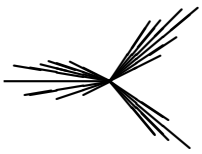


Instituto de Física Teórica
Universidade Estadual Paulista

Notes for a Course on

CLASSICAL FIELDS

R. Aldrovandi and J. G. Pereira

March - June / 2008

Contents

1 Special Relativity: A Recall **1**

1.1 Introduction . . . 1
1.2 Classical Mechanics . . . 2
1.3 Hints Toward Relativity . . . 9
1.4 Relativistic Spacetime . . . 11
1.5 Lorentz Vectors and Tensors . . . 23
1.6 Particle Dynamics . . . 26

2 Transformations **33**

2.1 Transformation Groups . . . 33
2.2 Orthogonal Transformations . . . 37
2.3 The Group of Rotations . . . 42
2.4 The Poincaré Group . . . 49
2.5 The Lorentz Group . . . 50

3 Introducing Fields **59**

3.1 The Standard Prototype . . . 60
3.2 Non-Material Fields . . . 66
3.2.1 Optional reading: the Quantum Line . . . 67
3.3 Wavefields . . . 69
3.4 Internal Transformations . . . 70

4 General Formalism **74**

4.1 Lagrangian Approach . . . 77
4.1.1 Relativistic Lagrangians . . . 77
4.1.2 Simplified Treatment . . . 79
4.1.3 Rules of Functional Calculus . . . 81
4.1.4 Variations . . . 83
4.2 The First Noether Theorem . . . 86

4.2.1 Symmetries and Conserved Charges . . . 88
4.2.2 The Basic Spacetime Symmetries . . . 90
4.2.3 Internal Symmetries . . . 94
4.3 The Second Noether Theorem . . . 96
4.4 Topological Conservation Laws . . . 98

5 Vector Fields **102**
5.1 Real Vector Fields . . . 103
5.2 Complex Vector Fields . . . 105

6 Electromagnetic Field **108**
6.1 Maxwell's Equations . . . 108
6.2 Transformations of $\vec{E}$ and $\vec{H}$ . . . 111
6.3 Covariant Form of Maxwell's Equations . . . 115
6.4 Lagrangian, Spin, Energy . . . 118
6.5 Motion of a Charged Particle . . . 121
6.6 Electrostatics and Magnetostatics . . . 126
6.7 Electromagnetic Waves . . . 129

7 Scalar Fields **136**
7.1 Real Scalar Fields . . . 136
7.2 Complex Scalar Fields . . . 138

8 Dirac Fields **144**
8.1 Dirac Equation . . . 144
8.2 Non-Relativistic Limit: Pauli Equation . . . 149
8.3 Covariance . . . 150
8.4 Lagrangian Formalism . . . 157
8.5 Parity . . . 158
8.6 Charge Conjugation . . . 160
8.7 Time Reversal and $\mathcal{CPT}$ . . . 161

9 Gauge Fields **164**
9.1 Introduction . . . 164
9.2 The Notion of Gauge Symmetry . . . 166
9.3 Global Transformations . . . 168
9.4 Local Transformations . . . 169
9.5 Local Noether Theorem . . . 170
9.6 Field Strength and Bianchi Identity . . . 173

9.7 Gauge Lagrangian and Field Equation 174
9.8 Final Remarks 177

10 Gravitational Field **181**
10.1 General Concepts 181
10.2 The Equivalence Principle 182
10.3 Pseudo-Riemannian Metric 184
10.4 The Notion of Connection 185
10.5 Curvature and Torsion 186
10.6 The Levi-Civita Connection 187
10.7 Geodesics 188
10.8 Bianchi Identities 188
10.9 Einstein's Field Equations 190
10.10 The Schwarzschild Solution 191

Chapter 1

Special Relativity: A Recall

1.1 Introduction

The results of measurements made by an observer depend on the reference frame of that observer. There is, however, a preferred class of frames, in which all measurements give the same results, the so-called *inertial* frames. Such frames are characterized by the following property:

a particle not subject to any force moves with constant velocity.

This is not true if the particle is looked at from an accelerated frame. Accelerated frames are *non-inertial* frames. It is possible to give to the laws of Physics invariant expressions that hold in any frame, accelerated or inertial, but the fact remains that measurements made in different general frames give different results.

Inertial frames are consequently very special, and are used as the basic frames. Physicists do their best to put themselves in frames which are as near as possible to such frames, so that the lack of inertiality produce negligible effects. This is not always realizable, not even always desirable. Any object on Earth's surface will have accelerations (centrifugal, Coriolis, etc). And we may have to calculate what an astronaut in some accelerated rocket would see.

Most of our Physics is first written for inertial frames and then, when necessary, adapted to the special frame actually used. These notes will be exclusively concerned with *Physics on inertial frames*.

We have been very loose in our language, using words with the meanings they have for the man-in-the-street. It is better to start that way. We shall make the meanings more precise little by little, while discussing what is involved in each concept. For example, in the defining property of inertial frames given above, the expression "moves with a constant velocity" is a vector statement: also the velocity

direction is fixed. A straight line is a curve keeping a constant direction, so that the property can be rephrased as

a free particle follows a straight line.

But then we could ask: in which space? It must be a space on which vectors are well defined. Further, measurements involve fundamentally distances and time intervals. The notion of distance presupposes that of a metric. The concept of metric will suppose a structure of differentiable manifold — on which, by the way, derivatives and vectors are well defined. And so on, each question leading to another question. The best gate into all these questions is an examination of what happens in Classical Mechanics.

1.2 Classical Mechanics

§ 1.1 Consider an inertial frame K in which points are attributed cartesian coordinates $\mathbf{x} = (x^1, x^2, x^3)$ and the variable t is used to indicate time. It is usual to introduce unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ along the axes Ox, Oy and Oz with origin O, so that $\mathbf{x} = x^1\,\mathbf{i} + x^2\,\mathbf{j} + x^3\,\mathbf{k}$. Suppose that another frame K' coincides with K initially (at $t = 0$), but is moving with constant velocity $\mathbf{u}$ with respect to that frame. Seen from K', the values of the positions and time variable will be (see Figure 1.1)

$$\mathbf{x}' = \mathbf{x} - \mathbf{u}\,t \tag{1.1}$$

$$t' = t\,. \tag{1.2}$$

These transformations deserve comments and addenda:

1. they imply a simple law for the composition of velocities: if an object moves with velocity $\mathbf{v} = (v^1, v^2, v^3) = (\frac{dx^1}{dt}, \frac{dx^2}{dt}, \frac{dx^3}{dt})$ when observed from frame K, it will have velocity
$$\mathbf{v}' = \mathbf{v} - \mathbf{u} \tag{1.3}$$
when seen from K'.

2. as $\mathbf{u}$ is a constant vector, a constant $\mathbf{v}$ implies that **v'** is also constant, so that K' will be equally inertial; a point which is fixed in K (for example, its origin $\mathbf{x} = 0$) will move along a straight line in K'; and vice-versa: K moves with constant velocity $= -\,\mathbf{u}$ in K'; if a third frame K'' displaces itself with constant velocity with respect to K', it will move with constant velocity with respect to K and will be inertial also; being inertial is a reflexive, symmetric and transitive property; in this logical sense, all inertial frames are equivalent.

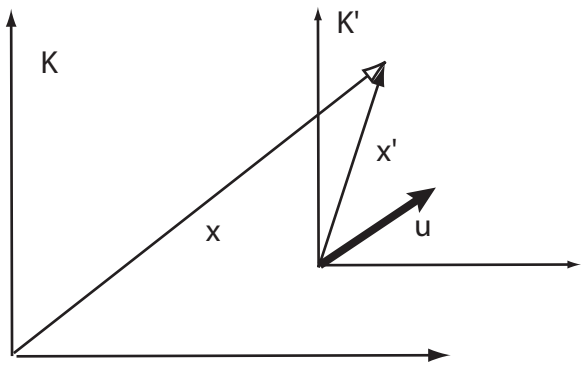


Figure 1.1: *Comparison of two frames.*

3. Newton's force law holds in both reference systems; in fact, its expression in K,

$$m\,\frac{dv^k}{dt} = m\,\frac{d^2x^k}{dt^2} = F^k, \tag{1.4}$$

implies

$$m\,\frac{dv'^k}{dt'} = m\,\frac{d^2x'^k}{dt'^2} = F'^k = F^k\ .$$

A force has the same value if measured in K or in K'. Measuring a force in two distinct inertial frames gives the same result. It is consequently impossible to distinguish inertial frames by making such measurements. Also in this physical sense all inertial frames are equivalent. Of course, the free cases $\mathbf{F}' = \mathbf{F} = 0$ give the equation for a straight line in both frames.

4. equation (1.2), put into words, states that *time is absolute*; given two events, the clocks in K and K' give the same value for the interval of time lapsing between them.

5. transformation (1.1) is actually a particular case. If a rotation of a fixed angle is performed around any axis, the relation between the coordinates will be given by a rotation operator R,

$$\mathbf{x}' = R\,\mathbf{x}\ . \tag{1.5}$$

Rotations are best represented in matrix language. Take the space coordinates

as a column-vector $\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$ and the 3×3 rotation matrix

$$R = \begin{pmatrix} R^1{}_1 & R^1{}_2 & R^1{}_3 \\ R^2{}_1 & R^2{}_2 & R^2{}_3 \\ R^3{}_1 & R^3{}_2 & R^3{}_3 \end{pmatrix} . \tag{1.6}$$

Equation (1.5) becomes

$$\begin{pmatrix} x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} R^1{}_1 & R^1{}_2 & R^1{}_3 \\ R^2{}_1 & R^2{}_2 & R^2{}_3 \\ R^3{}_1 & R^3{}_2 & R^3{}_3 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} . \tag{1.7}$$

The velocity and the force will rotate accordingly; with analogous vector columns for the velocites and forces, $\mathbf{v}' = R\ \mathbf{v}$ and $\mathbf{F}' = R\ \mathbf{F}$. With the transformed values, Newton's law will again keep holding. Recall that a general constant rotation requires 3 parameters (for example, the Euler angles) to be completely specified.

6. a comment on what is meant by "measurements give the same values" is worthwhile. Suppose we measure the force between two astronomical objects. Under a rotation, the force changes its components, and so does the position vectors, etc. The number obtained for the value of the force (that is, the modulus of the force vector) will, however, be the same for a rotated observer and for an unrotated observer.

7. Newton's law is also preserved by translations in space and by changes in the origin of time:

$$\mathbf{x}' = \mathbf{x} - \mathbf{a} \tag{1.8}$$

$$t' = t - a^0 , \tag{1.9}$$

with constant $\mathbf{a}$ and a^0. Eq.(1.8) represents a change in the origin of space. These transformations can be put into a matrix form as follows: add the time coordinate to those of space, in a 5-component vector column $\begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix}$. The transformations are then written

$$\begin{pmatrix} t' \\ x'^1 \\ x'^2 \\ x'^3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & -a^0 \\ 0 & 1 & 0 & 0 & -a^1 \\ 0 & 0 & 1 & 0 & -a^2 \\ 0 & 0 & 0 & 1 & -a^3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} . \tag{1.10}$$

As a 5×5 matrix, the rotation (1.6) takes the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & R^1{}_1 & R^1{}_2 & R^1{}_3 & 0 \\ 0 & R^2{}_1 & R^2{}_2 & R^2{}_3 & 0 \\ 0 & R^3{}_1 & R^3{}_2 & R^3{}_3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.11}$$

8. transformation (1.1) is usually called a pure Galilei transformation, or a galilean *boost*; it can be represented as

$$\begin{pmatrix} t' \\ x'^1 \\ x'^2 \\ x'^3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -u^1 & 1 & 0 & 0 & 0 \\ -u^2 & 0 & 1 & 0 & 0 \\ -u^3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} . \tag{1.12}$$

§ 1.2 Transformations (1.1), (1.2), (1.5), (1.8) and (1.9) can be composed at will, giving other transformations which preserve the laws of classical mechanics. The composition of two transformations produces another admissible transformation, and is represented by the product of the corresponding matrices. There is clearly the possibility of doing no transformation at all, that is, of performing the identity transformation

$$\begin{pmatrix} t' \\ x'^1 \\ x'^2 \\ x'^3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} = \begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} . \tag{1.13}$$

If a transformation is possible, so is its inverse — all matrices above are invertible. Finally, the composition of three transformations obey the associativity law, as the matrix product does. The set of all such transformations constitute, consequently, a group. This is the *Galilei group*. For a general transformation to be completely specified, the values of ten parameters must be given (three for $\mathbf{a}$, three for $\mathbf{u}$, three angles for R, and a^0). The transformations can be performed in different orders: you can, for example, first translate the origins and then rotate, or do it in the inverse order. Each order leads to different results. In matrix language, this is to say that the matrices do not commute. The Galilei group is, consequently, a rather involved non-abelian group. There are many different ways to parameterize a general transformation. Notice that other vectors, such as velocities and forces, can also be attributed 5-component columns and will follow analogous rules.

§ 1.3 The notions of vector and tensor presuppose a group. In current language, when we say that V is a vector in euclidean space, we mean a vector under rotations. That is, V transforms under a rotation R according to

$$V'^i = R^i{}_j V^j.$$

In this expression the so-called "Einstein convention" has been supposed: repeated upper-lower indices are summed over the all the values they can assume. This convention will be used throughout this text. Notice that $i, j, k, ... = 1, 2, 3$. When we say that T is a second-order tensor, we mean that is transforms under rotations according to

$$T'^{ij} = R^i{}_m R^j{}_n T^{mn},$$

and so on for higher-order tensors.

§ 1.4 The notation used above suggests a new concept. The set of columns

$$\begin{pmatrix} t \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix}$$

constitute a vector space, whose members represent all possible positions and times. That vector space is the *classical spacetime*. The concept of spacetime only acquires its full interest in Special Relativity, because this spacetime of Classical Mechanics is constituted of two independent pieces: space itself, and time. It would be tempting, always inspired by the notation, to write $t = x^0$ for the first component, but there is a problem: all components in a column-vector should have the same dimension, which is not the case here. To get dimensional uniformity, it would be necessary to multiply t by some velocity. In Classical Physics, all velocities change in the same way, and so that the 0-th component would have strange transformation properties. In Special Relativity there exists a universal velocity, the velocity of light c, which is the same in every reference frame. It is then possible to define $x^0 = ct$ and build up a space of column-vectors whose components have a well-defined dimensionality.

§ 1.5 We have said that the laws of Physics can be written as expressions which are the same in any frame. This invariant form requires some mathematics, in special the formalism of differential forms. Though it is comfortable to know that laws are frame-independent even if measurements are not, the invariant language is not widely used. The reason is not ignorance of that language. Physics is an experimental science and every time a physicist prepares his apparatuses to take data, (s)he is forced to employ some particular frame, and some particular coordinate system. (S)he must, consequently, know the expressions the laws involved assume in that

particular frame and coordinate system. The laws acquire different expressions in different frames because, seen from each particular frame, they express relationships between *components* of vectors, tensors and the like. In terms of components the secret of inertial frames is that the laws are, seen from them, *covariant*: An equality will have the right hand side and the left hand side changing in the same way under transformations between them.

§ 1.6 The principles of Classical Mechanics can be summed up* in the following statements:

- There are reference systems (or "frames"), called by definition "inertial systems", which are preferential, because
- the laws of nature are the same in all of them (galilean relativity).
- Given an initial inertial system, all the other inertial systems are in uniform rectilinear motion with respect to it (inertia).
- The motion of a physical system is completely determined by its initial state, that is, by the positions and velocities of all its elements for some initial time (classical determinism).
- The basic law says that acceleration, defined as the second time–derivative of the cartesian coordinates, equals the applied force per unit mass: $\ddot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \dot{\mathbf{x}}, t)$ (Newton's law).

As we can detect the acceleration of a system on which we are placed by making measurements, the initial inertial system can be taken as any one with vanishing acceleration.

§ 1.7 Transformations can be introduced in two ways. In the so-called "passive" way, the frames are transformed and then it is found what the coordinates are in the new frames. In the alternative way, called "active", only the coordinates are transformed.

§ 1.8 Let us sum up what has been said, with some signs changed for the sake of elegance. The transformations taking one into another the classical inertial frames are:

* For a splendid discussion, see V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, 1968.

(i) rotations $R(\boldsymbol{\omega})$ of the coordinate axis as in (1.11), where $\boldsymbol{\omega}$ represents the set of three angles necessary to determine a rotation;

(ii) translations of the origins in space and in time: $\mathbf{x}' = \mathbf{x} + \mathbf{a}$ and $t' = t + a^o$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & a^0 \\ 0 & 1 & 0 & 0 & a^1 \\ 0 & 0 & 1 & 0 & a^2 \\ 0 & 0 & 0 & 1 & a^3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} ; \tag{1.14}$$

(iii) uniform motions (galilean boosts) with velocity $\mathbf{u}$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ u^1 & 1 & 0 & 0 & 0 \\ u^2 & 0 & 1 & 0 & 0 \\ u^3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.15}$$

The generic element of the Galilei group can be represented as

$$G(\boldsymbol{\omega}, \mathbf{u}, \mathbf{a}) = \begin{pmatrix} 1 & 0 & 0 & 0 & a^0 \\ u^1 & R^1{}_1 & R^1{}_2 & R^1{}_3 & a^1 \\ u^2 & R^2{}_1 & R^2{}_2 & R^2{}_3 & a^2 \\ u^3 & R^3{}_1 & R^3{}_2 & R^3{}_3 & a^3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.16}$$

Exercise 1.1 This is a particular way of representing a generic element of the Galilei group. Compare it with that obtained by

1. multiplying a rotation and a boost;
2. the same, but in inverse order;
3. performing first a rotation, then a translation;
4. the same, but in inverse order.

