiv \frac{d\bar{R}}{d\tau} \tag{10.4}$$

is a timelike unit vector. It is the *tangent* to the world line at the event $\bar{R}(\tau)$ on the line. For the special case of a straight line (on which $d\bar{R}/d\tau$ is constant) $\hat{V}$ is obviously the four-velocity. For a curved world line we *define* the four-velocity to be $\hat{V}$.

The *four-acceleration* is defined through

$$\bar{A} \equiv \frac{d\hat{V}}{d\tau} = \frac{d^2\bar{R}}{d\tau^2}. \tag{10.5}$$

Differentiate the equation $\hat{V}^2 = -1$ with respect to τ and you obtain

$$\frac{d\hat{V}}{d\tau} \cdot \hat{V} + \hat{V} \cdot \frac{d\hat{V}}{d\tau} = 0, \tag{10.6}$$

i.e.,

$$\bar{A} \cdot \hat{V} = 0. \tag{10.7}$$

This tells us that $\bar{A}$ according to the rule at the end of Section 3.1 is a spacelike vector. It is not in general a unit vector.

Let us now discuss how an accelerated world line is judged from an unaccelerated observer. We shall choose the one whose world line touches the curved world line at a certain point $\bar{R}(\tau_0)$ on the line (see Figure 10.1). We choose the origin to lie on the observer world line and perform a split with respect to the four-velocity $\hat{U}$ of the observer

$$\bar{R}(\tau) = t\hat{U} + \bar{r} \tag{10.8}$$

where $\bar{r} \cdot \hat{U} = 0$. Differentiate (10.8)

$$d\bar{R} = dt\hat{U} + d\bar{r} \tag{10.9}$$

and square

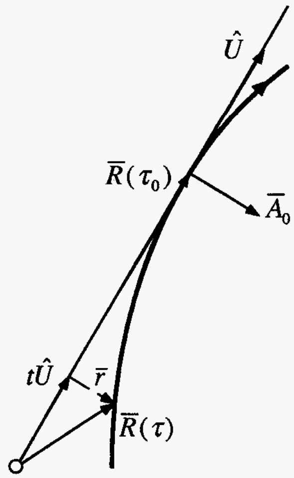

Figure 10.1: A curved world line $\bar{R} = \bar{R}(\tau)$ is measured in the standard way by an observer on the straight world line touching the curved line at some point $\bar{R}(\tau_0)$. The origin is chosen to lie on the observer world line.

$$-d\tau^2 = -(dt)^2 + (d\bar{r})^2. \tag{10.10}$$

Hence

$$d\tau = dt\sqrt{1 - v^2}. \tag{10.11}$$

This gives

$$\hat{V} \equiv \frac{d\bar{R}}{d\tau} = \frac{1}{\sqrt{1-v^2}}\frac{d\bar{R}}{dt} = \frac{1}{\sqrt{1-v^2}}\frac{d}{dt}(t\hat{U} + \bar{r}) = \frac{1}{\sqrt{1-v^2}}(\hat{U} + \bar{v}) \tag{10.12}$$

where $\bar{v} \equiv d\bar{r}/dt$.

At the point $\bar{R}(\tau_0)$ we have $\hat{V} = \hat{U}$ so that $\bar{v} = 0$. The τ-derivative of $1/\sqrt{1-v^2}$ therefore vanishes at this point and we obtain

$$\bar{A}_0 = (d\hat{V}/d\tau)_0 = (d\bar{v}/dt)_0 \equiv (\bar{a})_0 \tag{10.13}$$

where $(\bar{a})_0$ is the 'ordinary', Newtonian acceleration measured by an observer who is momentarily at rest relative to the accelerated particle. Hence if the 'particle' is a space ship the vector $\bar{A}$, at a certain point on its world line, is *the acceleration which the crew experiences at that point.*

10.2 Examples

10.2.1 Constant Acceleration

To begin as simply as possible let us consider a two-dimensional space-time curve which could be the history of a space ship travelling in one direction as seen from an observer. Only one scalar condition is required to determine such a curve and we pick the simple equation

$$(\bar{R})^2 = R^2 \tag{10.14}$$

where $\bar{R}$ is the position vector and *R is a constant.* The origin is supposed to be some fixed event.

Differentiating (10.14) we obtain

$$\bar{R} \cdot \hat{V} = 0. \tag{10.15}$$

Differentiating once more we obtain

$$\bar{R} \cdot \bar{A} = 1 \tag{10.16}$$

where $\hat{V}^2 = -1$ was used.

All the vectors $\bar{R}$, $\hat{V}$, and $\bar{A}$ lie in a 2-plane. This together with the equations (10.7) and (10.15) tells us that the vectors $\bar{R}$ and $\bar{A}$ must be parallel (or antiparallel). The equations (10.14) and (10.16) then tell us that

$$\bar{A} = \bar{R}/R^2. \tag{10.17}$$

Squaring this we obtain

$$\bar{A}^2 = 1/R^2, \tag{10.18}$$

showing that the acceleration is constant. In fact the world line we have studied could be the world line of a space ship which is running its engines so that the crew feels a constant push from the floor.

Let $\bar{R}^2 = (R_1)^2$ and $\bar{R}^2 = (R_2)^2$ be two curves like (10.14) with the same origin and in the same 2-plane. According to (10.15) any straight line which goes through the origin cuts both curves orthogonally. Therefore $R_1 - R_2$ is the orthogonal distance between the curves. In fact one of the lines could be the front world-line and the other one the tail world-line of a space ship of constant length. We see that the head and the tail must be accelerated unequally, $1/R_1$ and $1/R_2$ respectively, to keep the length of the ship unaltered.

Introducing some arbitrary observer with four-velocity $\hat{U}$ we can split the vector $\bar{R}$

$$\bar{R} = t\hat{U} + \bar{r} \text{ with } \bar{r} \cdot \hat{U} = 0 \tag{10.19}$$

and (10.14) runs

$$r^2 - t^2 = R^2. \tag{10.20}$$

This representation of the spacetime curves has been employed in the upper Figure 10.2. The fact that the curves are hyperbolae is the reason why motion under constant acceleration is sometimes called *hyperbolic motion.* Important properties of the spacetime curves are hidden in this figure. It is far from obvious from the figure that all points on such a curve are equivalent and that every observer gets the same picture. Neither can one see easily that the lines through the origin are orthogonal to the curves and that the distance between the lines is constant. The failure of the figure in these respects stems of course from the fact that we have drawn spacetime curves in a Euclidean space. As we can see from the lower Figure 10.2 the corresponding situation in Euclidean space has actually all the mentioned properties. We see also from this figure that it is quite natural that the curvature must be different on two curved lines which are at a constant orthogonal distance from one another, corresponding in the spacetime case to different acceleration. We must, however, not expect that the Euclidean figure in all respects is a reliable representation of the spacetime situation. The most obvious difference is that the circles in the Euclidean case are

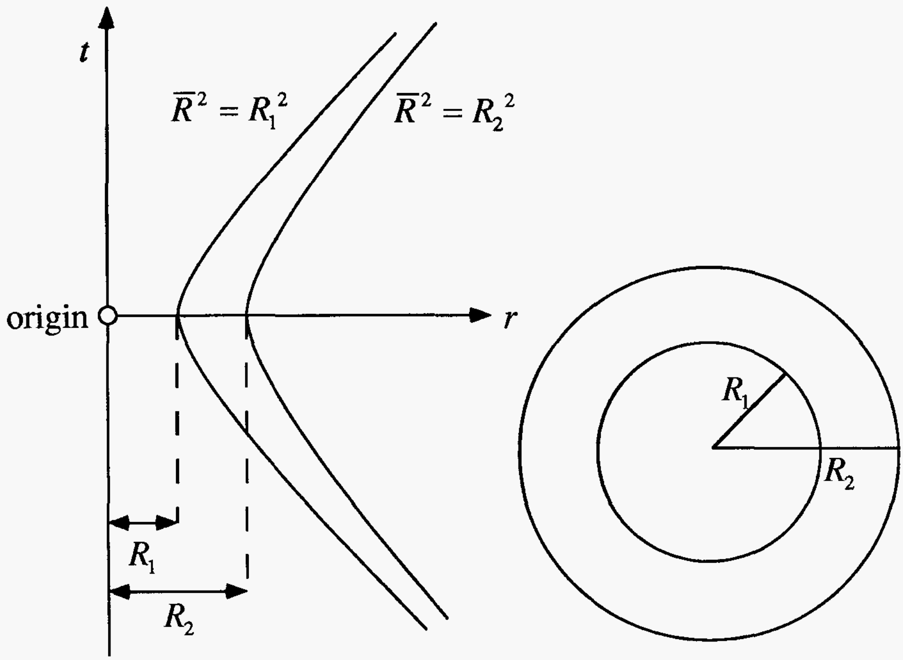

Figure 10.2: Illustration of two flat world lines $\bar{R}^2 = (R_1)^2$ and $\bar{R}^2 = (R_2)^2$ with a common origin. In the left figure a split $\bar{R} = t\hat{U} + \bar{r}$ has been made with respect to some observer and the figure is an (r,t)-diagram. In the right figure the corresponding curves in Euclidean space are drawn. We notice that a number of general features of the spacetime world lines are nicely understood from this figure although the circles are closed contrary to the actual world lines.

closed curves and this can certainly not be so for the spacetime world lines. Any timelike closed curve violates causality (cf. the discussion in Section 6.4).

10.2.2 Fitting a Car into a Garage

There are a number of problems in relativity which sometimes are said to be paradoxes but which actually are quite easy to analyze and turn

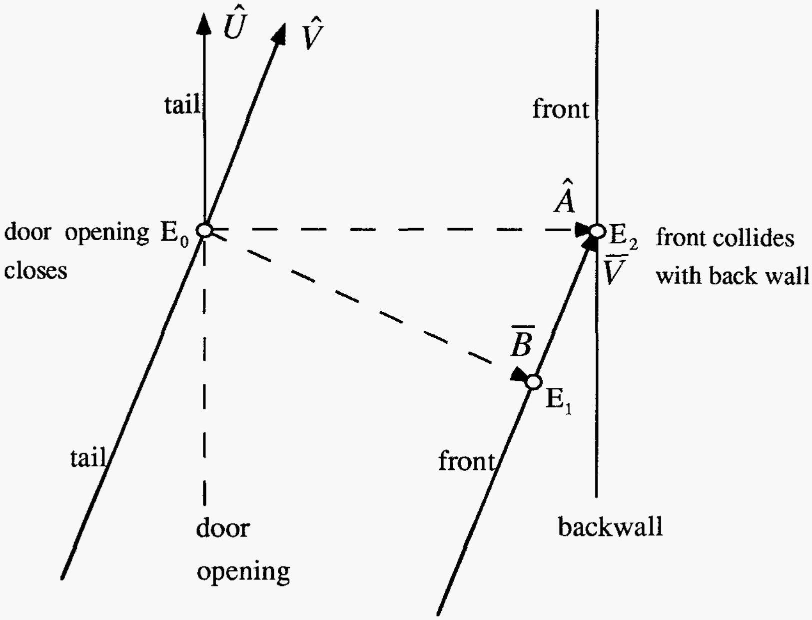

Figure 10.3: Euclidean correspondence to a spacetime development in which a car is driven with high speed into a garage where it is stopped. Three events play important rôles. At E_2 the front of the car hits the back wall of the garage. At E_0 the tail of the car reaches the door opening of the garage. These two events are simultaneous judged from the garage. Finally E_1 is the event which is simultaneous with E_0 judged from the car. The argument in the text concerns the triangle formed by these events.

out to contain no contradictions. One of the simplest is the following: I have acquired a very fast car which unfortunately happens to be just a trifle too long for my garage. Can I exploit the Lorentz contraction (see Section 7.3) to get my car into the garage? Driving the car at high speed this phenomenon makes the car shorter and still the car notices no deformation. Could I simply drive the car into the garage and stop all parts of it simultaneously and suddenly after quickly shutting the door behind it?

We are of course not interested in practical difficulties but only in questions of principle. In Figure 10.3 we have sketched the corresponding situation in Euclidean space. Let us say that $\hat{U}$ is the four-velocity of the garage and $\hat{V}$ the four-velocity of the car driven into the garage. Let us further adopt the length of the garage as the unit of length and $\hat{A}$ as orthogonal to $\hat{U}$ and stretching from the door opening to the back wall of the garage. Similarly $\bar{B}$ is assumed to be orthogonal to $\hat{V}$ and stretching from the tail to the front of the car. We idealize the situation and assume that the front of the car continues unaccelerated until it hits the back wall of the garage where it is suddenly stopped and that the tail is unaccelerated until it reaches the door opening. At that event the tail is suddenly stopped.

If v is the absolute value of the velocity of the car relative to the garage the following Lorentz transformation is valid (see Section 7.1)

$$\hat{V} = \gamma(\hat{U} + v\hat{A}) \tag{10.21}$$

$$\hat{B} = \gamma(\hat{A} + v\hat{U}) \tag{10.22}$$

and also the following inverse transformation

$$\hat{U} = \gamma(\hat{V} - v\hat{B}) \tag{10.23}$$

$$\hat{A} = \gamma(\hat{B} - v\hat{V}). \tag{10.24}$$

It is the last one of these equations which interests us now. Let us introduce an event E_1 on the front world-line which is simultaneous, judged from the car, with the event in which the tail reaches the door opening. Introduce further the vector $\bar{V}$ stretching from E_1 to the event E_2 in which the front hits the wall. All this should be clear from Figure 10.3. Equation (10.24) gives immediately

$$\bar{B} = \gamma\hat{B} \tag{10.25}$$

$$\bar{V} = -v\gamma\hat{V}. \tag{10.26}$$

From this we may draw the conclusion that the car is longer than the garage (remember that $\gamma > 1$). The situation is thus the one we are interested in. Further we see that $\bar{V}$ is negative relative to $\hat{V}$, which means that judged from the car its front actually strikes the back wall *before* the tail reaches the door opening, although these two events are judged as simultaneous from the garage. Both $\hat{A}$ and $\bar{B}$ are spacelike, which means that no signal telling about the frontal collision can reach the back of the car until the tail reaches the door opening. Therefore no material strength can stop the car from being compressed (as judged from the car itself) and you can shut the door behind it. You may have to be quick, however, if the car is very elastic and soon tends to regain its original length. We have in this example an illustration of the fact that no perfectly rigid bodies exist.

Though the method discussed here guarantees success fitting your car into your garage I would not recommend it.

10.2.3 Rotating Wheel

In a resting wheel all the material particles have straight world-lines. When you start spinning the wheel the world lines become tilted and form helices in spacetime. Only the centre has still a straight world-line. Using the centre as observer and regarding a ring of matter at a distance ρ from it we realize that such a ring must have the circumference $2\pi\rho$ whether the wheel is rotating or not. In a sense the Lorentz contraction is prevented from taking place. Instead the tilting of the world lines has the effect that the orthogonal distance between them increases, i.e., *a deformation takes place.*

This is an inevitable deformation which always is connected with the rotation. How much energy is required to effectuate the deformation depends, however, on the stresses in the wheel. We thus see that the moment of inertia of a wheel depends on the stresses within it. From a deep point of view this is an important consequence of the theory of relativity although it plays a negligible rôle under every day circumstances.

Part II

TENSORS

Chapter 11

Definition and Examples

11.1 Definition

Just as scalars may be functions of other scalars, scalars may also be functions of vectors and vectors may be functions of vectors. Tensors are such functions. An example is the famous electromagnetic field tensor which gives the four-acceleration of a charged particle as a function of the four-velocity of the particle. We shall discuss this in Chapter 16. Already in the next section (Section 11.2) we will, however, give examples of tensors in ordinary Euclidean space applying the tensor concept to situations which are more familiar to you.

In general the physical conditions are different in different spacetime points so that for instance the relation between four-velocity and four-acceleration of a charged particle can vary in spacetime which means that the electromagnetic field tensor changes from point to point. It is a major task for physics to investigate the laws of nature which govern the development of tensors in spacetime.

Let T be a function delivering a scalar for each vector we supply it with

$$S = T[\bar{V}]. \tag{11.1}$$

The square bracket [ ] introduced here is meant to signal the one and only property which a function of vectors must satisfy to qualify as a 'tensor'. The property in question is *linearity*, which is defined as follows:

$$T[a\bar{V}' + b\bar{V}''] = aT[\bar{V}'] + bT[\bar{V}'']. \tag{11.2}$$

This must be valid for arbitrary vectors $\bar{V}'$ and $\bar{V}''$ and for arbitrary real numbers a and b. Nothing prevents a tensor from having more than one argument but it must be linear in each of them, e.g.,

$$\begin{aligned} &T[a\bar{V}_1' + b\bar{V}_1'', \bar{V}_2, \ldots \bar{V}_n] = \\ &= aT[\bar{V}_1', \bar{V}_2, \ldots \bar{V}_n] + bT[\bar{V}_1'', \bar{V}_2, \ldots \bar{V}_n] \end{aligned} \tag{11.3}$$

etcetera.

We physicists don't always express ourselves quite clearly when speaking about functions. It is common to say and write something like the following: 'The pressure p is a function $p = p(h)$ of the height h above sea level'. Here 'pressure' plays a double rôle. It is both a number and a mapping from one set of numbers to another set of numbers. Most physicists use this double talk and understand it pretty well. We actually used it ourselves in Chapter 10. From a more systematic point of view the notation is, however, unfortunate. The function (or mapping) $p()$ has as much to do with the height as with the pressure and therefore one should not refer just to the pressure in the notation. It is better either to introduce both p and h somehow in the notation of the function (which one seldom does) or to use an entirely different notation, e.g., $p = f(h)$ and define $f(\ )$ as the *procedure* (experimental or mathematical) through which p is obtained from h.

It is important to make this very clear when dealing with tensors. A tensor is a mapping and nothing else. Thus $T[\ ]$ is a tensor but when you supply it with a vector $\bar{V}$ the result $T[\bar{V}]$ is a scalar. As we have already indicated, a tensor can be either a mapping of vectors on scalars or a mapping of vectors on vectors. Later on we shall generalize this but for the moment we are content with these two types which we denote $T[\ ,\ ,\ldots]$ and $\bar{T}[\ ,\ ,\ldots]$ respectively. Sometimes we drop the empty brackets and write simply T and $\bar{T}$. It is then important to have memorized that they are tensors so that they are not mistaken to be a scalar and a vector, respectively.

11.2 Examples

At this stage we cannot give many examples of tensors from relativity. For this reason we turn to classical non-relativistic physics using 'ordinary' three dimensional vectors in Euclidean space.

Our first example is taken from classical mechanics. The angular velocity $\bar{\Omega}$ and the angular momentum $\bar{L}$ of a rotating rigid body are not always parallel but there is a linear relation between them, so that we may write

$$\bar{L} = \bar{I}[\bar{\Omega}]. \tag{11.4}$$

Given the mass distribution within the rigid body there is a unique angular momentum corresponding to each angular velocity. Therefore the mapping — the tensor $\bar{I}[\,]$ — is well defined. It is known under the name of the *moment of inertia tensor.*

A second example is given by the relation between electric field strength $\bar{E}$ and current density $\bar{\jmath}$ in a conductor. In many cases Ohm's law is valid so that the two vectors are parallel $\bar{\jmath} = C\bar{E}$. The proportionality factor C is called the *conductivity.* Provided that C is independent of $\bar{E}$ multiplication with C is a special case of tensor — linearity is then trivially satisfied. You should notice the exact wording here. If one wishes to be precise one should *not* say that the tensor is the number C. The tensor is *multiplication* by C, which is a mapping.

Ohm's law is not always satisfied in its simplest form. The conductor may not be isotropic, so that the current density and the electric field are not parallel. Plasmas in magnetic fields offer extreme examples of this. Nevertheless the current density may at least approximately be a linear function of the electric field, so that we may write

$$\bar{\jmath} = \bar{C}[\bar{E}] \tag{11.5}$$

where $\bar{C}[\,]$ is a tensor called the *conductivity tensor.* This tensor — this relation between $\bar{E}$ and $\bar{\jmath}$ — depends on the physical conditions in the conductor and can vary from point to point. Notice, however, that the assumption of linearity contains the assumption that $\bar{C}$ is *independent of* $\bar{E}$. In a still more general case a non-linear relation between the electric field and the current density can occur. Such a relation cannot be written in the form (11.5) at least not with a $\bar{C}$ which is independent of $\bar{E}$.

Our third example concerns stresses in a solid. Consider a solid in a given situation when it has been deformed somehow so that stresses have developed (see Figure 11.1). We choose a small surface element $\overline{dA}$ somewhere in the solid and ask what the force $\overline{dF}$ is with which the molecules on the side 1 of the surface element are influenced by the molecules on side 2. In a liquid these two vectors are parallel and the proportionality factor is the *pressure*. In solids this rule is replaced by the rule that there is a linear relation between the two vectors

$$\overline{dF} = \bar{\sigma}[\overline{dA}] \tag{11.6}$$

where $\bar{\sigma}[\ ]$ is a tensor named the *stress tensor*. Like the conductivity tensor the stress tensor may vary from point to point within the solid. The stress tensor must in fact satisfy certain differential equations which in principle and sometimes also in practice makes it possible to calculate it with the knowledge of material properties and boundary conditions. How this type of equations can be formulated will hopefully become clear when we discuss the corresponding problem in electrodynamics in Chapter 16. One further remark: The stress tensor is in a way the oldest tensor. The term 'tensor' derives from the study of 'tensions' in solids.

As a fourth example consider a particle in a static situation. The energy δW required to move the particle the small distance $\delta\bar{r}$ to a new static situation is a linear function of $\delta\bar{r}$. Hence we may write

$$\delta W = F[\delta\bar{r}]. \tag{11.7}$$

The function $F[\ ]$ thus defined is a tensor called *the force tensor* acting on the particle. Dividing both sides of (11.7) by some small parameter $\delta\lambda$ and using the linearity,

$$\delta W/\delta\lambda = F[\delta\bar{r}/\delta\lambda], \tag{11.8}$$

F is seen to be defined not only for small but for arbitrary vector arguments.

To look at the force as a tensor is actually more fundamental than to regard it as a vector essentially because the former point of view

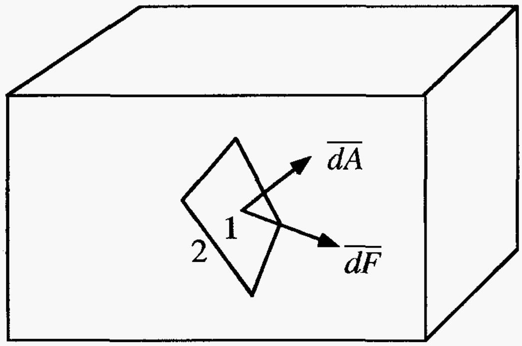

Figure 11.1: When a solid body is deformed stresses develop within it. Let $\overline{dA}$ be a surface element and $\overline{dF}$ be the force which the matter on side 2 exercises on the matter on side 1. The stress tensor is the mapping of $\overline{dA}$-vectors on $\overline{dF}$-vectors.

requires no use of the scalar product. We will now take this thought a little bit further. For a given $F[\,]$ the equation

$$F[\bar{v}] = 0 \tag{11.9}$$

is a linear equation for $\bar{v}$. If two or more vectors satisfy the equation then all linear combinations of them must also satisfy the equation. Equation (11.9) is therefore satisfied by the vectors in an entire *linear subspace*. In three dimensions there exist only three types of such linear subspaces: The null vector, parallel vectors, and vectors parallel to a plane. Now Equation (11.9) gives just one scalar restriction on the vector $\bar{v}$. Hence it gives a plane to which $\bar{v}$ must be parallel.

If you like to form a picture of $F[\,]$ you should not see it as an arrow but rather as a succession of uniformly spaced parallel planes (see Figure 11.2) determined as just mentioned. The spacing must be chosen so that if you place an arbitrary vector $\bar{v}$ among the planes, $F[\bar{v}]$ is equal to the number of planes penetrated by the vector. It should be kept in mind that this picture reflects the situation *at one point* in space — the position of the particle on which the force is acting.

Tensors $F[\,]$ can like vectors be linearly combined. Multiplication by a number corresponds to multiplication of the density of planes by the

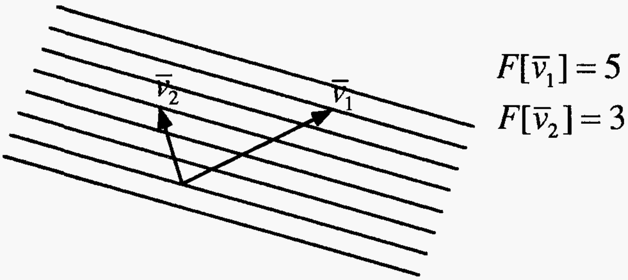

Figure 11.2: A tensor $F[\ ]$ can be pictured as a succession of uniformly spaced parallel planes. The number $F[\bar{v}]$ is given by the number of planes which the vector $\bar{v}$ penetrates. As this is a figure in two dimensions the planes are replaced by straight lines.

number in question. Addition of tensors is illustrated in Figure 11.3.

Regarding the force as a *vector* $\bar{F}$ you must employ the scalar product and write $\delta W = \bar{F} \cdot \delta\bar{r}$. In this formulation the scalar δW appears as a linear function of *two* vectors $\bar{F}$ and $\delta\bar{r}$. Hence

$$\delta W = \bar{F} \cdot \delta\bar{r} = g[\bar{F}, \delta\bar{r}] \tag{11.10}$$

where $g[\ ,\ ]$ is a tensor defined as the scalar product. This tensor is named the *metric tensor*. As the scalar product between two vectors has the same value independent of the position in space, the metric tensor is a constant tensor.

Also the cross product is a tensor. Consider, e.g., the Lorentz force on a particle with charge e

$$\frac{d}{dt}\bar{p} = e\bar{v} \times \bar{B} \equiv e\bar{\varepsilon}[\bar{v}, \bar{B}]. \tag{11.11}$$

Here $\bar{v}$ is the velocity of the particle and $\bar{B}$ is the magnetic field strength. The tensor $\bar{\varepsilon}[\ ,\ ]$ which is defined as the cross product is called the *Levi-Civita tensor* or the *volume tensor.*

Tensors are still more important in spacetime than they are in ordinary Euclidean space and we will now leave the non-relativistic examples to resume our discussion of spacetime.

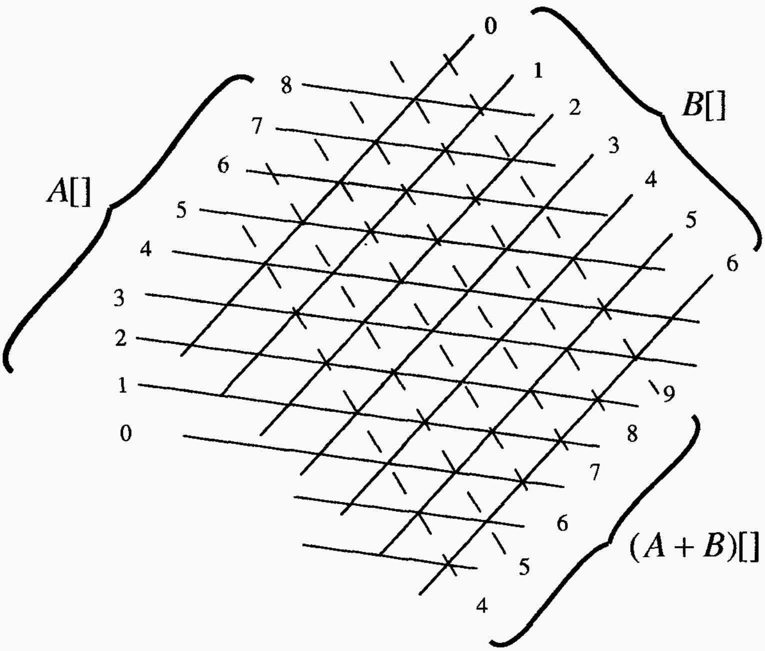

Figure 11.3: Addition of tensors in two dimensions. The set of A-lines corresponds to one tensor and the set of B-lines to another. The sum of these is a tensor corresponding to the broken lines. Notice how they run through intersections between A-lines and B-lines. This has the consequence that following one of the resulting lines we count just as quickly forward among the A-lines as we count backwards among the B-lines.

Chapter 12

Algebraic Properties

12.1 First Rank Tensors

This is where we start a more systematic formal study of tensors. Taking one thing at a time contributes to clarity and we are going to first study general properties of tensors without assuming that a scalar product exists. The only remaining assumptions are that vectors can be combined linearly thus obtaining new vectors and that all vectors can be obtained as linear combinations of four (but not fewer) selected vectors, i.e., that the vectors form a four-dimensional *vector space* in the mathematical sense. In Section 12.6 we will discuss the properties which depend on the existence of a scalar product.

The simplest non-trivial type of tensor is the *first rank tensor* $T[\ ]$ which maps single vectors on scalars. This entire section is devoted to such tensors. If you like to have a picture in mind while you read this fairly abstract section please return to the previous section (Section 11.2) where a first rank tensor in three dimensions was discussed from this point of view. Essentially the same picture can be used in four dimensions.

First an almost trivial remark: Every tensor $T[\ ]$ maps the null-vector on zero. This follows immediately from the linearity which gives $T[0] = T[-0] = -T[0]$, hence $T[0] = 0$.

A special case of first rank tensor is the *null-tensor*, which maps *every* vector on zero, i.e., $T[\bar{V}] = 0$ for all $\bar{V}$. This defines a tensor

— the linearity is obvious. It is not necessary to introduce a special notation for the null-tensor. We simply write $T[\,] = 0$.

The *linear combination* of two tensors $S[\ ]$ and $T[\ ]$ is defined as follows

$$(aS + bT)[\bar{V}] \equiv aS[\bar{V}] + bT[\bar{V}] \tag{12.1}$$

This must be valid for all vectors $\bar{V}$. The numbers a and b are arbitrary. Notice that the right hand side is linear in $\bar{V}$ so that $(aS+bT)$ actually is a tensor.

A number of tensors $A[\ ]$, $B[\ ]$, $\ldots X[\ ]$ are said to be *linearly independent* if the equation

$$aA + bB + \ldots + xX = 0 \tag{12.2}$$

has the consequence

$$a = b = \ldots = x = 0. \tag{12.3}$$

You should compare this with the definition of linear independence of vectors in Chapter 5.

Introducing four linearly independent four-vectors $\bar{A}$, $\bar{B}$, $\bar{C}$, and $\bar{D}$ we can define a first rank tensor $R[\ ]$ in the following way. Split an arbitrary vector $\bar{V}$ as follows

$$\bar{V} = a\bar{A} + \bar{Q} \tag{12.4}$$

where a is a number and the vector $\bar{Q}$ is a linear combination of $\bar{B}$, $\bar{C}$, and $\bar{D}$. Then define $R[\,]$ by

$$R[\bar{V}] \equiv a. \tag{12.5}$$

The right hand side is linear in $\bar{V}$ so that $R[\,]$ thus defined is in fact a tensor. By letting $\bar{B}$, $\bar{C}$, and $\bar{D}$ successively exchange rôles with $\bar{A}$ three more tensors, say $S[\,]$, $T[\,]$, and $U[\ ]$ are defined in the same manner. This is a standard method to define four first rank tensors from four linearly independent vectors and it is much used both in proofs and applications. We shall use it immediately in an important proof.

Consider an arbitrary first rank tensor $J[\ ]$ and let it work on an arbitrary vector $\bar{V}$ which we split as follows

$$\bar{V} = a\bar{A} + b\bar{B} + c\bar{C} + d\bar{D}. \tag{12.6}$$

The result is

$$\begin{aligned} &J[\bar{V}] = aJ[\bar{A}] + bJ[\bar{B}] + cJ[\bar{C}] + dJ[\bar{D}] = \\ &J[\bar{A}]R[\bar{V}] + J[\bar{B}]S[\bar{V}] + J[\bar{C}]T[\bar{V}] + J[\bar{D}]U[\bar{V}] = \\ &(J[\bar{A}]R + J[\bar{B}]S + J[\bar{C}]T + J[\bar{D}]U)[\bar{V}] \end{aligned} \tag{12.7}$$

or since $\bar{V}$ is arbitrary

$$J = J[\bar{A}]R + J[\bar{B}]S + J[\bar{C}]T + J[\bar{D}]U. \tag{12.8}$$

Thus we see that an arbitrary first rank tensor can be expanded in the just defined tensors R, S, T, and U. Moreover we see that the sum on the right hand side can be zero only if all the coefficients vanish. Hence R, S, T, and U are linearly independent. This shows that also the first rank tensors form a four-dimensional vector space in the mathematical sense but it is not the same space in which the original vectors live — you must not mix vectors and tensors linearly.

The tensor defined in (12.5) gives $R[\bar{A}] = 1$. But the only restriction on $\bar{A}$ (for use as a part of four linearly independent vectors) is $\bar{A} \neq 0$. Hence picking an arbitrary vector different from the null-vector, there exists a tensor which gives unity when working on this particular vector. An important corollary is that if $T[\bar{V}] = 0$ *for all tensors* T then $\bar{V} = 0$. We will soon have use for this fact.

Knowing that the first rank tensors form a four-dimensional vector space we may turn the interpretation of $T[\bar{V}]$ around and for a fixed $\bar{V}$ see it as a linear mapping of first rank tensors (i.e. as a mapping of mappings) on scalars. The linearity is seen in (12.1). Each vector $\bar{V}$ causes a mapping. Linear combination of mappings correspond to linear combination of vectors. The zero mapping is obtained for $\bar{V} = 0$ and for $\bar{V} = 0$ only (see the theorem a few lines above). Two different vectors can therefore not correspond to the same mapping and there is hence a one-to-one correspondence between vectors and mappings.

To summarize the results of this section we have found that there is a remarkable symmetry between vectors and first rank tensors. Both form four-dimensional vector spaces in the mathematical sense that any linear combination of objects forms objects of the same kind and that all objects of one kind can be constructed linearly from four (but not fewer) selected objects. The tensors are linear mappings of vectors on scalars but vectors can also function as linear mappings of tensors on scalars. Usually this is expressed by saying that there is a *dual* relationship between vectors and first rank tensors and very often first rank tensors are referred to as *dual vectors.*

To find an example we still have at this stage to turn to non-relativistic physics. In Section 11.2 the work δW was regarded as a linear function of the vector of transfer $\delta\bar{r}$. This function $F[\,]$ is known as the force tensor.

According to the definition of linear combination (12.1) δW may also be regarded as a linear function of $F[\,]$ for a given $\delta\bar{r}$ and the main result of this section tells us that $F[\,]$ can be written as a linear combination of three given linearly independent first rank tensors. In Section 12.6 we shall see that due to the metric there is a force *vector* corresponding to $F[\,]$ (see also equation (11.10)).

The rule (12.1) as such does *not* tell us that the linear combination of two force tensors gives another force tensor although it certainly gives a first rank tensor. In statics (keeping to that part of mechanics) it is, however, assumed that two or more forces can be applied simultaneously to a particle and that the resulting force is the sum of those. With this extra assumption it is obvious that any linear combination of two forces *is* a force. Linear combination is then a *closed operation* for forces and forces form a linear space, which coincides with the space of first rank tensors.

12.2 Generalized Tensors

12.2.1 Symmetries

In Chapter 11 tensors were defined and discussed according to their effect on vectors. The tensor $T[\;,\;]$ is defined by the numbers $T[\bar{U},\bar{V}]$ obtained with all possible pairs of vectors $(\bar{U},\bar{V})$. The *order* in which the vectors are presented is important, because in general the number $T[\bar{U},\bar{V}]$ is different from the number $T[\bar{V},\bar{U}]$. Given an arbitrary tensor $T[\;,\;]$ we may always define the *transpose* $\tilde{T}[\;,\;]$ of the tensor by the following equation which must be valid for all vectors $\bar{U}$ and $\bar{V}$

$$\tilde{T}[\bar{U},\bar{V}] \equiv T[\bar{V},\bar{U}]. \tag{12.9}$$

You should notice that $\tilde{T}$ according to this definition is linear in its arguments and consequently is a tensor.

There is a special class of tensors called *symmetric tensors.* They satisfy the equation

$$\tilde{S} = S. \tag{12.10}$$

Tensors satisfying[1]

$$\tilde{A} = -A \tag{12.11}$$

are called *antisymmetric.* Both symmetric and antisymmetric tensors are important in physics.

12.2.2 Abstract Indices

For tensors with only two entries the 'tilde' (i.e. $\tilde{}$) is adequate in tensor equations describing symmetries. For more arguments you have to invent an awkward number of different symbols to stand for the different orderings of the vector arguments if you would like to follow the same route as in Section 12.2.1. To escape this a marker system has been invented, which we will use extensively in the rest of this book. Instead of inserting the vector arguments in the order in which they are presented we mark them as well as the tensor entries with the same set of markers writing, e.g.,

$$T_{ab}[\ ,\ ]\bar{U}_a\bar{V}_b = T_{ab}[\ ,\ ]\bar{V}_b\bar{U}_a = T[\bar{U},\bar{V}] \tag{12.12}$$

and

$$T_{ba}[\ ,\ ]\bar{U}_a\bar{V}_b = T[\bar{V},\bar{U}]. \tag{12.13}$$

The condition that S is symmetric can now be written

$$S_{ba} = S_{ab} \tag{12.14}$$

and the condition that A is antisymmetric can be written

[1]The minus sign on the right hand side has the meaning that whatever vectors you supply $-T$ with you obtain minus the value of T supplied with the same vectors.

$$A_{ba} = -A_{ab}. \tag{12.15}$$

In such equations it will always be understood that the two members of the equation are to be presented the same vector arguments with the same markings and that these markers are the same as the tensor markers. The equation is therefore meaningless unless the same markers appear on both sides. The kind of markers used is entirely irrelevant, but usually we will use lower case Latin letters from the beginning of the alphabet. A change of markers in for instance equation (12.14) does not change its meaning. We could, e.g., just as well write it $S_{cd} = S_{dc}$.

To make the notation neater we will omit both the tensor bracket and the vector bar. To signal which is which we will instead put the vector markers 'upstairs' and the tensor markers 'downstairs'. Thus we write, e.g.,

$$T[\bar{U}, \bar{V}] = T_{ab}U^a V^b. \tag{12.16}$$

With this step the symmetry between vectors and first rank tensors which we found in Section 12.1 becomes manifest in the notation. We write, e.g.,

$$T[\bar{V}] = T_a V^a. \tag{12.17}$$

We will assume that nothing is changed in the meaning of the right hand side of this equation if we change the order, writing

$$T[\bar{V}] = V^a T_a. \tag{12.18}$$

This gives rise to the thought that we could generalize the concept of tensor so that it has entries not only for vectors but also some entries for dual vectors (first rank tensors). Linearity in all arguments is still required. In Section 12.1 we have seen that $T[\bar{V}] = T_a V^a$ for a given vector $\bar{V}$ can be regarded as a linear mapping of first rank tensors on scalars. Hence vectors can be regarded as a kind of tensors in the generalized sense. A tensor of the type $\bar{T}[\ ,\ ,\ldots]$ (notice the bar) introduced in Section 11.1 can now also be regarded as a mapping on scalars but with one entry which accepts dual vectors. The form of such a tensor is $T^a{}_{bc...}$ and the the scalar corresponding to the dual vector $A[\,]$ and the vectors $\bar{B}$, $\bar{C}$, etc is given by

$$S = T^a{}_{bc...}A_aB^bC^c\ldots. \tag{12.19}$$

The indices which we have introduced in this section and described as 'markers' are usually called *abstract indices.* They were only fairly recently suggested and then in a more advanced context[2]. The total number of indices on a tensor is referred to as the *rank* of the tensor. There is no firmly established term for the number of upstairs indices and the number of downstairs indices, but let us call the former the *uprank* and the latter the *downrank.*

As a matter of principle it is not quite satisfactory to write 'the tensor $T_a{}^{bc}$ ' instead of 'the tensor T ' in a piece of text. The specific markers chosen do in no way help to characterize the tensor. They should strictly speaking be used exclusively as pointers showing where to insert specified arguments in the tensor T. Nevertheless you will often in the literature see phrases like the one criticized here and still worse we will use such phrases ourselves in the following. The reason is that it is quite practical with a notation describing the *index pattern* of a tensor (the order in which up and down indices occur). If you are aware that this is the only content of this mode of writing it should not lead to any misunderstandings.

12.3 Tensor Algebra

Multiplication of a tensor by a real number and *addition of tensors* with the same uprank and the same downrank is defined as follows[3]

$$\begin{aligned}&S^a{}_{bc}A_aB^bC^c \equiv (xX^a{}_{bc} + yY^a{}_{bc})A_aB^bC^c \equiv\\ &xX^a{}_{bc}A_aB^bC^c + yY^a{}_{bc}A_aB^bC^c.\end{aligned} \tag{12.20}$$

We have here taken a third rank tensor with one index up and two down as example. The tensor

[2]This was done 1986 in a book 'Spinors and Spacetime' by R. Penrose and W. Rindler.

[3]For first rank tensors this occurs already in Section 12.1.

$$S^a{}_{bc} = xX^a{}_{bc} + yY^a{}_{bc} \tag{12.21}$$

is through (12.20) defined in terms of the tensors $X^a{}_{bc}$ and $Y^a{}_{bc}$. Notice that the right hand side is linear in A_a, B^b, and C^c so that $S^a{}_{bc}$ actually is a tensor. Equation (12.20) must be an identity in the sense that it is valid for each triplet ($A[\,]$,$\bar{B}$, $\bar{C}$). The factors x and y are two arbitrary real numbers. The markers a, b, and c have two purposes according to Section 12.2.2. First their positions up or down tell us the characters of entries one, two, and three. They admit as arguments dual vector, vector, and vector, respectively. Secondly the markers tell us to which entries the arguments are supposed to go when we compare the terms with one another. As an example the same argument must be placed in the second entry in all three terms in (12.21). You should compare (12.21) with the following equation:

$$T^a{}_{bc} = xX^a{}_{bc} + yY^a{}_{cb} \tag{12.22}$$

where b and c have been exchanged in the last term. In this case the argument which goes to the second entry in the first two terms must go to the third entry in the last term and vice versa.

The *tensor product* between two tensors, say $X^a{}_b$ and $Y^c{}_{de}$, is defined as follows:

$$\begin{aligned} &P^a{}_b{}^c{}_{de} A_a B^b C_c D^d E^e \equiv X^a{}_b Y^c{}_{de} A_a B^b C_c D^d E^e \equiv \\ &(X^a{}_b A_a B^b)(Y^c{}_{de} C_c D^d E^e). \end{aligned} \tag{12.23}$$

The generalization to arbitrary tensors is obvious. Notice the linearity on the right hand side which makes the product P a tensor under the assumption that X and Y are tensors. The uprank for the product tensor is the sum of the upranks for the factors. The same is true for the downrank.

From these definitions it follows immediately that provided you let the markers follow unchanged with their respective tensors you can formally handle tensors as real numbers as long as you employ addition and multiplication only. The associative, distributive, and commutative laws are valid. As an example we can write

$$V^aT_b{}^c(K^d + L^d) = T_b{}^cK^dV^a + V^aL^dT_b{}^c. \tag{12.24}$$

12.4 Expansion

We know from Chapter 5 that an arbitrary vector can be written as a linear combination of any set of four linearly independent vectors and we know from Section 12.1 that the corresponding proposition is true for dual vectors. The set of vectors or dual vectors in which one expands is usually called the *basis* and the expansion coefficients are called *components.*

Let us now consider an arbitrary second rank mixed (one index down one up) tensor $T_a{}^b$ and try to expand it using some basis of four linearly independent vectors $(V_1)^a$, $(V_2)^a$, $(V_3)^a$, and $(V_4)^a$. If we supply $T_a{}^b$ with some arbitrary vector $\bar{U}$ we obtain a vector $U^aT_a{}^b$ which we can expand

$$U^aT_a{}^b = x_1(V_1)^b + x_2(V_2)^b + x_3(V_3)^b + x_4(V_4)^b \tag{12.25}$$

Here x_1, x_2, x_3, and x_4 are four numbers which depend on $\bar{U}$ as well as on T. They are in fact *linear functions of* $\bar{U}$ which follows immediately from the linearity of the left hand side of (12.25) and the uniqueness of the coefficients (see Chapter 5, page 46). Hence each coefficient can be written as a first rank tensor with $\bar{U}$ inserted, e.g., $x_1 = (S_1)_aU^a$. The arbitrary vector $\bar{U}$ can then be omitted from both sides of (12.25), leaving the equation

$$T_a{}^b = (S_1)_a(V_1)^b + (S_2)_a(V_2)^b + (S_3)_a(V_3)^b + (S_4)_a(V_4)^b. \tag{12.26}$$

Thus we see that an arbitrary tensor $T_a{}^b$ can be expanded as a linear combination of four arbitrarily chosen linearly independent vectors. The coefficients are tensors of a rank one. Given the tensor $T_a{}^b$ and the four vectors the coefficients are obviously unique.

To make the notation more compact the previous equation is usually written

$$T_a{}^b = \sum_{r=1}^{4} (S_r)_a (V_r)^b. \qquad (12.27)$$

This compact way of writing is still more practical if you have to perform a double summation, which occurs for instance if you would like to compare expansions with two different bases of vectors. Suppose that $\bar{V}_r$ (with $r = 1, 2, 3$, or 4) is one such basis and that $\bar{V}'_s$ (with $s = 1, 2, 3$, or 4) is another such basis. Each vector $\bar{V}_r$ can then be expanded as linear combination of the vectors $\bar{V}'_s$

$$(V_r)^b = \sum_{s=1}^{4} C_{rs} (V'_s)^b \qquad (12.28)$$

where C_{rs} are 16 expansion coefficients. Introducing this in the previous equation you obtain the expansion of $T_a{}^b$ in terms of the $\bar{V}'_s$ vectors

$$T_a{}^b = \sum_{r=1}^{4} \sum_{s=1}^{4} (S_r)_a C_{rs} (V'_s)^b = \sum_{s=1}^{4} (S'_s)_a (V'_s)^b \qquad (12.29)$$

where

$$(S'_s)_a = \sum_{r=1}^{4} (S_r)_a C_{rs}. \qquad (12.30)$$

These are the new coefficients corresponding to the basis consisting of the vectors $\bar{V}'_s$.

It is trivial to generalize the expansion of the tensor $T_a{}^b$ to any kind of tensor. Every tensor can be expanded with respect to any one of its indices. For an upstairs index you need a basis of four linearly independent vectors while for a downstairs index you need a basis of four linearly independent dual vectors to accomplish the expansion. In the former case the coefficients are tensors of uprank one less than the uprank of the original tensor. In the latter case the coefficients are tensors of downrank one less than the downrank of the original tensor. These coefficients are uniquely determined by the original tensor, which index you choose to expand with respect to, and the basis.

If you like you may continue the expansion by expanding the tensor coefficients obtaining multiple sums. You can carry this so far that you finally end up with a multiple sum in which each term consists of a product of only vectors and dual vectors.

Two examples of expansion can be found at the end of the next section (Section 12.5).

12.5 Contraction

From a dual vector $T[\ ]$ and a vector $\bar{V}$ we can form the scalar $T[\bar{V}] = T_a V^a$. This can be expressed in a slightly different way: We can first form the second rank tensor $T_a V^b$ and then *contract* this tensor. Contraction means putting two markers equal, in this case obtaining $T_a V^a$. One of the markers must be upstairs and the other one downstairs. In this way a tensor is formed with a rank which is two lower than the rank of the original tensor. The new tensor has lost the ability to admit vectors (or dual vectors) to the contracted indices.

So far we have given a meaning to the term 'contraction' only for a special kind of tensors, namely tensors of the form $T_{a\ldots}{}^{b\ldots}V^c$. In this case it is clear what we mean by a contraction in which c is involved.

Using expansion we may, however, generalize contraction to be meaningful for every tensor which contains both upstairs and downstairs indices. The recipe is simply to expand the original tensor with respect to the upstairs index we want to involve in the contraction using some vector basis and then *make the chosen contraction in each term.* An example is the following: To find $T_a{}^{ba}$ we must expand the tensor $T_a{}^{bc}$

$$T_a{}^{bc} = \sum_{r=1}^{4} (S_r)_a{}^b (V_r)^c \tag{12.31}$$

and then contract in each term obtaining

$$T_a{}^{ba} = \sum_{r=1}^{4} (S_r)_a{}^b (V_r)^a. \tag{12.32}$$

The right hand side here has a well defined meaning.

So far we do not know how the result of the contraction changes when the vector basis is changed. We shall now prove that it does not change at all. In fact holding on to our example and using (12.28) and (12.30) we have

$$\sum_{r=1}^{4}(S_r)_a{}^b(V_r)^a = \sum_{r=1}^{4}\sum_{s=1}^{4}(S_r)_a{}^b C_{rs}(V'_s)^a = \sum_{s=1}^{4}(S'_s)_a{}^b(V'_s)^a. \quad (12.33)$$

It is trivial to generalize this proof to a case in which S has any number of tensor markers in addition to a.

What markers you use to indicate a contraction is irrelevant. So is, e.g., $T_aV^a = T_bV^b$. You must use this freedom to make sure that the marker you use for a contraction is not already used for some other purpose. Otherwise misunderstandings are apt to arise. For instance you must never write $V^a = T^a{}_aU^a$. Writing such an illegal expression you don't make clear if you actually mean $V^a = T^b{}_bU^a$ or $V^a = T^a{}_bV^b$.

Simultaneous contraction over more than one index occurs frequently. A special case of this is a contraction of a symmetric with an antisymmetric tensor. Let S_{ab} be symmetric and A^{ab} antisymmetric. Then

$$S_{ab}A^{ab} = S_{ba}A^{ab} = -S_{ba}A^{ba} = -S_{ab}A^{ab}. \quad (12.34)$$

The last equation follows from the fact that no importance is attached to the particular marker used to indicate a contraction. Hence we may change the marker 'c' to 'd' and the marker 'd' to 'c' without changing the meaning of the expression. From (12.34) follows that

$$S_{ab}A^{ab} = 0. \quad (12.35)$$

An expression vanishes if it contains a contraction of a symmetric pair of indices with an antisymmetric pair of indices. This rule is worth memorizing.

Contraction is an important operation, which occurs in many physical situations. Roughly speaking it effectuates a summation over all directions. Divided by the dimension of the vector space the contraction gives an average over all directions. We shall illustrate this by the moment of inertia tensor I defined in Section 11.2. Let us expand it in terms of an arbitrary triplet of linearly independent vectors $(\bar{X}, \bar{Y}, \bar{Z})$

$$I^a{}_b = X^a(I_x)_b + Y^a(I_y)_b + Z^a(I_z)_b \quad (12.36)$$

where $(I_x)_b$, etc. are the expansion coefficients. Remember now the physical interpretation of I. If we contract I with an angular velocity vector $\bar{\Omega}$ we obtain the angular momentum

$$L^a = I^a{}_b\Omega^b = \{(I_x)_b\Omega^b\}X^a + \{(I_y)_b\Omega^b\}Y^a + \{(I_z)_b\Omega^b\}Z^a. \quad (12.37)$$

The right hand side here is obviously an expansion of the angular momentum in terms of the vectors $(\bar{X}, \bar{Y}, \bar{Z})$. This tells us the meaning of the scalar coefficients $(I_x)_b\Omega^b$, etc. So is $(I_x)_b\Omega^b$ the component of the angular momentum along $\bar{X}$ obtained from the angular velocity $\bar{\Omega}$. The quantity $(I_x)_bX^b$ is therefore the component of the angular momentum along $\bar{X}$ if you choose the angular velocity to be $\bar{X}$. The contraction of I is according to (12.36) the sum of three such quantities

$$I^a{}_a = (I_x)_aX^a + (I_y)_aY^a + (I_z)_aZ^a. \quad (12.38)$$

We know that the value of this expression is independent of how you choose the vectors $(\bar{X}, \bar{Y}, \bar{Z})$ and you may in particular choose these vectors to be three orthogonal unit vectors. The first term can then be said to be the 'the moment of inertia in the x-direction', etc. and the contraction divided by three can be said to be the 'average of the moment of inertia over the x-,y-, and z-directions'.