Do boosts commute with each other? ■

In terms of the components, the general transformation can be written

$$\begin{aligned} x'^i &= R^i{}_j x^j + u^i t + a^i \\ t' &= t + a^0. \end{aligned} \tag{1.17}$$

In the first expression, we insist, the Einstein convention has been used.

Exercise 1.2 With this notation, compare what results from:

1. performing first a rotation then a boost;
2. the same, but in inverse order;
3. performing first a rotation, then a translation;
4. the same, but in inverse order.

■

The general form of a Galilei transformation is rather complicated. It is usual to leave rotations aside and examine the remaining transformations in separate space directions:

$$x' = x + u^1 t + a^1 \tag{1.18}$$
$$y' = y + u^2 t + a^2 \tag{1.19}$$
$$z' = z + u^3 t + a^3 \tag{1.20}$$
$$t' = t + a^0. \tag{1.21}$$

1.3 Hints Toward Relativity

§ 1.9 In classical physics interactions are given by the potential energy V, which usually depends only on the space coordinates: in various notations, $\mathbf{F}$ = - grad V = $-\nabla V$, or $F^k = -\partial^k V$. Forces on a given particle, caused by all the others, depend only on their position at a given instant; a change in position changes the force *instantaneously*. This instantaneous propagation effect violates experimental evidence. That evidence says two things:

(i) no effect can propagate faster than the velocity of light c and

(ii) that velocity c is a frame-independent universal constant.

This is in clear contradiction with the law of composition of velocities (1.3). This is a first problem with galilean Physics.

§ 1.10 There is another problem. Classical Mechanics has galilean invariance, but Electromagnetism has not. In effect, Maxwell's equations would be different in frames K and K'. They are invariant under rotations and changes of origin in space and time, but not under transformations (1.1) and (1.2). To make things simpler, take the relative velocity $\mathbf{u}$ along the axis Ox of K. Instead of

$$x' = x - u^1 t \tag{1.22}$$
$$t' = t\ , \tag{1.23}$$

Maxwell's equations are invariant under the transformations

$$x' = \frac{x - u^1 t}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{1.24}$$

$$t' = \frac{t - \frac{u}{c^2} x}{\sqrt{1 - \frac{u^2}{c^2}}} \,, \tag{1.25}$$

where c is the velocity of light. These equations call for some comments:

- time is no more absolute;
- u cannot be larger than c;
- they reduce to (1.22) and (1.23) when $\frac{u}{c} \to 0$;
- all experiments confirming predictions of Classical Mechanics consider velocities which are, actually, much smaller that c, whose experimental value is $\approx 2.997 \times 10^8$ m/sec;
- Maxwell's equations, on the other hand, do deal with phenomena propagating at the velocity of light (as light itself).

§ 1.11 These considerations suggest an interesting possibility: that (1.24) and (1.25) be the real symmetries of Nature, with Classical Mechanics as a low-velocity limit. This is precisely the claim of Special Relativity, superbly corroborated by an overwhelming experimental evidence.[†] In particular, the Michelson (1881) experiment showed that the value of c was independent of the direction of light propagation. The light velocity c is then supposed to be a universal constant, which is further the upper limit for the velocity of propagation of any disturbance. This leads to the Poincaré principle of relativity, which supersedes Galilei's. There is a high price to pay: the notion of potential must be abandoned and Mechanics has to be entirely rebuilt, with some other group taking the role of the Galilei group. It is clear, furthermore, that the composition of velocities (1.3) cannot hold if some velocity exists which is the same in every frame.

[†] See, for instance, A.P. French, *Special Relativity*, W.W. Norton, New York, 1968. Or R.K. Pathria, *The Theory of Relativity*, 2nd. edition, Dover, New York, of 1974 but reprinted in 2003. Or still the recent appraisal by Yuan Zhong Zhang, *Special Relativity and its Experimental Foundations*, World Scientific, Singapore, 1997.

1.4 Relativistic Spacetime

§ 1.12 Special Relativity has been built up by Fitzgerald, Lorentz, Poincaré and Einstein through an extensive examination of transformations (1.24), (1.25) and their generalization. The task is to modify the classical structure in some way, keeping the pieces confirmed by experiments involving high-velocity bodies. In particular, the new group should contain the rotation group. After Poincaré and Minkowski introduced the notion of spacetime, a much simpler road was open. We shall approach the subject from the modern point of view, in which that notion play the central role (as an aside: it plays a still more essential role in General Relativity). We have above introduced classical spacetime. That concept was created after special-relativistic spacetime, in order to make comparisons easier. And classical spacetime is, as said in §1.4, a rather artificial construct, because time remains quite independent of space.

§ 1.13 We have said that some other group should take the place of the Galilei group, but that rotations should remain, as they preserve Maxwell's equations. Thus, the group of rotations should be a common subgroup of the new group and the Galilei group. Rotations preserve distances between two points in space. If these points have cartesian coordinates $\mathbf{x} = (x^1, x^2, x^3)$ and $\mathbf{y} = (y^1, y^2, y^3)$, their distance will be

$$d(\mathbf{x}, \mathbf{y}) = \left[(x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2\right]^{1/2}. \tag{1.26}$$

That distance comes from a metric, the Euclidean metric. Metrics are usually defined for infinitesimal distances. The Euclidean 3-dimensional metric defines the distance

$$d\mathbf{l}^2 = \delta_{ij} dx^i dx^j$$

between two infinitesimally close points whose cartesian coordinates differ by $d\mathbf{x} = (dx^1, dx^2, dx^3)$.

A metric, represented by the components g_{ij}, can be represented by an invertible symmetric matrix whose entries are precisely these components. The Euclidean metric is the simplest conceivable one,

$$(\delta_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.27}$$

in cartesian coordinates. This will change, of course, if other coordinate systems are used. Endowed with this metric, the set $\mathbb{R}^3$ of 3-uples of real numbers becomes a metric space. This is the Euclidean 3-dimensional space $\mathbb{E}^3$.

Exercise 1.3 Metric (1.27) is trivial in cartesian coordinates. In particular, it is equal to its own inverse. Look for the expression of that metric (using dl^2) in spherical coordinates, which are given by

$$\left.\begin{array}{c} x = r\cos\theta\cos\phi \\ y = r\cos\theta\sin\phi \\ z = r\sin\theta. \end{array}\right\} \tag{1.28}$$

■

§ 1.14 A metric g_{ij} defines:

a scalar product of two 3-uples $\mathbf{u}$ and $\mathbf{v}$, by

$$\mathbf{u}\cdot\mathbf{v} = g_{ij}u^i v^j \ ; \tag{1.29}$$

several notations are current: $\mathbf{u}\cdot\mathbf{v} = (\mathbf{u},\mathbf{v}) = <\mathbf{u},\mathbf{v}>$;

orthogonality: $\mathbf{u}\perp\mathbf{v}$ when $\mathbf{u}\cdot\mathbf{v} = 0$;

the norm $|\mathbf{v}|$ of a vector $\mathbf{v}$ (or modulus of $\mathbf{v}$) by

$$|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v} = g_{ij}v^i v^j \ ; \tag{1.30}$$

the distance between two points $\mathbf{x}$ and $\mathbf{y}$, defined as

$$d(\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}| \ . \tag{1.31}$$

The scalar product, and consequently the norms and distances, are invariant under rotations and translations. Notice that

1. equation (1.26) is just that euclidean distance.
2. in the euclidean case, because the metric is positive-definite (has all the eigenvalues with the same sign), $d(\mathbf{x},\mathbf{y}) = 0 \Longleftrightarrow \mathbf{x} = \mathbf{y}$;
3. in (1.31) it is supposed that we know what the difference between the two points is; the Euclidean space is also a vector space, in which such a difference is well-defined; alternatively, and equivalently, the differences between the cartesian coordinates, as in Eq.(1.26), can be taken.

§ 1.15 When rotations are the only transformations — in particular, when time is not changed — we would like to preserve the above distance. We have seen that (t, x^1, x^2, x^3) is not dimensionally acceptable. Now, however, with a universal constant c, we can give a try to $x = (ct, x^1, x^2, x^3)$.

It is tempting to inspect a 4-dimensional metric like

$$(\eta_{\alpha\beta}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{1.32}$$

With points of spacetime indicated as $x = (ct, x^1, x^2, x^3)$ that metric would lead to the infinitesimal distance

$$ds^2 = c^2dt^2 - d\mathbf{l}^2$$

and to the finite distance

$$d(x,y) = |x-y| = \left[(ct^1 - ct^2)^2 - (x^1 - y^1)^2 - (x^2 - y^2)^2 - (x^3 - y^3)^2\right]^{1/2} . \tag{1.33}$$

This is invariant under rotations, transformations (1.24), (1.25) and their generalizations. Actually, the *Lorentz metric*

$$ds^2 = c^2dt^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \tag{1.34}$$

turns out to be the metric of relativistic spacetime. It is usual to define the variable $x^0 = ct$ as the 0-th (or 4-th) coordinate of spacetime, so that

$$ds^2 = \eta_{\alpha\beta}dx^\alpha dx^\beta = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 . \tag{1.35}$$

The Lorentz metric defines a scalar product which is relativistically invariant, as well as the other notions defined by any metric as seen in § 1.14. Of course, things have been that easy because we knew the final result, painfully obtained by our forefathers. Minkowski space (actually, spacetime) is the set $\mathbb{R}^4$ of ordered 4-uples with the metric (1.35) supposed.

The overall sign is a matter of convention. The relative sign is, however, of fundamental importance, and makes a lot of difference with respect to a positive-definite metric. In particular, $d(x,y) = 0$ no more implies $x = y$. It shows also *how* the time variable differs from the space coordinates. We shall examine its consequences little by little in what follows.

§ 1.16 A metric is used to lower indices. Thus, a variable x_i is defined by $x_i = g_{ij}x^j$. We have insisted that the metric be represented by an invertible matrix. The entries of the inverse to a metric g_{ij} are always represented by the notation g^{ij}. The inverse metric is used to raise indices: $x^i = g^{ij}x_j$. We are, as announced, using Einstein's notation for repeated indices. In Euclidean spaces described in cartesian

coordinates, upper and lower indices do not make any real difference. But they do make a great difference in other coordinate systems.

Points $x = (x^\alpha) = (x^0, x^1, x^2, x^3)$ of relativistic spacetime are called *events*. The ds in (1.35) is called the *interval*. The conventional overall sign chosen above is mostly used by people working on Field Theory. It has one clear disadvantage: upper and lower indices in cartesian coordinates on 3-space differ by a sign. Notice that we use upper-indexed notation for coordinates and some other objects (such as velocities v^k, u^k and forces F^k). Another point of notation: position vectors in 3-space are indicated by boldfaced ($\mathbf{x}$) or arrowed ($\vec{x}$) letters, while a point in spacetime is indicated by simple letters (x).

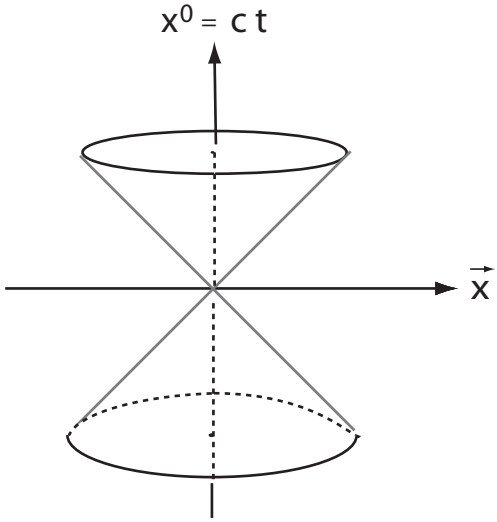


Figure 1.2: *Light cone of an event at the origin.*

§ 1.17 The light cone Expression (1.33) is not a real distance, of course. It is a "pseudo-distance". Distinct events can be at a zero pseudo-distance of each other. Fix the point $y = (y^0, y^1, y^2, y^3)$ and consider the set of points x at a vanishing pseudo-distance of y. The condition for that,

$\eta_{\alpha\beta}(x^\alpha - y^\alpha)(x^\beta - y^\beta) =$

$$(x^0 - y^0)^2 - (x^1 - y^1)^2 - (x^2 - y^2)^2 - (x^3 - y^3)^2 = 0 \tag{1.36}$$

is the equation of a cone (a 3-dimensional conic hypersurface). Take y at the origin (the cone apex) and use (1.35): $c^2dt^2 = d\mathbf{x}^2$. A particle on the cone will consequently have velocity $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ satisfying $\mathbf{v}^2 = c^2$, so that $|\mathbf{v}| = c$. This cone is called the *light*

cone of event y. Any light ray going through $y = 0$ will stay on that hypersurface, as will any particle going through $y = 0$ and traveling at the velocity of light. The situation is depicted in Figure 1.2, with the axis $x^0 = ct$ as the cone axis.

§ 1.18 Causality Notice that particles with velocities $v < c$ stay inside the cone. As no perturbation can travel faster than c, any perturbance at the cone apex will affect only events inside the upper half of the cone (called the *future cone*). On the other hand, the apex event can only be affected by incidents taking place in the events inside the lower cone (the *past cone*). This is the main role of the Lorentz metric: to give a precise formulation of *causality* in Special Relativity. Notice that causality somehow organizes spacetime. If point Q lies inside the (future) light cone of point P, then P lies in the (past) light cone of Q. Points P and Q are *causally related*. Nevertheless, a disturbance in Q will not affect P. In mathematical terms, the past-future relationship is a *partial ordering*, "partial" because not every two points are in the cones of each other.

The horizontal line in Figure 1.2 stands for the present 3-space. Its points lie outside the cone and cannot be affected by whatever happens at the apex. The reason is clear: it takes time for a disturbance to attain any other point. Only points in the future can be affected. Classical Mechanics should be obtained the limit $c \to \infty$. We approach more and the classical vision by opening the cone solid angle. If we open the cone progressively to get closer and closer the classical case, the number of 3-points in the possible future (and possible past) increases more and more. In the limit, the present is included in the future and in the past: instantaneous communication becomes possible.

§ 1.19 Types of interval The above discussion leads to a classification of intervals between two points P and Q. If one is inside the light cone of the other, so that one of them can influence the other, then their interval is positive. That kind of interval is said to be *timelike*. Negative intervals, separating points which are causally unconnected, are called *spacelike*. And vanishing intervals, lying on the light cones of both points, are *null*. A real particle passing through an event will follow a line inside the future line cone of that event, which is called its *world line*. Real world lines are composed of timelike or null intervals. To strengthen the statement that no real distances are defined by the Lorentz metric, let us notice that there is always a "zero-length" path between any two points in Minkowski space. In order to see it, (i) draw the complete light cones of both points (ii) look for their intersection and (iii) choose a path joining the points while staying on the light cones.

§ 1.20 Proper time Let us go back to the interval (1.35) separating two nearby events. Suppose two events at the same position in 3-space, so that $d\mathbf{l}^2 = d\vec{x}^2 = 0$. They are the same point of 3-space at different times, and their interval reduces to

$$ds^2 = (dx^0)^2 = c^2 dt^2 \ . \tag{1.37}$$

An observer fixed in 3-space will have that interval, which is pure coordinate time. (S)he will be a "pure clock". This time measured by a fixed observer is its *proper time*. Infinitesimal proper time is just ds. Let us now attribute coordinates x' to this clock in its own frame, so that $ds = cdt'$, and compare with what is seen by a nearby observer, with respect to which the clock will be moving and will have coordinates x (including a clock). Interval invariance will give

$$ds^2 = c^2 dt'^2 = c^2 dt^2 - d\mathbf{x}^2 = c^2 dt^2 (1 - \frac{d\mathbf{x}^2}{c^2 dt^2}),$$

or

$$dt' = dt \left(1 - \frac{v^2}{c^2}\right)^{1/2} \tag{1.38}$$

with the velocity $\mathbf{v} = \frac{d\mathbf{x}}{dt}$. By integrating this expression, we can get the relationship between a finite time interval measured by the fixed clock and the same interval measured by the moving clock:

$$t'_2 - t'_1 = \int_{t_1}^{t_2} dt \left(1 - \frac{v^2}{c^2}\right)^{1/2} \ . \tag{1.39}$$

If both observers are inertial $\mathbf{v}$ is constant and the relationship between a finite proper time lapse $\Delta t'$ and the same lapse measured by the moving clock is

$$\Delta t' = \left(1 - \frac{v^2}{c^2}\right)^{1/2} \Delta t \ \leq \ \Delta t \ . \tag{1.40}$$

Proper time is smaller than any other time. This is a most remarkable, non-intuitive result, leading to some of the most astounding confirmations of the theory. It predicts that "time runs slower" in a moving clock. An astronaut will age less than (his) her untravelling twin brother (the *twin paradox*). A decaying particle moving fast will have a longer lifetime when looked at from a fixed clock (*time dilatation*, or *time dilation*).

Exercise 1.4 Consider a meson μ. Take for its mean lifetime $2.2 \times 10^{-6} s$ in its own rest system. Suppose it comes from the high atmosphere down to Earth with a velocity $v = 0.9c$. What will be its lifetime from the point of view of an observer at rest on Earth ? ■

§ 1.21 We have arrived at the Lorentz metric by generalizing the Euclidean metric to 4-dimensional spacetime. Only the sign in the time coordinate differs from an Euclidean metric in 4-dimensional space. That kind of metric is said to be "pseudo-Euclidean". The group of rotations preserves the Euclidean metric of $\mathbb{E}^3$. The group generalizing the rotation group so as to preserve the pseudo-Euclidean Lorentz metric is a group of "pseudo-rotations" in 4-dimensional space, the *Lorentz group*. There are 3 independent rotations in 3-space: that on the plane xy, that on plane yz and that on plane zx. In four space with an extra variable $\tau = ct$, we should add the rotations in planes $x\tau$, $y\tau$ and $z\tau$. Due to the relative minus sign, these transformations are pseudo-rotations, or rotations with imaginary angles. Instead of sines and cosines, hyperbolic functions turn up. The transformation in plane $x\tau$ which preserves $\tau^2 - x^2$ is

$$x' = x\cosh\phi + \tau\sinh\phi \quad ; \quad \tau' = x\sinh\phi + \tau\cosh\phi. \tag{1.41}$$

Indeed, as $\cosh^2\phi - \sinh^2\phi = 1$, $\tau'^2 - x'^2 = \tau^2 - x^2$. In order to find ϕ, consider in frame K the motion of a particular point, the origin of frame K' moving with velocity $u = x/t$. From $x' = 0$ we obtain $\tanh\phi = -u/c$, $\cosh\phi = \left(1 - \frac{u^2}{c^2}\right)^{-1/2}$ and $\sinh\phi = -\frac{u}{c}\left(1 - \frac{u^2}{c^2}\right)^{-1/2}$. Inserting these values in the transformation expression, we find just (1.24) and (1.25),

$$x' = \frac{x - ut}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{1.42}$$

$$t' = \frac{t - \frac{u}{c^2}x}{\sqrt{1 - \frac{u^2}{c^2}}}\,, \tag{1.43}$$

which are the *Lorentz transformations* of the variables x and t. Such transformations, involving one space variable and time, are called *pure Lorentz transformations*, or *boosts*. The group generalizing the rotations of $\mathbb{E}^3$ to 4-dimensional spacetime, the Lorentz group, includes 3 transformations of this kind and the 3 rotations. To these we should add the translations in 4-space, representing changes in the origins of the four coordinates. The 10 transformations thus obtained constitute the group which replaces the Galilei group in Special Relativity, the *Lorentz inhomogeneous group* or *Poincaré group*.

§ 1.22 Rotations in a d-dimensional space are represented by orthogonal $d \times d$ matrices with determinant $= +1$. The group of orthogonal $d \times d$ matrices is indicated by the symbol $O(d)$. They include transformations preserving the euclidean metric in d dimensions (this will be seen below, in section 2.2). Those with determinant $= +1$

are called "special" because they are continuously connected to the identity matrix. They are indicated by $SO(d)$. Thus, the rotations of $\mathbb{E}^3$ form the group $SO(3)$. This nomenclature is extended to pseudo-orthogonal groups, which preserve pseudo-euclidean metrics. The pseudo-orthogonal transformations preserving a pseudo-euclidean metric with p terms with one sign and $d-p$ opposite signs is labeled $SO(p, d-p)$. The Lorentz group is $SO(3,1)$. The group of translations in such spaces is variously denoted as T^d or $T^{p,d-p}$. For spacetime the notations T^4 and $T^{3,1}$ are used. Translations do not commute with rotations or pseudo-rotations. If they did, the group of Special Relativistic transformations would be the direct product of $SO(3,1)$ and $T^{3,1}$. The Poincaré group is a semi-direct product, indicated $\mathcal{P} = SO(3,1) \oslash T^{3,1}$.

§ 1.23 The inverses to transformations (1.42) and (1.43) are

$$x = \frac{x' + ut'}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{1.44}$$

$$\begin{aligned} y' &= y \\ z' &= z \end{aligned}$$

$$t = \frac{t' + \frac{u}{c^2}x'}{\sqrt{1 - \frac{u^2}{c^2}}} \ . \tag{1.45}$$

Exercise 1.5 Show it. ■

§ 1.24 Take again a clock at rest in K', and consider two events at the same point (x', y', z') in K', separated by a time interval $\Delta t' = t'_2 - t'_1$. What will be their time separation Δt in K ? From (1.45), we have

$$t_1 = \frac{t'_1 + \frac{u}{c^2}x'}{\sqrt{1 - \frac{u^2}{c^2}}} \ ; \ \ t_2 = \frac{t'_2 + \frac{u}{c^2}x'}{\sqrt{1 - \frac{u^2}{c^2}}} \ ,$$

whose difference gives just (1.40):

$$\Delta t = t_2 - t_1 = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} \Delta t' \ . \tag{1.46}$$

§ 1.25 Lorentz contraction Take now a measuring rod at rest in K, disposed along the axis Ox. Let x_2 and x_1 be the values of the x coordinates of its extremities at a given time t, and $\Delta x = x_2 - x_1$ its length. This length $l_0 = \Delta x$, measured in its own rest frame, is called *proper length*. What would be that length seen from K', also at a fixed time t' ? Equation (1.44) gives

$$x_1 = \frac{x'_1 + ut'}{\sqrt{1 - \frac{u^2}{c^2}}} \ ; \ \ x_2 = \frac{x'_2 + ut'}{\sqrt{1 - \frac{u^2}{c^2}}},$$

whose difference is

$$\Delta x = x_2 - x_1 = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} \Delta x' \ . \tag{1.47}$$

Thus, seen from a frame in motion, the length $l = \Delta x'$ is always smaller than the proper length:

$$l = l_0 \ \left(1 - \frac{u^2}{c^2}\right)^{1/2} \ \leq \ l_0 \ . \tag{1.48}$$

Proper length is larger than any other. This is the *Lorentz contraction*, which turns up for space lengths. The proper length is the largest length a rod can have in any frame. The ubiquitous expression $\left(1 - \frac{u^2}{c^2}\right)^{1/2}$ is called the *Lorentz contraction factor*. It inverse is indicated by

$$\gamma = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} \ . \tag{1.49}$$

This notation is almost universal, and so much so that the factor is called the *gamma factor*. Equations (1.44), (1.45), (1.46) and (1.48) acquire simpler aspects,

$$\begin{aligned} x &= \gamma(x' + ut') && (1.50)\\ t &= \gamma(t' + \frac{u}{c^2}x') && (1.51)\\ \Delta t &= \gamma \ \Delta t_0 && (1.52)\\ l_0 &= \gamma \ l \ , && (1.53) \end{aligned}$$

and so do most of the expressions found up to now. Also almost universal is the notation

$$\beta = \frac{u}{c} \ , \tag{1.54}$$

such that $\gamma = \frac{1}{\sqrt{1-\beta^2}}$.

§ 1.26 What happens to a volume in motion ? As the displacement takes place along one sole direction, a volume in motion is contracted according to (1.48), that is:

$$V = V_0 \ \left(1 - \frac{u^2}{c^2}\right)^{1/2} \quad or \quad V_0 = \gamma \ V \ . \tag{1.55}$$

§ 1.27 Composition of velocities Let us now go back to the composition of velocities. We have said that, with the universality of light speed, it was impossible to retain the simple rule of galilean mechanics. Take the differentials of (1.50) and (1.51), including the other variables:

$$\begin{aligned} dx &= \gamma(dx' + udt') \\ dy &= dy'; \quad dz = dz' \\ dt &= \gamma(dt' + \frac{u}{c^2}dx') \ . \end{aligned}$$

Dividing the first 3 equations by the last,

$$\frac{dx}{dt} = \frac{dx' + udt'}{dt' + \frac{u}{c^2}dx'} \; ; \quad \frac{dy}{dt} = dy' \frac{\sqrt{1 - \frac{u^2}{c^2}}}{dt' + \frac{u}{c^2}dx'}; \quad \frac{dz}{dt} = dz' \frac{\sqrt{1 - \frac{u^2}{c^2}}}{dt' + \frac{u}{c^2}dx'} \; .$$

Now factor dt' out in the right hand side denominators:

$$v_x = \frac{v'_x + u}{1 + \frac{u}{c^2}v'_x} \quad ; \quad v_y = v'_y \frac{\sqrt{1 - \frac{u^2}{c^2}}}{1 + \frac{u}{c^2}v'_x} \quad ; \quad v_z = v'_z \frac{\sqrt{1 - \frac{u^2}{c^2}}}{1 + \frac{u}{c^2}v'_x} \quad . \tag{1.56}$$

These are the composition laws for velocities. Recall that the velocity u is supposed to point along the Ox axis. If also the particle moves only along the Ox axis ($v_x \parallel u$, $v_y = v_z = 0$), the above formulae reduce to

$$v = \frac{v' + u}{1 + \frac{uv'}{c^2}} \; . \tag{1.57}$$

The galilean case (1.3) is recovered in the limit $u/c \to 0$. Notice that we have been forced to use all the velocity components in the above discussion. The reason lies in a deep difference between the Lorentz group and the Galilei group: Lorentz boosts in different directions, unlike galilean boosts, do not commute. This happens because, though contraction is only felt along the transformation direction, time dilatation affects all velocities — and, consequently, the angles they form with each other.

§ 1.28 Angles and aberration Let us see what happens to angles. In the case above, choose coordinate axis such that the particle velocity lies on plane xy. In systems K and K', it will have components $v_x = v\cos\theta$; $v_y = v\sin\theta$ and $v'_x = v'\cos\theta'$; $v'_y = v'\sin\theta'$, with obvious choices of angles. We obtain then from (1.56)

$$\tan\theta = \frac{v'\sin\theta'}{u + v'\cos\theta'}\sqrt{1 - \frac{u^2}{c^2}} \; . \tag{1.58}$$

Thus, also the velocity directions are modified by a change of frame. In the case of light propagation, $v = v' = c$ and

$$\tan\theta = \frac{\sin\theta'}{u/c + \cos\theta'}\sqrt{1 - \frac{u^2}{c^2}} \; . \tag{1.59}$$

This is the formula for *light aberration*. The aberration angle $\Delta\theta = \theta' - \theta$ has a rather intricate expression which tends, in the limit $u/c \to 0$, to the classical formula

$$\Delta\theta = \frac{u}{c}\sin\theta' \; . \tag{1.60}$$

Exercise 1.6 Show it, using eventually

$$\tan(\theta' - \theta) = \frac{\tan\theta' - \tan\theta}{1 + \tan\theta' \tan\theta} \ .$$

■

§ 1.29 Four-vectors We have seen that the column (ct, x, y, z) transforms in a well-defined way under Lorentz transformations. That way of transforming defines a *Lorentz vector*: any set $V = (V^0, V^2, V^2, V^3)$ of four quantities transforming like (ct, x, y, z) is a Lorentz vector, or *four-vector*. By (1.50) and (1.51), they will have the behavior

$$V^1 = \gamma(V'^1 + \frac{u}{c} V'^0) \tag{1.61}$$

$$V^2 = V'^2 \ ; \ V^3 = V'^3 \tag{1.62}$$

$$V^0 = \gamma(V'^0 + \frac{u}{c} V'^1) \ . \tag{1.63}$$

It is usual to call V^0 the "time component" of V, and the V^k's, "space components".

§ 1.30 The classification discussed in § 1.19, there concerned with space and time coordinates, can be extended to four-vectors. A four-vector V is timelike if $|V|^2 = \eta_{\alpha\beta} V^\alpha V^\beta$ is > 0; V is spacelike if $|V|^2 < 0$; and a null vector if $|V|^2 = 0$. Real velocities, for example, must be timelike or null. But we have beforehand to say what we understand by a velocity in 4-dimensional spacetime.

§ 1.31 The **four-velocity** of a massive particle is defined as the position variation with proper time:

$$u^\alpha = \frac{dx^\alpha}{ds} \ . \tag{1.64}$$

Writing

$$ds = \frac{cdt}{\gamma} \ , \tag{1.65}$$

we see that

$$u^1 = \gamma \frac{v^x}{c} \quad ; \quad u^2 = \gamma \frac{v^y}{c} \quad ; \quad u^3 = \gamma \frac{v^z}{c} \ , \tag{1.66}$$

with $v^x = \frac{dx}{dt}$, etc, and $\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$. In the same way we find the fourth component, simply

$$u^0 = \gamma \ . \tag{1.67}$$

The four-velocity is, therefore, the four-vector

$$u = \gamma \left(1, \frac{v^x}{c}, \frac{v^y}{c}, \frac{v^z}{c}\right) \ . \tag{1.68}$$

This velocity has a few special features. First, it has dimension zero. The usual dimension can be recovered by multiplying it by c, but it is a common practice to leave it so. Second, its components are not independent. Indeed, it is immediate from (1.66) and (1.67) that u has unit modulus (or unit norm): $u^2 = (u^0)^2 - \vec{u}^2 = \gamma^2 - \gamma^2 \frac{\vec{v}^2}{c^2} = \gamma^2(1 - \frac{\vec{v}^2}{c^2})$.

$$\therefore \; u^2 = \eta_{\alpha\beta} \, u^\alpha u^\beta = 1 \; . \tag{1.69}$$

Four-velocities lie, consequently, on a hyperbolic space.

Acceleration is defined as

$$a^\beta = \frac{d^2 x^\beta}{ds^2} = \frac{du^\beta}{ds} \; . \tag{1.70}$$

Taking the derivative $\frac{d}{ds}$ of (1.69), it is found that velocity and acceleration are always orthogonal to each other:

$$a \cdot u = a^\alpha u_\alpha = \eta_{\alpha\beta} \, a^\alpha u^\beta = 0 \; . \tag{1.71}$$

Only for emphasis: we have said in §1.14 that a metric defines a scalar product, orthogonality, norm, etc. Both the above scalar product and the modulus (1.69) are those defined by the metric η.

§ 1.32 Quantities directly related to velocities are extended to 4-dimensional spaces in a simple way. Suppose a particle with electrical charge e moves with a velocity $\vec{v}$. Its current will be $\vec{j} = e\,\vec{v}$. A four-vector current is defined as

$$j^\alpha = e \, u^\alpha \; , \tag{1.72}$$

or

$$j = e\,\gamma \; \left(1, \frac{v^x}{c}, \frac{v^y}{c}, \frac{v^z}{c} \right) \; . \tag{1.73}$$

Electromagnetism can be written in terms of the scalar potential ϕ and the vector potential $\vec{A}$. They are put together into the four-vector potential

$$A = (\phi, \vec{A}) = (\phi, A^x, A^y, A^z) \; . \tag{1.74}$$

Invariants turn up as scalar products of four-vectors. The interaction of a current with an electromagnetic field, appearing in the classical Lagrangean, is of the "current-potential" type, $\vec{j} \cdot \vec{A}$. The interaction of a charge e with a static electromagnetic field is $e\phi$. These forms of interaction are put together in the scalar

$$j \cdot A = e \; A_\alpha u^\alpha = e \; A_\alpha \frac{dx^\alpha}{ds} \; . \tag{1.75}$$

§ 1.33 The results of §1.8 are adapted accordingly. A 4×4 matrix Λ will represent a Lorentz transformation which, acting on a 4-component column x, gives the transformed x'. Equations (1.17) are replaced by

$$x'^{\alpha} = \Lambda^{\alpha}{}_{\beta} x^{\beta} + a^{\alpha} \tag{1.76}$$

The boosts are now integrated into the (pseudo-)orthogonal group, of which Λ is a member. Translations in space and time are included in the four-vector a. As for the Galilei group, a 5×5 matrix is necessary to put pseudo-rotations and translations together. The matrix expression of the general Poincaré transformation (1.76) has the form

$$x' = L\,x = \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \\ 1 \end{pmatrix} = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 & a^0 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 & a^1 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 & a^2 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 & a^3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} . \tag{1.77}$$

These transformations constitute a group, the *Poincaré group*. The Lorentz transformations are obtained by putting all the translation parameters $a^{\alpha} = 0$. In this no-translations case, a 4×4 version suffices. The complete, general transformation matrix is highly complicated, and furthermore depends on the parameterization chosen. In practice, we decompose it in a product of rotations and boosts, which is always possible. The boosts, also called "pure Lorentz transformations", establish the relationship between unrotated frames which have a relative velocity $\vec{v} = v\vec{n}$ along the unit vector $\vec{n}$. They are given by

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\frac{v}{c}\,n^1 & -\gamma\frac{v}{c}\,n^2 & -\gamma\frac{v}{c}\,n^3 \\ -\gamma\frac{v}{c}\,n^1 & 1+(\gamma-1)n^1n^1 & (\gamma-1)n^1n^2 & (\gamma-1)n^1n^3 \\ -\gamma\frac{v}{c}\,n^2 & (\gamma-1)n^2n^1 & 1+(\gamma-1)n^2n^2 & (\gamma-1)n^2n^3 \\ -\gamma\frac{v}{c}\,n^3 & (\gamma-1)n^3n^1 & (\gamma-1)n^3n^2 & 1+(\gamma-1)n^3n^3 \end{pmatrix} , \tag{1.78}$$

where, as usual, $\gamma = (1 - v^2/c^2)^{-1/2}$.