As one more example let us consider the *identity tensor* δ^a_b defined by the identity

$$V^a = \delta^a_bV^b \quad (12.39)$$

valid for all $\bar{V}$. Expand this tensor in some set of vectors $(\bar{A}, \bar{B}, \bar{C}, \bar{D})$ and you obtain

$$\delta^a_b = A^aR_b + B^aS_b + C^aT_b + D^aU_b. \quad (12.40)$$

The tensors (R, S, T, U) appearing here have a very special relation to the vector basis. For $\delta^a_bA^b$ to be equal to A^a the following must be true

$$R_aA^a = 1, \; S_aA^a = 0, \; T_aA^a = 0, \; U_aA^a = 0 \quad (12.41)$$

and there are three similar sets of equations concerning $\bar{B}$, $\bar{C}$, and $\bar{D}$. Expressed in words *each one of the tensors gives unity when supplied* *with 'its own' vector and zero when supplied with any one of the other vectors.* Compare this with Section 12.1, equation (12.5), and you will find that the 'standard method' described there gives the current tensors R, S, T, and U. In Section 12.1 it was proven that these tensors are linearly independent and form a tensor basis. A vector basis and a tensor basis which are related as described here are said to be bases *dual* to one another. Equation (12.40) can be quite useful at times.

The dual base property emphasized above gives immediately the contraction of the identity tensor. Contracting on the right hand side of (12.40) four terms are obtained, each of them equal to unity. Hence

$$\delta^a_a = 4. \tag{12.42}$$

12.6 Tensors and the Scalar Product

12.6.1 Relation between Vectors and Dual Vectors

Developing the subject of tensors we have so far avoided using the scalar product between vectors. Now, however, the time has come to introduce it into the subject.

Using the scalar product one can always find a first rank tensor corresponding to each vector: Let $\bar{U}$ be some given vector. Corresponding to this vector there is a tensor $U[\,]$ defined by

$$U[\bar{V}] \equiv \bar{U} \cdot \bar{V} \tag{12.43}$$

where $\bar{V}$ is arbitrary. Notice that the right hand side is linear in $\bar{V}$ telling us that $U[\,]$ actually is a tensor. Sometimes the notation $\bar{U}\cdot$ (with a dot) is used for this tensor.

To avoid misunderstanding this method of defining a tensor should be contrasted to the method used in Section 12.1 where we defined *four* first rank tensors from *four* linearly independent vectors. That method does not require the use of a scalar product, but it requires four linearly independent vectors even if we would wish to construct only one first rank tensor and it gives no connection pairwise between vectors and tensors.

We shall now prove that *if a number of $\bar{U}$ vectors are linearly independent then the corresponding $\bar{U}\cdot$ tensors are also linearly independent*: Suppose that $\bar{U}_s$ are N linearly independent vectors and that c_s are N numbers such that

$$\sum_{s=1}^{N} c_s \bar{U}_s \cdot = 0. \tag{12.44}$$

As a consequence

$$(\sum_{s=1}^{N} c_s \bar{U}_s) \cdot \bar{A} = 0 \tag{12.45}$$

for an arbitrary vector $\bar{A}$. This and the rule (5.8) tells us that

$$\sum_{s=1}^{N} c_s \bar{U}_s = 0. \tag{12.46}$$

From this and the supposed linear independence of the vectors $\bar{U}_s$ follows that all the c-coefficients vanish, $c_s = 0$, and if we look back at (12.44) it becomes obvious that the N tensors $\bar{U}_s\cdot$ have to be linearly independent. Q.E.D.

This tells us that four linearly independent tensors $\bar{U}_s\cdot$ can be found and that therefore *every tensor can be written in the form $\bar{U}\cdot$. The vector $\bar{U}$ corresponding to some given tensor $\bar{U}[\ ]$ is unique* which is easily realized: Suppose that $\bar{U}_1$ and $\bar{U}_2$ are two vectors corresponding to the same tensor, i.e.,

$$\bar{U}_1 \cdot = \bar{U}_2 \cdot . \tag{12.47}$$

This means that

$$(\bar{U}_2 - \bar{U}_1) \cdot \bar{V} = 0 \tag{12.48}$$

for all vectors $\bar{V}$. According to (5.8) this implies that $\bar{U}_2 = \bar{U}_1$. Q.E.D.

In summary the relation $U[\ ] = \bar{U}\cdot$ gives a one-to-one linear relation between vectors and dual vectors: To each vector there exists one and only one dual vector and conversely.

12.6.2 The Metric Tensor

The scalar product can be regarded as a symmetric tensor, the *metric tensor* $g[\ ,\ ]$ defined by

$$g[\bar{A}, \bar{B}] \equiv \bar{A} \cdot \bar{B} \tag{12.49}$$

for arbitrary vectors $\bar{A}$ and $\bar{B}$. Using markers as in Section 12.2 this can be written

$$g_{ab} A^a B^b = \bar{A} \cdot \bar{B} \tag{12.50}$$

and (12.43) can be written

$$U_a V^a = g_{ab} V^a U^b. \tag{12.51}$$

As $\bar{V}$ is arbitrary we may also write

$$U_a = g_{ab} U^b. \tag{12.52}$$

The metric tensor is thus used to *lower an index* changing a vector into a dual vector. Any tensor of the same character as the metric does perform the same trick but the metric is of such a central importance that it has been agreed to use the same main letter (in our example U) to denote the vector and the corresponding dual vector when the metric is used to lower the index.

From Section 12.6.1 we know that (12.52) has a unique inverse changing dual vectors into vectors. Due to the linearity this must also be effected by a tensor, the *inverse metric tensor* g^{ab}, which thus is defined by

$$U^a = g^{ab} U_b. \tag{12.53}$$

The inverse metric tensor *raises an index.* Multiplying (12.53) by an arbitrary tensor V_a an equation is obtained whose left hand side is a scalar product and thus indifferent to the exchange of $\bar{U}$ and $\bar{V}$. We conclude that *g^{ab} is a symmetric tensor.*

Any index in any tensor can be raised or lowered using g^{ab} and g_{ab} respectively. Here are two examples

$$T^a{}_{bc} = g_{bd}T^{ad}{}_c \tag{12.54}$$

$$T^a{}_{bc} = g^{ad}T_{dbc}. \tag{12.55}$$

Raising an index of the metric itself obviously gives the identity tensor

$$g^{ac}g_{cb} = \delta^a_b. \tag{12.56}$$

Chapter 13

Tensor Fields

13.1 Scalar Fields. Gradients

You have certainly met a number of *scalar fields* in ordinary Euclidean space. Pressure and temperature which vary with position are two examples. The height above sea level is another.

What characterizes a scalar field also in spacetime is that it attaches a scalar to each point in the space considered

$$\psi = \Psi(\bar{R}). \tag{13.1}$$

Here $\bar{R}$ is the position vector and ψ is the scalar attached to $\bar{R}$. The *scalar field* is the *function* Ψ, sometimes written $\Psi(\ )$. Notice the difference in typography between the scalar ψ and the scalar field Ψ.

Consider now positions on a straight line which passes through the point given by $\bar{R}$ and has the same direction as a given vector $\bar{V}$. The position vectors for points on this line are given by $\bar{R} + \lambda\bar{V}$ where λ is a parameter (see Figure 13.1). With $\bar{R}$ and $\bar{V}$ given we have reduced Ψ to a function of λ.

We are in particular interested in the derivative with respect to λ at the point $\bar{R}$, i.e., at $\lambda = 0$. This derivative is denoted $D_{\bar{V}}\Psi(\bar{R})$. Hence

$$D_{\bar{V}}\Psi(\bar{R}) \equiv \frac{d}{d\lambda}\Psi(\bar{R} + \lambda\bar{V})\bigg|_{\lambda=0} = \lim_{\lambda\to 0}[\Psi(\bar{R} + \lambda\bar{V}) - \Psi(\bar{R})]/\lambda \tag{13.2}$$

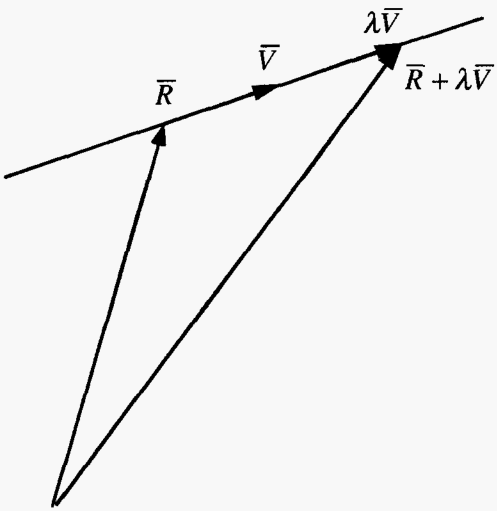

Figure 13.1: By restriction to a straight line $\bar{R} + \lambda\bar{V}$ a scalar field is reduced to a function of the parameter λ.

where we have assumed that the derivative exists and is continuous. For each given vector $\bar{V}$, $D_{\bar{V}}\Psi$ is a new function of $\bar{R}$, i.e., a new scalar field.

Suppose you would like to calculate how fast the temperature varies with the height in the atmosphere. What you do is to investigate how much it varies over a number of vertical metre sticks and then you divide by the number of sticks. Finally you go to the limit of zero number of sticks. In this way you obtain the temperature change per metre. In equation (13.2) the same calculation is performed with the difference that the vertical metre stick is generalized to a vector $\bar{V}$ which for instance could be a yard stick placed in some arbitrary direction. The quantity $D_{\bar{V}}\Psi(\bar{R})$ is thus the change of the scalar ψ *per vector* $\bar{V}$ at the point given by $\bar{R}$. This type of derivative is called *directional derivative.*

We will (in this chapter) assume that all directional derivatives of Ψ exist everywhere within the region of spacetime which we are considering. This is usually expressed as *differentiability* of Ψ. We will also assume that $D_{\bar{V}}\Psi$ (for an arbitrary vector $\bar{V}$) is differentiable, i.e., that $D_{\bar{U}}D_{\bar{V}}\Psi$ exists for arbitrary vectors $\bar{U}$ and $\bar{V}$. Furthermore we will assume that all these scalar fields are continuous.

Written out the double derivative runs

$$D_{\bar{U}}D_{\bar{V}}\Psi(\bar{R}) \equiv \frac{\partial}{\partial\mu}\left[\frac{\partial}{\partial\lambda}\Psi(\bar{R}+\mu\bar{U}+\lambda\bar{V})\right]\Bigg|_{\lambda=\mu=0}. \tag{13.3}$$

In the differentiation $\partial/\partial\lambda$ the parameter μ is treated as a constant and in the differentiation $\partial/\partial\mu$ the parameter λ is treated as a constant. We could very well contemplate further differentiation, but as a rule we are content with first and second order derivatives.

Let us now calculate

$$D_{(\bar{U}+\bar{V})}\Psi(\bar{R}) = \frac{d}{d\lambda}\Psi(\bar{R}+\lambda\bar{U}+\lambda\bar{V})\Bigg|_{\lambda=0}. \tag{13.4}$$

On the right hand side we are told to differentiate with respect to a variable λ which occurs twice in the function to be differentiated. From analysis we know that such a derivative can be written as a sum of two terms. The first one is the derivative with respect to the first λ, the second λ being treated as a constant (in our case $\lambda = 0$). In the second term the rôles are reversed. Hence

$$\begin{aligned}
&D_{(\bar{U}+\bar{V})}\Psi(\bar{R}) = \\
&\frac{d}{d\lambda}\Psi(\bar{R}+\lambda\bar{U})\Bigg|_{\lambda=0} + \frac{d}{d\lambda}\Psi(\bar{R}+\lambda\bar{V})\Bigg|_{\lambda=0} = \\
&D_{\bar{U}}\Psi(\bar{R}) + D_{\bar{V}}\Psi(\bar{R}).
\end{aligned} \tag{13.5}$$

Moreover we can calculate $D_{x\bar{V}}\Psi$ for some number x

$$\begin{aligned}
D_{x\bar{V}}\Psi(\bar{R}) &= \frac{d}{d\lambda}\Psi(\bar{R}+\lambda x\bar{V})\Bigg|_{\lambda=0} = \\
x\,\frac{d}{d\zeta}\Psi(\bar{R}+\zeta\bar{V})\Bigg|_{\zeta=0} &= xD_{\bar{V}}\Psi(\bar{R}).
\end{aligned} \tag{13.6}$$

The equations (13.5) and (13.6) show that $D_{\bar{V}}\Psi$ *is linear in* $\bar{V}$. Hence we may introduce a *tensor field*[1] $D_a\Psi$ defined through

$$V^a D_a \Psi \equiv D_{\bar{V}}\Psi \tag{13.7}$$

for all $\bar{V}$. This tensor field is usually called the *gradient* of Ψ.

Let us now demonstrate the calculation of a directional derivative in a simple example. Consider the scalar field Ψ given by

$$\psi = \Psi(\bar{R}) = \bar{R}^2 + 1/(\bar{K}\cdot\bar{R}) \tag{13.8}$$

where $\bar{K}$ is a given vector and $\bar{K}\cdot\bar{R} \neq 0$.

$$\begin{aligned} D_{\bar{V}}\Psi(\bar{R}) = \frac{d}{d\lambda}[(\bar{R}+\lambda\bar{V})^2 + 1/(\bar{K}\cdot(\bar{R}+\lambda\bar{V}))]\Big|_{\lambda=0} = \\ \frac{d}{d\lambda}[\bar{R}^2 + 2\lambda\bar{R}\cdot\bar{V} + \lambda^2\bar{V}^2 + 1/(\bar{K}\cdot\bar{R}+\lambda\bar{K}\cdot\bar{V})]\Big|_{\lambda=0} = \\ 2\bar{R}\cdot\bar{V} - (\bar{K}\cdot\bar{V})/(\bar{K}\cdot\bar{R})^2. \end{aligned} \tag{13.9}$$

Using this we obtain

$$\begin{aligned} D_{\bar{U}}D_{\bar{V}}\Psi(\bar{R}) = D_{\bar{U}}[2\bar{R}\cdot\bar{V} - (\bar{K}\cdot\bar{V})/(\bar{K}\cdot\bar{R})^2] = \\ \frac{d}{d\mu}[2(\bar{R}+\mu\bar{U})\cdot\bar{V} - (\bar{K}\cdot\bar{V})/(\bar{K}\cdot(\bar{R}+\mu\bar{U}))^2]\Big|_{\mu=0} = \\ 2\bar{U}\cdot\bar{V} + 2(\bar{K}\cdot\bar{V})(\bar{K}\cdot\bar{U})/(\bar{K}\cdot\bar{R})^3. \end{aligned} \tag{13.10}$$

The vectors $\bar{V}$ and $\bar{U}$ are arbitrary. These two results (13.9) and (13.10) can therefore be written

$$D_a\Psi(\bar{R}) = 2R_a - K_a/(\bar{K}\cdot\bar{R})^2 \tag{13.11}$$

and

[1] A tensor field attaches a tensor of some given rank to each point in spacetime. See further Section 13.2.

$$D_a D_b \Psi(\bar{R}) = 2g_{ab} + 2K_a K_b / (\bar{K} \cdot \bar{R})^3 \tag{13.12}$$

respectively.

If you are uncertain it may be a good idea to calculate directional derivatives and then translate them to gradients just as we have done it here. The calculations can, however, be shortened if you use gradients all the way. The following rules are then useful. They follow immediately from the definition of $D_{\bar{V}}\Psi(\bar{R})$ as a derivative and they are actually nothing but the well-known rules for differentiation.

- Differentiation of a sum:

$$D_a(\Psi_1 + \Psi_2) = D_a\Psi_1 + D_a\Psi_2. \tag{13.13}$$

- Differentiation of a product (Leibniz rule):

$$D_a\Psi_1\Psi_2 = \Psi_1 D_a\Psi_2 + \Psi_2 D_a\Psi_1. \tag{13.14}$$

- Differentiation of a function of a function:

$$D_a f(\Psi) = f' D_a \Psi \tag{13.15}$$

  where f' is the derivative of f with respect to its argument.

- Change of order:

$$D_b D_a \Psi = D_a(D_b \Psi). \tag{13.16}$$

Let us investigate the form that the gradient takes from the point of view of a particular observer with, say, the four-velocity $\hat{W}$. As the observer counts his time t in the direction $\hat{W}$ we have

$$D_{\hat{W}}\Psi(\bar{R}) = \frac{d}{dt}\Psi(\bar{R} + t\hat{W})|_{t=0} = \frac{\partial}{\partial t}\Psi(\bar{R}). \tag{13.17}$$

The meaning of the partial time derivative is that the position in the observer's orthogonal space is kept constant.

Using the four-velocity $\hat{W}$ we may split the vector $\Delta\bar{R}$ between two spacetime positions in the standard way (cf. (4.9))

$$\Delta\bar{R} = \Delta t\hat{W} + \Delta\bar{r} \tag{13.18}$$

where $\Delta\bar{r}\cdot\hat{W} = 0$. Using this we may define a new differentiation ∇ as follows

$$\nabla_{\Delta\bar{R}}\Psi \equiv D_{\Delta\bar{r}}\Psi. \tag{13.19}$$

On account of (13.18) the right hand side is linear i $\Delta\bar{R}$ (keeping $\hat{W}$ constant) so that a tensor $\nabla_a\Psi$ exists.

Notice that the definition (13.19) has the following consequences

$$\nabla_{\hat{W}} = 0 \quad \text{and} \quad \nabla_{\Delta\bar{r}} = D_{\Delta\bar{r}}. \tag{13.20}$$

Using the split (13.18) we obtain

$$\Delta R^a D_a\Psi = \Delta t\frac{\partial}{\partial t}\Psi + \Delta r^a\nabla_a\Psi \tag{13.21}$$

which shows that $D_a\Psi$ itself can be split as follows

$$D_a\Psi = -W_a\frac{\partial}{\partial t}\Psi + \nabla_a\Psi. \tag{13.22}$$

13.2 Tensor Fields

A *tensor field* attaches a tensor to each position vector $\bar{R}$ in spacetime, for instance

$$T^a{}_b = \mathcal{T}^a{}_b(\bar{R}) \tag{13.23}$$

where $\mathcal{T}^a{}_b(\ )$ is the tensor field. Notice the use of a script letter $\mathcal{T}$ to denote the field. Empty brackets are clumsy and they will usually not be written out in the following. It is, however, convenient, especially in calculating derivatives, to see clearly if an object is a tensor or a tensor field. For this reason we will make it a habit to use script notation for tensor fields. Due to trivial typographical limitations we cannot stick consistently to this convention. This is true in particular for scalar fields.

Consider a first rank tensor field $\mathcal{T}_a$. For each given vector K^a it gives a scalar field $\mathcal{T}_a K^a$. If this scalar field is differentiable (or continuously differentiable) to a certain order *for all* K^a we say that the tensor field $\mathcal{T}_a$ itself has this property.

The directional derivative of the scalar field $\mathcal{T}_a K^a$ with respect to some vector $\bar{V}$ is

$$(D_{\bar{V}} \mathcal{T}_a K^a)(\bar{R}) = \lim_{\lambda \to 0} [\mathcal{T}_a(\bar{R} + \lambda \bar{V}) K^a - \mathcal{T}_a(\bar{R}) K^a]/\lambda. \qquad (13.24)$$

This expression is obviously linear in $\bar{K}$. Hence we can define a new tensor field $D_{\bar{V}} \mathcal{T}_a$ through

$$(D_{\bar{V}} \mathcal{T})_a K^a \equiv D_{\bar{V}} (\mathcal{T}_a K^a) \qquad (13.25)$$

for an arbitrary vector $\bar{K}$. As found in Section 13.1 the right hand side here is linear in $\bar{V}$ and a second-rank tensor field $D_a \mathcal{T}_b$ can be defined through

$$D_a \mathcal{T}_b \, V^a K^b \equiv D_{\bar{V}} (\mathcal{T}_b K^b) \qquad (13.26)$$

for arbitrary vectors $\bar{V}$ and $\bar{K}$. This equation can almost trivially be generalized to tensors of arbitrary rank. Let us somewhat symbolically denote a general tensor as $T^{a\dots}_{b\dots}$. Then

$$\begin{aligned} &(D_c \mathcal{T})^{a\dots}_{b\dots} \, V^c W_a \dots K^b \dots = \\ &D_{\bar{V}} (\mathcal{T}^{a\dots}_{b\dots} W_a \dots K^b \dots) \end{aligned} \qquad (13.27)$$

where $K^a \dots$ are arbitrary vectors and $W_a \dots$ are arbitrary dual vectors so many in number that $\mathcal{T}$ is completely 'filled up'. The tensor field $(D_c \mathcal{T})^{a\dots}_{b\dots}$ is known as the *gradient* of $\mathcal{T}$.

This definition delivers immediately a general technique for calculating the directional derivative and the gradient of a tensor field: *Introduce a number of arbitrary auxiliary vectors and dual vectors converting* *the tensor field to a scalar field which you know how to differentiate.* You can eliminate these arbitrary vectors and dual vectors from the final result using (13.27) and obtain the required gradient. To follow this instruction to the letter may be unnecessarily tedious. There are a few observations which can make life easier for you and we turn to them now.

• *The Leibniz rule.* In the following example (13.27) has been used twice

$$\begin{aligned}&\{D_c(\mathcal{A}^a\mathcal{B}_b)\}V^cM_aN^b = D_{\bar{V}}\{(\mathcal{A}^aM_a)(\mathcal{B}_bN^b)\} =\\&(\mathcal{A}^aM_a)D_{\bar{V}}(\mathcal{B}_bN^b) + (\mathcal{B}_bN^b)D_{\bar{V}}(\mathcal{A}^aM_a) =\\&\{\mathcal{A}^aD_c\mathcal{B}_b + \mathcal{B}_bD_c\mathcal{A}^a\}V^cM_aN^b.\end{aligned} \tag{13.28}$$

Here V^c, M_a, and N_b are arbitrary. Hence

$$D_c(\mathcal{A}^a\mathcal{B}_b) = \mathcal{A}^aD_c\mathcal{B}_b + \mathcal{B}_bD_c\mathcal{A}^a. \tag{13.29}$$

This indicates that there is a shortcut for differentiation of a product of tensors. There is in fact an obvious generalization of (13.29): $\mathcal{A}$ and $\mathcal{B}$ may be replaced by tensors of arbitrary rank provided the indices follow unchanged with their tensorial factors in all the terms of the equation. We express this by stating that *in calculating the directional derivative or the gradient of a product of scalar and tensor fields the Leibniz rule may be used.* Compare (13.14) and (13.29).

• *Commutation of differentiation with contraction.* In order to investigate the behaviour of contraction under differentiation consider the directional derivative of a second-rank tensor field $\mathcal{T}^a{}_b$

$$\begin{aligned}&D_{\bar{V}}\mathcal{T}^a{}_b(\bar{R})S_aB^b =\\&\lim_{\lambda\to 0}[(\mathcal{T})^a{}_b(\bar{R}+\lambda\bar{V})S_aB^b - (\mathcal{T})^a{}_b(\bar{R})S_aB^b]/\lambda.\end{aligned} \tag{13.30}$$

Remember now equation (12.40) and choose S_a and B^b in (13.30) successively as the four pairs of dual vector and vector appearing on the right hand side of (12.40). Then add the four equations so obtained.

This obviously leads to an equation which is the same as (13.30) except that $S_a B^b$ has been replaced by δ^b_a everywhere. Equation (12.39) then tells us that δ^a_b can be omitted if we replace the index b by a. Thus we obtain

$$(D_{\bar V}\mathcal{T})^a{}_a(\bar R) = \lim_{\lambda\to 0}[\mathcal{T}^a{}_a(\bar R+\lambda\bar V)-\mathcal{T}^a{}_a(\bar R)]/\lambda = D_{\bar V}(\mathcal{T}^a{}_a)(\bar R). \tag{13.31}$$

The conclusion is that *the order between differentiation and contraction is of no consequence.*

• *Commutation of differentiation with the raising and lowering of indices.* With the two vectors M^a and N^a given the scalar $g_{ab}M^aN^b = \bar M\cdot\bar N$ is a given constant. Hence $D_c g_{ab}M^aN^b = 0$ for arbitrary vectors M^a and N^b, which gives

$$D_c g_{ab} = 0. \tag{13.32}$$

The metric and also the inverse metric are constant tensors. For this reason you may freely move indices up and down inside and outside of differentiations. For instance

$$\mathcal{A}^a D_c\mathcal{B}_a = \mathcal{A}_a D_c\mathcal{B}^a. \tag{13.33}$$

A *divergence* is a gradient in which a contraction has been formed involving the gradient index. The differential operator D^aD_a is known as the *d'Alembertian* and denoted $D^aD_a = \Box$.

Chapter 14

Spacetime Volumes

14.1 Volume Without Metric

14.1.1 Basic Concepts. Parallelepipeds

Consider some homogeneous radioactive material and a spacetime region within the world tube of this material. A count of the number of decays within this region gives a measure of the *spacetime volume* of the region in question in comparison to other regions within the same world tube. As long as we consider only the relative size of volumes it is unnecessary to employ the metric and we will find that the theory of spacetime volumes is exactly like the theory of volumes in ordinary space except for the number of dimensions. For this reason we shall postpone using the metric until Section 14.2.

In most of the present section we shall investigate the volumes of parallelepipeds[1]. In spacetime a parallelepiped is specified by four spacetime vectors. Let us choose a standard parallelepiped specified by the linearly independent vectors $\bar{A}$, $\bar{B}$, $\bar{C}$, and $\bar{D}$ and assume that the spacetime volume of this parallelepiped is V, measured for instance as the number of decays.