1.5 Lorentz Vectors and Tensors

§ 1.34 We have said in § 1.3 that vectors and tensors always refer to a group. They actually ignore translations. Vectors are differences between points (technically, in an affine space), and when you do a translation, you change both its end-points, so that its components do not change. A Lorentz vector obeys

$$V'^{\alpha} = \Lambda^{\alpha}{}_{\beta} V^{\beta} . \tag{1.79}$$

A 2nd-order Lorentz tensor transforms like the product of two vectors:

$$T'^{\alpha\beta} = \Lambda^{\alpha}{}_{\gamma}\Lambda^{\beta}{}_{\delta}T^{\gamma\delta} \ . \tag{1.80}$$

A 3rd-order tensor will transform like the product of 3 vectors, with three Λ-factors, and so on. Such vectors and tensors are *contravariant* vectors and tensors, which is indicated by the higher indices. Lower indices signal *covariant* objects. This is a rather unfortunate terminology sanctioned by universal established use. It should not be mistaken by the same word "covariant" employed in the wider sense of "equally variant". A covariant vector, or covector, transforms according to

$$V'_{\alpha} = \Lambda_{\alpha}{}^{\beta}V_{\beta} \ . \tag{1.81}$$

The matrix with entries $\Lambda_{\alpha}{}^{\beta}$ is the inverse of the previous matrix Λ. This notation will be better justified later. For the time being let us notice that, as indices are lowered and raised by η and η^{-1}, we have

$$V'_{\alpha}V'^{\alpha} = \eta_{\alpha\beta}V'^{\beta}V'^{\alpha} = \eta_{\alpha\beta}\Lambda^{\beta}{}_{\gamma}\Lambda^{\alpha}{}_{\delta}V^{\gamma}V^{\delta} = \Lambda_{\alpha\gamma}\Lambda^{\alpha}{}_{\delta}V^{\gamma}V^{\delta} = \Lambda_{\alpha}{}^{\gamma}\Lambda^{\alpha}{}_{\delta}V_{\gamma}V^{\delta} \ ,$$

so that we must have $\Lambda_{\alpha}{}^{\gamma}\Lambda^{\alpha}{}_{\delta} = \delta^{\gamma}_{\delta}$ in order to preserve the value of the norm. Therefore, $(\Lambda^{-1})^{\gamma}{}_{\alpha} = \Lambda_{\alpha}{}^{\gamma}$ and (1.81) is actually

$$V'_{\alpha} = \Lambda_{\alpha}{}^{\beta}V_{\beta} = V_{\beta}\ (\Lambda^{-1})^{\beta}{}_{\alpha} \ , \tag{1.82}$$

with the last matrix acting from the right. A good picture of what happens comes as follows: conceive a (contravariant) vector u as column $\begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix}$ with four entries and a covector v as a row $\begin{pmatrix} v^0, v^1, v^2, v^3 \end{pmatrix}$. Matrices will act from the left on columns, and act from the right on rows. The scalar product $v \cdot u$ will be a row-column product,

$$\begin{pmatrix} v^0, v^1, v^2, v^3 \end{pmatrix} \begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix} .$$

The preservation of the scalar product is then trivial, as

$$v' \cdot u' = \begin{pmatrix} v^0, v^1, v^2, v^3 \end{pmatrix} \Lambda^{-1}\Lambda \begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix} = \begin{pmatrix} v^0, v^1, v^2, v^3 \end{pmatrix} \begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix} = v \cdot u. \tag{1.83}$$

Summing up, covariant vectors transform by the inverse Lorentz matrix. As in the contravariant case, a covariant tensor transforms like the product of covariant vectors, etc.

The transformation of a mixed tensor will have one Λ-factor for each higher index, one Λ^{-1}-factor for each lower index. For example,

$$T'^{\alpha}{}_{\beta}{}^{\gamma} = \Lambda^{\alpha}{}_{\delta}\ \Lambda_{\beta}{}^{\epsilon}\ \Lambda^{\gamma}{}_{\phi}\ T^{\delta}{}_{\epsilon}{}^{\phi} \ . \tag{1.84}$$

§ 1.35 We shall later consider vector *fields*, which are point-dependent (that is, event-dependent) vectors $V = V(x)$. They will describe the states of systems with infinite degrees of freedom, one for each point, or event. In that case, a Lorentz transformation will affect both the vector itself and its argument:

$$V'^{\alpha}(x') = \Lambda^{\alpha}{}_{\beta}\, V^{\beta}(x) \ , \tag{1.85}$$

where $x'^{\alpha} = \Lambda^{\alpha}{}_{\beta} x^{\beta}$. Tensor fields will follow suit.

Comment 1.1 The Galilei group element (1.16) is a limit when $v/c \to 0$ of the generic group element (1.77) of the Poincaré group. We have said "a" limit, not "the" limit, because some redefinitions of the transformation parameters are necessary for the limit to make sense. The procedure is called a Inönü–Wigner contraction.[‡]

Comment 1.2 Unlike the case of Special Relativity, there is no metric on the complete 4-dimensional spacetime which is invariant under Galilei transformations. For this reason people think twice before talking about "spacetime" in the classical case. There is space, and there is time. Only within Special Relativity, in the words of one of the inventors of spacetime,"... space by itself, and time by itself, are condemned to fade away into mere shadows, and only a kind of union of the two preserves an independent reality".[§]

[‡] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications*, J.Wiley, New York, 1974.

[§] H. Minkowski, "Space and time", in *The Principle of Relativity*, New York, Dover, 1923. From the 80th Assembly of German Natural Scientists and Physicians, Cologne, 1908.

1.6 Particle Dynamics

§ 1.36 Let us go back to Eq.(1.39). Use of Eq.(1.65) shows that time, as indicated by a clock, is

$$\frac{1}{c}\int_\alpha ds\ ,$$

the integral being taken along the clock's worldline α. From the expression $ds = \sqrt{c^2dt^2 - d\vec{x}^2}$ we see that each infinitesimal contribution ds is maximal when $d\vec{x}^2 = 0$, that is, along the pure-time straight line, or the cone axis. We have said in § 1.19 that there are always zero-length paths between any two events. These paths are formed with contributions $ds = 0$ and stand on the light-cones. The farther a path α stays from the light cone, the larger will be the integral above. The largest value of $\int_\alpha$ will be attained for $\alpha =$ the pure-time straight line going through each point.

Hamilton's minimal-action principle is a mechanical version of Fermat's optical minimal-time principle. Both are unified in the relativistic context, but $\int_\alpha$ is a *maximal* time or length. Let us only retain that the integral $\int_\alpha$ is an extremal for a particle moving along a straight line in 4-dimensional spacetime. "Moving along a straight line" is just the kind of motion a free particle should have in an inertial frame. If we want to obtain its equation of motion from an action principle, the action should be proportional to $\int_\alpha ds$. The good choice for the action related to a motion from point P to point Q is

$$S = -\,mc\int_P^Q ds\ . \tag{1.86}$$

The factor mc is introduced for later convenience. The sign, to make of the action principle a minimal (and not a maximal) principle.

Let us see how to use such a principle to get the equation of motion of a free particle. Take two points P and Q in Minkowski spacetime, and consider the integral

$$\int_P^Q ds = \int_P^Q \sqrt{\eta_{\alpha\beta}dx^\alpha dx^\beta}\ .$$

Its value depends on the path chosen. It is actually a functional on the space of paths between P and Q,

$$\mathcal{F}[\alpha_{PQ}] = \int_{\alpha_{PQ}} ds. \tag{1.87}$$

An extremal of this functional would be a curve α such that $\delta S[\alpha] = \int \delta ds = 0$. Now,

$$\delta ds^2 = 2\ ds\ \delta ds = 2\ \eta_{\alpha\beta}\ dx^\alpha \delta dx^\beta,$$

so that

$$\delta ds = \eta_{\alpha\beta}\ \frac{dx^\alpha}{ds}\ \delta dx^\beta\ .$$

Thus, commuting the differential d and the variation δ and integrating by parts,

$$\delta S[\alpha] = \int_P^Q \eta_{\alpha\beta}\ \frac{dx^\alpha}{ds}\ \frac{d\delta x^\beta}{ds}\ ds = -\ \int_P^Q \eta_{\alpha\beta}\ \frac{d}{ds}\ \frac{dx^\alpha}{ds}\ \delta x^\beta\ ds$$

$$= -\ \int_P^Q \eta_{\alpha\beta}\ \frac{d}{ds}\ u^\alpha\ \delta x^\beta\ ds.$$

The variations δx^β are arbitrary. If we want to have $\delta S[\alpha] = 0$ for arbitrary δx^β, the integrand must vanish. Thus, an extremal of the action (1.86) will satisfy

$$mc\ \frac{d}{ds}\ u^\alpha = mc\ \frac{d^2x^\alpha}{ds^2} = 0. \tag{1.88}$$

This is the equation of a straight line, and — as it has the aspect of Newton's second law — shows the coherence of the velocity definition (1.64) with the action (1.86). The solution of this differential equation is fixed once initial conditions are given. We learn in this way that a vanishing acceleration is related to an extremal of $S[\alpha_{PQ}]$. In the presence of some external force F^α, this should lead to a force law like

$$mc\ \ \frac{du^\alpha}{ds} = F^\alpha. \tag{1.89}$$

§ 1.37 To establish comparison with Classical Mechanics, let us write

$$S = \int_P^Q L dt\ , \tag{1.90}$$

with L the Lagrangian. By (1.65), we have

$$L = -\ mc^2\sqrt{1 - \frac{v^2}{c^2}}\ . \tag{1.91}$$

Notice that, for small values of $\frac{v^2}{c^2}$,

$$L \approx -\ mc^2(1 - \frac{v^2}{2c^2}) \approx -\ mc^2 + \frac{mv^2}{2}\ , \tag{1.92}$$

the classical Lagrangian with the constant mc^2 extracted.

§ 1.38 The momentum is defined, as in Classical Mechanics, by $p_k = \frac{\delta L}{\delta v^k}$, which gives

$$\mathbf{p} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\ m\mathbf{v} = \gamma\ m\mathbf{v} \tag{1.93}$$

which, of course, reduces to the classical $\mathbf{p} = m\mathbf{v}$ for small velocities.

The energy, again as in Classical Mechanics, is defined as $\mathcal{E} = \mathbf{p} \cdot \mathbf{v} - L$, which gives the celebrated expression

$$\mathcal{E} = \gamma\, mc^2 = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \ . \tag{1.94}$$

This shows that, unlike what happens in the classical case, a particle at rest has the energy

$$\mathcal{E} = mc^2 \tag{1.95}$$

and justifies its subtraction to arrive at the classical Lagrangian in (1.92). Relativistic energy includes the mass contribution. Notice that both the energy and the momentum would become infinite for a massive particle of velocity $v = c$. That velocity is consequently unattainable for a massive particle.

Equations (1.93) and (1.94) lead to two other important formulae. The first is

$$\mathbf{p} = \frac{\mathcal{E}}{c^2}\, \mathbf{v}. \tag{1.96}$$

The infinities mentioned above cancel out in this formula, which holds also for massless particles traveling with velocity c. In that case it gives

$$|\mathbf{p}| = \frac{\mathcal{E}}{c} \ . \tag{1.97}$$

The second formula comes from taking the squares of both equations. It is

$$\mathcal{E}^2 = \mathbf{p}^2 c^2 + m^2 c^4 \tag{1.98}$$

and leads to the Hamiltonian

$$H = c\sqrt{\mathbf{p}^2 + m^2 c^2} \ . \tag{1.99}$$

Now, we can form the four-momentum

$$p = mcu = (\mathcal{E}/c, \mathbf{p}) = (\mathcal{E}/c, \gamma m\mathbf{v}) = \gamma(mc, m\mathbf{v}) \ , \tag{1.100}$$

whose square is

$$p^2 = m^2 c^2 \ . \tag{1.101}$$

§ 1.39 Force, if defined as the derivative of p with respect to proper time, will give

$$F = \frac{d}{ds}\, p = mc\, \frac{du}{ds} \ ,$$

just Eq.(1.89). Because u is dimensionless the quantity F, defined in this way, has not the mechanical dimension of a force (Fc would have).

Exercise 1.7 The Kronecker completely antisymmetric symbol ϵ_{ijk} is defined by

$$\epsilon_{ijk} = \begin{cases} 1 \text{ if } ijk \text{ is an even permutation of } 123 \\ -1 \text{ if } ijk \text{ is an odd permutation of } 123 \\ 0 \text{ otherwise} \end{cases} \tag{1.102}$$

The starting value is $\epsilon_{123} = 1$. A useful determinant form is

$$\epsilon_{ijk} = \begin{vmatrix} \delta^1_i & \delta^1_j & \delta^1_k \\ \delta^2_i & \delta^2_j & \delta^2_k \\ \delta^3_i & \delta^3_j & \delta^3_k \end{vmatrix} . \tag{1.103}$$

Indices are here raised and lowered with the euclidean metric.

A Verify the following statements:

1. the component k of the vector product of v and u is $(\vec{v} \times \vec{u})^k = \epsilon^k{}_{ij}\ v^i u^j$
2. in euclidean 3-dimensional space, an antisymmetric matrix with entries M_{ij} is equivalent to a vector $v^k = \frac{1}{2}\ \epsilon^{kij}\ M_{ij}$
3. the inverse formula is $M_{ij} = \frac{1}{2}\ \epsilon_{ijk}\ v^k$.

B Calculate

1. $\epsilon_{ijk}\epsilon^{imn}$
2. $\epsilon_{ijk}\epsilon^{ijn}$
3. $\epsilon_{ijk}\epsilon^{ijk}$.

■

Exercise 1.8 (Facultative: supposes some knowledge of electromagnetism and vector calculus) Tensor (6.65) is *Maxwell's tensor*, or electromagnetic *field strength*. If we compare with the expressions of the electric field $\vec{E}$ and the magnetic field $\vec{B}$ in vacuum, we see that

$$F_{0i} = \partial_0 A_i - \ \partial_i A_0 = \partial_{ct} A_i - \ \partial_i \phi = E_i \ ;$$

$$F_{ij} = \partial_i A_j - \ \partial_j A_i = \epsilon_{ijk} (rot \vec{A})_k = \epsilon_{ijk} B_k \ ,$$

where ϵ_{ijk} is the completely antisymmetric Kronecker symbol. Consequently,

$$F^i{}_\beta\ u^\beta = F^i{}_0\ u^0 + F^i{}_j\ u^j = E^i\ \gamma + \epsilon^i{}_{jk} B_k\ \gamma \frac{v^j}{c} = \gamma \left(E^i\ + \frac{1}{c} \epsilon^i{}_{jk} v^j B_k \right)$$

$$= \gamma \left(\vec{E} + \frac{1}{c}\ \vec{v} \times \vec{B} \right)^i . \tag{1.104}$$

Thus, for the space components, Eq.(6.67) is

$$\frac{d}{ds}\ \vec{p} = \frac{e}{c}\ \gamma \left(\vec{E} + \frac{1}{c}\ \vec{v} \times \vec{B} \right) .$$

Using Eq.(1.65),

$$\vec{F} = \frac{d}{dt}\ \vec{p} = e\ \left(\vec{E} + \frac{1}{c}\ \vec{v} \times \vec{B} \right) , \tag{1.105}$$

which is the usual form of the Lorentz force felt by a particle of charge e in a electromagnetic field. The time component gives the time variation of the energy:

$$mc\ \frac{d}{ds}\ u^0 = \frac{e}{c}\ F^0{}_i\ u^i$$

$$\therefore\ \frac{d}{ds}\ (\gamma mc^2) = \gamma\ \frac{d}{dt}\ \mathcal{E} = e\ \gamma\ \vec{E}\cdot\vec{v}\ \therefore\ \frac{d}{dt}\ \mathcal{E} = e\ \vec{E}\cdot\vec{v}. \tag{1.106}$$

■

Exercise 1.9 Let us modify the action of a free particle in §1.36 to

$$S = -\int mc\ ds,$$

admitting now the possibility of a mass which changes along the path (think of a rocket spending its fuel along its trajectory).

1. Taking into account the possible variation of m, find, by the same procedure previously used, the new equation of force:

$$\frac{d}{ds}\ [mc\ u_\alpha] = \frac{\partial}{\partial x^\alpha}\ mc, \tag{1.107}$$

   with a force given by the mass gradient turning up;

2. show that this is equivalent to

$$mc\ \frac{d}{ds}\ u_\alpha = \left(\delta^\beta_\alpha - u^\beta u_\alpha\right)\frac{\partial}{\partial x^\beta}\ (mc). \tag{1.108}$$

   Show furthermore, using (1.69), that

3. the force is orthogonal to the path, that is, to its velocity at each point;

4. the matrix P of entries $P^\beta{}_\alpha = \delta^\beta_\alpha - u^\beta u_\alpha$ is a projector, that is, $P^2 = P$.

   At each point along the path, P projects on a 3-dimensional plane orthogonal to the 4-velocity.

■

§ 1.40 Summing up Before going further, let us make a short *resumé* on the general notions used up to now, in a language loose enough to make them valid both in the relativistic and the non-relativistic cases (so as to generalize the classical notions of § 1.6).

Reference frame A reference frame is a "coordinate system for space positions, to which a clock is bound". The coordinate system has two pieces:

(i) a fixed set of vectors, like the $\mathbf{i}, \mathbf{j}, \mathbf{k}$ usually employed in our ambient 3-dimensional space;

(ii) a set of coordinate functions, as the usual cartesian, spherical or cylindrical coordinates. The clock provides a coordinate in the 1-dimensional time axis.

Inertial frame a reference frame such that free (that is, in the absence to any forces) motion takes place with constant velocity is an *inertial frame*;

(a) in Classical Physics, Newton's force law in an inertial frame is

$$m\,\frac{dv^k}{dt} = F^k;$$

(b) in Special Relativity, the force law in an inertial frame is

$$mc\,\frac{du^\alpha}{ds} = F^\alpha.$$

Incidentally, we are stuck to cartesian coordinates to discuss forces: the second time-derivative of a coordinate is an acceleration only if that coordinate is cartesian.

Transitivity of inertia a reference frame moving with constant velocity with respect to an inertial frame is also an inertial frame. Measurements made at two distinct inertial frames give the same results. It is consequently impossible to distinguish inertial frames by making measurements.