Consider now the parallelepiped specified by $\bar{A}'$, $\bar{B}$, $\bar{C}$, and $\bar{D}$. Notice that all the vectors except the first one are the same as in the

[1]Strictly speaking the term should rather be 'hyperparallelepiped' in the four-dimensional case. We will, however, use the term 'parallelepiped' generically, valid for all dimensions.

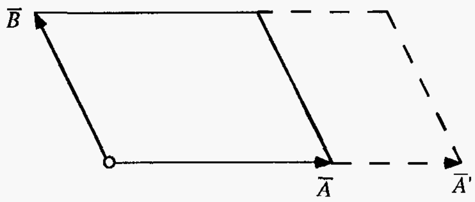

Figure 14.1: Keeping three of the edge vectors in a parallelepiped unchanged and multiplying the fourth one by a multiplies the volume of the parallelepiped by a. Two of the unchanged edge vectors are suppressed in the figure.

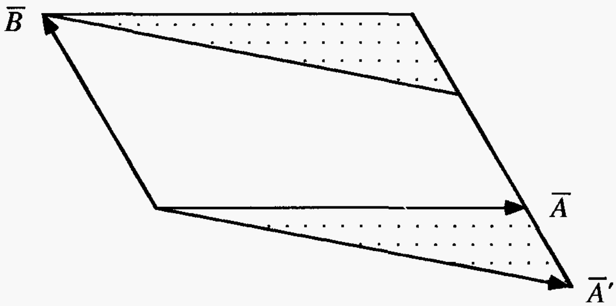

Figure 14.2: Keeping three of the edge vectors of a parallelepiped constant and adding a number times one of them to the fourth edge vector keeps the volume unchanged, because equal volumes have been added and subtracted. The two edge vectors which do not take part in this operation are suppressed in the figure.

former set. In Figure 14.1 $\bar{A}'$ has the form $\bar{A}' = a\bar{A}$. In this case the volume is obviously aV. In Figure 14.2 $\bar{A}'$ has the form $\bar{A}' = \bar{A} + b\bar{B}$. In this case the volume is obviously unchanged equal to V. Continue this reasoning and you will find that if $\bar{A}'$ has the form

$$\bar{A}' = a\bar{A} + \bar{Q} \tag{14.1}$$

where $\bar{Q}$ is a linear combination of $\bar{B}$, $\bar{C}$, and $\bar{D}$, then the volume is equal to aV. Thus the volume is a linear function of the vector $\bar{A}'$ everything else kept constant. As a matter of fact we recognize it as proportional to the standard first rank tensor in (12.5). In the same fashion the volume is found to be linear in the other three edge vectors. Hence *the volume is obtained from a fourth rank tensor.* We will denote this tensor as $\varepsilon[\;,\;,\;,\;]$ so that, e.g.,

$$V = \varepsilon[\bar{A}, \bar{B}, \bar{C}, \bar{D}]. \tag{14.2}$$

In this section we are not interested in determining ε (and V) more than up to a scalar factor. Such a factor will always cancel when we calculate *ratios* between volumes.

The linearity of the volume tensor tells us among other things that the volume changes sign if one vector among its arguments changes sign. Taking the absolute value would destroy the linearity and complicate the theory considerably. We must therefore accept the occurrence of negative volumes. This should not be interpreted to mean that the number of radioactive decays mentioned above could be negative. It simply means that formally adding a negative volume is the same as subtracting a positive volume.

Choosing $\bar{A}' = \bar{B}$ equation (14.1) gives $a = 0$ and thus the volume equal to zero. This tells us that *if two edge vectors are equal the volume vanishes*, e.g.,

$$\varepsilon_{abcd} U^a U^c = 0. \tag{14.3}$$

This is true for all vectors $\bar{U}$. Now choose $\bar{U}$ in particular to be $\bar{U} = \bar{K} + \bar{L}$ where $\bar{K}$ and $\bar{L}$ are two arbitrary vectors. Inserting this into (14.3) utilizing the linearity and then using (14.3) again we find

$$\varepsilon_{abcd}\left(K^a L^c + L^a K^c\right) = 0 \tag{14.4}$$

or since $\bar{K}$ and $\bar{L}$ are arbitrary

$$\varepsilon_{abcd} = -\varepsilon_{cbad}. \tag{14.5}$$

Obviously ε_{abcd} is antisymmetric in every pair of indices, it is *totally antisymmetric.* This is the same symmetry as a determinant exhibits

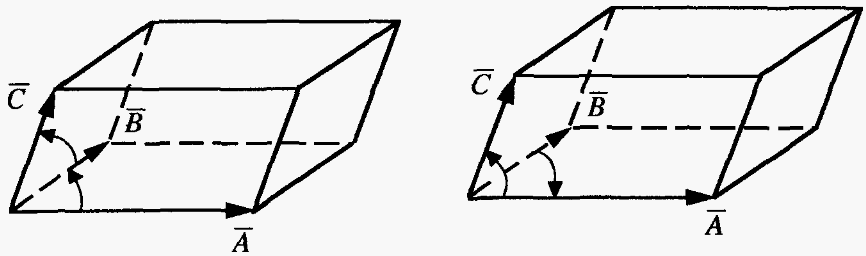

Figure 14.3: Two different ordered parallelepipeds specified by the vectors $\bar{A}$, $\bar{B}$, $\bar{C}$ and $\bar{B}$, $\bar{A}$, $\bar{C}$ respectively. Read the vectors in the order indicated by the small hooked arrows.

under exchanges of rows and exchanges of columns. Notice that conversely (14.3) follows from the antisymmetry. Assuming antisymmetry the left hand side of (14.3) must both be unchanged and change sign when the $\bar{U}$-vectors are interchanged. This can be true only if it is zero.

The fact that the volume changes sign when two edge vectors are interchanged shows that *the order between the edge vectors is important.* When the order between the edge vectors is recorded we will speak about an *ordered parallelepiped.* If you wish to note the ordering in a figure you may do so with small hooked arrows. See Figure 14.3.

14.1.2 Calculation of Volumes

The actual calculation of the ratio between two volumes can proceed as follows: Assume that you are interested in the relative volumes of two ordered parallelepipedes specified by $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$ and $\bar{A}'$, $\bar{B}'$, $\bar{C}'$, $\bar{D}'$. Then express the primed vectors as linear combinations of the unprimed vectors (assuming that these are linearly independent) and insert the expressions in the volume $V' = \varepsilon[\bar{A}', \bar{B}', \bar{C}', \bar{D}']$. Expand this using the linearity and you obtain a sum of terms each containing ε with the vectors $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$ inserted in some order. In each of these you must rearrange the vectors by pairwise exchanges so that you obtain $\varepsilon[\bar{A}, \bar{B}, \bar{C}, \bar{D}]$, which is V. Performing these exchanges you must keep track of the sign changes. The final result is obviously proportional to V. It can often save quite a lot of labour using the following rule one or more times in the process: *A vector occurring in one of the entries of*

ε can be added with an arbitrary coefficient to any other entry without changing the value, e.g.,

$$\varepsilon_{abcd}U^aV^bX^cY^d = \varepsilon_{abcd}U^aV^bX^c(Y^d + kX^d). \tag{14.6}$$

This rule follows immediately from linearity and (14.3). A corollary is that *$\varepsilon[\bar{U},\bar{V},\bar{X},\bar{Y}] = 0$ is the necessary and sufficient condition for the vectors $\bar{U},\bar{V},\bar{X},\bar{Y}$ to be linearly dependent* (provided that ε is not the null tensor, a case which is pretty trivial). In essence you probably know this already from the theory of determinants and below you can find a short discussion of this and other consequences of (14.6).

Consider the set of vectors $\bar{U},\bar{V},\bar{X},\bar{Y}$. A number of procedures like (14.6) changes this to a set $\bar{U},\bar{V},\bar{X},\bar{Z}$ with $\bar{Z} = \bar{Y} + u\bar{U} + v\bar{V} + x\bar{X}$.

- Obviously linear dependence and independence is not changed. If the vectors in the first set are linearly dependent (independent) then the vectors in the second set are also linearly dependent (independent).
- Assume linear independence and divide the vectors into two groups. Using the above procedure we may then replace one of the vectors in one of the groups with a vector which is parallel to any given vector which is linearly independent of the vectors in the other group. An example should be sufficient to convince you: Assume that $\bar{K}$ is linearly independent of $\bar{U}$ and $\bar{V}$. In the expansion $\bar{K} = u\bar{U} + v\bar{V} + x\bar{X} + y\bar{Y}$ the coefficients x and y cannot then both be zero. Assume that $y \neq 0$. We may then replace $\bar{Y}$ with $\bar{Y} + u/y\,\bar{U} + v/y\,\bar{V} + x/y\,\bar{X}$ which is parallel to $\bar{K}$.
- With a series of such replacements we can obtain

  $$\varepsilon[\bar{U},\bar{V},\bar{X},\bar{Y}] = k\varepsilon[\bar{A},\bar{B},\bar{C},\bar{D}]$$

  where the two sets of vectors are arbitrary sets of linearly independent vectors and k is a finite non-zero number.
- Assuming linear dependence $\varepsilon[\bar{U},\bar{V},\bar{X},\bar{Y}] = 0$, because the vectors may be combined to the null vector.
- Provided that ε is not the null-tensor it is obvious from the points three and four that $\varepsilon[\bar{U},\bar{V},\bar{X},\bar{Y}] = 0$ is the necessary and sufficient condition for the set of vectors to be linearly dependent.

You should notice that in the directions for use given above no other properties of ε are utilized than linearity and antisymmetry. Hence any completely antisymmetric non-null tensor of rank four could serve as volume tensor in spacetime as long as we are interested only in ratios between volumes. In three dimensions the volume tensor must of course be of rank three. A completely antisymmetric tensor of rank three in ordinary space is the triple scalar product. It is therefore a legitimate volume tensor.

The following is a more systematic but often less practical way to express the above directions for calculating volumes. Let $\bar{A}_0$, $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ (in this order) specify an ordered parallelepiped and the vectors $\bar{A}'_0$, $\bar{A}'_1$, $\bar{A}'_2$, $\bar{A}'_3$ another one. The vectors within each set are supposed to be linearly independent. Hence the vectors $\bar{A}'_r$ can be written as linear combinations of the vectors $\bar{A}_s$ (and conversely)

$$\bar{A}'_r = \sum_{r=0}^{3} C_{rs}\bar{A}_s. \tag{14.7}$$

Here the coefficients C_{rs} form a 4 by 4 non-singular matrix. They are uniquely given by the two ordered parallelepipeds. The volume of the primed one is

$$V' = \varepsilon[\bar{A}'_0, \bar{A}'_1, \bar{A}'_2, \bar{A}'_3] = \sum_{m=0}^{3}\sum_{n=0}^{3}\sum_{p=0}^{3}\sum_{q=0}^{3} C_{0m}C_{1n}C_{2p}C_{3q}\varepsilon[\bar{A}_m, \bar{A}_n, \bar{A}_p, \bar{A}_q]. \tag{14.8}$$

The only terms surviving on the right hand side are those for which m, n, p, and q are all different. The ε-factor for each of them is either V or $-V$ depending on if m, n, p, q is an even or an odd permutation of $0, 1, 2, 3$. Hence the definition of the determinant appears on the right hand side and we obtain

$$V' = |C_{rs}|\ V. \tag{14.9}$$

14.1.3 Sums of Volumes

One of the most important properties of the volume is that the volume of a region is the sum of the volumes of its parts, assuming that all the volumes are relatively positive. No matter how you divide a region into parallelepipeds their volumes always add up to the same value. This property is very well-known from ordinary space, in fact it is usually taken for granted. It is valid also in spacetime but we shall spare you a detailed proof presenting only a sketch.

One way to approach this theorem is to define the volume primarily for triangles, tetrahedrons, and hypertetrahedrons (in two, three, and four dimensions respectively) rather than for parallelepipedes. Such an object is generically known as a *simplex*. At each vertex of a simplex d edge vectors meet, where d is the number of dimensions. Consider some vertex and insert the corresponding edge vectors into the ε-tensor. The so obtained number multiplied by $1/d!$ is by definition the volume of the simplex. It can then be seen that

- this definition is consistent in the sense that when you divide a simplex into simplices the volume is additive.
- when you divide a parallelepiped into simplices the sum of the volumes comes out right.
- any polyhedron can be divided into simplices and if you have divided a polyhedron in two different ways you can always find a division which is a finer division of both the former ones. Therefore the volume of a polyhedron is unique.

14.1.4 Volumes and Coordinates

Coordinates are scalar fields used to label space (or spacetime) points. For a complete labelling as many fields are required as the space has dimensions. On a *coordinate line* all the coordinates except one are constant. The coordinate lines form a mesh dividing the space into infinitesimal parallelepipeds. To be able to draw simple figures let us study the two-dimensional case. The generalization to more dimensions is almost trivial.

In Figure 14.4 we consider an infinitesimal parallelogram specified by the vectors $\overline{dx}$ in the direction of constant y and $\overline{dy}$ in the direction of constant x. In two dimensions the volume tensor $\varepsilon[\ ,\ ]$ is of rank two and the volume (called 'area' in two dimensions) of the parallelogram is

$$dV = \varepsilon[\overline{dx}, \overline{dy}] = \varepsilon[\bar{\partial}_x, \bar{\partial}_y] dx\, dy \tag{14.10}$$

where dx and dy are the coordinate differences between head and tail of the vectors $\overline{dx}$ and $\overline{dy}$ respectively. The vector fields $\bar{\partial}_x$ and $\bar{\partial}_y$ are

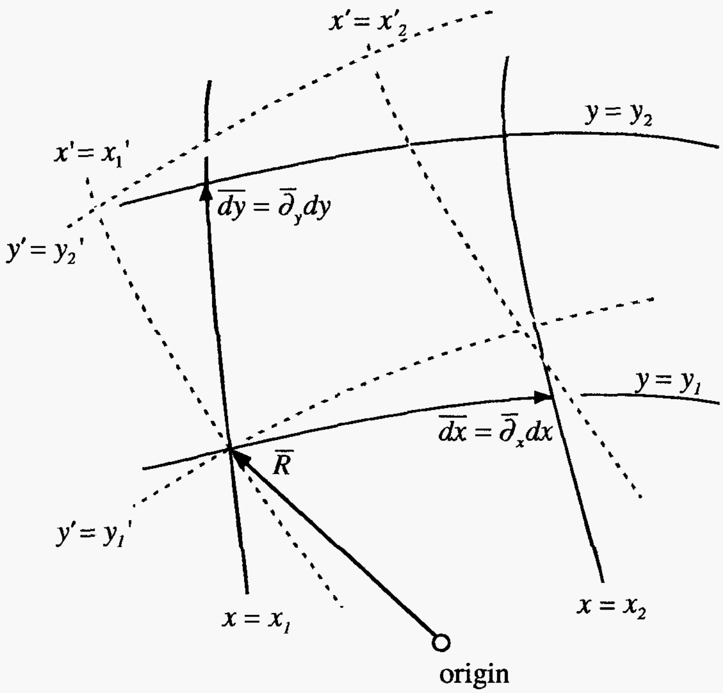

Figure 14.4: Two sets of coordinate lines in two-dimensional space.

defined through $\overline{dx} = \bar{\partial}_x dx$ and $\overline{dy} = \bar{\partial}_y dy$. They are obviously uniquely defined by the two scalar fields x and y. At each point of space they form a vector basis known as the *coordinate basis*. Please notice that the indices x and y have a different meaning from the abstract indices introduced in Section 12.2.2. You may regard the choice of the notation $\bar{\partial}_x$ and $\bar{\partial}_y$ as a mnemonic for remembering how the basis is changed when the coordinates are changed. Working with the coordinates ξ and η as well as x and y you have

$$\bar{\partial}_x = \frac{\partial \xi}{\partial x}\bar{\partial}_\xi + \frac{\partial \eta}{\partial x}\bar{\partial}_\eta \tag{14.11}$$

The partial derivatives occurring here are scalar fields. It is as usual understood that taking the derivative $\partial\xi/\partial x$ (at some point $\bar{R}$) y is kept constant and taking $\partial\eta/\partial y$ x is kept constant. The equation (14.11) looks very much like the relation (14.14) well known from analysis. The latter one is in fact obtained if you contract (14.11) with D_a. A more detailed account of this is given at the end of the present subsection.[2]

Quite often one wishes to convert a volume integral to a multiple integral over coordinates. This is done with (14.10). To calculate $\varepsilon[\bar{\partial}_x, \bar{\partial}_y]$ it is often best to express $\bar{\partial}_x$ and $\bar{\partial}_y$ in terms of some standard set of vectors using e.g. (14.11) and then use the rules in Section 14.1.2.

If Ψ is an arbitrary scalar field we obtain

$$\frac{\partial \Psi(\bar{R})}{\partial x} = \frac{1}{dx}\frac{\partial \Psi(\bar{R})}{\partial x}dx = \frac{1}{dx}D_{\bar{dx}}\Psi(\bar{R}) = D_{\bar{\partial}_x}\Psi(\bar{R}) \tag{14.12}$$

and the corresponding equation for $\partial\Psi(\bar{R})/\partial y$. As Ψ is an arbitrary scalar field these equations can also be written

$$\partial/\partial x = \partial_x^a D_a, \quad \partial/\partial y = \partial_y^a D_a. \tag{14.13}$$

Suppose that you employ one more set of coordinates ξ, η and that x, y are functions of these. Then

$$\frac{\partial}{\partial x} = \frac{\partial \xi}{\partial x}\frac{\partial}{\partial \xi} + \frac{\partial \eta}{\partial x}\frac{\partial}{\partial \eta}. \tag{14.14}$$

Consulting (14.13) we see that this can be written

$$\partial_x^a D_a = \frac{\partial \xi}{\partial x}\partial_\xi^a D_a + \frac{\partial \eta}{\partial x}\partial_\eta^a D_a. \tag{14.15}$$

Hence

$$\bar{\partial}_x = \frac{\partial \xi}{\partial x}\bar{\partial}_\xi + \frac{\partial \eta}{\partial x}\bar{\partial}_\eta \tag{14.16}$$

and the corresponding equation

$$\bar{\partial}_y = \frac{\partial \xi}{\partial y}\bar{\partial}_\xi + \frac{\partial \eta}{\partial y}\bar{\partial}_\eta. \tag{14.17}$$

According to (14.9) the ratio between the two volumes $\varepsilon[\bar{\partial}_x, \bar{\partial}_y]$ and $\varepsilon[\bar{\partial}_\xi, \bar{\partial}_\eta]$ is thus

$$\varepsilon[\bar{\partial}_x, \bar{\partial}_y]/\varepsilon[\bar{\partial}_\xi, \bar{\partial}_\eta] = J \tag{14.18}$$

where J is the determinant

[2] In more advanced differential geometry the vector field $\bar{\partial}_x$ is identified with the differential operator $\partial/\partial x$. It requires some experience before one gets used to this identification and we refrain from doing it in this course.

$$J = \begin{vmatrix} \partial\xi/\partial x & \partial\eta/\partial x \\ \partial\xi/\partial y & \partial\eta/\partial y \end{vmatrix} \tag{14.19}$$

known as the *Jacobian*. For the parallelogram given by $\overline{d\xi}$, $\overline{d\eta}$ to have the same volume as $\overline{dx}$, $\overline{dy}$, i.e.,

$$dV = \varepsilon[\overline{dx}, \overline{dy}] = \varepsilon[\overline{d\xi}, \overline{d\eta}] \tag{14.20}$$

we must according to (14.10) choose $d\xi$, $d\eta$ so that

$$d\xi d\eta = J dx dy. \tag{14.21}$$

This is the well-known rule for translating a volume integral from one set of coordinates to another set of coordinates. Notice that if J happens to be negative one or more of the coordinate differentials have to be negative.

14.2 Volume and Metric. Orientation

Consider equation (14.7) once more, forming all the scalar products between the primed $\bar{A}$-vectors

$$\bar{A}'_r \cdot \bar{A}'_s = \sum_{m=0}^{3} \sum_{n=0}^{3} C_{rm} C_{sn} \bar{A}_m \cdot \bar{A}_n. \tag{14.22}$$

Defining G' to be the 4 by 4 matrix whose elements are given by the left hand side of this equation and G as the corresponding matrix containing the unprimed vectors, the equation can be written as a matrix equation $G' = CG\tilde{C}$. Here C is the matrix with the elements C_{rs} and $\tilde{C}$ is the transpose of this matrix. Using the rule which says that the determinant of a product of matrices is the product of the corresponding determinants we thus conclude that $|G'| = |C|^2|G|$. We have here also used the fact that the determinant of a transposed matrix is the determinant of the matrix itself. Remembering (14.9) we thus obtain the equation

$$V'^2/|\bar{A}'_r \cdot \bar{A}'_s| = V^2/|\bar{A}_r \cdot \bar{A}_s|. \tag{14.23}$$

This equation shows that all determinants of the form $|\bar{A}_r \cdot \bar{A}_s|$ must have the same sign assuming only that $\bar{A}_r$ form a set of four linearly independent vectors. To investigate which sign it is let us choose a particular set of vectors $\bar{E}_r$ consisting of one timelike unit vector and three spacelike vectors orthogonal to it, assuming that the latter ones are orthogonal unit vectors. With this choice the matrix $(\bar{E}_r \cdot \bar{E}_s)$ is

$$(\bar{E}_r \cdot \bar{E}_s) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14.24}$$

with the corresponding determinant equal to -1. Thus $|\bar{A}_r \cdot \bar{A}_s|$ is always negative (or zero, cf. Chapter 5). In Section 14.1 we considered only ratios between volumes. We can now normalize the volume by requiring the members of equation (14.23) to be equal to minus unity, i.e.,

$$V^2 = -|\bar{A}_r \cdot \bar{A}_s|. \tag{14.25}$$

The minus sign here is characteristic for spacetime.

This important equation is actually a special case of an equation which can be obtained if we consider *three* sets of vectors. Apart from the unprimed and the primed sets we now also introduce a double-primed set. The volumes are denoted V, V', and V'' respectively. We construct the quantities $\bar{A}_r'' \cdot \bar{A}_s'$, expand both $\bar{A}_r''$ and $\bar{A}_s'$ in the set $\bar{A}_m$, and proceed just like before now obtaining

$$V'V''/|\bar{A}_r'' \cdot \bar{A}_s'| = V^2/|\bar{A}_r \cdot \bar{A}_s| \tag{14.26}$$

instead of (14.23). With the normalization (14.25) this gives

$$V'V'' = -|\bar{A}_r' \cdot \bar{A}_s''|. \tag{14.27}$$

This is the generalization of (14.25) mentioned a few lines above.

The only restriction assumed on the vectors occurring in (14.27) is the one on linear independence. It is, however, obvious that if a linear dependence would occur in either set of vectors in (14.27) both members of the equation would vanish so that the equation would be trivially satisfied in such a case. It is therefore valid for eight arbitrary vectors $\bar{A}_r'$ and $\bar{A}_r''$. We may therefore eliminate the vectors and write

(14.27) as a tensor equation using (14.2) on the left hand side and using the definition of determinants on the right hand side. The result is

$$\varepsilon_{abcd}\varepsilon_{efgh} = -(g_{ae}g_{bf}g_{cg}g_{dh} + \ldots) \tag{14.28}$$

where the term written out on the right hand side is only the first of 24 terms. The others are obtained from the first one by permuting the indices (efgh) in all possible ways writing a plus sign in the case of an even and a minus sign in case of an odd permutation. This equation and contractions of it are often used in applications. Contraction over one pair of indices and use of (12.56) and (12.42) gives the following equation

$$\varepsilon^{abcd}\varepsilon_{afgh} = \\ -(\delta^b_f\delta^c_g\delta^d_h + \delta^b_h\delta^c_f\delta^d_g + \delta^b_g\delta^c_h\delta^d_f - \delta^b_g\delta^c_f\delta^d_h - \delta^b_h\delta^c_g\delta^d_f - \delta^b_f\delta^c_h\delta^d_g). \tag{14.29}$$

It may pay to study the right hand side and memorize how the total antisymmetry over (fgh) is achieved. A second contraction gives

$$\varepsilon^{abcd}\varepsilon_{abgh} = -2(\delta^c_g\delta^d_h - \delta^c_h\delta^d_g), \tag{14.30}$$

a third contraction gives

$$\varepsilon^{abcd}\varepsilon_{abch} = -6\delta^d_h, \tag{14.31}$$

and a fourth contraction gives

$$\varepsilon^{abcd}\varepsilon_{abcd} = -24. \tag{14.32}$$

The normalization (14.25) of the volume leaves the sign of the volume, i.e., the sign of the volume tensor, undetermined. Therefore only the relative sign between volumes has a meaning so far. We can, however, arbitrarily choose a reference — one particular body, e.g., a man, with a future directed vector pointing along its world tube and three vectors in the direction of sticks or extremities pointing out from the body — and define the corresponding volume to be positive. This will remove the last indeterminacy from ε. From a physical point of view such a procedure is trivial. It is obviously tantamount to deciding that the 'absolute' sign is the sign relative to the reference. From a mathematical point of view a new structure is introduced with the reference. It is known as *orientation* of spacetime. The content of this structure is that a sign is given to the volume of each ordered parallelepiped, making it meaningful to speak about positive and negative orientation of parallelepipeds. In the following chapters an orientation of spacetime will be understood.

The volume of a parallelepiped with linearly independent edge vectors is different from zero (see Section 14.1.2). If you therefore change an ordered parallelepiped *continuously* keeping the edge vectors linearly independent the volume will not change sign. This can be utilized to spread an orientation reference among observers just physically accelerating and rotating a reference body from one state to another. When you compare different persons this way you can reach an agreement on the meaning of right and left and you make the remarkable discovery that the great majority of men and women prefer to throw a ball with the same hand (called the right hand). Mankind is therefore not orientation symmetric. A physically interesting question is if there exists a deep asymmetry of this kind in nature but that is not a main issue of this course. Only in connection with electrodynamics will we have a short discussion on this point (see Section 16.3).

14.3 Hypersurface Volumes

In the study of three dimensions one often considers the area given by two edge vectors. Similarly in spacetime problems one is often interested in the volume given by three edge vectors, say $\bar{A}$, $\bar{B}$, and $\bar{C}$. The *hypersurface element* given by these vectors is defined as the dual vector

$$\Sigma_a \equiv \varepsilon_{abcd} A^b B^c C^d. \tag{14.33}$$

The hypersurface volume squared is defined as $|g^{ab}\Sigma_a\Sigma_b|$. In the case that all three vectors $\bar{A}$, $\bar{B}$, and $\bar{C}$ are orthogonal to one and the same timelike direction it is also possible to generally define the sign of the hypersurface volume. In that case we define the volume as

$$a \equiv U^a \Sigma_a = U^a \varepsilon_{abcd} A^b B^c C^d \tag{14.34}$$

where U^a is the future directed unit vector orthogonal to the vectors $\bar{A}$, $\bar{B}$, and $\bar{C}$. Also Σ^a is orthogonal to these vectors. Therefore $\hat{U}$ and $\bar{\Sigma}$ are parallel and (14.34) can alternatively be written

$$U_a = -\frac{1}{a}\Sigma_a = -\frac{1}{a}\varepsilon_{abcd}A^b B^c C^d. \tag{14.35}$$

Notice the sign.

Another consequence of (14.34) is that

$$\varepsilon_{bcd} \equiv U^a \varepsilon_{abcd} \tag{14.36}$$

functions as the volume tensor in the space orthogonal to $\hat{U}$. On the left hand side a reference to $\hat{U}$ is understood. Using ε_{bcd} we can form not only the volume in three-dimensional space but also the cross product: $\bar{c} = \bar{a} \times \bar{b}$ is given by $c_a = \varepsilon_{abc}a^b b^c$.

Chapter 15

Currents

15.1 Particle Flow. Four-Current Density

Consider a great number of particles moving with respect to each other but so that neighbour particles are (almost) at rest relative to one another. When the microscopic structure can be neglected this may be regarded as continuous flowing matter. A certain collection of neighbouring particles form a tube of world lines (see Figure 15.1).

If no creation or destruction of particles takes place the number of particle world-lines is the same through every section of the world tube. In this case the number of particles is said to be *conserved.*

The same argument can be applied to a situation in which particles close to each other can have high velocities but collide very often. In such a case we are interested in the average positions and average velocities of the particles — the *macroscopic flow.*

An observer interprets the number of lines penetrating a surface in typically two different ways depending on the type of surface. In Figure 15.2 is illustrated how a world tube penetrates two typical surfaces defined by the observer with four-velocity $\hat{U}$. One of the surfaces S_1 is the observers orthogonal space at the time $t = 0$. The other one S_2 is $x = 0$ which is a plane containing the vector $\hat{U}$ itself. Actually these planes are three-dimensional but we have (of course) had to suppress one dimension in the figure. With suitable reconnaissance the observer

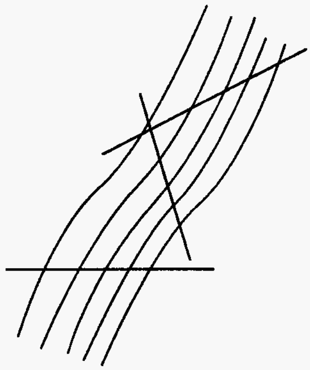

Figure 15.1: A collection of world lines form a tube in spacetime. No matter how this tube is cut, the number of lines through the section is the same, if particles are neither destroyed nor created.

interprets the number of world lines crossing S_1 as the number of particles in his three-dimensional orthogonal space at time $t = 0$. On the other hand he interprets the number of world lines crossing S_2 as the number of particles passing the surface $x = 0$ in his orthogonal space irrespective of at what time they pass.

Consider now a small parallelepiped somewhere in a tube of particle world lines (see Figure 15.3). One of its edge vectors is chosen to be $\lambda\hat{W}$ where λ is a small positive number and $\hat{W}$ is the four-velocity of the particles at the site of the parallelepiped. The remaining three edge vectors are denoted $\bar{A}$, $\bar{B}$, and $\bar{C}$. They are assumed to be small and ordered so that the volume V of the parallelepiped is positive, i.e.,

$$\varepsilon[\hat{W}, \bar{A}, \bar{B}, \bar{C}] > 0. \tag{15.1}$$

Imagine just for the sake of the argument that each particle twinkles once per unit proper time. The number of twinkles N_T in our parallelepiped is then obviously equal to λ times the number of particle lines N, i.e., $N_T = \lambda N$. The number of twinkles inside the parallelepiped is of course proportional to the volume V and we may introduce a spacetime density of twinkles

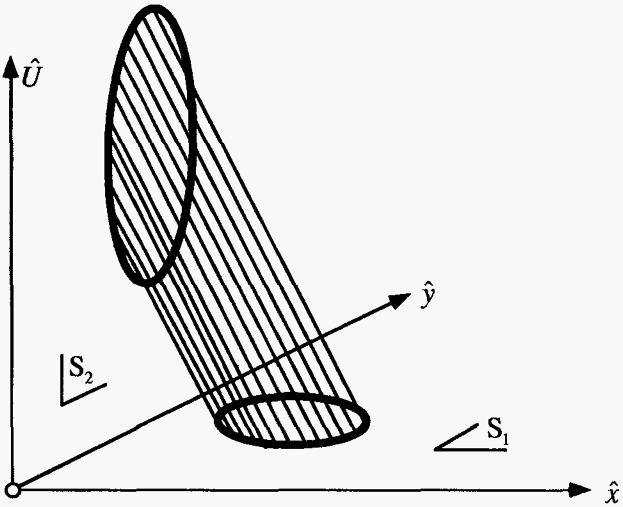

Figure 15.2: The observer with the four-velocity $\hat{U}$ interprets the flux of particle world lines in different ways depending on the direction of the surface through which the lines pass. The flux through the orthogonal space S_1 at time $t = 0$ is interpreted as the number of particles present at that time. The flux through S_2, i.e., $x = 0$, is interpreted as the number of particles passing through the surface $x = 0$ in the observer's orthogonal space, irrespective of at what time they pass.

$$\rho \equiv N_T/V = N/\varepsilon[\hat{W}, \bar{A}, \bar{B}, \bar{C}]. \tag{15.2}$$

As the number N of particle lines is positive ρ is also positive.

Introducing the *four-current density*

$$\bar{J} \equiv \rho \hat{W} \tag{15.3}$$

equation (15.2) can be written

$$N = \varepsilon[\bar{J}, \bar{A}, \bar{B}, \bar{C}] = J^a d\Sigma_a \tag{15.4}$$

where $d\Sigma_a$ is the *hypersurface element*

$$d\Sigma_a = \varepsilon_{abcd} A^b B^c C^d. \tag{15.5}$$

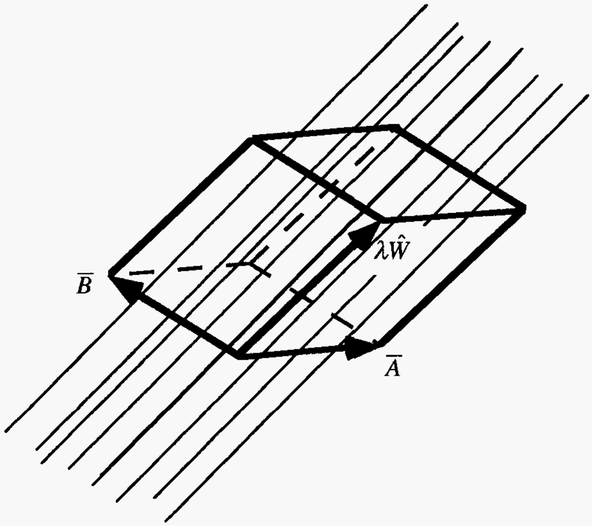

Figure 15.3: A small parallelepiped is placed in a particle flow. One of its edges is in the direction of the local particle four-velocity. The orientation is chosen so that the volume is positive.

The four-current density is determined entirely by the particle flow while the surface element is independent of the flow except for its sign. According to (15.1) we have chosen the sign of the hypersurface element to be such that $J^a d\Sigma_a > 0$. With this proviso (15.4) tells us that *the number of particle world lines through a hypersurface element is given by the contraction of the hypersurface element with the four-current density.*

A special case of this is that $\varrho_U \equiv -\hat{U} \cdot \bar{J}$ *is the density of lines, i.e., lines per volume, through the space orthogonal to* $\hat{U}$, where $\hat{U}$ is a four-velocity. Use equation (14.35) to verify this. The observer with the four-velocity $\hat{U}$ interprets ϱ_U as the density of particles in his orthogonal space. Please notice the difference between ρ and ϱ. The two quantities thus denoted are closely related but there is also a distinct difference. The first notation ρ is used only in connexion with a simple particle flow where there is at each point a four-velocity $\hat{W}$ defined by the system itself.[1] Then $\rho = -\hat{W} \cdot \bar{J}$ as can be seen by (15.3). The second notation ϱ refers to an observer. Sometimes the four-velocity of the observer is indicated with an index as in ϱ_U but the index is often left out in cases when it is clear from the context to what observer ϱ is referred.

15.2 Gauss's Theorem

Assume that one vertex of a parallelepiped in spacetime has the position vector $\bar{R}_0$ and that the edge vectors with their feet in this vertex are $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$, and $\bar{A}_4$. One side of the parallelepiped is then given by the equation

$$\bar{R} = \bar{R}_0 + x\bar{A}_2 + y\bar{A}_3 + z\bar{A}_4 \tag{15.6}$$

where the numbers x, y, and z are between 0 and 1. This side will be denoted (1−). The side given by

$$\bar{R} = \bar{R}_0 + \bar{A}_1 + x\bar{A}_2 + y\bar{A}_3 + z\bar{A}_4 \tag{15.7}$$

will be denoted (1+). There are eight sides in all: (1±), (2±), (3±), and (4±). The corresponding situation in three dimensions is illustrated in Figure 15.4.

We shall conventionally define the corresponding surface elements to be $\pm\varepsilon[\ ,\bar{A}_2,\bar{A}_3,\bar{A}_4]$, $\pm\varepsilon[\bar{A}_1,\ ,\bar{A}_3,\bar{A}_4]$, $\pm\varepsilon[\bar{A}_1,\bar{A}_2,\ ,\bar{A}_4]$, and $\pm\varepsilon[\bar{A}_1,\bar{A}_2,\bar{A}_3,\ ]$.

Consider now some vector field $\bar{\mathcal{F}}(\ )$ and the integration of this field over the sides. As an example let us choose the integration over (1+)

$$\int_{(1+)} \mathcal{F}^a dS_a = \int_0^1\int_0^1\int_0^1 \varepsilon[\bar{\mathcal{F}}(\bar{R}_0 + \bar{A}_1 + x\bar{A}_2 + y\bar{A}_3 + z\bar{A}_4), \bar{A}_2, \bar{A}_3, \bar{A}_4]dxdydz \tag{15.8}$$

where $dS_a = \varepsilon_{abcd}A_2^b dx\ A_3^c dy\ A_4^d dz$ was used (cf. Subsection 14.1.4).

[1]This is not always the case as will be seen in Section 15.3.

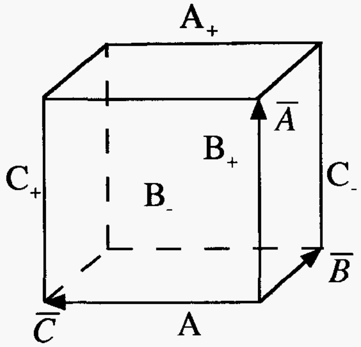

Figure 15.4: A parallelepiped in three dimensions has six sides: $A_\pm$, $B_\pm$, and $C_\pm$. The corresponding surface elements are conventionally defined to be $\pm\varepsilon[\ ,\bar{B},\bar{C}]$, $\pm\varepsilon[\bar{A},\ ,\bar{C}]$, and $\pm\varepsilon[\bar{A},\bar{B},\ ]$.

We will now focus our attention to a *small* parallelepiped. To be able to go to this limit in a convenient manner let us replace the edge vectors $\bar{A}_r$ by $\lambda\bar{A}_r$. Remembering the difference in sign in the definition of the surface element on (1+) and (1−) we then obtain

$$\lim_{\lambda\to 0}\{\int_{(1+)}\mathcal{F}^a dS_a + \int_{(1-)}\mathcal{F}^a dS_a\}/\lambda^4 =$$
$$\lim_{\lambda\to 0}\int_0^1\int_0^1\int_0^1\{\varepsilon[\frac{1}{\lambda}\bar{\mathcal{F}}(\bar{R}_0+\lambda\bar{A}_1+x\lambda\bar{A}_2+y\lambda\bar{A}_3+z\lambda\bar{A}_4),\bar{A}_2,\bar{A}_3,\bar{A}_4] -$$
$$\varepsilon[\frac{1}{\lambda}\bar{\mathcal{F}}(\bar{R}_0+x\lambda\bar{A}_2+y\lambda\bar{A}_3+z\lambda\bar{A}_4),\bar{A}_2,\bar{A}_3,\bar{A}_4]\}dxdydz =$$
$$\lim_{\lambda\to 0}\int_0^1\int_0^1\int_0^1 \varepsilon[D_{\bar{A}_1}\bar{\mathcal{F}}(\bar{R}_0+x\lambda\bar{A}_2+y\lambda\bar{A}_3+z\lambda\bar{A}_4),\bar{A}_2,\bar{A}_3,\bar{A}_4]dxdydz$$
$$=\varepsilon[D_{\bar{A}_1}\bar{\mathcal{F}}(\bar{R}_0),\bar{A}_2,\bar{A}_3,\bar{A}_4]. \qquad (15.9)$$

The limit $\lambda\to 0$ of the integral over the entire boundary of the parallelepiped is thus

$$\lim_{\lambda\to 0}\oint\mathcal{F}^a dS_a/\lambda^4 =$$
$$\varepsilon[D_{\bar{A}_1}\bar{\mathcal{F}}(\bar{R}_0),\bar{A}_2,\bar{A}_3,\bar{A}_4]+\varepsilon[\bar{A}_1,D_{\bar{A}_2}\bar{\mathcal{F}}(\bar{R}_0),\bar{A}_3,\bar{A}_4]+$$
$$\varepsilon[\bar{A}_1,\bar{A}_2,D_{\bar{A}_3}\bar{\mathcal{F}}(\bar{R}_0),\bar{A}_4]+\varepsilon[\bar{A}_1,\bar{A}_2,\bar{A}_3,D_{\bar{A}_4}\bar{\mathcal{F}}(\bar{R}_0)]. \qquad (15.10)$$

Expand the vector field in the edge vectors

$$\bar{\mathcal{F}} = \sum_{r=1}^{4} \mathcal{G}_r \bar{A}_r \tag{15.11}$$

where $\mathcal{G}_r$ are four scalar fields. Equation (15.10) can then be written

$$\lim_{\lambda \to 0} \oint \mathcal{F}^a dS_a / \lambda^4 = \sum_{r=1}^{4} D_{\bar{A}_r} \mathcal{G}_r \, \varepsilon[\bar{A}_1, \bar{A}_2, \bar{A}_3, \bar{A}_4] =$$
$$D_a \sum_{r=1}^{4} A_r^a \mathcal{G}_r \, \varepsilon[\bar{A}_1, \bar{A}_2, \bar{A}_3, \bar{A}_4] = D_a \mathcal{F}^a \varepsilon[\bar{A}_1, \bar{A}_2, \bar{A}_3, \bar{A}_4] \tag{15.12}$$

or for the small parallelepiped

$$\oint \mathcal{F}^a dS_a = D_a \mathcal{F}^a dV. \tag{15.13}$$

The expression $D_a \mathcal{F}^a$ occurring on the right hand side is the *divergence* of the vector field $\mathcal{F}^a$ (see the end of Section 13.2).

Remember the rule on page 147 for choosing surface elements of a parallelepiped. According to this rule *a surface element contracted on an outward directed vector gives a quantity of the same sign as the volume of the parallelepiped.* Hence if we divide some region of space-time into tightly stacked parallelepipeds all with volumes of the same sign two coinciding surfaces will have surface elements of opposite sign. In the integral of a vector field over all the parallelepiped boundaries the parts of the integral over all the internal boundaries will therefore cancel. What remains is the boundary of the entire region. Thus

$$\oint \mathcal{F}^a dS_a = \int D_a \mathcal{F}^a dV. \tag{15.14}$$

The sign rule which has to be followed here runs: *If dV is everywhere positive the sign of dS_a must be chosen so that the contraction of dS_a with an outward directed vector is positive. The opposite sign must be chosen if dV is everywhere negative.* Equation (15.14) is *Gauss's theorem.*

15.3 Different Kinds of Currents

Let us now return to particle currents. In general the four-current density depends on the position so that it is a vector *field*. For this reason we should denote it with a script letter $\bar{\mathcal{J}}$. Consider a region of spacetime on the boundary of which the surface element is chosen to be such that the contraction with an outward directed vector is everywhere positive. The quantity

$$dn = \mathcal{J}^a dS_a \tag{15.15}$$

is then according to (15.4) the number of particle lines at points of the boundary where $\bar{\mathcal{J}}$ is outward directed and minus the number of lines where $\bar{\mathcal{J}}$ is inward directed. For this reason the following quantity

$$\oint dn = \oint \mathcal{J}^a dS_a = \int D_a \mathcal{J}^a dV \tag{15.16}$$

is the difference between the number of outgoing and the number of ingoing lines, e.g., the net number of particles created in the spacetime region. If the number of particles is conserved the equation $\oint dn = 0$ is valid for every spacetime region so that the divergence of the four-current density vanishes

$$D_a \mathcal{J}^a = 0. \tag{15.17}$$

Particles can have electric charge. If all the particles have the same charge e we can multiply (15.16) by this charge and define new quantities $q \equiv en$ and $\mathcal{I}^a \equiv e\mathcal{J}^a$, the *electric charge* and the *electric four-current density.* Usually there are two or more systems of particles with different particle charges in the same spacetime region. They can for instance be protons and electrons. The total charge and the total electric four-current density are then defined as the sum of the partial charges and the sum of the partial electric four-current densities, respectively

$$q = q_1 + q_2, \quad \mathcal{I}^a = \mathcal{I}_1^a + \mathcal{I}_2^a. \tag{15.18}$$

Conservation of electric charge is an important law of nature. It states that although the number of particles and the partial charges may not be conserved *the total charge is always conserved.* Hence,

$$D_a\mathcal{I}^a = D_a\mathcal{I}_1^a + D_a\mathcal{I}_2^a = 0. \tag{15.19}$$

Protons and electrons may recombine to neutral hydrogen atoms or hydrogen atoms may be ionized to protons and electrons. In all such processes the total charge is conserved.

Let us now leave the electric currents and consider again one single system of particles this time assuming that the macroscopic four-velocity is the same as the particle four-velocity, i.e., that the particles do not perform any irregular motions. The particles are then said to form a *dust cloud.* The net four-momentum leaving a spacetime region is now given by[2]

$$\oint \mathcal{P}_a dn = \oint \mathcal{P}_a \mathcal{J}^b dS_b = \int D_b(\mathcal{P}_a\mathcal{J}^b)dV \tag{15.20}$$

where the vector field $\mathcal{P}_a$ is defined as the particle four-momentum, which generally varies from point to point. The tensor

$$\mathcal{T}_a{}^b \equiv \mathcal{P}_a\mathcal{J}^b \tag{15.21}$$

is known as the *energy-momentum tensor for dust.* If there are more than one dust cloud in the same region of spacetime their energy-momentum tensors add to form the total energy-momentum tensor. With no external influence on the collection of dust clouds the total four-momentum is conserved, which means that the total energy-momentum tensor is divergence free[3]

$$D_b\mathcal{T}_a{}^b = 0. \tag{15.22}$$

Gauss' theorem is not confined to particle currents. In fact every divergence free field corresponds to a conservation law.

---

[2]If you are in doubt as to whether this is a legitimate procedure multiply by an *arbitrary* (constant) vector K^a. The equation is then seen to be a direct application of Gauss' theorem.

[3]This treatment of dust must be modified if the particles perform irregular motions. The reason is that the energy-momentum tensor is quadratic in the particle four-velocities and therefore obtains non-vanishing contributions from irregular motions when an averaging is performed.

Part III

ELECTRODYNAMICS

Introduction

The study of electric and magnetic phenomena played the leading part in the initial development of the theory of relativity. After decades of fruitless speculation concerning the world ether in which the electromagnetic phenomena were supposed to be anchored the realisation dawned at the beginning of the 20th century that the scene is not the world ether in a Newtonian space and time but rather a unified spacetime. As is well known it was Albert Einstein's work on electromagnetism in moving media which initiated the development of these new ideas. They were hard to digest at the time but to the reader of this book the concept of spacetime should by now be no great mystery and there is no reason to motivate and develop the theory of relativity starting from the Maxwellian theory of electrodynamics. To us it is more natural to present electrodynamics directly as a field theory in spacetime and then show that with the introduction of an observer the theory agrees with the Maxwellian theory.

In Chapter 9 we studied particle reactions but without mentioning the dynamics behind them — only conservation laws were utilized. In classical Newtonian mechanics the motion of a body is typically governed by forces and these are determined by the positions of other bodies. It is understood that the influence of a body is instantaneously transmitted to other bodies and this is as we know against what we have learnt from the theory of relativity. In relativity the influence between bodies cannot propagate faster than with the characteristic velocity (which we have set to unity). The classical picture must therefore be replaced by *local* action of particle on particle or of fields in *local* interaction with particles or with other fields. Typically a body locally causes a disturbance in a field and this disturbance travels away and meets another body where a local interaction causes an effect. A field which has such effects must have a dynamics of its own, i.e., its own equations of motion. For this reason particles and fields are much more on a par with one another in relativity than particles and forces are in classical mechanics. In relativistic quantum theory the difference has actually disappeared entirely.

Keeping to classical — non-quantum — physics there are very few phenomena in nature which may be described as the kind of field we have described here. The only candidates are electromagnetism and gravitation. Attempts to treat gravitation within special relativity lead, however, to inconsistencies which require a more general concept of spacetime than the one offered in special relativity. This is done in Einstein's theory of general relativity which is outside the scope of this book. Thus electromagnetism is the only field theory which remains to us. This field is on the other hand of enormous importance: It is responsible for the properties of atoms and molecules and all chemical reactions, light is an electromagnetic phenomenon, as well as radio waves, and on the surface of stars as well as in interstellar space electromagnetic plasma phenomena play a leading part. These applications of electromagnetism cannot be covered in a book like this one. They require in fact several text books of their own. Our only ambition is to present the electromagnetic field in such a way that it is obvious how nicely it fits into spacetime.

Chapter 16

Sourcefree Electromagnetism

16.1 The Wave Equation

In Section 8.1 we studied plane waves of the form (8.2)

$$\psi = \Psi(\bar{R}) = A\sin(\bar{K}\cdot\bar{R} + F_0) \tag{16.1}$$

where A and F_0 are constant numbers, $\bar{K}$ is a constant vector and $\bar{R}$ is the position vector in spacetime. Using the rule for differentiating a function of a function we can calculate the gradient of Ψ in (16.1)

$$D_a\Psi(\bar{R}) = A\cos(\bar{K}\cdot\bar{R} + F_0)D_a(\bar{K}\cdot\bar{R}) = A\cos(\bar{K}\cdot\bar{R} + F_0)K_a. \tag{16.2}$$

Now calculate the gradient of this expression obtaining

$$D_bD_a\Psi(\bar{R}) = -A\sin(\bar{K}\cdot\bar{R} + F_0)K_bK_a. \tag{16.3}$$

Finally raise one of the indices and contract

$$\Box\Psi(\bar{R}) = D^aD_a\Psi(\bar{R}) = -A\sin(\bar{K}\cdot\bar{R} + F_0)\bar{K}^2. \tag{16.4}$$

Remembering (8.14) we thus find

$$\Box\Psi = m^2\Psi. \tag{16.5}$$

where m is a number. This is the *Klein-Gordon equation.* We shall concentrate our interest on the case when the group velocity is unity. As we know from Subsection 8.2 this means that $\bar{K}$ is null-like and $m = 0$. Hence

$$\Box\Psi = 0. \tag{16.6}$$

This is the famous *wave equation.*

You may ask what we have accomplished by replacing the explicit expression in (16.1) by the more opaque differential equation (16.6). The gain is that *all the parameters* $\bar{K}$, A, *and* F_0 *have disappeared.* Hence the same equation (16.6) is satisfied by all plane waves with null-like $\bar{K}$. And what is more, all linear superpositions of such waves do also satisfy (16.6). There is a converse to this theorem: All solutions to (16.6) can be written as superpositions of plane waves with null-like $\bar{K}$.[1]

As should be clear from the introduction to the present part the main theme for relativistic theory (both classical and quantum) is fields. These may be scalar fields or tensor fields (or a sort of generalized tensor fields known as spinor fields which fall outside the scope of this book). Each field may exist as a free field propagating in spacetime and it may interact locally with particles and with other fields. It is through such interactions that the field acts on our measuring devices. *The free field not interacting with anything is assumed to be a superposition of plane waves. Hence the wave equation* (16.6) *characterizes a free* Ψ*-wave.* The electromagnetic field is *not* a scalar field as we shall soon see, but the discussion of the wave equation is precisely the same for the electromagnetic field as for the scalar field.

Introducing an observer and using (13.22) we find the following form of the wave equation

$$\{\nabla^2 - \partial^2/\partial t^2\}\Psi = 0. \tag{16.7}$$

The operator ∇^2 is the *Laplacian* often denoted Δ. We avoid this latter notation as it could be misinterpreted as 'increment'.

[1] This is a consequence of the well-known *Fourier's theorem*, which we will not go further into here.