Causality in non-relativistic classical physics the interactions are given by the potential energy, which usually depends only on the space coordinates; forces on a given particle, caused by all the others, depend only on their position at a given instant; a change in position changes the force instantaneously; this instantaneous propagation effect — or action-at-a-distance — is a typically classical, non-relativistic feature; it violates special-relativistic causality, which says that no effect can propagate faster than the velocity of light.

Relativity the laws of Physics can be written in a form which is invariant under change of frame. In particular, all the laws of nature are the same in all inertial frames; or, alternatively, the equations describing them are invariant under the transformations (of space coordinates and time) taking one inertial frame into the other; or still, the equations describing the laws of Nature in terms of space coordinates and time keep their forms in different inertial frames; this "principle of relativity" is an experimental fact; there are three known "Relativities":

(1) Galilean Relativity, which holds in non-relativistic classical physics; the transformations between inertial frames belong to the Galilei group;

(2) Special Relativity, which is our subject; transformations between inertial frames belong to the Poincaré group;

(3) General Relativity, involved with non-inertial frames and the so-called inertial forces, including gravitation. If you look at things from an accelerated frame, those things will seem to be subject to a force, which has however a very special characteristic: it is the same for all things. Of course, that force is only an effect of your own acceleration, but it has in common with gravity that universal character. Locally — that is, in a small enough domain of space a gravitational force cannot be distinguished from that kind of "inertial" force.

§ 1.41 There have been tentatives to preserve action-at-a-distance in a relativistic context, but a simpler way to consider interactions while respecting Special Relativity is of common use in field theory: interactions are mediated by a field, which has a well-defined behavior under transformations; and disturbances propagate with finite velocities, with the velocity of light as an upper bound.

Chapter 2

Transformations

2.1 Transformation Groups

§ 2.1 We can use changes in spacetime to illustrate the main aspects of transformation groups. Transformations are then seen as the effect of acting with matrices on spacetime column–vectors. The null transformation on spacetime, for example, will be given by the identity matrix

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.1}$$

The parity transformation is the inversion of all the space components of every vector,

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} . \tag{2.2}$$

Its effect on the position vector will be

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} x^0 \\ -x^1 \\ -x^2 \\ -x^3 \end{pmatrix}$$

One can also conceive the specular inversion of only one of the coordinates, as the x-inversion or the x–and–y-inversion:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad ; \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.3}$$

The time reversal transformation will be given by

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.4}$$

Composition of transformations is then represented by the matrix product. The so–called PT transformation, which inverts all the space and time components, is given by the product

$$PT = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} . \tag{2.5}$$

§ 2.2 Transformation groups As most of transformations appearing in Physics are members of groups, let us formalize it a bit, adapting the algebraic definition of a group. A set of transformations is organized into a group G, or constitute a group if:

(i) given two of them, say T_1 and T_2, their composition $T_1 \cdot T_2$ is also a transformation which is a member of the set;

(ii) the identity transformation (I such that $I \cdot T_k = T_k \cdot I = T_k$, for all $T_k \in G$) belongs to the set;

(iii) to each transformation T corresponds an inverse T^{-1}, which is such that $T^{-1}T = TT^{-1} = I$ and is also a member;

(iv) associativity holds: $(T_1 \cdot T_2) \cdot T_3 = T_1 \cdot (T_2 \cdot T_3)$ for all triples $\{T_1, T_2, T_3\}$ of members.

Notice that, in general, $T_i \cdot T_j \neq T_j \cdot T_i$. When $T_i \cdot T_j = T_j \cdot T_i$, we say that T_i and T_j commute. If $T_i \cdot T_j = T_j \cdot T_i$ is true for all pairs of members of G, G is said to be a commutative, or abelian group. A subgroup of G is a subset H of elements of G satisfying the same rules.

§ 2.3 The transformations P and T above do form a group, with the composition represented by the matrix product. P and T, if applied twice, give the identity, which shows that they are their own inverses. They are actually quite independent and in reality constitute two independent (and rather trivial) groups. They have, however, something else in common: they cannot be obtained by a step–by–step addition of infinitesimal transformations. They are "discrete" transformations, in contraposition to the "continuous" transformations, which are those that can be obtained by composing infinitesimal transformations step–by–step. Notice that the determinants of the matrices representing P and T are -1. The determinant of the identity is $+1$. Adding an infinitesimal contribution to the identity will give a matrix with determinant near to $+1$. Groups of transformations which can be obtained in this way from the identity, by adding infinitesimal contributions, are said to be "continuous" and "connected to the identity". P and T are not connected to the identity.

§ 2.4 The continuous transformations appearing in Physics are *a priori* supposed to belong to some Lie group, that is, a continuous smooth group.

Lie groups are typically represented by matrices. When a member of a continuous group G is close to the identity, it will be given by a matrix like $I + \delta W$, where δW is a small matrix, that is, a matrix with small entries. Actually, a very general characteristic of a matrix M belonging to a Lie group is the following: M can be written in the form of an exponential,

$$M = e^W = I + W + \frac{1}{2!}\, W^2 + \frac{1}{3!}\, W^3 + \cdots \ .$$

Consider, for example, the effect of acting on the triple $\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$ with the matrix

$$I + \delta W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -\delta\phi & 0 \\ \delta\phi & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -\delta\phi & 0 \\ \delta\phi & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ .$$

It gives an infinitesimal rotation in the plane (x^1, x^2):

$$\begin{pmatrix} x^{1'} \\ x^{2'} \\ x^{3'} \end{pmatrix} = \begin{pmatrix} 1 & -\delta\phi & 0 \\ \delta\phi & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} x^1 - \delta\phi\, x^2 \\ x^2 + \delta\phi\, x^1 \\ x^3 \end{pmatrix} \ .$$

We suppose $\delta\phi$ to be very small, so that this is a transformation close to the unity.

Exercise 2.1 Take the matrix

$$W = \begin{pmatrix} 0 & -\phi & 0 \\ \phi & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ .$$

Exponentiate it, and find the finite version of a rotation:

$$\begin{pmatrix} x^{1'} \\ x^{2'} \\ x^{3'} \end{pmatrix} = e^W \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$$
$$= \begin{pmatrix} x^1\cos\phi - x^2\sin\phi \\ x^2\cos\phi + x^1\sin\phi \\ x^3 \end{pmatrix} .$$

■

§ 2.5 But there is more. The set of $N \times N$ matrices, for any integer N, forms also a vector space. In a vector space of matrices, we can always choose a linear base, a set $\{J_a\}$ of matrices linearly independent of each other. Any other matrix can be written as a linear combination of the J_a's: $W = w_a J_a$. We shall *suppose* that all the elements of a matrix Lie group G can be written in the form $M = \exp[w_a J_a]$, with a fixed and limited number of J_a. The matrices J_a are called the generators of G. They constitute an algebra with the operation defined by the commutator, which is the Lie algebra of G.

2.2 Orthogonal Transformations

§ 2.6 We have said in § 1.22 that a group of continuous transformations preserving a symmetric bilinear form η on a vector space is an orthogonal group or, if the form is not positive–definite, a pseudo–orthogonal group.

Rotations preserve the distance $d(\mathbf{x}, \mathbf{y})$ of $\mathbf{E}^3$ because $R(\omega)$ is an orthogonal matrix. Let us see how this happens. Given a transformation represented by a matrix M,

$$x^{i'} = \sum_j M^{i'}{}_j \, x^j \ ,$$

the condition for preserving the distance will be

$$\sum_{i'} x^{i'} y^{i'} = \sum_{i'} M^{i'}{}_j \, M^{i'}{}_k \, x^j \, y^k = \sum_i x^i y^i \ ;$$

that is, with M^T the transpose of M,

$$\sum_{i'} M^{i'}{}_j \, M^{i'}{}_k = \sum_{i'} \left(M^T\right)^j{}_{i'} \, M^{i'}{}_k = \left(M^T M\right)^j{}_k = \delta^j{}_k \ ,$$

which means that M is an orthogonal matrix: $M^T M = I$. This is indeed an orthogonality condition, saying that the columns of M are orthonormal to each other.

Given the transformation $x^{\alpha'} = \Lambda^{\alpha'}{}_\alpha \ x^\alpha$, to say that "$\eta$ is preserved" is to say that the distance calculated in the primed frame and the distance calculated in the unprimed frame are the same. Take the squared distance in the primed frame, $\eta_{\alpha'\beta'} \ x^{\alpha'} x^{\beta'}$, and replace $x^{\alpha'}$ and $x^{\beta'}$ by their transformation expressions. We must have

$$\eta_{\alpha'\beta'} \ x^{\alpha'} x^{\beta'} = \eta_{\alpha'\beta'} \ \Lambda^{\alpha'}{}_\alpha \, \Lambda^{\beta'}{}_\beta \ x^\alpha x^\beta = \eta_{\alpha\beta} \ x^\alpha x^\beta \ , \quad \forall x \ . \tag{2.6}$$

This is the group–defining property, a condition on the $\Lambda^{\alpha'}{}_\alpha$'s. We see that it is necessary that

$$\eta_{\alpha\beta} = \eta_{\alpha'\beta'} \, \Lambda^{\alpha'}{}_\alpha \, \Lambda^{\beta'}{}_\beta = \Lambda^{\alpha'}{}_\alpha \, \eta_{\alpha'\beta'} \, \Lambda^{\beta'}{}_\beta .$$

The matrix form of this condition is, for each group element Λ,

$$\Lambda^T \ \eta \ \Lambda = \eta \ , \tag{2.7}$$

where Λ^T is the transpose of Λ. It follows clearly that $\det \Lambda = \pm 1$.

When η is the Lorentz metric, the above condition defines the Λ belonging to the Lorentz group.

Comment 2.1 Consider the transformation

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} x^0 + x^3 \\ x^1 \\ x^2 \\ x^0 - x^3 \end{pmatrix} .$$

It is a transformation of coordinates of a rather special type. As the determinant is $= -1$, it cannot be obtained as a continuous deformation of the identity. Of course, it does not represent the passage from one inertial frame to another. The equation $x_- = x^0 - x^3 = 0$ says that the point (x^0, x^1, x^2, x^3) is on the light-cone. The remaining variables $x_+ = x^0 + x^3, x^1$ and x^2 represent points on the cone. The coordinates x'^α are called "light-cone coordinates".

Exercise 2.2 Show that the matrix inverse to Λ can be written as

$$(\Lambda^{-1})^\alpha{}_{\beta'} = \Lambda_{\beta'}{}^\alpha \ .$$

■

There is a corresponding condition on the members of the group Lie algebra. For each member A of that algebra, there will exist a group member Λ such that $\Lambda = e^A$. Taking $\Lambda = I + A + \frac{1}{2}A^2 + \ldots$ and $\Lambda^T = I + A^T + \frac{1}{2}(A^T)^2 + \ldots$ in the above condition and comparing order by order, we find that A must satisfy

$$A^T = -\ \eta\ A\ \eta^{-1} = -\ \eta^{-1}\ A\ \eta \tag{2.8}$$

and will consequently have vanishing trace: tr A = tr A^T = - tr $(\eta^{-1}\ A\ \eta)$ = - tr $(\eta\eta^{-1}\ A)$ = - tr A $\therefore$ tr $A = 0$.

Comment 2.2 Actually, it follows from the formal identity $\det M = \exp[\mathrm{tr} \ln M]$ that $\det M = 1 \Rightarrow \mathrm{tr} \ln M = 0$.

We shall need some notions of algebra, Lie groups and Lie algebras. They are introduced through examples in what follows, in a rather circular and repetitive way, as if we were learning a mere language.

§ 2.7 The invertible $N \times N$ real matrices constitute the real linear group $GL(N, R)$. Members of this group can be obtained as the exponential of some $K \in gl(N, R)$, the set of all real $N \times N$ matrices. $GL(N, R)$ is thus a Lie group, of which $gl(N, R)$ is the Lie algebra. The generators of the Lie algebra are also called, by extension, generators of the Lie group.

Consider the set of $N \times N$ matrices. This set is, among other things, a vector space. The simplest matrices will be those $\Delta_\alpha{}^\beta$ whose entries are all zero except for that of the α-th row and β-th column, which is 1:

$$(\Delta_\alpha{}^\beta)^\delta{}_\gamma = \delta^\delta{}_\alpha\, \delta^\beta{}_\gamma \ . \tag{2.9}$$

An arbitrary $N \times N$ matrix K can be written $K = K^\alpha{}_\beta\, \Delta_\alpha{}^\beta$. Thus, for example, $P = \Delta_0{}^0 - \Delta_1{}^1 - \Delta_2{}^2 - \Delta_3{}^3$, and $T = -\Delta_0{}^0 + \Delta_1{}^1 + \Delta_2{}^2 + \Delta_3{}^3$.

The $\Delta_\alpha{}^\beta$'s have one great quality: they are linearly independent (none can be written as linear combinations of the other). Thus, the set $\{\Delta_\alpha{}^\beta\}$ constitutes a basis

(the "canonical basis") for the vector space of the $N \times N$ matrices. An advantage of this basis is that the components of a member K as a vector written in basis $\{\Delta_\alpha{}^\beta\}$ are the very matrix elements: $(K)^\alpha{}_\beta = K^\alpha{}_\beta$. Consider now the product of matrices: it takes each pair (A, B) of matrices into another matrix AB. In our notation, a matrix product is performed coupling lower–right indices to higher–left indices, as in

$$\begin{aligned}\left(\Delta_\alpha{}^\beta\, \Delta_\phi{}^\xi\right)^\delta{}_\epsilon &= \left(\Delta_\alpha{}^\beta\right)^\delta{}_\gamma\, \left(\Delta_\phi{}^\xi\right)^\gamma{}_\epsilon \\ &= \delta^\beta{}_\phi\, \left(\Delta_\alpha{}^\xi\right)^\delta{}_\epsilon \end{aligned} \tag{2.10}$$

where, in $(\Delta_\alpha{}^\beta)^\delta{}_\gamma$, γ is the column index.

Exercise 2.3 Use (2.9),

$$(\Delta_\alpha{}^\beta)^\delta{}_\gamma = \delta^\delta{}_\alpha\, \delta^\beta{}_\gamma,$$

to show that (2.10) is true. ■

§ 2.8 Algebra This type of operation, taking two members of a set into a third member of the same set, is called a binary internal operation. A binary internal operation defined on a vector space V makes of V an *algebra*. The matrix product defines an algebra on the vector space of the $N \times N$ matrices, called the product algebra. Take now the operation defined by the commutator: it is another binary internal operation, taking each pair (A, B) into the matrix $[A, B] = AB - BA$. Thus, the commutator turns the vector space of the $N \times N$ matrices into another algebra. But, unlike the simple product, the commutator defines a very special kind on algebra, a Lie algebra.

§ 2.9 Lie algebra A Lie algebra comes up when, in a vector space, there is an operation which is antisymmetric and satisfies the Jacobi identity. This is what happens here, because $[A, B] = -[B, A]$ and

$$[[A, B], C] + [[C, A], B] + [[B, C], A] = 0.$$

This Lie algebra, of the $N \times N$ real matrices with the operation defined by the commutator, is called the real N-linear algebra, denoted $gl(N, R)$. A theorem (Ado's) states that any Lie algebra can be seen as a subalgebra of $gl(N, R)$, for some N.

The members of a vector base for the underlying vector space of a Lie algebra are the generators of the Lie algebra. $\{\Delta_\alpha{}^\beta\}$ is called the *canonical base* for $gl(N, R)$. A Lie algebra is summarized by its commutation table. For $gl(N, R)$, the commutation relations are

$$\left[\Delta_\alpha{}^\beta, \Delta_\phi{}^\zeta\right] = f_{\binom{\alpha}{\beta}\binom{\phi}{\zeta}}{}^{\binom{\delta}{\gamma}}\, \Delta_\gamma{}^\delta\ . \tag{2.11}$$

The constants appearing in the right-hand side are the *structure coefficients*, whose values in the present case are

$$f_{(\alpha}{}_{\beta)(}{}^{\phi}{}_{\zeta)}{}^{(\delta}{}_{\gamma)} = \delta_{\phi}{}^{\beta}\,\delta_{\alpha}{}^{\gamma}\,\delta_{\delta}{}^{\zeta} - \delta_{\alpha}{}^{\zeta}\,\delta_{\phi}{}^{\gamma}\,\delta_{\delta}{}^{\beta}\ . \tag{2.12}$$

§ 2.10 A group is a matrix group when it is a subgroup of $GL(N,R)$ for some value of N. A Lie group G can be isomorphic to matrix groups with many different values of N. Each one of these "copies" is a linear *representation* of G. A finite linear transformation with parameters $w^{\alpha}{}_{\beta}$ is given by the matrix $M = \exp[w] = \exp[w^{\alpha}{}_{\beta}\Delta_{\alpha}{}^{\beta}]$. Then,

$$\begin{aligned} w^2 &= w^{\alpha}{}_{\beta}\,w^{\phi}{}_{\xi}\,\Delta_{\alpha}{}^{\beta}\,\Delta_{\phi}{}^{\xi} = w^{\alpha}{}_{\beta}\,w^{\beta}{}_{\xi}\,\Delta_{\alpha}{}^{\xi} = (w^2)^{\alpha}{}_{\xi}\,\Delta_{\alpha}{}^{\xi} \\ w^3 &= w^{\alpha}{}_{\beta}\,w^{\phi}{}_{\xi}\,w^{\gamma}{}_{\delta}\,\Delta_{\alpha}{}^{\beta}\,\Delta_{\phi}{}^{\xi}\,\Delta_{\gamma}{}^{\delta} = (w^3)^{\alpha}{}_{\xi}\,\Delta_{\alpha}{}^{\xi}, \text{etc}, \end{aligned}$$

and M will have entries

$$M^{r'}{}_{s} = \left(\mathrm{e}^{w^{\alpha}{}_{\beta}\Delta_{\alpha}{}^{\beta}}\right)^{r'}{}_{s} = \left(\sum_{n=0}^{\infty}\frac{w^n}{n!}\right)^{r'}{}_{s} = \sum_{n=0}^{\infty}\frac{1}{n!}\,(w^n)^{r'}{}_{s}\ .$$

To first order in the parameters,

$$M^{r'}{}_{s} \approx \delta^{r'}{}_{s} + w^{r'}{}_{s}\ .$$

If a metric η is defined on an N-dimensional space, the Lie algebras $so(\eta)$ of the orthogonal or pseudo–orthogonal groups will be subalgebras of $gl(N,R)$. Given an algebra $so(\eta)$, both basis and entry indices can be lowered and raised with the help of η. We define new matrices $\Delta_{\alpha\beta}$ by lowering labels with η : $(\Delta_{\alpha\beta})^{\delta}{}_{\gamma} = \delta^{\delta}{}_{\alpha}\eta_{\beta\gamma}$. Their commutation relations become

$$[\Delta_{\alpha\beta}, \Delta_{\gamma\delta}] = \eta_{\beta\gamma}\,\Delta_{\alpha\delta} - \eta_{\alpha\delta}\,\Delta_{\gamma\beta}\ . \tag{2.13}$$

Exercise 2.4 Use Exercise 2.3 to prove (2.13). ■

The generators of $so(\eta)$ will then be $J_{\alpha\beta} = \Delta_{\alpha\beta}$ - $\Delta_{\beta\alpha}$, with commutation relations

$$[J_{\alpha\beta}, J_{\gamma\delta}] = \eta_{\alpha\delta}J_{\beta\gamma} + \eta_{\beta\gamma}J_{\alpha\delta} - \eta_{\beta\delta}J_{\alpha\gamma} - \eta_{\alpha\gamma}J_{\beta\delta}\ . \tag{2.14}$$

These are the general commutation relations for the generators of orthogonal and pseudo–orthogonal groups. We shall meet many cases in what follows. Given η, the algebra is fixed up to conventions. The usual group of rotations in the 3-dimensional Euclidean space is the special orthogonal group, denoted by $SO(3)$. Being "special" means connected to the identity, that is, represented by 3×3 matrices of determinant $= +1$.