16.2 The Field Tensor

It is well known that an electromagnetic wave, e.g., a light wave, affects a particle in proportion to its electric charge[2] e. The effect is restricted to a change of the four-velocity, while the number of particles and their masses are unchanged. We should therefore look for an equation giving the four-acceleration, or perhaps still more natural the proper time derivative of the four-momentum $d\bar{P}/d\tau$, in terms of the field. According to the principle of local interaction which was discussed in the introduction to the present part it should be the field *at the spacetime location of a particle* which determines $d\bar{P}/d\tau$.

A first guess is that the field is a vector field $\bar{\mathcal{F}}$ and that the wanted equation is

$$d\bar{P}/d\tau|_{\bar{R}} = md\hat{U}/d\tau|_{\bar{R}} = e\bar{\mathcal{F}}(\bar{R}) \tag{16.8}$$

where m is the mass and $\hat{U}$ is the four-velocity of the particle. The particle is assumed to be a test particle which does not influence the field and $\bar{\mathcal{F}}$ is assumed to be independent of the state of the particle.

We know from (10.7) that the four-velocity and the four-acceleration are always orthogonal. Hence the equation (16.8) leads to

$$(\hat{U}|_{\bar{R}}) \cdot \bar{\mathcal{F}}(\bar{R}) = 0. \tag{16.9}$$

Independently of the field we may however choose any four-velocity of the particle. Hence the field is orthogonal to every timelike vector. An arbitrary vector can, however, be written as a linear combination of timelike vectors. The field is therefore orthogonal to all vectors which means that the field vanishes: $\bar{\mathcal{F}} = 0$. A vanishing field does not affect charges at all and is obviously not what we are looking for so (16.8) does not work.

At this stage it seems pretty obvious that we must allow the expression for $d\bar{P}/d\tau$ to contain more information about the particle itself. A scalar property of the particle won't help us. The only generally available vectors are the four-velocity and its proper-time derivatives. It seems pointless, however, to involve the four-acceleration or derivatives

[2]The notation e is often used for the elementary charge. Here it stands for any charge of a particle.

of it as it is the four-acceleration we want to calculate. What remains to us is the four-velocity. The next guess and the one which will turn out to lead to electrodynamics is that *$d\bar{P}/d\tau$ is linearly dependent on the particle vector $e\hat{U}$* (and vanishes when this vector vanishes), i.e.,

$$m dU_a/d\tau = e\mathcal{F}_{ab}U^b \qquad (16.10)$$

where $\mathcal{F}$ is a second rank tensor field which is independent of the state of the test particle. It is known as the *electromagnetic field tensor*. From a purely systematic point of view the term is not quite a happy one. 'Electromagnetic tensor field' would have been more appropriate. To avoid unnecessary loading down of the equation '$\bar{R}$' has been omitted. It is understood that both members are taken at one and the same point in spacetime.

We must now again use (10.7) which together with (16.10) gives

$$\mathcal{F}_{ab}U^aU^b = 0. \qquad (16.11)$$

This is true with $\hat{U}$ as an arbitrary timelike future directed vector. The combination $\hat{U} + \hat{V}$ is such a vector if $\hat{U}$ and $\hat{V}$ are such vectors (see the end of Section 6.4). Hence

$$0 = \mathcal{F}_{ab}(U^a + V^a)(U^b + V^b) = (\mathcal{F}_{ab} + \mathcal{F}_{ba})U^aV^b. \qquad (16.12)$$

Any vector can be written as a linear combination of timelike future directed vectors. Hence (16.12) is true for arbitrary vectors $\bar{U}$ and $\bar{V}$. This implies that *the electromagnetic field tensor is antisymmetric*

$$\mathcal{F}_{ab} = -\mathcal{F}_{ba}. \qquad (16.13)$$

16.3 The Field Equations

16.3.1 General Considerations

The field equations for a free scalar field are according to Section 16.1 $\Box\Psi = 0$. The same argument as in that section leads us to the field equation

$$\Box \mathcal{F}_{ab} = 0 \tag{16.14}$$

if $\mathcal{F}_{ab}$ is a free light-like field which we in this chapter will assume that it is.

This is, however, not the whole story. There are a few general principles which one wishes to follow in the construction of field equations for free fields. They run as follows:

- The equations should be *linear* and *homogeneous*.
- No 'structure' should be present except the field itself and the metric.
- The set of equations should be *maximal*.

The first point here is equivalent to the principle of superposition which we met already at the end of Section 8.1: The linear combination (with arbitrary constant coefficients) of two solutions to the equations must itself be a solution to the equations.

The second point tells us among other things that the equations must not contain any tensors (including vectors) or tensor fields except the field we are setting up equations for and the metric tensor g_{ab}. In particular there must be no observer four-velocity involved in the equations.

We have included only the metric not the volume tensor in the basic structure. The reason is that the only further information contained in the volume tensor is the orientation (see Section 14.2) and we will make the physical assumption that the phenomena we study are *orientation symmetric* as discussed in Section 14.2. The field equations governing these phenomena must then also be orientation symmetric, i.e., they must look the same irrespective of the orientation of spacetime. If we had included the volume tensor this assumption implies that we would have had to construct the field equations so that they did not change when the sign of the volume tensor was changed. Then there would have had to occur either an odd or an even number of ε-factors in every term, and they could have been eliminated from the field equations using (14.28) – (14.32).

The third point expresses the stipulation that the field equations determine the field 'as much as possible', without excluding plane wave solutions like (16.1), with $K_a \neq 0$ and $A \neq 0$. Each field equation is a condition on the solutions and reduces the class of fields which are solutions. The requirement of maximality tells us that the set of field equations should be such that no further equation can be added to it — that no further reduction in the class of solutions can be made without eliminating the plane wave solutions. It may happen that more than one maximal set of equations can be constructed. In such a case we say that each possibility constitutes a separate 'theory'.

16.3.2 The Four-Potential Equations

It is easy to supplement (16.14) with other equations as we shall soon see. Equation (16.14) alone is therefore not maximal. Let us start a search from scratch for a maximal set of equations applying the general considerations of Section 16.3.1. This can be done in two different ways which turn out to lead to identical results. The most direct route is simply to investigate differential equations for $\mathcal{F}_{ab}$ itself. We will, however, follow the second route in which one starts by asking: Could we as one of the conditions on $\mathcal{F}_{ab}$ require that it can be constructed from some simpler field? It is not possible to construct $\mathcal{F}_{ab}$ from a scalar field Ψ. The only second rank tensor which can be constructed in a linear way from Ψ is $D_a D_b \Psi$ and it is symmetric not antisymmetric. From a *first rank tensor field* $\mathcal{A}_a$ it is on the other hand easy to construct an antisymmetric tensor field. Hence a possible condition on $\mathcal{F}_{ab}$ is that it can be written as

$$\mathcal{F}_{ab} = D_a \mathcal{A}_b - D_b \mathcal{A}_a. \tag{16.15}$$

This condition on $\mathcal{F}_{ab}$ is still not maximal because we may put restrictions on $\mathcal{A}_a$. One obvious choice of restriction is

$$\Box \mathcal{A}_a = 0 \tag{16.16}$$

leading immediately to (16.14).

We are not quite through yet because one scalar field $D_a \mathcal{A}^a$ (the divergence of $\mathcal{A}^a$) can be formed and we may construct one more equation by putting it equal to zero

$$D^a \mathcal{A}_a = 0. \tag{16.17}$$

We have now reached the limit. The equations (16.15), (16.16), and (16.17) constitute a maximal set. We can check this by writing the general equation for a plane $\mathcal{A}_a$-wave and the corresponding $\mathcal{F}_{ab}$-wave

$$\mathcal{A}_a(\bar{R}) = A_a \cos(K_b R^b + F_0) \tag{16.18}$$

and

$$\mathcal{F}_{ab}(\bar{R}) = -F_{ab} \sin(K_c R^c + F_0) \tag{16.19}$$

where A_a and F_{ab} are constant tensors, the latter one being antisymmetric. The field equations (16.15),(16.16), and (16.17) give

$$F_{ab} = K_a A_b - K_b A_a, \tag{16.20}$$

$$K_a K^a = 0, \tag{16.21}$$

and

$$K_a A^a = 0 \tag{16.22}$$

respectively. Equation (16.21) tells us that $\bar{K}$ is null-like. Equation (16.22) then tells us that $\bar{A}$ is either spacelike or parallel to $\bar{K}$ or vanishes (see Section 6.3). In the latter two cases the relation (16.20) gives $F_{ab} = 0$. Hence $\bar{A}$ has to be spacelike in an acceptable plane wave solution. In fact choosing $\bar{A}$ as any spacelike vector orthogonal to $\bar{K}$ the expression (16.20) is obviously non-vanishing. Thus the field equations we have contemplated so far do not exclude plane wave solutions.

Let us now supplement (16.20), (16.21), and (16.22) with one more equation and show that this necessarily excludes all plane wave solutions. Such a supplementary equation must have the form

$$K_a K_b \ldots\ A_i + (\text{permutations}) = 0 \tag{16.23}$$

where the bracket denotes a sum of terms. Each term is some number times the first term with an exchange between the index i and some of the other indices. Notice that every attempt to contract over a pair of indices in (16.23) leads to $0 = 0$ due to (16.21), and (16.22). Multiplying (16.23) by A^i and contracting over i all terms vanish due to (16.22) except the first one and we obtain $\bar{A}^2 = 0$. This excludes spacelike $\bar{A}$-vectors and plane waves. No equation (16.23) can therefore be accepted and the equations (16.15), (16.16), and (16.17) form a maximal set. Q.E.D.

The first rank tensor field $\mathcal{A}$ introduced here is known as the *four-vector potential* sometimes shortened to *four-potential.*

16.3.3 The Field Tensor Equations

We will in this subsection investigate the differential equations for $\mathcal{F}_{ab}$ which result from equations (16.15), (16.16), and (16.17) of the previous subsection.

Let us first calculate the divergence of $\mathcal{F}_{ab}$

$$D_a\mathcal{F}^a{}_b = D_a(D^a\mathcal{A}_b - D_b\mathcal{A}^a) = \Box\mathcal{A}_b - D_bD^a\mathcal{A}_a = 0. \tag{16.24}$$

The divergence of $\mathcal{F}_{ab}$ vanishes.

Let us further calculate the divergence of $\varepsilon^{abcd}\mathcal{F}_{cd}$

$$D_a\varepsilon^{abcd}\mathcal{F}_{cd} = \varepsilon^{abcd}D_a(D_c\mathcal{A}_d - D_d\mathcal{A}_c) = 0. \tag{16.25}$$

The last equation follows simply from the fact that the volume tensor is completely antisymmetric while the double gradient is symmetric (see equation (12.35)). Using (14.29) we may reshape (16.25). Multiply by ε_{fbgh} and six terms are obtained. They are, however, pairwise equal due to the antisymmetry of $\mathcal{F}_{cd}$. After contractions utilizing the fact that δ is the identity tensor the result runs as follows

$$D_f\mathcal{F}_{gh} + D_h\mathcal{F}_{fg} + D_g\mathcal{F}_{hf} = 0. \tag{16.26}$$

Notice how the indices are permuted cyclically from term to term.

It is an interesting fact that (16.25) (and (16.26)) is a consequence of (16.15) alone — the equations (16.16) and (16.17) were not used in the derivation (16.25). The converse is in fact also true so that

(16.26) *is the necessary and sufficient condition for* $\mathcal{F}_{ab}$ *to be of the form* $\mathcal{F}_{ab} = D_a\mathcal{A}_b - D_b\mathcal{A}_a$. The proof follows the same lines as the proof in three dimensions that a divergence-free vector field can be written as a curl and you can find the details below.

We will here show how (16.15) can be solved under the condition (16.26). The solution requires the introduction of one arbitrary dual vector $W[\,]$ and one vector $\bar{V}$ which is arbitrary except for the normalization condition $W[\bar{V}] = 1$. A scalar field $t(\,)$ can now be defined through $t(\bar{R}) = W[\bar{R}]$. It has obviously the property $D_a t = W_a$ and hence $D_{\bar{V}} t = 1$. In Figure 16.1 we have illustrated the situation. The *projection* $\bar{r}$ of $\bar{R}$ has by definition the properties $\bar{R} = k\bar{V} + \bar{r}$ and $W[\bar{r}] = 0$, i.e.,

$$r^a = R^a - V^a W_b R^b = \Pi^a_b R^b \tag{16.27}$$

where Π^a_b is the *projection tensor*

$$\Pi^a_b \equiv \delta^a_b - V^a W_b. \tag{16.28}$$

Notice that

$$D_{\bar{V}}\bar{r} = D_{\bar{V}}(\bar{R} - t(\bar{R})\bar{V}) = 0. \tag{16.29}$$

Using he projection tensor we may define the differential operator

$$\nabla_a \equiv \Pi^b_a D_b = D_a - W_a D_{\bar{V}} \tag{16.30}$$

with the properties $V^a\nabla_a = 0$ and $\nabla_a t = 0$. We may also define a new antisymmetric tensor field

$$f_{ab} \equiv \Pi^c_a \Pi^d_b \mathcal{F}_{cd} = \mathcal{F}_{ab} - W_a V^c \mathcal{F}_{cb} - W_b V^c \mathcal{F}_{ac} \tag{16.31}$$

with the property $V^a f_{ab} = 0$ and

$$\nabla_a f_{bc} + \nabla_c f_{ab} + \nabla_b f_{ca} = 0. \tag{16.32}$$

The latter equation is obtained simply by equipping (16.26) with three projection tensors, one for each index.

After these preliminaries we may now give the promised solution

$$\mathcal{A}_a(\bar{R}) = -\int_0^{t(\bar{R})} \mathcal{F}_{ab}(\bar{r} + t'\bar{V})V^b dt' \; + a_a \tag{16.33}$$

where $\bar{r} = \bar{R} - t(\bar{R})\bar{V}$ (see Figure 16.1). The dual vector field a_a is required to have the following properties

$$a_a V^a = 0 \tag{16.34}$$

$$D_{\bar V} a_a = 0 \tag{16.35}$$

$$\nabla_a a_b - \nabla_b a_a = f_{ab} \quad \text{for } t = 0. \tag{16.36}$$

Remembering (16.32) we see that solving a_a poses exactly the same problem as the original problem of solving $\mathcal{A}_a$ apart from the fact that the a_a-problem is bound to the hypersurface $t = 0$ which is three-dimensional. In our discussion neither the dimensionality nor the metric was used. Hence we may repeat it and trap down the dimension until we have a complete solution of $\mathcal{A}_a$.

What now remains is to prove that (16.33) is a solution to (16.15). We will do this by a straight forward calculation. We start by splitting D_a

$$(D_a\mathcal{A}_b - D_b\mathcal{A}_a)(\bar R) = (\nabla_a\mathcal{A}_b - \nabla_b\mathcal{A}_a)(\bar R) + D_{\bar V}(W_a\mathcal{A}_b - W_b\mathcal{A}_a)(\bar R). \tag{16.37}$$

We now insert (16.33) here and take care of the second term first. Due to $D_{\bar V}\bar r = 0$ we obtain a contribution only from differentiation on the upper limit of the integral. Thus from $D_{\bar V} t = 1$ and $D_{\bar V} a_a = 0$

$$D_{\bar V}(W_a\mathcal{A}_b - W_b\mathcal{A}_a)(\bar R) = -(W_a\mathcal{F}_{bc} - W_b\mathcal{F}_{ac})(\bar R)V^c. \tag{16.38}$$

Due to $\nabla t = 0$ the first term in (16.37) gives no contribution from differentiation of the integration limit and

$$\begin{aligned}&(\nabla_a\mathcal{A}_b - \nabla_b\mathcal{A}_a)(\bar R) = \\ &-\int_0^{t(\bar R)} (\nabla_a\mathcal{F}_{bc} - \nabla_b\mathcal{F}_{ac})(\bar r + t'\bar V)V^c dt' \;+\; (\nabla_a a_b - \nabla_b a_a)(\bar r).\end{aligned} \tag{16.39}$$

In the last term the argument '$\bar R$' has been changed to '$\bar r$'. This can be done due to $D_{\bar V} a_a = 0$. In the integral we now write ∇_a as $\nabla_a = D_a - W_a D_{\bar V}$. Employing (16.26) we then obtain

$$\begin{aligned}&\int_0^{t(\bar R)} (\nabla_a\mathcal{F}_{bc} - \nabla_b\mathcal{F}_{ac})(\bar r + t'\bar V)V^c dt' = \\ &-\int_0^{t(\bar R)} D_c\mathcal{F}_{ab}(\bar r + t'\bar V)V^c dt' - \int_0^{t(\bar R)} D_{\bar V}(W_a\mathcal{F}_{bc} - W_b\mathcal{F}_{ac})(\bar r + t'\bar V)V^c dt'.\end{aligned} \tag{16.40}$$

The integrals can be regarded as line integrals along $\bar V$ and both the integrals can therefore be evaluated giving

$$(\nabla_a\mathcal{A}_b - \nabla_b\mathcal{A}_a)(\bar R) = [\mathcal{F}_{ab}(\bar R) + (W_a\mathcal{F}_{bc} - W_b\mathcal{F}_{ac})V^c]_{\bar r}^{\bar R}. \tag{16.41}$$

Collecting everything and using (16.31) we obtain

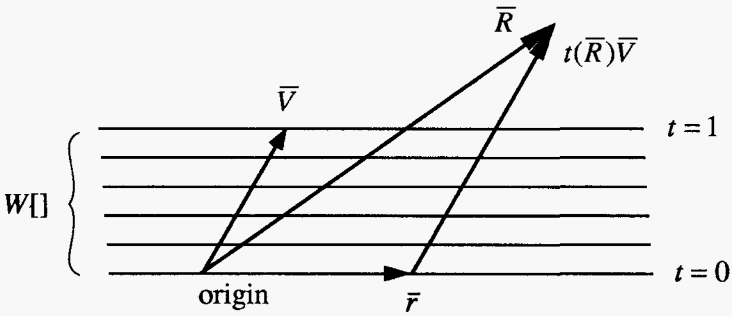

Figure 16.1: The horizontal lines in this figure represent three-dimensional planes $W[\bar{R}] = const$ in spacetime. The vector $\bar{V}$ is normalized through $W[\bar{V}] = 1$. The vector $\bar{r}$ is the projection of $\bar{R}$ on the subspace $t = 0$, i.e., $W[\bar{r}] = 0$.

$$(D_a\mathcal{A}_b - D_b\mathcal{A}_a)(\bar{R}) = \mathcal{F}_{ab}(\bar{R}) + (\nabla_a a_b - \nabla_b a_a)(\bar{r}) - f_{ab}(\bar{r}) \tag{16.42}$$

and using (16.36) here we finally obtain

$$D_a\mathcal{A}_b - D_b\mathcal{A}_a = \mathcal{F}_{ab}. \tag{16.43}$$

Q.E.D.

16.3.4 Gauge

The electromagnetic field is observed through its effect on charges which is governed by (16.10). What we measure is therefore the field tensor $\mathcal{F}_{ab}$.[3]

Let us now concentrate our attention on the relation (16.15) between the four-potential and the field tensor. Through this relation the field tensor is uniquely determined by the four-potential but the converse is not true. One sees directly that the four-potential $\mathcal{A}_a$ and the four-potential $\mathcal{A}_a + D_a\chi$ give the same field tensor, due to the symmetry of

[3]This is strictly true in classical physics but has to be modified in quantum physics.

the double gradient, and that they therefore are physically equivalent. Here χ is an arbitrary scalar field. A part of the four-potential is therefore not physically relevant so that there is a 'loose fit' of the potential to the field tensor. It is known as the *gauge freedom.* Don't be confused by the term on linguistic grounds. There is a rather special historic reason for it.

Suppose that $D_a\mathcal{A}^a = 0$. Then $D_a(\mathcal{A}^a + D^a\chi) = \Box\chi$. Utilizing the gauge freedom we may thus determine the divergence of the potential to be any desired scalar field[4]. The choice $D_a\mathcal{A}^a = 0$ in (16.17) is actually arbitrary from this point of view. It defines, as one says, a particular *gauge* which is known as *Lorentz gauge.* Although it is arbitrary from a physical point of view the Lorentz gauge turns out to be quite practical in many applications and it is very much used.

The field tensor is unaffected by gauge changes and so are of course also the equations for the field tensor, (16.24) and (16.25). Assuming $D_a\mathcal{F}^{ab} = 0$ the following equation is obtained for the four-potential (see the derivation (16.24))

$$\Box\mathcal{A}_a - D_a(D_b\mathcal{A}^b) = 0. \tag{16.44}$$

You can easily check that this equation is unchanged as it should when a gauge change is made in $\mathcal{A}_a$. This is the field equation you must use rather than $\Box\mathcal{A}_a = 0$ when the gauge is not decided or decided to be different from the Lorentz gauge.

The equations (16.15) and (16.44) are equivalent to the equations (16.24) and (16.25).

16.4 Maxwell's Equations

16.4.1 The Electric and Magnetic Fields

The electromagnetic field equations are written with no reference to any observer and from a purely physical and mathematical point of view they are best kept that way. We human beings are, however, so used

[4]It will be shown in Chapter 18 that the equation $\Box\chi = \rho$ can be solved for any scalar field ρ.

to interpret the world in terms of space and time rather than in terms of spacetime that the electromagnetic equations are very often written with explicit reference to an observer although this is unnecessary from a deeper point of view. It was also in this form that the equations were developed for the first time by Clerk Maxwell. In this section we shall obtain his equations by using an observer four-velocity W^a to rewrite the electromagnetic equations.

Before we start in earnest just a notational detail. The field $\varepsilon_{abcd}\mathcal{F}^{cd}$ has already appeared in the previous section (see (16.25)). It is customary to introduce a special notation for multiplication by ε_{abcd} and contracting over one or more of the indices. For the moment we will keep to contraction over two indices writing

$$F^*{}_{ab} \equiv \frac{1}{2}\varepsilon_{abcd}F^{cd}. \tag{16.45}$$

This 'star operation' can be repeated and (14.30) gives

$$F^{**}{}_{ab} = -F_{ab}. \tag{16.46}$$

The tensor F^* is known as the *dual* of F. This is an entirely different definition of 'dual' from the one introduced at the end of Section 12.1. It is a bit unfortunate that the term 'dual' has acquired two different meanings but you will have to live with it and keep them apart by context.

With recourse to the four-velocity W^a we need just a vector field to construct an antisymmetric tensor field. We will in fact introduce *two* vector fields $\mathcal{E}^a$ and $\mathcal{B}^a$ *both assumed to be orthogonal to* W^a. It is then possible to construct two antisymmetric tensor fields

$$\mathcal{K}_{ab} \equiv W_a\mathcal{E}_b - W_b\mathcal{E}_a \tag{16.47}$$

and

$$\mathcal{L}_{ab} \equiv W_a\mathcal{B}_b - W_b\mathcal{B}_a \tag{16.48}$$

and from these an antisymmetric tensor field

$$\mathcal{F}_{ab} = \mathcal{K}_{ab} + \mathcal{L}^*{}_{ab} \tag{16.49}$$

or

$$\mathcal{F}_{ab} = W_a\mathcal{E}_b - W_b\mathcal{E}_a + \varepsilon_{abcd}W^c\mathcal{B}^d. \tag{16.50}$$

The equation (16.49) can easily be solved for the fields $\mathcal{E}_a$ and $\mathcal{B}_a$. Multiply it by W^b and the second term cancels while the first one gives $\mathcal{E}_a$ (remember that $\mathcal{E}^a$ and W^a are assumed to be orthogonal) so that

$$\mathcal{E}_a = \mathcal{F}_{ab}W^b. \tag{16.51}$$

Performing the dual operation on (16.49) and using (16.46) an equation is obtained which can be solved for $\mathcal{B}_a$ giving

$$\mathcal{B}_a = -\mathcal{F}^*{}_{ab}W^b. \tag{16.52}$$

What we have now shown is that the equations (16.51) and (16.52) are necessary consequences of (16.50). By inserting the two former equations into the last one it is shown that (16.51) and (16.52) are also sufficient for (16.50) to be valid (see the detailed proof below). Hence the two fields $\mathcal{E}^a$ and $\mathcal{B}^a$ together carry the same information as the field $\mathcal{F}_{ab}$. (The vector W^a contains no information at all concerning the field.)

We will here show through a direct calculation that (16.51) and (16.52) satisfy (16.50). We start by inserting (16.52) into the last term of (16.50) and using (14.29)

$$\varepsilon_{abcd}W^c\mathcal{B}^d = -(1/2)\varepsilon_{abcd}W^c(\varepsilon^{dfgh}W_f\mathcal{F}_{gh}) = \\ -(\delta^f_a\delta^g_b\delta^h_c + \delta^h_a\delta^f_b\delta^g_c + \delta^g_a\delta^h_b\delta^f_c)W^cW_f\mathcal{F}_{gh}. \tag{16.53}$$

Due to the antisymmetry of $\mathcal{F}_{gh}$ we have here been able to combine six δ-terms pairwise thus obtaining only three terms. Performing the contractions we obtain

$$\varepsilon_{abcd}W^c\mathcal{B}^d = -W^cW_a\mathcal{F}_{bc} - W^cW_b\mathcal{F}_{ca} + \mathcal{F}_{ab} = -W_a\mathcal{F}_{bc}W^c + W_b\mathcal{F}_{ac}W^c + \mathcal{F}_{ab}. \tag{16.54}$$

The first two terms here are obtained with opposite sign when (16.51) is inserted into the first two terms of (16.50). Therefore the end result of inserting (16.51) and (16.52) into (16.50) is $\mathcal{F}_{ab}$. Q.E.D.