Exercise 2.5 When η is the Lorentz metric, (2.14) is the commutation table for the generators of the Lorentz group. Use Exercise 2.4 to prove (2.14). ■

§ 2.11 The group $O(N)$ is formed by the orthogonal $N \times N$ real matrices. The group $U(N)$ is the group of unitary $N \times N$ complex matrices. $SO(N)$ is formed by all the matrices of $O(N)$ which have determinant $= +1$. $SU(N)$ is formed by all the matrices of $U(N)$ which have determinant $= +1$. In particular, the group $O(3)$ is formed by the orthogonal 3×3 real matrices. The group $U(2)$ is the group of unitary 2×2 complex matrices. $SO(3)$ is formed by all the matrices of $O(3)$ which have determinant $= +1$. $SU(2)$ is formed by all the matrices of $U(2)$ which have determinant $= +1$.

Comment 2.3 If a group $SO(p,q)$ preserves $\eta(p,q)$, so does the corresponding affine group, which includes the translations. We should be clear on this point. When we write x^j, for example, we mean $x^j - 0$, that is, the coordinate is counted from the origin. The translation indicated by a^j is a global translation of the whole space — of the point and of the origin together. $x^j - 0$ goes into $(x^j + a^j)$ - $(0 + a^j)$. Vectors are not changed. All points are changed in the same way, so that differences remain invariant. This would perhaps become more evident if we noticed that the (squared) distances between two point $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{E}^3$ are actually always $\eta(\mathbf{x}' - \mathbf{x}, \mathbf{x}' - \mathbf{x}) = (x' - x)^2 + (y' - y)^2 + (z' - z)^2$. Translations lead simultaneously $\mathbf{x} \to \mathbf{x} + \mathbf{a}$ and $\mathbf{x}' \to \mathbf{x}' + \mathbf{a}$. The affine group related to $SO(r,s)$ is frequently called "inhomogeneous $SO(p,q)$" and denoted $ISO(r,s)$. In cases given below, the Euclidean group on $\mathbb{E}^3$ can be indicated also by $ISO(3)$, and the Poincaré group, also called inhomogeneous Lorentz group, by $ISO(3,1)$.

The simplest and best–known non–trivial Lie group is that formed by rotations in 3-dimensional Euclidean space. We shall use it as a spearhead to introduce and illustrate the main ideas on more general Lie groups.

2.3 The Group of Rotations

§ 2.12 The rotation group in Euclidean 3-dimensional space can actually be taken either as the special orthogonal group $SO(3)$ or as the special unitary group $SU(2)$. Their Lie algebras are the same. The distinction lies in their manifolds, and has important consequences. For example, $SU(2)$ has twice more representations than $SO(3)$. In particular, the Pauli representation given below (Eq. 2.23) is a representation of $SU(2)$, but not of $SO(3)$. As far as space only in concerned, it is $SO(3)$ which is at work, but some fields — spinors — can "feel" the difference. In this section we shall be using the ordinary metric $g_{ij} = \delta_{ij}$ for the Euclidean space, so that there will be no difference at all between raised and lowered indices.

Comment 2.4 Lie groups are differentiable manifolds. The $SU(2)$ manifold is a 3-dimensional sphere S^3, while the $SO(3)$ manifold is like half the sphere S^3. The topology of $SU(2)$ is consequently simpler. It all happens as if $SU(2)$ "covered" $SO(3)$ twice. Technically, $SU(2)$ is indeed the double covering of $SO(3)$.

§ 2.13 The fundamental representation (that of lowest dimension) of $SO(3)$ is given precisely by the orthogonal 3×3 matrices which have given it its name. The most convenient basis for the 3 generators J_a of the Lie algebra, which satisfy the commutation relations

$$[J_b, J_c] = i f^a{}_{bc} J_a \tag{2.15}$$

will be that in which their matrix elements are just the structure coefficients,

$$(J_b)_{ac} = i f^a{}_{bc} \, . \tag{2.16}$$

Given a Lie algebra, the representation whose matrix elements are just the structure coefficients is called the "adjoint representation" of the algebra. It is more deeply concerned with the group geometry than any other representation. For $SO(3)$, this representation coincides with the fundamental one, which is not the case for most groups.

For the rotation group, the structure constants are given by $f^a{}_{bc} = \epsilon_{abc}$, the Kronecker completely antisymmetric symbol given in Exercise 1.7. Thus, (2.15) is just the usual table of commutators

$$[J_1, J_2] = i\, J_3 \quad ; \quad [J_2, J_3] = i\, J_1 \quad ; \quad [J_3, J_1] = i\, J_2 \, ; \tag{2.17}$$

(if we take hermitian matrices for the J_k's; if antihermitian, just drop the i factors). The matrices are

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad ; \quad J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \quad ; \quad J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \, . \tag{2.18}$$

Without the i factors, the generators constitute also a basis for the underlying vector space of the Lie algebra. In terms of the group parameters, which can be taken as 3 angles collectively represented by the vector $\boldsymbol{\omega} = (w_1, w_2, w_3)$, a generic member of the Lie algebra will be

$$W = -\,i\,J_a\,w_a = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} .$$

The general group element $R = \mathrm{e}^W$ can be obtained without much ado by direct exponentiation. Using the notation $\omega = |\boldsymbol{\omega}|$ and defining in the parameter space the unit vector $\mathbf{u} = \boldsymbol{\omega}/|\boldsymbol{\omega}|$, its explicit form is

$$R(\boldsymbol{\omega}) = \begin{pmatrix} \mathcal{C} + (1-\mathcal{C})u_1u_1 & (1-\mathcal{C})u_1u_2 - \mathcal{S}\,u_3 & (1-\mathcal{C})u_1u_3 + \mathcal{S}\,u_2 \\ (1-\mathcal{C})u_1u_2 + \mathcal{S}\,u_3 & \mathcal{C} + (1-\mathcal{C})u_2u_2 & (1-\mathcal{C})u_2u_3 - \mathcal{S}\,u_1 \\ (1-\mathcal{C})u_1u_3 - \mathcal{S}\,u_2 & (1-\mathcal{C})u_2u_3 + \mathcal{S}\,u_1 & \mathcal{C} + (1-\mathcal{C})u_3u_3 \end{pmatrix} \quad (2.19)$$

where $\mathcal{C} = \cos\omega$ and $\mathcal{S} = \sin\omega$. The general expressions of the matrix elements, as in

$$\begin{aligned} W_{ij} &= -\,\epsilon_{ijk}\omega_k \quad ; \quad W_{ij}^2 = w_i\,w_j - \delta_{ij}\,\omega^2\ ; \\ R_{ij} &= \delta_{ij} - \epsilon_{ijk}\,u_k\,\sin\omega + (u_iu_j - \delta_{ij})(1-\cos\omega)\ , \end{aligned}$$

are particularly useful for computations. The intuitive meaning of the exponential matrix becomes clear when we apply it to a vector $\mathbf{x}$:

$$\begin{aligned} R(\boldsymbol{\omega})\,\mathbf{x} &= (\mathbf{u}\cdot\mathbf{x})\mathbf{u} + (\mathbf{u}\times\mathbf{x})\times\mathbf{u}\cos\omega + (\mathbf{u}\times\mathbf{x})\sin\omega \\ &= \mathbf{x}\cos\omega + (\mathbf{u}\times\mathbf{x})\sin\omega + (1-\cos\omega)(\mathbf{u}\cdot\mathbf{x})\mathbf{u}\ , \end{aligned}$$

or still, decomposing $\mathbf{x}$ into a piece parallel to $\mathbf{u}$

$$\mathbf{x}^{\parallel} = (\mathbf{u}\cdot\mathbf{x})\mathbf{u}$$

and a piece orthogonal to $\mathbf{u}$,

$$\mathbf{x}^{\perp} = (\mathbf{u}\times\mathbf{x})\times\mathbf{u}\cos w = x - (\mathbf{u}\cdot\mathbf{x})\mathbf{u},$$

$$R(\boldsymbol{\omega})\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}\cos\omega + (\mathbf{u}\times\mathbf{x})\sin\omega\ . \quad (2.20)$$

Thus, $R(\boldsymbol{\omega})$ represents a rotation around $\mathbf{u}$ or $\boldsymbol{\omega}$. The unit vector $\mathbf{u}$, which is an eigenvector of $R(\boldsymbol{\omega})$ with eigenvalue 1, defines the *axis of rotation*. In particular, a rotation of an angle ϕ around the 3^{rd} axis $\mathbf{k}$ (see §1.1) is given by $\omega = (0, 0, \phi)$ and has the effect

$$R(0,0,\phi)\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}\cos\phi + (\mathbf{k}\times\mathbf{x})\sin\phi\ . \quad (2.21)$$

A complete rotation around the 3^{rd} axis is given by $\omega = (0, 0, 2\pi)$. It leads to $R(0, 0, 2\pi)\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp} = \mathbf{x}$. As expected, a complete rotation around an axis corresponds to the identity transformation: $R(0, 0, 2\pi) = I$. We shall see below, in the discussion of the Pauli representation, that this result is not as trivial as it may seem.

Notice that these transformations can be seen as acting on the "position vector" $\mathbf{x}$, or any other vector. We can give here an important step towards abstraction: to define a vector as an object transforming as above. Still better: an object transforming as above is said to belong to the "vector representation" of the group.

§ 2.14 Casimir operator We can write also the explicit form of W^2:

$$W^2 = \begin{pmatrix} -{\omega_3}^2 - {\omega_2}^2 & \omega_1\omega_2 & \omega_1\omega_3 \\ \omega_1\omega_2 & -{\omega_3}^2 - {\omega_1}^2 & \omega_2\omega_3 \\ \omega_1\omega_3 & \omega_2\omega_3 & -{\omega_2}^2 - {\omega_1}^2 \end{pmatrix} .$$

By taking the trace, we find that

$$-\frac{1}{2}\,\mathrm{tr}W^2 = -\mathrm{tr}(J_aJ_b)w^aw^b = -\frac{1}{2}\,\omega^a\omega^b\epsilon_{cad}\epsilon_{dbc} = \delta_{ab}\omega^a\omega^b = w^2 .$$

The matrix making its appearance here, the bilinear form with entries $\gamma_{ab} = \frac{1}{2}\,\mathrm{tr}(J_aJ_b)$, is an invariant. It gives always the same matrix if acted upon by the group transformations: $R^{-1}\gamma R = \gamma$. Or, equivalently, it commutes with all the generators. The operator $J^2 = \gamma_{ab}J^aJ^b$ is also an invariant, the squared angular momentum. This kind of invariant operator, obtained by using an invariant bilinear form on the generators, is called a Casimir invariant. It has a fixed eigenvalue in a fixed representation.

Comment 2.5 γ_{ab} is actually a metric on the group manifold, called the Killing–Cartan metric [more about that far below, see Eq.(9.64)]. We have chosen the factors so that, for the rotation group, it coincides with the Euclidean metric (1.27) in cartesian coordinates,

$$(\delta_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{2.22}$$

Exercise 2.6 For the squared angular momentum, it is usual to write the eigenvalue in the form $J^2 = j(j+1)$. Show that, for the vector representation given by matrices (2.18), $J^2 = 2$ and $j = 1$. ■

§ 2.15 For general Lie groups and algebras, more invariants can be found, each one related to an invariant multilinear form. The number of independent invariants will

be the rank of the Lie algebra. The rotation group has rank 1: the above invariant is the only one. The Euclidean group, the group $SU(3)$, the Lorentz group and the Poincaré group are rank-2 groups and will have two independent invariants. As the invariants commute with every operator, they have the same eigenvalues for all the states of a given representation. An important result from group theory is the following: a representation is completely characterized by the eigenvalues of the independent invariant operators. The number of independent invariant operators is the *rank* of the group. Their eigenvalues are consequently used to label the representations. Of course, any function of the invariants is itself invariant. It is then possible to choose the independent invariants to be used — with obvious preference for those with a clear physical meaning.

§ 2.16 The action of the group transformations on its own Lie algebra is called the adjoint representation of the group. The algebra members are transformed by similarity: given any element $M = J_a M^a$ in the algebra, it will transform according to $M \to M' = g^{-1} M g$. Because of its role in the differential structure of the Lie group, this is the most important representation, on which all the other representations somehow mirror themselves. Here, for an arbitrary element M,

$$M' = R^{-1} M R = M \cos\omega + [M, W] \frac{\sin\omega}{\omega} + W(\mathbf{M} \cdot \boldsymbol{\omega}) \frac{1 - \cos\omega}{\omega^2} \; .$$

Comment 2.6 We see easily why $\gamma_{ab} = \frac{1}{2} \, \mathrm{tr}(J_a J_b)$ is an invariant: it is a trace, and traces are invariant under similarities because $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and consequently $\mathrm{tr}M' = \mathrm{tr}(R^{-1} M R) = \mathrm{tr}(R R^{-1} M) = \mathrm{tr}M$.

Comment 2.7 The form (2.19) is one of many possible different parameterizations for a given rotation. It has been used because it gives special emphasis to the role of the Lie algebra and the related adjoint representation. The Euler angles, much used in the problem of the rigid body, provide another. Lie groups are manifolds of a very special kind and, roughly speaking, parameters are coordinates on the group manifold. Changing parameterizations corresponds to changing coordinate systems. As with coordinates in general, a special parameterization can ease the approach to a particular problem. The Cayley-Klein parameters, for example, are more convenient to solve some gyroscope problems.

§ 2.17 For $SU(2)$, the fundamental representation is generated by the 2×2 Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad ; \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad ; \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \; . \tag{2.23}$$

It is easily checked that the generators of the Lie algebra, satisfying (2.17), are actually $\{\frac{1}{2} \, \sigma_a\}$. The general element W in the Lie algebra representation is now

given by

$$W = \tfrac{1}{2}\,\sigma_a\,\omega_a = \tfrac{1}{2}\begin{pmatrix} \omega_3 & \omega_1 - i\omega_2 \\ \omega_1 + i\omega_2 & -\omega_3 \end{pmatrix} .$$

The group element can again be found by direct exponentiation:

$$R(\omega) = e^{iW} = I\cos\frac{\omega}{2} + i(\boldsymbol{\sigma}\cdot\frac{\boldsymbol{\omega}}{\omega})\sin\frac{\omega}{2} . \tag{2.24}$$

§ 2.18 As a rule, the value of j specifies the "spin value". For the spinor representation, $J^2 = 3/4$ and $j = 1/2$. We say that "spin one–half" systems are described. The vector representation is attributed "spin–one". We shall see much more about spinors when discussing Dirac fields.

Exercise 2.7 In a way analogous to Exercise 2.6, show that for the spinor representation given by the Pauli matrices (2.23), the values are indeed $J^2 = 3/4$ and $j = 1/2$. ■

§ 2.19 We have seen above that $R(0,0,2\pi) = I$ for the vector representation. Here, however, we find immediately that $R(0,0,2\pi) = I\cos\pi = -I$. A complete rotation around an axis does not lead back to the original point. Only a double complete rotation, like $R(0,0,4\pi)$, does. Objects belonging to this representation are deeply different from vectors. They are called (Weyl or Pauli) spinors. The column vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, eigenvectors of σ_3, can be used as a basis: any such spinor can be written in the form

$$\psi = \psi_\uparrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \psi_\downarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \psi_\uparrow \\ \psi_\downarrow \end{pmatrix} . \tag{2.25}$$

Comment 2.8 The reason for the "arrow" notation is the following: σ_3 will appear later as the operator giving the spin eigenvalues along the axis $0z$: eigenvalue +1 for spin "up", eigenvalue -1 for spin "down". The general Pauli spinor will be a superposition of both.

Comment 2.9 Non–matrix representations can be of great help in some cases. For example, take the functions $\Psi(\mathbf{x})$ defined on the Euclidean 3-dimensional space. Acting on them, the set of differential operators $J_m = i\epsilon_{mrs}x^r\partial^s$ satisfy the rules (2.17) and lead to the invariant value $j = 1$. They provide thus a representation equivalent to that of matrices (2.18). Two distinct kinds of operators, acting on different spaces but with the same value of the invariants are said to provide different "realizations" of the corresponding representation.

§ 2.20 We have above defined the vector representation of the group. A vector, or a member of the vector representation, is any object $\mathbf{v} = (v^1, v^2, v^3)^T$ transforming according to $\mathbf{v} = R(\boldsymbol{\omega})\,\mathbf{v}$, or , $(v^{1'}, v^{2'}, v^{3'})^T = R(\boldsymbol{\omega})\,(v^1, v^2, v^3)^T$. We shall use the compact version

$$v^{i'} = R^{i'}{}_j\, v^j ,$$

giving each component. The position vector is a particular case, with $x^{i'} = R^{i'}{}_j\, x^j$. A question may come to the mind: how would a composite object, such as one with components like $x^i x^j$, transform? Well, each piece will transform separately, and $x^{i'} x^{j'} = R^{i'}{}_m R^{j'}{}_n x^m x^n$. Any object transforming in this way,

$$T^{i'j'} = R^{i'}{}_m R^{j'}{}_n T^{mn} \ ,$$

is said to be a second–order tensor, or to belong to a tensor representation of the group. Notice that there is no need at all that it be a product like $x^i x^j$: the only thing which matters is the way it undergoes the transformations of interest. In the same way, higher–order tensors are defined: a tensor of order r is any object transforming according to

$$T^{i_1' i_2' i_3' \ \cdots \ i_r'} = R^{i_1'}{}_{j_1} R^{i_2'}{}_{j_2} R^{i_3'}{}_{j_3} \ \ldots R^{i_r'}{}_{j_r} T^{j_1 j_2 j_3 \ \cdots \ j_r} \ .$$

We have thus an infinite series of tensor representations of the rotation group. Different representations mean different spaces: $T^{i_1 i_2}$, which can be represented by a 3×3 matrix, belongs to a space quite different from that to which a vector v^i belongs. It is enough to look at the dimensions to see that each tensorial order leads to different spaces. But there is more. As is well known in the case of rotations, higher–dimensional vectors spaces, on whose members the groups acts by higher–dimensional matrices, provide new representations. Each one corresponds to some value of $\mathbf{J}^2 = j(j+1)$. It is proven that there is one representation of $SO(3)$ for each non–vanishing integer value of j, and one representation of $SU(2)$ for each non–vanishing half–integer or integer value of j. Thus, $SU(2)$ has all the representations $j = 1, 2, 3, \ldots$ of $SO(3)$, plus those with $j = 1/2, 3/2, 5/2, \ldots$ To these we can add a $j = 0$ representation, whose members have one component only and are called invariants, scalars or singlets of both groups.

§ 2.21 The $N \times N$ matrices representing $SO(3)$ constitute a group, isomorphic to our original $SO(3)$. And we arrive thus to the fundamental notion of abstract group. It is far more convenient to look at all these "copies" of $SO(3)$ as mere representations of one same abstract group, which by historical and intuitive reasons we call ... again $SO(3)$. Abstract groups coming up in Physics have this thing in common, they are named after the first representation that has been found. $SU(2)$ is no more simply the group of special unitary complex 2×2 matrices: it is the abstract group isomorphic to that one, with an infinite number of representations given by higher–order matrices.

When we talk currently about "vectors", we usually mean just the above vectors, those of $SO(3)$. We shall, however, see other groups below. Each one will have its

own vectors, tensors, etc. It should be clear that, whenever we use the expressions "vectors" and "tensors", we should specify the group.

2.4 The Poincaré Group

§ 2.22 Different observers, as said, are supposed to be attached to different inertial reference frames. The field equations must be valid and the same in any inertial reference frame. This means that they must be written as equalities with the right– and left–hand sides transforming in exactly the same way. We say then that they are "in covariant form". To obtain covariant equations, it is enough that they come as extrema of an action functional which is invariant under the group of transformations of inertial frames.

The transformations of the Poincaré group are, by definition, those continuous transformations under which is invariant the interval between two events, or points on Minkowski spacetime.

The Poincaré transformations are of two types, each constituting a sub–group: translations and Lorentz transformations. The latter leave the above expression invariant because they are the (pseudo) rotations in Minkowski space, on which the interval represents a (pseudo) distance. Spacetime translations constitute the so–called inhomogeneous part of the Poincaré group. They leave the interval invariant because they change both the origin and other points in the same way. Mathematically speaking, this means that spacetime should be seen not as a vector space, but as an affine space: Minkowski spacetime has no preferred origin. The interval is also invariant under some discrete transformations: inversions (2.2), (2.3) of the space axis and of the time axis (2.4). Some authors round up these transformations and those constituting the Lorentz group into the "full Lorentz group". Other withdraw from this denomination the time inversion. We shall consider only the continuous transformations here, and consider a Poincaré transformation as a Lorentz transformation plus a translation. Given cartesian coordinates on spacetime,

$$x'^{\alpha} = \Lambda^{\alpha}{}_{\beta}\, x^{\beta} + a^{\alpha} \ . \tag{2.26}$$

The interval will remain invariant if

$$\eta_{\alpha\beta}\, \Lambda^{\alpha}{}_{\gamma}\, \Lambda^{\beta}{}_{\delta} = \eta_{\gamma\delta} \ . \tag{2.27}$$

These pseudo–orthogonality conditions on the matrix $\Lambda = (\Lambda^{\alpha}{}_{\beta})$ can be seen, as already said, as the defining relations of the Lorentz subgroup. It is usual to indicate (2.26) by the compact notation

$$L = (\Lambda, a) \ . \tag{2.28}$$

Pure translations and Lorentz transformations are given respectively by (I, a) and $(\Lambda, 0)$. The composition of two Poincaré transformations comes immediately from (2.26):

$$(\Lambda', a')(\Lambda, a) = (\Lambda'\Lambda, \Lambda' a + a') \ . \tag{2.29}$$

Comment 2.10 In (2.26), a Lorentz transformation is performed first, and then translations are applied: $(\Lambda, a) = (1, a) \cdot (\Lambda, 0)$. This is the most widely–used convention. Another is possible: we could have written $x'^a = \Lambda^\alpha{}_\beta (x^\beta + a^\beta)$ and $(\Lambda, a) = (\Lambda, 0) \cdot (1, a)$. In the latter parameterization, a Lorentz transformation is applied to the already translated point.

A general Poincaré transformation is conveniently represented by the 5×5 matrix appearing in (1.77).

2.5 The Lorentz Group

§ 2.23 The general Lorentz transformation, including rotations, is far more complicated than that given in Eq.(1.78). In practice, we decompose it in a product of rotations and boosts, which is always possible. Boosts can be seen as "rotations" of imaginary angles: it is enough to take $\cosh \alpha = \gamma$ and $\mathrm{tgh}\, \alpha = v/c$. As to the translations, they form an abelian subgroup.

Let us consider now the fundamental question: given the fields necessary to the description of the physical system, how do we obtain an invariant Lagrangian $\mathcal{L}(\vec{\phi}, \partial_\mu \vec{\phi})$? It is a preliminary requisite that the fields have a well–defined behavior when submitted to transformations. Fields — used to describe systems with a continuous infinity of degrees of freedom — belong to certain function spaces, which we could call "configuration spaces" — just the spaces of those degrees of freedom. Under a transformation, a field ϕ is taken into some other field ϕ' belonging to the same space of ϕ and which represents, in the new frame, the same configuration represented by ϕ in the original frame. It is in this sense that a configuration is invariant in Special Relativity: it is impossible to discover, by the sole analysis of the states, on which frame the system stands. Thus, given an L as in (1.77), there exists acting on the configuration space an operator $U(L)$ such that

$$\phi' = U(L)\, \phi \ .$$

Suppose then that, acting on the configuration space, there are operators $U(L_1), U(L_2), U(L_3)$, etc, corresponding to the transformations L_1, L_2, L_3, etc. And suppose furthermore that those operators respect the group conditions:

(i) $U(L_1 L_2) = U(L_1)\ U(L_2)$ (composition)

(ii) $U(L_0)$ is the identity operator if L_0 is the group identity: $U(L_0)\phi = \phi$

(iii) $U(L^{-1})\,U(L) = U(L)\,U(L^{-1}) = U(L_0)$

(iv) $U(L_1L_2)\,U(L_3) = U(L_1)\,U(L_2L_3)$.

We say then that $U(L)$ "represents" L, and that the set of all $U(L)$ constitutes a representation of the group on the configuration space.

Comment 2.11 This scheme is quite general, holding for any group acting on some space, though not every space accepts a representation of a given group.

We are now in position to state what is understood by "well–defined" behavior. A field will have a well–defined behavior if it belongs to a space on which a representation is defined. It is usual to say, for simplicity, that "ϕ belongs to the representation". Thus, the admissible fields (the "relativistic fields") are those belonging to the representations of the Poincaré group.

There is a last step. A field can be given by an individual function ϕ, or by several functions ϕ_i which can be seen as components of a single field, arranged for example in a column–vector $(\phi_1, \phi_2, \dots)^T$ on which the operators $U(L)$ will act as matrices. Well, but not only vectors have components. Tensors have them, which generalize vectors. Fields can be tensors, on whose indices the matrix of the vector representation will act one at a time. Configuration space would then consist of tensors. But, even then, the $U(L)$ would provide linear transformations. They are said to constitute a *linear representation*. Notice that non–linear representations are quite possible, but the formalism they are involved in is far more complicated. For simplicity, we shall devote ourselves almost exclusively to linear representations, which are realized by fields with components.

As said, we should specify the group whenever we use the expressions "vectors" and "tensors". The tensors used in General Relativity are tensors of the group of general coordinate transformations. We shall be interested, by now, in the Lorentz group tensors (which are unaffected by translations). A tensor of a generic order N (or: a field belonging to a tensor representation) will transform according to

$$T^{\mu'_1\mu'_2\dots\mu'_N} = \Lambda^{\mu'_1}{}_{\nu_1}\,\Lambda^{\mu'_2}{}_{\nu_2}\dots\Lambda^{\mu'_N}{}_{\nu_N}\,T^{\nu_1\nu_2\dots\nu_N} \tag{2.30}$$

§ 2.24 A first important particular case is the scalar field which, because it has only one component, is invariant:

$$\phi'(x') = \phi(x)\ . \tag{2.31}$$

It represents a single infinity of degrees of freedom.

Comment 2.12 When a field belongs to a representation but has only one component, we say that it is a singlet. This terminology holds for other groups: whenever a field ignores a symmetry group, we put it into a singlet representation. The Lagrangian of any theory is a scalar with respect to all the symmetries.

A second important particular case is the vector field, which transforms like x^μ:

$$V^{\mu'}(x') = \Lambda^{\mu'}{}_{\nu}\, V^{\nu}(x). \tag{2.32}$$

A second-order tensor will have the behavior

$$T^{\mu'\nu'}(x') = \Lambda^{\mu'}{}_{\rho}\, \Lambda^{\nu'}{}_{\sigma} T^{\rho\sigma}(x), \tag{2.33}$$

and so on for higher-order tensors. The word "tensor" can be used in a larger sense, so as to encompass scalars (0-th order tensors) and vectors (first order). We shall later on examine the main cases of physical interest. But it should be clear that it is the Lorentz group which fixes the terminology. When talking about relativistic fields, we classify them as scalars, vectors, 2nd-order tensors, spinors, etc. All these names refer to their behavior under the Lorentz group.

> Relativistic fields are defined according to their behavior under Lorentz transformations, that is, according to the Lorentz group representation they belong to.

§ 2.25 A surprise comes out in this story: the tensor representations do not cover all the linear representations. This is due to the fact that the mappings $\Lambda \to U(\Lambda)$ defining the representations are not necessarily single–valued. We have seen that this happens for the $SU(2)$ Pauli representation: a rotation of an angle 2π, which in $\mathbb{E}^3$ is the same as a rotation of 4π, is taken into +1 or -1.

Comment 2.13 For the time being, only particles corresponding to low order tensors and spinors (that is, small values of j) have been discovered in Nature.

§ 2.26 Let us make a parenthesis on Quantum Mechanics. The fields $\Psi(\mathbf{x}, t) = \Psi(x)$ turning up are, ultimately, wavefunctions. Take two of them, say $\phi(x)$ and $\psi(x)$ and submit them to a change of frame: $\phi'(x') = U(L)\phi(x)$, $\psi'(x') = U(L)\psi(x)$. The invariance of physical measurements will require

$$| < \phi(x)|\psi(x) > |^2 = | < \phi'(x')|\psi'(x') > |^2 = | < \phi(x)|U^\dagger(L)U(L)|\psi(x) > |^2 \,,$$

that is,

$$U^\dagger(L)\; U(L) = \pm 1 \,.$$

The operators $U(L)$ must, consequently, be either unitary or antiunitary — which, by the way, is true for any symmetry. In the Hamiltonian case, for example, two systems are equivalent when related by a canonical transformation. Here, two systems are equivalent when related by a unitary or antiunitary transformation. A discrete transformation can, in principle, be represented by an antiunitary operator. For continuous transformations, however, which can, when small enough, be seen as infinitesimally close to the identity, the operators must be actually unitary, as for small transformations both U and $U^\dagger$ are both close to the identity. Consequently, an $U(L)$ connected to the identity will have the form

$$U(L) = \mathrm{e}^{iJ} \ ,$$

where J is a hermitian operator (or $U(L) = \mathrm{e}^{J}$, where J is antihermitian.).
In the representation $U : (\Lambda, a) \to U(\Lambda, a)$, the composition rule is

$$U(\Lambda', a')\ U(\Lambda, a) = U(\Lambda'\Lambda, \Lambda' a + a') \ . \tag{2.34}$$

The operators will have the forms

$$U(a) = \exp[i a_\mu P^\mu] \tag{2.35}$$

$$U(\Lambda) = \exp[\tfrac{i}{2}\, \alpha_{\mu\nu} M^{\mu\nu}] \ , \tag{2.36}$$

where a_μ and $\alpha_{\mu\nu} = -\,\alpha_{\nu\mu}$ are the parameters of the transformations and P^μ, $M^{\mu\nu}$ their generators, matrices in the linear representations. The commutation relations of the generators are characteristic of the group itself and, consequently, independent of the representation. The generators themselves belong to a vector space and, with the operation defined by the commutator, constitute an algebra. The commutators are antisymmetric and satisfy the Jacobi identity, defining the Lie algebra of the group. The Poincaré commutation rules are

$$[P^\mu, P^\nu] = 0 \tag{2.37}$$

$$\left[M^{\mu\nu}, P^\lambda\right] = -\,i\ \left(P^\mu\, \eta^{\nu\lambda} - P^\nu\, \eta^{\mu\lambda}\right) \tag{2.38}$$

$$[M^{\mu\nu}, M^{\rho\sigma}] = i\ (M^{\mu\rho}\, \eta^{\nu\sigma} + M^{\nu\sigma}\, \eta^{\mu\rho} - M^{\nu\rho}\, \eta^{\mu\sigma} - M^{\mu\sigma}\, \eta^{\nu\rho}) \, . \tag{2.39}$$

It is usual to forget good (mathematical) manners and call this commutation table "the group algebra". Equation (2.37) says that translations are independent of each other; (2.38) says that their generators transform as vectors under Lorentz transformations; and (2.39), that the generators of Lorentz transformations behave as second–order tensors. This can be directly verified by using the "kinematic" realization

$$P^\mu = i\ \partial^\mu \tag{2.40}$$

$$M^{\mu\nu} = -\,i\ (x^\mu \partial^\nu - x^\nu \partial^\mu) \tag{2.41}$$

whose operators act on functions defined on spacetime (when no place for confusion exists, we shall be using the notation $\partial_\alpha = \frac{\partial}{\partial x^\alpha}$). We have said that generators transform like vectors under Lorentz transformations. It is immediate to find that property (2.38) is satisfied if P^μ is replaced by x^μ. Any object satisfying it is, by definition, a vector of the Lorentz Lie algebra.

Exercise 2.8 Verify that the operators (2.40) and (2.41) really satisfy Eqs.(2.37-2.39). ■

The action functional should be invariant. As the modulus of the Jacobian determinant of these transformations is 1, the integration measure $d^4x = dx^0dx^1dx^2dx^3$ is invariant, and consequently the density must also be invariant.

Expression (2.41) generalizes to spacetime the non–relativistic orbital angular momentum $M^{ij} = -i(x^i\partial^j - x^j\partial^i) = M_{ij} = x_ip_j - x_jp_i$, related to the usual angular momentum $L_k = (\mathbf{r} \times \mathbf{p})_k$ by $L_k = (1/2)\epsilon_{kij}M_{ij}$. The component M^{12} of angular momentum is the generator of rotations on the plane (12), and so on. Let us recall that the spin of a particle is the angular momentum which remains in its rest frame, an effect of quantum nature: in $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, operators $\mathbf{r}$ and $\mathbf{p}$ cannot have both well–determined simultaneous values in Quantum Mechanics. In the classical case, it can be accounted for by adding to (2.41) an operator $s^{\mu\rho}$ commuting with $M^{\rho\sigma}$ and also satisfying (2.39). It is, in this way, "put in by hand".