Splitting the four-velocity $\hat{U}$ according to (9.6) and (6.2)

$$m\hat{U} = \bar{P} = E\hat{W} + \bar{p} \tag{16.55}$$

$$\hat{U} = \gamma(\hat{W} + \bar{u}) \tag{16.56}$$

(with $\bar{p} \cdot \hat{W} = \bar{u} \cdot \hat{W} = 0$) the equation of motion (16.10) can be written

$$\frac{d}{d\tau}(EW_a + p_a) = e(W_a\mathcal{E}_b - W_b\mathcal{E}_a + \varepsilon_{abcd}W^c\mathcal{B}^d)\gamma(W^b + u^b) =$$
$$e\gamma(W_a\mathcal{E}_b u^b + \mathcal{E}_a + W^c\varepsilon_{cabd}u^b\mathcal{B}^d). \tag{16.57}$$

This equation can be divided into two equations, one of them contains the terms parallel to $\hat{W}$ and the other one contains the terms orthogonal to $\hat{W}$. Let us also make use of $dt = \gamma d\tau$ (see e.g. Section 10.1) and we obtain

$$dE/dt = e\bar{\mathcal{E}} \cdot \bar{u} \tag{16.58}$$

$$d\bar{p}/dt = e(\bar{\mathcal{E}} + \bar{u} \times \bar{\mathcal{B}}). \tag{16.59}$$

In the last equation above we have made use of the fact that $\varepsilon_{abd} \equiv W^c\varepsilon_{cabd}$ is the volume tensor in the three-dimensional orthogonal space. In a three-dimensional space the cross product $\bar{a} = \bar{b} \times \bar{c}$ is defined by $a_a \equiv \varepsilon_{abc}b^bc^c$ (see Section 14.3).

You may recognize the equations obtained as the *Lorentz equations of motion* for a charged particle in an electromagnetic field *with $\bar{\mathcal{E}}$ playing the rôle of the electric and $\bar{\mathcal{B}}$ playing the rôle of the magnetic field.*

16.4.2 The Field Equations

Introducing (16.50) and the split (13.22) into the field equation (16.24) you obtain the following equation

$$(-W^a\frac{\partial}{\partial t} + \nabla^a)(W_a\mathcal{E}_b - W_b\mathcal{E}_a + \varepsilon_{abcd}W^c\mathcal{B}^d) = 0 \tag{16.60}$$

or

$$\frac{\partial}{\partial t}\mathcal{E}_b - W_b\nabla^a\mathcal{E}_a + \varepsilon_{abcd}\nabla^a W^c \mathcal{B}^d = 0. \tag{16.61}$$

This can be written as two equations. One of them contains the terms which are parallel to $\hat{W}$ and the other the terms orthogonal to $\hat{W}$ (for future reference notice that the sign of the left hand side is changed)

$$\nabla_a\mathcal{E}^a = 0 \tag{16.62}$$

$$W^c\varepsilon_{cbad}\nabla^a\mathcal{B}^d - \frac{\partial}{\partial t}\mathcal{E}_b = 0. \tag{16.63}$$

In three-dimensional language this runs

$$\text{div } \bar{\mathcal{E}} = 0 \tag{16.64}$$

$$\text{curl } \bar{\mathcal{B}} - \frac{\partial}{\partial t}\bar{\mathcal{E}} = 0 \tag{16.65}$$

where curl $\bar{\mathcal{B}} \equiv \bar{\nabla} \times \bar{\mathcal{B}}$. In (16.64) and (16.65) you probably recognize a part of the Maxwell field equations. The remaining equations are obtained from (16.25) which may be written as $D_a\mathcal{F}^{*ab} = 0$. According to (16.49) and (16.46) the starring of $\mathcal{F}$ means the same as exchanging $\mathcal{E}$ and $\mathcal{B}$ and changing the sign of $\mathcal{E}$. Hence we obtain the following equations

$$\text{div } \bar{\mathcal{B}} = 0 \tag{16.66}$$

$$\text{curl } \bar{\mathcal{E}} + \frac{\partial}{\partial t}\bar{\mathcal{B}} = 0. \tag{16.67}$$

These equations together with (16.64) and (16.65) form the complete set of Maxwell's equations. We have therefore shown that the field theory we have developed is equivalent to the Maxwell theory.

To round this off let us study also the four-potential from the observer point of view. We thus split the four-potential

$$\bar{A} = \phi\hat{W} + \bar{a} \tag{16.68}$$

where $\bar{a} \cdot \hat{W} = 0$. You have to remember that $\bar{a}$ as well as ϕ is a vector *field* though this is not evident from the notation. Introducing (16.68), (13.22), and (16.50) in (16.15) we obtain

$$W_a \mathcal{E}_b - W_b \mathcal{E}_a + \varepsilon_{abcd} W^c \mathcal{B}^d =$$
$$(-W_a \frac{\partial}{\partial t} + \nabla_a)(\phi W_b + a_b) - (-W_b \frac{\partial}{\partial t} + \nabla_b)(\phi W_a + a_a). \tag{16.69}$$

The terms here are of different types. In one type both the factor with index 'a' and the one with index 'b' are (after raising of the indices) proportional to $\hat{W}$, in another type the a-factor is proportional to $\hat{W}$ and the b-factor is orthogonal to $\hat{W}$ etc. Thus the equation can be taken to pieces and we obtain

$$\mathcal{E}_b = -\frac{\partial}{\partial t} a_b - \nabla_b \phi \tag{16.70}$$

and

$$\varepsilon_{abcd} W^c \mathcal{B}^d = \nabla_a a_b - \nabla_b a_a. \tag{16.71}$$

Multiply the latter equation by ε^{abef} and use (14.30) thus obtaining

$$-2(W^e \mathcal{B}^f - W^f \mathcal{B}^e) = 2\varepsilon^{abef} \nabla_a a_b. \tag{16.72}$$

Multiplying this equation by W_e we obtain

$$\mathcal{B}^f = W_e \varepsilon^{efab} \nabla_a a_b. \tag{16.73}$$

The equations (16.70) and (16.73) may be written in three-dimensional form

$$\bar{\mathcal{E}} = -\text{grad } \phi - \frac{\partial}{\partial t} \bar{a} \tag{16.74}$$

and

$$\bar{\mathcal{B}} = \text{curl } \bar{a}. \tag{16.75}$$

In these equations you recognize the relations in the Maxwellian theory between the scalar potential ϕ and the vector potential $\bar{a}$ on one hand and the electric and magnetic fields on the other hand.

Using (13.22) and (16.68) the divergence of $\bar{\mathcal{A}}$ may be written

$$D^a \mathcal{A}_a = (-W^a \frac{\partial}{\partial t} + \nabla^a)(\phi W_a + a_a) = \frac{\partial}{\partial t}\phi + \nabla^a a_a = \frac{\partial}{\partial t}\phi + \text{div } \bar{a}. \tag{16.76}$$

Thus the Lorentz condition takes the well-known form.

The wave equation (16.16) obviously splits into two equations

$$\Box\phi = 0 \tag{16.77}$$

and

$$\Box\bar{a} = 0. \tag{16.78}$$

Chapter 17

Electro-Magnetism with Sources

17.1 The Field Equations

17.1.1 Sources

The influence of the electromagnetic field on charges is given by (16.10) but so far we have not considered the converse, i.e., how the electromagnetic field is influenced by charges. You should by now have at least a vague notion that there must be such a back action as the action of one physical system on another one is always accompanied by a reaction which goes the other way. In Section 17.2 this will be more precisely formulated.

The equations of motion for charges will be kept as they are. In fact the discussion leading us to them (se Section 16.2) is unchanged by the fact that we now take into account that the field is influenced by the charges and there is not much of a choice if we stick with the assumption that the four-acceleration of a charge is linear in the electric four-current $e\hat{U}$ of the charge (and vanishes when this four-current vanishes).

The field equations (16.24) and (16.25) must, however, be modified. Let us write the new equations as

$$D_b\mathcal{F}^{ba} = \mathcal{M}^a \tag{17.1}$$

and

$$D_b \mathcal{F}^{*ba} = \mathcal{N}^a. \tag{17.2}$$

This is obviously a completely general but also completely non-informational modification as far as the vector fields $\mathcal{M}^a$ and $\mathcal{N}^a$ are unspecified. They are known as the *sources* of the electromagnetic field.

From (17.1) and (17.2) the conclusion can immediately be drawn that that $\bar{\mathcal{M}}$ as well as $\bar{\mathcal{N}}$ is divergence free, i.e.,

$$D_a \mathcal{M}^a = D_a \mathcal{N}^a = 0. \tag{17.3}$$

According to Section 15.3 there is thus a conserved quantity corresponding to each of the vector fields. To learn more about them we shall in this section resort to the a posteriori method of referring to some of Maxwell's equations although there are deeper ways investigating the interaction between charges and fields (see Section 17.2). Introducing an observer with four-velocity $\hat{W}$, multiplying (17.1) and (17.2) by $\hat{W}$, using (13.22), and using the two equations (16.51) and (16.52) we obtain

$$\mathrm{div}\bar{\mathcal{E}} \equiv \nabla_a \mathcal{E}^a = W^a \mathcal{M}_a \tag{17.4}$$

$$\mathrm{div}\bar{\mathcal{B}} \equiv \nabla_a \mathcal{B}^a = -W^a \mathcal{N}_a. \tag{17.5}$$

Two of Maxwell's equations run[1] $\mathrm{div}\bar{\mathcal{E}} = 4\pi\varrho = -4\pi\hat{W}\cdot\bar{\mathcal{I}}$ and $\mathrm{div}\bar{\mathcal{B}} = 0$. Here $\bar{\mathcal{I}}$ is the electrical four-current density and we have used the rule at the end of Section 15.1 to obtain the relation between $\bar{\mathcal{I}}$ and the electric charge density ϱ, which now is considered to be a field though without change of notation. The observer four-velocity $\hat{W}$ is arbitrary. Hence $\bar{\mathcal{M}} = -4\pi\bar{\mathcal{I}}$ and $\bar{\mathcal{N}} = 0$. Thus we see that *the conserved quantity corresponding to $\bar{\mathcal{M}}$ is the electric charge and the conserved quantity corresponding to $\bar{\mathcal{N}}$ is identically zero.* The final field equations are

$$D_b \mathcal{F}^{ba} = -4\pi \mathcal{I}^a \tag{17.6}$$

[1]The constant '4π' occurring here is conventional as will be discussed in Section 17.2. The choice of constant is not completely standardized.

and

$$D_b\mathcal{F}^{*ba} = 0. \tag{17.7}$$

17.1.2 Maxwell's Equations

Introducing a four-velocity $\hat{W}$ you can split the electric four-current density as follows

$$\bar{\mathcal{I}} = \varrho\hat{W} + \bar{\imath} \tag{17.8}$$

where $\hat{W} \cdot \bar{\imath} = 0$. From the rule at the end of Section 15.1 it is seen that ϱ is the charge density measured by the observer with the four-velocity $\hat{W}$. The vector field $\bar{\imath}$ is the the *electric current density* measured by the same observer. Using (13.22) we see that the equation $D_a\mathcal{I}^a = 0$ takes the form

$$\frac{\partial}{\partial t}\varrho + \mathrm{div}\bar{\imath} = 0. \tag{17.9}$$

In this equation we recognize the the continuity equation for electric charge. Using the calculations of Section 16.4.2 the complete set of Maxwell's equations is easily obtained

$$\mathrm{div}\ \bar{\mathcal{E}} = 4\pi\varrho \tag{17.10}$$

$$\mathrm{curl}\ \bar{\mathcal{B}} - \frac{\partial}{\partial t}\bar{\mathcal{E}} = 4\pi\bar{\imath} \tag{17.11}$$

$$\mathrm{div}\ \bar{\mathcal{B}} = 0 \tag{17.12}$$

$$\mathrm{curl}\ \bar{\mathcal{E}} + \frac{\partial}{\partial t}\bar{\mathcal{B}} = 0. \tag{17.13}$$

17.1.3 The Potential Equations

Equation (17.7) tells us that the equation (16.25) or alternatively (16.26) is unchanged in the presence of charges. This guarantees that the field tensor can be written in the form (16.15)

$$\mathcal{F}_{ab} = D_a \mathcal{A}_b - D_b \mathcal{A}_a. \tag{17.14}$$

The equations (17.6) and (16.24) give us

$$\Box \mathcal{A}_b - D_b D^a \mathcal{A}_a = -4\pi \bar{\mathcal{I}}. \tag{17.15}$$

The possibility to gauge $\bar{\mathcal{A}}$ is unchanged and we may choose the Lorentz gauge $D^a \mathcal{A}_a = 0$ obtaining

$$\Box \bar{\mathcal{A}} = -4\pi \bar{\mathcal{I}}. \tag{17.16}$$

Using (16.68) and (17.8) this may be split into two equations

$$\Box \phi = -4\pi \varrho \tag{17.17}$$

$$\Box \bar{a} = -4\pi \bar{\imath}. \tag{17.18}$$

Remembering that $\Box = \nabla^2 - \partial^2/\partial t^2$ we recognize in (17.17) and (17.18) the potential equations of the Maxwell theory.

As we shall see in Chapter 18 the equation (17.16) can always be solved for $\bar{\mathcal{A}}$ when the source $\bar{\mathcal{I}}$ is known. It is therefore often advantageous, technically, to calculate the four-potential first and then the field tensor through (17.14).

17.2 Energy-Momentum

17.2.1 Dust

So far we have considered a single test particle in an electromagnetic field. We can reshape the equation of motion for a particle into field equations if we consider instead a cloud of tiny particles, a *dust cloud*, as we did in Section 15.3. The definition of dust contains the assumption that the macroscopic motion is the same as the particle motion, in other words that the particles do not perform any irregular motions.

For simplicity we also assume for the present that all the particles have the same mass and the same charge.

The particle four-velocity now turns into a field $\hat{\mathcal{U}}$. The derivative $d/d\tau$ in equation (16.10) stands for the change per unit proper time along the four-velocity itself. Hence

$$\frac{d}{d\tau}\Psi = D_{\hat{\mathcal{U}}}\Psi = \mathcal{U}^a D_a \Psi \tag{17.19}$$

so that (16.10) turns into

$$m\rho\,\mathcal{U}^b D_b \mathcal{U}_a = \mathcal{F}_{ab}\mathcal{I}^b. \tag{17.20}$$

We have multiplied by the particle density ρ from (15.2) which refers to the four-velocity $\hat{U}$ of the particle flow itself. This density is now considered to be a scalar field. The electric four-current density of the cloud

$$\bar{\mathcal{I}} \equiv e\rho\hat{\mathcal{U}} \tag{17.21}$$

is a vector field. According to Section 15.3 the particle four-current density $\bar{\mathcal{J}} = \rho\hat{\mathcal{U}}$ is divergence free

$$D_a(\rho\,\mathcal{U}^a) = 0. \tag{17.22}$$

Equation (17.20) can therefore be rewritten in the following form

$$D^b \mathcal{T}_{\mathrm{D}ab} = \mathcal{F}_{ab}\,\mathcal{I}^b \tag{17.23}$$

where

$$\mathcal{T}_{\mathrm{D}ab} \equiv m\rho\,\mathcal{U}_a\,\mathcal{U}_b = \mathcal{P}_a\,\mathcal{J}_b \tag{17.24}$$

is *the energy-momentum tensor for dust* introduced in Section 15.3.

Let us consider an observer with four-velocity $\hat{W}$. According to the end of Section 15.1 he will interpret the quantity

$$W^a W^b \mathcal{T}_{\mathrm{D}ab} = (\hat{W}\cdot\bar{\mathcal{P}})(\hat{W}\cdot\bar{\mathcal{J}}) \tag{17.25}$$

as particle energy times particle density, i.e., as energy density. From (17.24) follows immediately that this quantity is positive (remember that $\rho > 0$).

There can be more than one dust cloud in one and the same region of spacetime or the particles we consider may perform irregular motions. In both cases the distribution of particle velocities may locally at each point of spacetime be considered as arising from a distribution of dust clouds. The total four-momentum current density is obtained simply by adding the contributions from all the clouds, in other words by adding the energy-momentum tensors

$$\mathcal{T}_{ab} = \sum_i (\mathcal{T}_{\mathrm{D}i})_{ab}. \tag{17.26}$$

The sum here runs over all the dust clouds. In a special case the distribution of clouds is *isotropic* by which means that for each spacetime point $\bar{R}$ there exists a four-velocity $\hat{V} = \hat{\mathcal{V}}(\bar{R})$ such that an observer with this four-velocity measures dust particles hitting him equally from all directions. Splitting the dust particle four-velocities $\hat{U}_i$ with respect to this four-velocity $\hat{U}_i = \gamma_i(\hat{V} + \bar{v}_i)$ with $\hat{V} \cdot \bar{v}_i = 0$ we obtain

$$\mathcal{T}_{ab} = \sum_i m_i \rho_i \gamma_i^2 (V_a + v_{ia})(V_b + v_{ib}) =$$
$$\sum_i m_i \rho_i \gamma_i^2 (V_a V_b + V_a v_{ib} + v_{ia} V_b + v_{ia} v_{ib}). \tag{17.27}$$

The middle terms here vanish when the sum over all the clouds is performed because the $\bar{v}_i$ are distributed equally over all directions. Let us define the sum over the last term as

$$t_{ab} \equiv \sum_i m_i \rho_i \gamma_i^2 v_{ia} v_{ib}. \tag{17.28}$$

If two vectors $\bar{a}$ and $\bar{b}$ are orthogonal we obtain $t_{ab} a^a b^b = 0$. We realize that this is true by considering a subsum in which all the $\bar{v}_i$-vectors give the same value to $\bar{a} \cdot \bar{v}_i$. In this collection there are pairs of $\bar{v}_i$-vectors giving the same absolute value but opposite signs to $\bar{b} \cdot \bar{v}_i$. If $\bar{a}$ is orthogonal to $\hat{V}$ we obtain $t_{ab} a^a a^b \propto a^a a_a$ due to the isotropy. Finally $t_{ab} V^a = 0$. These properties give a complete description of the tensor t_{ab} except for a scalar factor and we realize that

$$t_{ab} \propto g_{ab} + V_a V_b. \tag{17.29}$$

Summarizing we find that

$$\mathcal{T}_{ab} = (\epsilon + \wp) \mathcal{V}_a \mathcal{V}_b + \wp g_{ab}. \tag{17.30}$$

where the scalar field ϵ is the *total energy density* defined by $\epsilon \equiv \sum_i m_i \rho_i \gamma_i^2$ and $\wp$ is another scalar field, the *pressure*.

A material possessing the energy-momentum tensor (17.30) is known as a *perfect fluid*. It does not have the same simple form of a product between two (dual) vector fields as the dust cloud energy-momentum tensor. Nevertheless its interpretation is obviously the same: $\mathcal{T}_{ab}d\Sigma^b$ is the flux of four-momentum through the hypersurface element $d\Sigma^b$. Considering a hypersurface element of the type S_2 in Figure 15.2 relating to an observer with the four-velocity $\hat{\mathcal{V}}$ we find the flux

$$f_e\, dt = \mathcal{T}_e{}^b \varepsilon_{abcd} \mathcal{V}^a\, dt\, da^c\, db^d = \wp\, \varepsilon_{ecd}\, da^c\, db^d\, dt \qquad (17.31)$$

where the vectors $d\bar{a}$ and $d\bar{b}$ form a 2-surface in the observers orthogonal space. We may also write $\bar{f} = \wp\, d\bar{a} \times d\bar{b}$. This flux of momentum per unit time is composed of concurrent contributions from momentum carried by particles through the surface from one side to the other and particles *with opposite momentum* carrying it *in the opposite direction* through the surface. The momentum flux per unit time is a force and this force will manifest itself on a wall boundary of the the fluid on which the particles are bouncing. The wall takes up the combined momentum flux from the two contributions mentioned. From all said here it should be clear that the term 'pressure' is appropriate for the quantity $\wp$.

Obviously $\mathcal{T}_{ab}W^aW^b > 0$ for any timelike vector $\bar{W}$ because $\mathcal{T}_{ab}W^aW^b$ is a sum of positive terms. Behind this is of course the fact that the total energy always is positive. The stability of our world is actually founded on the fact that the *energy has a lower bound.* If this had not been the case processes could run amok tapping unlimited amounts of energy from some system which reaches lower and lower energies.

Another property which $\mathcal{T}$ inherits from $\mathcal{T}_{\mathrm{D}}$ is symmetry. Also the symmetry has a deep significance. Choosing an origin we may define the third rank tensor

$$\mathcal{M}_{abc} \equiv R_a \mathcal{T}_{bc} - R_b \mathcal{T}_{ac} \qquad (17.32)$$

where $\bar{R}$ is the position vector in spacetime. The tensor $\mathcal{M}_{abc}$ is known as the *angular momentum current density.* Thanks to the symmetry of $\mathcal{T}$ *this tensor is divergence free*

$$D_c \mathcal{M}_{ab}{}^c = g_{ca} \mathcal{T}_b{}^c - g_{cb} \mathcal{T}_a{}^c = \mathcal{T}_{ba} - \mathcal{T}_{ab} = 0 \qquad (17.33)$$

where we also have used the fact that $\mathcal{T}$ is divergence free.

Equation (17.23) shows that the energy-momentum tensor of the dust is not divergence free which means that there is a net change of four-momentum of the dust. If we wish to stick with the idea that the electromagnetic field is a physical system in its own right we should try to interpret this as a *transfer of four-momentum between the dust cloud and the electromagnetic field.* This will be done in the next subsection.

17.2.2 Electromagnetism

Our intention in this subsection is to search for an expression in terms of the electromagnetic field which can serve as electromagnetic energy-momentum tensor. To that end we will first derive quite a useful auxiliary equation. Assume that F_{ab} and H_{ab} are two antisymmetric tensors. Let us use (14.29) to express $H^*{}_{ac}F^{*bc}$ in terms of unstarred fields

$$H^*{}_{ac}F^{*bc} = \frac{1}{4}\varepsilon_{acde}\varepsilon^{bcfg}H^{de}F_{fg} =$$
$$-\frac{1}{2}(\delta^b_a\delta^f_d\delta^g_e + \delta^b_e\delta^f_a\delta^g_d + \delta^b_d\delta^f_e\delta^g_a)H^{de}F_{fg} =$$
$$-\frac{1}{2}(\delta^b_a H^{de}F_{de} + H^{db}F_{ad} + H^{be}F_{ea}) = H^{bc}F_{ac} - \frac{1}{2}\delta^b_a H^{cd}F_{cd}. \tag{17.34}$$

The equation (14.29) contains six δ-terms but in the above derivation the corresponding terms group themselves into pairs of identical terms due to the antisymmetry of H^{de}. A corollary to (17.34) is

$$H^*{}_{ac}F^{*ac} = -H_{ac}F^{ac}. \tag{17.35}$$

We can replace the tensors in (17.34) with tensor fields and we can take the gradient of the $\mathcal{F}$-field on both sides of the equation. Thus the following equation is obtained

$$\mathcal{H}^{bc}D_b\mathcal{F}_{ac} = \mathcal{H}^*{}_{ac}D_b\mathcal{F}^{*bc} + \frac{1}{2}\mathcal{H}^{cd}D_a\mathcal{F}_{cd}. \tag{17.36}$$

Using this relation together with (17.1) and (17.2) we can calculate the following divergence

$$D_b(\mathcal{F}_{ac}\mathcal{F}^{bc}) = (D_b\mathcal{F}_{ac})\mathcal{F}^{bc} + \mathcal{F}_{ac}D_b\mathcal{F}^{bc} =$$
$$(D_b\mathcal{F}^{*bc})\mathcal{F}^*{}_{ac} + \frac{1}{2}\mathcal{F}^{cd}D_a\mathcal{F}_{cd} + \mathcal{F}_{ac}D_b\mathcal{F}^{bc} =$$
$$\mathcal{N}^c\mathcal{F}^*{}_{ac} + \mathcal{M}^c\mathcal{F}_{ac} + \frac{1}{4}D_a(\mathcal{F}^{cd}\mathcal{F}_{cd}). \tag{17.37}$$

Moving the last term over to the left hand side we obtain a divergence equal to two 'source terms'. Alternatively we may write the equation in a more symmetric way: Obviously we obtain an equality if we in (17.37) replace $\mathcal{F}$ by $\mathcal{F}^*$ and exchange $\mathcal{M}$ and $\mathcal{N}$. Adding the resulting equation to (17.37) we obtain

$$\frac{1}{2}D_b(\mathcal{F}_{ac}\mathcal{F}^{bc} + \mathcal{F}^*{}_{ac}\mathcal{F}^{*bc}) = \mathcal{N}^c\mathcal{F}^*{}_{ac} + \mathcal{M}^c\mathcal{F}_{ac}. \tag{17.38}$$

Two terms have cancelled due to (17.35).

Using $\mathcal{N}^a = 0$ and $\mathcal{M}^a = -4\pi\mathcal{I}^a$ from Section 17.1.1 in (17.38) and combining it with (17.23) you obtain immediately

$$D^b(\mathcal{T}_{\mathrm{D}ab} + \mathcal{T}_{\mathrm{EM}ab}) = 0 \tag{17.39}$$

where the definition

$$\mathcal{T}_{\mathrm{EM}ab} \equiv \frac{1}{8\pi}(\mathcal{F}_{ac}\mathcal{F}^{bc} + \mathcal{F}^*{}_{ac}\mathcal{F}^{*bc}) \tag{17.40}$$

has been introduced. This tensor can obviously function as energy-momentum tensor for electromagnetism. Equation (17.39) shows that the four-momentum lost by the dust can be regarded as gained by the electromagnetic field and vice versa if we define the four-momentum current density by $\mathcal{T}_{\mathrm{EM}ab}$.

Judged only from the balance equation (17.39) $\mathcal{T}_{\mathrm{EM}}$ is not unique. The choice (17.40) of $\mathcal{T}_{\mathrm{EM}}$ has , however, a number of additional interesting properties. Here a few short comments about them.

- The tensor $\mathcal{T}_{\mathrm{EM}ab}$ is *symmetric* like $\mathcal{T}_{\mathrm{D}ab}$, i.e., $\mathcal{T}_{\mathrm{EM}ba} = \mathcal{T}_{\mathrm{EM}ab}$. It is therefore possible to define a conserved angular momentum. See Subsection 17.2.1.
- Let $\hat{W}$ be a four-velocity. Then $\mathcal{E}_a \equiv \mathcal{F}_{ab}W^b$ as well as $\mathcal{B}_a \equiv -\mathcal{F}^*{}_{ab}W^b$ is orthogonal to $\hat{W}$. They are therefore spacelike vectors. Hence from (17.40)

$$W^a\mathcal{T}_{\mathrm{EM}ab}W^b = \frac{1}{8\pi}(\bar{\mathcal{E}}^2 + \bar{\mathcal{B}}^2) > 0. \tag{17.41}$$

  The energy density inequality satisfied by $\mathcal{T}_{\mathrm{D}}$ is thus satisfied also by $\mathcal{T}_{\mathrm{EM}}$.
- The trace of the dust energy-momentum tensor is $\mathcal{T}_{\mathrm{D}}{}^a{}_a = -m\rho$. From (17.35) follows immediately that $\mathcal{T}_{\mathrm{EM}}{}^a{}_a = 0$. The electromagnetic energy-momentum tensor is *traceless*. We have here an indication that if electromagnetic radiation is considered to be a gas of particles (photons) as is done in quantum theory these particles are *mass less*.

- Here we have discussed the field equations first and then chosen $\mathcal{T}_{EM}$ guided by them. Sometimes it is more natural to do this in the opposite order, choosing energy-momentum tensor on the basis of the additional properties which we have discussed here or using a variational technique which is outside the framework of this course. If this is done the source terms in the field equations become dictated by the equations of motion through the requirement of four-momentum balance. To a certain degree this connection is seen also in our treatment: The energy-momentum tensor $\mathcal{T}_{EM}$ is symmetric against exchange of starred and unstarred quantities. The unsymmetrical appearance of the field equations (17.6) and (17.7) can therefore be seen entirely as a result of the equations of motion (16.10). This asymmetry has been much discussed and it has been suggested that another term containing a 'magnetic charge' be introduced in the equations of motion to make them symmetric. This would make also the field equations symmetric. So far this is mere speculation.

- Changing the unit of a quantity is the same as rescaling it. If the charge and the field tensor are rescaled as follows $e \to se$, $\mathcal{F} \to (1/s)\mathcal{F}$ the equations of motion (16.10) are unchanged. Here s is a constant scale factor. In the field equation (17.6) the conventional factor 4π must be replaced by a constant say k and k must be rescaled as follows $k \to s^2 k$ to keep the field equation unchanged. Alternatively, we may have a look at the energy-momentum tensor (17.40). We must there in the same way replace 4π by k and rescale k as above to keep the energy-momentum tensor unchanged. By this rescaling we may choose the absolute value $|k|$ at will but we cannot change the sign of k. In fact we already know that k must be positive for the energy density to be positive. The conventional choice $k = 4\pi$ is tantamount to a decision of the units of e and $\mathcal{F}$.

Chapter 18

Solution of the Wave Equation

18.1 The Green's Function

The importance of the wave equation should at this stage be obvious. We shall now see how to solve the wave equation with a given source, i.e.,

$$\Box\Psi = -4\pi\sigma. \tag{18.1}$$

Here σ is a given scalar field and we wish to solve for the scalar field Ψ. The symbol $\Box$ stands for $D^a D_a$. You may have met the solution as the usual 'retarded potential', an integral over three-dimensional space. Our aim throughout this course has been to use spacetime rather than space and time. We will do the same now and present the solution in a spacetime fashion.

In this chapter a certain familiarity with the Dirac δ-function is required. The δ-function is defined exclusively by its behaviour in integrals

$$\int_{-a}^{b} f(z)\delta(z)\,dz = f(0), \tag{18.2}$$

where the integration region contains the origin and the function f is supposed to 'behave well' at the origin.

We need also the four-dimensional δ^4-function with the property

$$\int \Psi(\bar{R})\delta^4(\bar{R})\,dV = \Psi(0), \tag{18.3}$$

where $\bar{R}$ is the position vector in spacetime, the integral is a volume integral over a spacetime region containing the origin, and Ψ is a scalar field which is well behaved at the origin. Here and in the following we assume that dV is everywhere chosen to be positive.

Suppose that a charge at some event gives rise to an electromagnetic disturbance. Choosing the event as the origin the electromagnetic disturbance propagates on the light cone $\bar{R}^2 = 0$. Let us therefore investigate the field

$$G(\bar{R}) \equiv \delta(\bar{R}^2) \tag{18.4}$$

which is zero everywhere except on the light cone. In particular we would like to calculate the result of the d'Alembertian acting on this scalar field, i.e., $\Box\delta(\bar{R}^2)$. Calculating this is the same as calculating the volume integral $\int \Box\delta(\bar{R}^2)\,dV$ over every region of spacetime. According to Gauss' theorem such an integral over the region V can be written as

$$\int_V \Box\delta(\bar{R}^2)\,dV = \oint D^a\delta(\bar{R}^2)d\Sigma_a = \oint 2\delta'(\bar{R}^2)R^a\,d\Sigma_a. \tag{18.5}$$

The surface integral is to be performed over the closed boundary hypersurface of V. The prime on the δ-function stands for derivative with respect to the argument which in the present case is $\bar{R}^2$. As $dV > 0$ the rules of Gauss' theorem tell us that $R^a\,d\Sigma_a > 0$ if $\bar{R}$ is outward directed and $R^a\,d\Sigma_a < 0$ if $\bar{R}$ is inward directed.

We will choose to integrate over the kind of spacetime volume illustrated in Figure 18.1. The boundary is made of two oppositely directed light cones with different apices.

Let us start calculating the hypersurface integral over L_2 illustrated in Figure 18.2. Obviously the integrand on this hypersurface is different from zero only where the hypersurface intersects the light cone F. For this reason a spacelike separation between the origin and the L_2 apex is uninteresting as there is no such intersection in that case. Let us

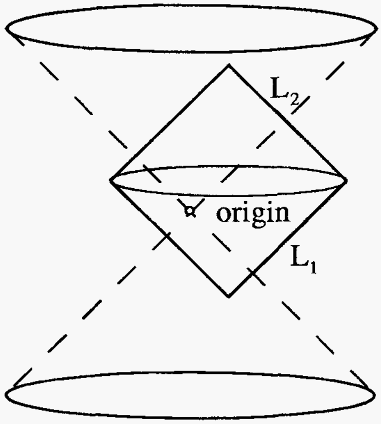

Figure 18.1: Spacetime figure illustrating a volume containing the origin. The boundary is made up of two oppositely directed light cones L_1 and L_2. The future light cone with apex at the origin cuts L_2 and the past light cone with that apex cuts L_1.

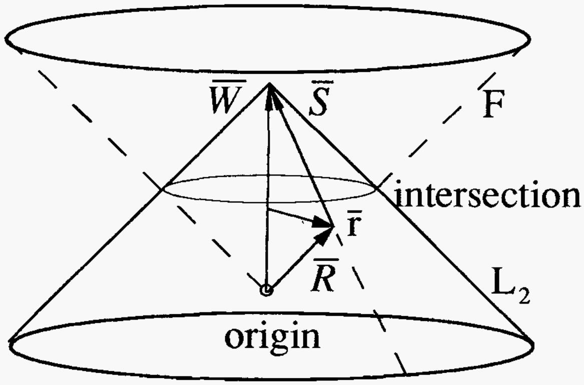

Figure 18.2: Spacetime figure illustrating how the hypersurface L_2 in Figure 18.1 with apex at $\bar{W}$ cuts the future light cone F with apex at the origin. Only in this intersection is the integrand we consider different from zero. The intersection appears as a circle in the figure but is actually a sphere.

therefore assume that the separation vector $\bar{W}$ is timelike. Notice that $\bar{R}$ is outward directed everywhere on L_2 so that $R^a\, d\Sigma_a > 0$.

Let $\hat{W}$ be the four-velocity parallel to $\bar{W}$ and split $\bar{R}$ in the standard manner $\bar{R} = t\hat{W} + r\hat{r}$ (where $\hat{r} \cdot \hat{W} = 0$). A point in spacetime is given by t, r, and the direction $\hat{r}$ in the orthogonal space to $\hat{W}$. The last one can in the well-known manner be given by two spherical polar angles θ and ϕ. At every point in spacetime the vector $\overline{d\theta}$ (in the direction for constant t, r, and ϕ) and $\overline{d\phi}$ (in the direction for constant t, r, and θ) are orthogonal to $\hat{W}$ and $\hat{r}$. This means according to (14.25) that the volume $\varepsilon[\bar{W}, \bar{r}, \overline{d\theta}, \overline{d\phi}]$ is a product of an area in $(\bar{W}, \bar{r})$-space and an area in $(\overline{d\theta}, \overline{d\phi})$-space. The latter one is the well-known factor $r^2 \sin\theta\, d\theta\, d\phi$ obtained in precisely the same way as in Euclidean geometry. To treat the former one it is practical for our purposes to introduce the vector $\bar{S}$ through

$$\bar{R} + \bar{S} = \bar{W} \tag{18.6}$$

and the scalar fields ξ, η, and ζ through

$$\xi(\bar{R}) \equiv \bar{R}^2 \quad \eta(\bar{R}) \equiv \bar{S}^2 \quad \zeta(\bar{R}) \equiv \bar{R} \cdot \bar{S}. \tag{18.7}$$

The square of the area spanned by $\bar{R}$ and $\bar{S}$ is according to (14.25)

$$A^2 = -\begin{vmatrix} \xi & \zeta \\ \zeta & \eta \end{vmatrix} = \zeta^2 - \xi\eta. \tag{18.8}$$

Thus

$$|\varepsilon[\bar{R}, \bar{S}, \bar{\partial}_\theta(\bar{R}), \bar{\partial}_\phi(\bar{R})]| = |A(\bar{R})| \{r(\bar{R})\}^2 \sin\theta. \tag{18.9}$$

Concerning the definition of the $\bar{\partial}$ vector fields see Section 14.1.4. In the following we will often omit the argument '$\bar{R}$ ' not to overload the formulae.

We are interested in performing our hypersurface integral over the hypersurface $\eta = 0$ and we would prefer to use the coordinates ξ, θ, and ϕ on that surface. Thus we would like to calculate

$$R^a d\Sigma_a = |R^a d\Sigma_a| = |\varepsilon[\bar{R}, \bar{\partial}_\xi, \bar{\partial}_\theta, \bar{\partial}_\phi]\, d\xi\, d\theta\, d\phi| \tag{18.10}$$

for $\eta = 0$. The vector field $\bar{\partial}_\xi$ occurring here is defined by the properties

$$\partial_\xi^a D_a \xi = 1 \quad \partial_\xi^a D_a \eta = 0 \tag{18.11}$$

(see Section 14.1.4). Now $D_a\xi(\bar{R}) = 2R_a$ and $D_a\eta(\bar{R}) = -2S_a$ so that $\bar{\partial}_\xi \cdot \bar{R} = 1/2$ and $\bar{\partial}_\xi \cdot \bar{S} = 0$. The solution to these equations is obviously

$$\bar{\partial}_\xi(\bar{R}) = \frac{1}{2A^2}(\zeta\bar{S} - \eta\bar{R}). \tag{18.12}$$

Thus using (18.9)

$$|\varepsilon[\bar{R}, \bar{\partial}_\xi, \bar{\partial}_\theta, \bar{\partial}_\phi]| = |\frac{\zeta}{2A^2}\varepsilon[\bar{R}, \bar{S}, \bar{\partial}_\theta, \bar{\partial}_\phi]| = \frac{|\zeta|}{2|A|}r^2 \sin\theta \underset{\eta\to 0}{=} (r^2/2)\sin\theta. \tag{18.13}$$

For $\eta = 0$ we obviously have $r = w - t$ (where w is defined through $\bar{W} = w\hat{W}$). Combining this with $\xi = r^2 - t^2$ we find

$$r = (1/2)(w + \xi/w). \tag{18.14}$$

Collecting everything we obtain

$$\begin{aligned} &2\int_{L_2} \delta'(\xi)\, R^a d\Sigma_a = \int_{L_2} \delta'(\xi) r^2 \sin\theta\, d\xi\, d\theta\, d\phi = \\ &\pi \int_{-w^2}^{w^2} \delta'(\xi)(w + \xi/w)^2 d\xi = -2\pi \int_{-w^2}^{w^2} \delta(\xi)\frac{1}{w}(w + \xi/w) d\xi \\ &= -2\pi. \end{aligned} \tag{18.15}$$

Here a partial integration has been performed. The 'out-integrated' part vanishes because $\delta(\xi)$ vanishes at the endpoints. The limits for the ξ-integral have been chosen somewhat arbitrarily. The only important thing is that zero is in the integration interval.

We have found that the integral over L_2 has the value -2π provided only that $\bar{W}$ is timelike. For a spacelike $\bar{W}$ the value of the integral is zero. The calculation of the integral over L_1 follows identically the same path as the calculation of the integral over L_2 and the result is the same. Considering the double-top hypersurface of Figure 18.1 we realize that the corresponding surface integral has the value -4π if the enclosed region contains the origin and the value zero if the origin is outside the region. Defining the volume integral as the corresponding hypersurface integral we thus conclude that

$$\Box G(\bar{R}) = \Box\delta(\bar{R}^2) = -4\pi\delta^4(\bar{R}). \tag{18.16}$$

This is a remarkably beautiful and also very useful result. Using it we can solve the wave equation with an arbitrary source term. Equation (18.1) is obviously solved by the following integral over the entire spacetime

$$\Psi(\bar{R}) = \int G(\bar{R} - \bar{R}')\sigma(\bar{R}')dV' \tag{18.17}$$

because

$$\begin{aligned} \Box \int G(\bar{R} - \bar{R}')\sigma(\bar{R}')dV' = \int \Box G(\bar{R} - \bar{R}')\sigma(\bar{R}')dV' = \\ -4\pi \int \delta^4(\bar{R} - \bar{R}')\sigma(\bar{R}')dV' = -4\pi\sigma(\bar{R}). \end{aligned} \tag{18.18}$$

The function G is said to be a *Green's function* for the wave equation. There are other Green's functions which give alternative solutions to (18.1). One of them is the *retarded* Green's function G_R which equals $2G$ everywhere except on the past light cone where it vanishes. It follows from the discussion in this section that

$$\Box G_R(\bar{R}) = -4\pi\delta^4(\bar{R}). \tag{18.19}$$

Using G_R a solution

$$\Psi_R(\bar{R}) = \int G_R(\bar{R} - \bar{R}')\sigma(\bar{R}')dV' \tag{18.20}$$

is obtained in which a disturbance associated with a source does not appear before the source. This solution therefore agrees with the notion of *causality*.

18.2 The Lienard-Wiechert Potential

Using the result of the previous section we can now solve the field equations for electromagnetism. The equations for the four-potential are

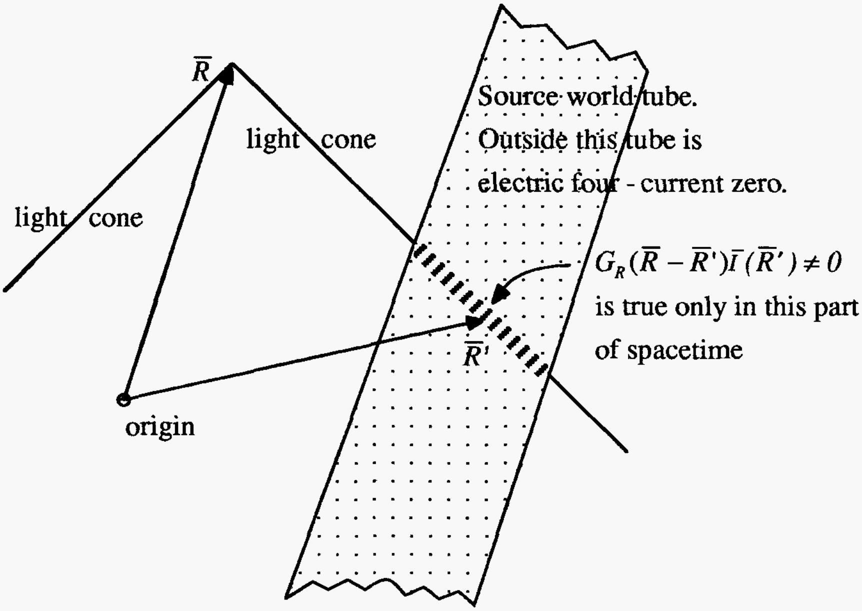

Figure 18.3: Spacetime figure showing the region of $\bar{R}'$ for which $G_R(\bar{R}-\bar{R}')\bar{\mathcal{I}}(\bar{R}')$ is different from zero when $\bar{R}$ is fixed. The source is assumed to be different from zero only within a world tube of finite width.

$$\Box\bar{\mathcal{A}} = -4\pi\bar{\mathcal{I}} \tag{18.21}$$

and

$$D_a\mathcal{A}^a = 0. \tag{18.22}$$

We assume that the four-current density $\mathcal{I}^a$ is a known vector field. The retarded solution is

$$\bar{\mathcal{A}}(\bar{R}) = \int G_R(\bar{R}-\bar{R}')\bar{\mathcal{I}}(\bar{R}')dV' \tag{18.23}$$

where the volume integral extends over all spacetime.

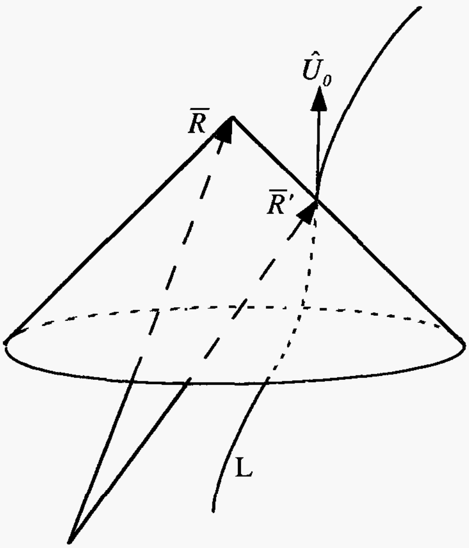

Figure 18.4: A particle world line L intersects the past light cone of the field point $\bar{R}$ at a specific point $\bar{R}_0$. $\hat{U}_0$ is the four-velocity of the particle at that point.

We know from the previous section that this satisfies the first of the above equations. Using the fact that the four-current density is divergence free it is easy to show that it satisfies also the second equation:

$$\begin{aligned} D_a\mathcal{A}^a = \int D_a G_R(\bar{R}-\bar{R}')\mathcal{I}^a(\bar{R}')dV' = \\ \int\{-D'_a G_R(\bar{R}-\bar{R}')\}\mathcal{I}^a(\bar{R}')dV' = \\ \int G_R(\bar{R}-\bar{R}')D'_a\mathcal{I}^a(\bar{R}')dV' = 0. \end{aligned} \tag{18.24}$$

We have here performed a partial integration using Gauss' theorem. The hypersurface term vanishes because we assume that the four-current is different from zero only in a world tube with finite width. For a given 'field point' $\bar{R}$ this makes $G_R(\bar{R}-\bar{R}')\bar{\mathcal{I}}(\bar{R}')$ vanish for $\bar{R}'$ outside a finite part of spacetime (see Figure 18.3).

Let us now put our attention to a special distribution of four-current density — the one arising from one single point charge. It can be written

$$\bar{\mathcal{I}}(\bar{R}') = e\int_L \delta^4(\bar{R}' - \bar{R}'')d\bar{R}'' \tag{18.25}$$

where the line integration is performed along the world line L of a particle and e is its charge. Substituting this into (18.23) we find

$$\bar{\mathcal{A}}(\bar{R}) = e\int_L G_R(\bar{R} - \bar{R}'')\, d\bar{R}''. \tag{18.26}$$

The integrand here is different from zero only on the past light cone with its apex in the field point $\bar{R}$ and the world line L intersects this light cone at *one specific* point say $\bar{R}_0$. This means that $\bar{\mathcal{A}}(\bar{R})$ obtains a contribution from one single point along the world line of the particle (see Figure 18.4). Denote the four-velocity of the particle at this point as $\hat{U}_0$ and introduce the notation $\bar{\rho} \equiv \bar{R} - \bar{R}''$ for convenience. We consider $\bar{R}$ and $\hat{U}_0$ as constants and obtain the following at the point $\bar{R}'' = \bar{R}_0$

$$d\bar{R}'' = -d\bar{\rho} = \hat{U}_0 d\tau \tag{18.27}$$

where τ is the proper time along the world line. The change of $\bar{\rho}^2$ along the world line at $\bar{R}'' = \bar{R}_0$ is

$$d(\bar{\rho}^2) = 2\bar{\rho}_0 \cdot d\bar{\rho} = -2\bar{\rho}_0 \cdot \hat{U}_0 d\tau \tag{18.28}$$

where $\bar{\rho}_0 = \bar{R} - \bar{R}_0$. Hence

$$\bar{\mathcal{A}}(\bar{R}) = e\int_L 2\delta(\bar{\rho}^2)\, d\bar{R}'' = -2e\int_L \delta(\bar{\rho}^2)\, \hat{U}_0/(2\bar{\rho}_0 \cdot \hat{U}_0)d(\bar{\rho}^2) \tag{18.29}$$

or

$$\bar{\mathcal{A}}(\bar{R}) = -\frac{e\hat{U}_0}{(\bar{R} - \bar{R}_0)\cdot\hat{U}_0}. \tag{18.30}$$

This is the famous *Lienard-Wiechert potential.* Notice that it is valid for arbitrary particle motion. The position vector $\bar{R}_0$ marks the point on the particle world line for which $(\bar{R}-\bar{R}_0)^2 = 0$ and $(\bar{R}-\bar{R}_0)\cdot\hat{U}_0 < 0$.

Index

aberration 67, 68
abstract indices 107
active transformation 59
addition
    of tensors 107
    of vectors 21
    of velocities 60
amplitude 63
angular frequency 64
antisymmetric tensors 105
antisymmetry 131
basis 109
causality 190
clock 9
    standard 10
    size of 10
collision 75
    elastic 76
commutativity
    of vector addition 22
components 109
conservation
    of four-momentum 73
    of particles 143
contraction of tensors 111
coordinate basis 136
coordinates 135
d'Alembertian 127
decay 75
delta function 185
    four-dimensional 185
density
    observed 146
    spacetime 144
dependence
    linear 45
derivative
    directional 120
differentiability 120
directional derivative 120
dispersion equation 65, 66
distance 39
divergence 127, 149
Doppler effect 67, 68
downrank 107
dual 104
dual bases 114
dual, star operation 169
dust 151, 179
eigentime 12
elastic collision 76
electric
    charge 176
    charge density 176
    current density 177
    field 171
    four-current 150
electromagnetic field tensor 160

elementary paral. displ. 19
energy 72
energy-momentum tensor
    for dust 151, 179
    for electromagnetism 182
equations of motion
    Lorentz 171
event 15, 16
    space 17
field tensor
    electromagnetic 160
field
    scalar 119
first rank tensors 101
flow
    macroscopic 143
force 104
force tensor 96
four-acceleration 82
four-current density 145
    electric 176
four-momentum 72
four-potential 164
four-vector potential 164
four-vectors 21
four-velocity 49
    for a curved world line 82
Fourier's theorem 158
free wave 158
future light cone 52
gauge 168
Gauss's theorem 149
general relativity 20
gradient 122
    of a tensor field 125
Green's function 190
group four-velocity 65
group velocity 66
hat denoting unit vector 41
homogeneity 26, 45, 161
hyperbolic motion 85
hypersurface element 141, 145
identity tensor 113, 117
independence
    linear 45
indices
    abstract 107
interpretation of sc. prod. 75
inverse metric tensor 116
Jacobian 138
Klein-Gordon equation 158
Laplacian 158
length 39
Levi-Civita tensor 99
Lienard-Wiechert potential 193
light cone 52
lightyears 41
linear
    combination of tensors 102
    dependence 45
    independence 45
    independence of tensors 102
linearity 45
    of tensors 93
    of equations 161
Lorentz
    contraction 62
    equations of motion 171
    gauge 168, 178
    transformation 59
        active 59
        passive 59
lowering index 116
macroscopic flow 143

magnetic field 171
mass 71
massless particles 74
maximal proper time, principle of 13
maximal set of equations 161
Maxwell's equations 169
metric tensor *98*, *116*
  inverse 116
modulations of a light wave 64
moment of inertia tensor 95, 113
momentum 72
multiplication of a tensor by a real number 107
Newtonian limit 71
null-like vector 33, 54
  split of 53
null-tensor 101
operational definition 9
ordered parallelepiped 132
orientation 141. 161
origin 21
orthogonal space 41
orthogonal vectors 30
own time 12
parallel
  displacement 20
  world lines 17
parallelepiped 129
  ordered 132
passive transformation 59
past light cone 52
perfect fluid 181
phase velocity 67
plane wave 63
position vector 21
potential
  four-vector 164
product
  scalar 25
  between tensors 108
projection tensor 165
proper time 12
raising index 117
rank *107*
relative velocity 49
relativity
  principle of 11
  theory of 11
rest energy 74
retarded Green's function 190
scalar field 119
scalar product 25, 26
simplex 135
simultaneity 37
sources 176
space travel 11
space
  orthogonal 41
spacelike 33, 54
spacetime 17
  density 144
  volume 129
special relativity 20
storage ring 14
stress tensor 96
sudden accelerations 12
superposition, principle of 64
symmetric tensors 105
tensor
  antisymmetric 105
  field 122, 124
  product 108
  Levi-Civita 99

contraction of 111
definition of 93
first rank 101
force 96
identity 113, 117
metric- 98, 116
null- 101
rank of 107
symmetric 105
volume- 99
tensors
linear combination of 102
linear independence of 102
threshold conditions 78
time 9
time
dilation 50
timelike 33, 54
transpose of a tensor 104
travel time 12
uprank 107
vector 21
vector
space 101
null-like 33
position- 21
spacelike 33
timelike 33
vectors
addition of 21
dual 104
orthogonal 30
velocity 49
of light 51
addition of 60
volume
tensor 99
in spacetime 129
equation 158
four-vector 63
vector 64
free 158
world line 16
straight 17
curved 81
world lines
parallel 17

2511 hc

ISBN 981-02-1929-6