§ 2.27 In order to classify the relativistic fields, we need a method to classify the representations of the Poincaré group. We have said in § 2.15, when talking about rotations, that invariant operators have fixed eigenvalues in a given representation, and can be used to classify representations. The number of independent, intercommuting invariant operators is the (algebra or group) rank. The rotation group has rank one, and the square angular momentum is the best choice to classify its representations. The Lorentz group and the Poincaré group have rank two. We need consequently two invariant operators of the algebra (2.37)-(2.39), two operators which commute with all the generators. It is easy to verify that $P_\mu P^\mu$, for example, does that. We shall want to classify particles, and the fields related to them, by the mass and the spin, and shall choose the two invariants which are nearer to them. These are special combinations of the momentum and the angular momentum. We shall see later how to obtain the momentum and angular momentum for fields (this is the role of the first Noether theorem), and here only illustrate the ideas in the case of a mechanical particle. The invariant best related to mass is precisely $P_\mu P^\mu$, whose eigenvalues are well–known from Special Relativity: $P_\mu P^\mu = m^2c^2$, m being the rest mass of the particle of 4-momentum P^μ. In realization (2.40), this invariant operator is (minus) the D'Alembertian. Another invariant operator is $W_\mu W^\mu$, where

W_μ is the Pauli–Lubanski operator

$$W_\sigma = -\tfrac{i}{2}\,\epsilon_{\mu\nu\rho\sigma}\,\left(M^{\mu\nu} + s^{\mu\nu}\right) P^\rho\ , \tag{2.42}$$

$\epsilon_{\mu\nu\rho\sigma}$ being the totally antisymmetric Levi–Civita (or Kronecker) symbol. This symbol generalizes to 4-dimensional space the symbol we have met in Exercise 1.7.

Comment 2.14 The totally antisymmetric Levi–Civita symbol, or 4-dimensional Kronecker symbol, is defined by

$$\epsilon_{\mu\nu\rho\sigma} = \begin{cases} 1 \text{ if } \mu\ \nu\ \rho\ \sigma \text{ is an even permutation of } 0123 \\ -1 \text{ if } \mu\ \nu\ \rho\ \sigma \text{ is an odd permutation of } 0123 \\ 0 \text{ otherwise} \end{cases} \tag{2.43}$$

The starting value is $\epsilon_{0123} = 1$. The determinant form is

$$\epsilon_{\mu\nu\rho\sigma} = \begin{vmatrix} \delta^0_\mu & \delta^0_\nu & \delta^0_\rho & \delta^0_\sigma \\ \delta^1_\mu & \delta^1_\nu & \delta^1_\rho & \delta^1_\sigma \\ \delta^2_\mu & \delta^2_\nu & \delta^2_\rho & \delta^2_\sigma \\ \delta^3_\mu & \delta^3_\nu & \delta^3_\rho & \delta^3_\sigma \end{vmatrix}. \tag{2.44}$$

Indices are here raised and lowered with the Lorentz metric, so that $\epsilon^{0123} = -\,1$. In 4-dimensional space there is no more a relationship between vectors and antisymmetric matrices. However, the symbol allows the definition of the dual: given a 2nd order antisymmetric tensor $F^{\rho\sigma}$, its dual is defined as

$$\tilde{F}_{\mu\nu} = \tfrac{1}{2}\ \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}. \tag{2.45}$$

Notice that, in order to prepare for the contraction, indices must be raised. The dual, consequently, depends on the metric. Some identities come out from the contractions of the symbols themselves:

1. once contracted:

$$\epsilon_{\mu\nu\rho\sigma}\epsilon^{\mu\alpha\beta\gamma} = \delta^\alpha_\nu\delta^\beta_\rho\delta^\gamma_\sigma - \delta^\alpha_\nu\delta^\beta_\sigma\delta^\gamma_\rho - \delta^\alpha_\sigma\delta^\beta_\rho\delta^\gamma_\nu + \delta^\alpha_\sigma\delta^\beta_\nu\delta^\gamma_\rho + \delta^\alpha_\rho\delta^\beta_\sigma\delta^\gamma_\nu - \delta^\alpha_\rho\delta^\beta_\nu\delta^\gamma_\sigma \tag{2.46}$$

2. twice contracted:

$$\epsilon_{\mu\nu\rho\sigma}\epsilon^{\mu\nu\beta\gamma} = 2\ (\delta^\beta_\rho\delta^\gamma_\sigma - \delta^\beta_\sigma\delta^\gamma_\rho) \tag{2.47}$$

3. thrice contracted:

$$\epsilon_{\mu\nu\rho\sigma}\epsilon^{\mu\nu\rho\gamma} = 3!\ \delta^\gamma_\sigma = 6\ \delta^\gamma_\sigma \tag{2.48}$$

4. totally contracted:

$$\epsilon_{\mu\nu\rho\sigma}\epsilon^{\mu\nu\rho\sigma} = 4!\ = 24\ . \tag{2.49}$$

Of course, other conventions concerning the signs are possible.

Operator $W_\mu W^\mu$ is of fundamental interest because it is intimately related to spin. Notice first that, as an invariant, it can be calculated in any reference frame. Suppose $m \neq 0$ and take the rest frame, in which the momentum is the 4-vector $\overset{\circ}{p} = (mc, 0)$. We see easily that $\overset{\circ}{W}_0 = 0$. The space components of W are

$$\overset{\circ}{W}_i = -\ \tfrac{i}{2}\,\epsilon_{lk0i}\ s^{lk}\ mc = -\ \tfrac{i}{2}\,mc\,\epsilon_{ilk}\ s^{lk} \equiv -\,mc\ \hat{S}_i\ . \tag{2.50}$$

We have reverted to the 3-dimensional Kronecker symbol, because $\epsilon_{ijk} = \epsilon_{0ijk}$. We have also profited to introduce the spin operator $\hat{S}$, a 3-vector: $\hat{S}_i = \frac{1}{2}\,\epsilon_{ilk}\; s^{lk}$. We now obtain $W_\mu W^\mu = W_i W^i = -\,W_i W_i = -\,m^2c^2\hat{S}_i\hat{S}_i = -\,m^2c^2\,\hat{\mathbf{S}}^2$, whose eigenvalues will be $-\,m^2c^2\,s(s+1)$. We have said above that the Poincaré group has rank two: the eigenvalues of two invariant operators are necessary to label a representation. If we choose P^2 and W^2 for that role, the representations will be classified by the values of two quantities of obvious physical meaning. A relativistic field will belong to a representation and be characterized by its spin and mass, besides other eventual invariants related to other symmetries. This is not true, of course, if $m = 0$. The spin operator is not defined in that case. We shall see later how an alternative quantity, the helicity, can be used in that case. Notice *en passant* that we need the whole Poincaré group in order to classify the elementary particles: the Lorentz group alone would provide no invariant related to the mass.

§ 2.28 If we want a matrix realization of the generic element of the Lorentz group, as in (1.78), we must beforehand introduce matrix generators. We can take, for the rotations, those of (2.18) transmuted into 4×4 matrices:

$$J_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}; J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}; J_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

and, for the "boosts",

$$K_1 = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; K_2 = -\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$K_3 = -\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

These six generators can be put together as matrices $J_{\alpha\beta}$ with elements

$$(J_{\alpha\beta})^\mu{}_\nu = i\,(\eta_{\alpha\nu}\,\delta^\mu{}_\beta - \eta_{\beta\nu}\,\delta^\mu{}_\alpha)\,. \tag{2.51}$$

The relationship is given by $J_k = \frac{1}{2}\,\epsilon_{kij}J_{ij}$; $J_{k0} = iK_k$; $J_{0k} = -\,iK_k$. The Lorentz matrices can then be written as exponentials,

$$\Lambda^\mu{}_\nu = \left(\exp[\tfrac{i}{2}\,\omega^{\alpha\beta}J_{\alpha\beta}]\right)^\mu{}_\nu\,,$$

where $\omega^{\alpha\beta} = -\,\omega^{\beta\alpha}$ are the transformation parameters. The matrices $J_{\alpha\beta} = -\,J_{\beta\alpha}$ satisfy (2.39): they provide, as expected, a representation (the vector representation) of the Lie algebra of the Lorentz group. The factor " $\frac{1}{2}$ " in the exponent accounts for double counting.

§ 2.29 Let us rewrite (2.32) for the vector fields:

$$\phi'^{\mu}(x') = \Lambda^{\mu}{}_{\nu}\,\phi^{\nu}(x) = \left(\exp[\tfrac{i}{2}\omega^{\alpha\beta}J_{\alpha\beta}]\right)^{\mu}{}_{\nu}\,\phi^{\nu}(x)\ .$$

Notice that the generators $J_{\alpha\beta}$ generate the complete Lorentz transformation, including the change $x \to x'$. In effect, it is impossible to effectuate a Lorentz transformation on the functional form ϕ^{ν} alone, as the argument is itself a Lorentz vector. The infinitesimal transformations, to first order in a small parameter $\delta\omega^{\alpha\beta}$, follow directly:

$$\phi'^{\mu}(x') = \phi^{\mu}(x) + \tfrac{i}{2}\,\delta\omega^{\alpha\beta}\,(J_{\alpha\beta})^{\mu}{}_{\nu}\,\phi^{\nu}(x) = \phi^{\mu}(x) + \delta\omega^{\mu}{}_{\nu}\,\phi^{\nu}(x)\ ,$$

that is,

$$\delta\phi^{\mu}(x) = \phi'^{\mu}(x') - \phi^{\mu}(x) = \tfrac{i}{2}\,\delta\omega^{\alpha\beta}\,(J_{\alpha\beta})^{\mu}{}_{\nu}\,\phi^{\nu}(x) = \delta\omega^{\mu}{}_{\nu}\,\phi^{\nu}(x)\ . \tag{2.52}$$

But we have also

$$\phi'^{\mu}(x') = \phi'^{\mu}(x+dx) = \phi'^{\mu}(x) + \partial_{\lambda}\phi^{\mu}\,dx^{\lambda}\ .$$

Equating the two expressions for $\phi'^{\mu}(x')$, we obtain the infinitesimal transformation at fixed point x:

$$\begin{aligned}\bar{\delta}\phi^{\mu}(x) = \phi'^{\mu}(x) - \phi^{\mu}(x) &= \delta\omega^{\alpha\beta}\,(J_{\alpha\beta})^{\mu}{}_{\nu}\,\phi^{\nu} - \partial_{\lambda}\phi^{\mu}\,dx^{\lambda} \\ &= \delta\phi^{\mu}(x) - \partial_{\lambda}\phi^{\mu}\,dx^{\lambda}\ .\end{aligned} \tag{2.53}$$

Notice that we have introduced the notation

$$\bar{\delta}\phi^{\mu}(x) = \phi'^{\mu}(x) - \phi^{\mu}(x) \tag{2.54}$$

for the fixed-point variation of ϕ. Let us rewrite the total variation:

$$\delta\phi^{\mu}(x) = \bar{\delta}\phi^{\mu}(x) + \partial_{\lambda}\phi^{\mu}\,dx^{\lambda}\ . \tag{2.55}$$

Comment 2.15 For the spacetime cartesian coordinates,

$$x'^{\epsilon} = x^{\epsilon} + \delta\omega^{\epsilon}{}_{\gamma}x^{\gamma} \quad ; \quad x^{\delta} = x'^{\delta} - \delta\omega^{\delta}{}_{\gamma}x'^{\gamma} = \left(\delta^{\delta}{}_{\gamma} - \delta\omega^{\delta}{}_{\gamma}\right)x'^{\gamma}\ . \tag{2.56}$$

This infinitesimal transformation corresponds, consequently, to a coordinate transformation characterized by

$$\frac{\partial x^{\delta}}{\partial x'^{\sigma}} = \delta^{\delta}{}_{\sigma} - \delta\omega^{\delta}{}_{\sigma}\ . \tag{2.57}$$

Chapter 3

Introducing Fields

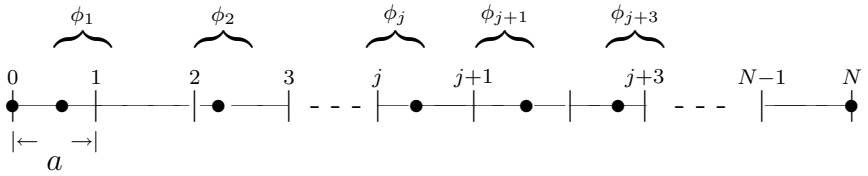
§ 3.1 In the Hamiltonian formulation of Classical Mechanics, each state of a physical system is represented by a point (q, p) on its phase space, where $q = (q^1, q^2, \cdots, q^N)$ indicates collectively its N degrees of freedom and $p = (p_1, p_2, \cdots, p_N)$ the respective conjugate momenta. The time evolution of the system is described by a line on that space, each point $(q(t), p(t))$ of which represents the set of values of all the degrees of freedom and corresponding momenta at a certain instant t. The motion along the line, supposed continuous, is described by Hamilton's equations. In the Lagrangian formulation of Classical Mechanics, states are made to correspond to points $(q(t), \dot{q}(t))$ in the so called μ-space, whose evolution is regulated by the Lagrange equations. Both the Hamiltonian and the Lagrangian formulations are refinements of Newton's approach. The theory is strictly deterministic: once the initial conditions are given, the state is fixed at every moment.

We are, of course, talking about the simplest mechanical systems, those with a finite number N of degrees of freedom (such as the harmonic oscillator, the pendulum, the system earth-sun, etc). The approaches are extended without much ado to systems with an infinite but countable number of degrees of freedom (as the classical gases, whose description requires an additional use of statistics to compensate for our ignorance of the detailed initial conditions).

The treatment can, finally, be also extended to systems with a continuous infinity of degrees of freedom. Such systems will be our concern here: a continuous infinity of degrees of freedom is what we shall call a **classical field**. Instead of the above q^k, the system is represented by a function $\phi(x)$, with ϕ replacing q and the continuous variable x replacing the tag k. The simplest among the systems of this kind is the so-called 1-dimensional solid, or better, the vibrating line. Each point of the continuous elastic solid takes part in the dynamics and the whole motion can only be described if all their positions are specified.

3.1 The Standard Prototype

§ 3.2 Let us describe a crude model for the vibrating line, which provides a paradigmatic example.* An intuitive approach starts with a discrete model and then examines the changes when a passage to the continuum is made. The system is conceived as a line of beads or "atoms" with longitudinal vibrations around their equilibrium positions, which are disposed at a distance a from their immediate neighbors. We shall suppose N "atoms" with the same mass m, so that the line has length $L = Na$.



The discrete vibrating line

Let ϕ_j be the displacement of the j-th atom with respect to its equilibrium position and suppose (here dynamics comes in) harmonic forces to be acting between (only) nearest neighbors, all with the same elastic constant K. This j-th atom will obey Newton's equation of motion

$$m\ddot{\phi}_j = K\left(\phi_{j+1} - \phi_j + \phi_{j-1} - \phi_j\right) \ , \tag{3.1}$$

with kinetic energy $T_j = (m\dot{\phi}_j^2)/2$ and potential energy

$$V_j = \tfrac{K}{4}\left[\left(\phi_{j+1} - \phi_j\right)^2 + \left(\phi_{j-1} - \phi_j\right)^2\right] \ . \tag{3.2}$$

The total Lagrangian function for all the atoms will be

$$\mathcal{L} = \tfrac{1}{2}\sum_{j=1}^{N}\left[m\dot{\phi}_j^2 - K\left(\phi_{j+1} - \phi_j\right)^2\right] \ , \tag{3.3}$$

leading to the action

$$S[\phi] = \int_0^T dt\mathcal{L} = \tfrac{1}{2}\int_0^T dt\sum_{j=1}^{N}\left[m\dot{\phi}_j^2 - K\left(\phi_{j+1} - \phi_j\right)^2\right] \ . \tag{3.4}$$

Each displacement ϕ_j is a degree of freedom, with conjugate momentum $\pi_j = \delta L/\delta\dot{\phi}_j = m\,\dot{\phi}_j$. The Lagrange equations turn out to be just (3.1).

* E. M. Henley and W. Thirring, *Elementary Quantum Field Theory*, McGraw-Hill, New York, 1962.

§ 3.3 The central actor in the Hamiltonian formulation is the Hamiltonian function

$$H = T + V = \sum_j \left[\frac{{\pi_j}^2}{2m} + \frac{m\,\omega^2}{2} \left(\phi_{j+1} - \phi_j\right)^2 \right] ,$$

where we have used the oscillator frequency $\omega = (K/m)^{1/2}$. Combining Hamilton's equations

$$\begin{aligned} \dot{\phi}_j &= \frac{\delta H}{\delta \pi_j} = \frac{\pi_j}{m} \\ \dot{\pi}_j &= -\frac{\delta H}{\delta \phi_j} = -\frac{\delta V}{\delta \phi_j} = K\left(\phi_{j+1} - \phi_j + \phi_{j-1} - \phi_j\right) \end{aligned}$$

we arrive at (3.1) again.

§ 3.4 The Hamilton equations are particular cases of the Liouville equation

$$\dot{F} = \{F, H\},$$

which governs the evolution of a general dynamical function $F(\phi, \pi)$. The curly bracket is the Poisson bracket:

$$\{A, B\} = \sum_j \left[\frac{\partial A}{\partial \phi_j} \frac{\partial B}{\partial \pi_j} - \frac{\partial A}{\partial \pi_j} \frac{\partial B}{\partial \phi_j} \right].$$

The only nonvanishing Poisson brackets involving the degrees of freedom and their momenta are $\{\phi_i, \pi_j\} = \delta_{ij}$.

We have thus the three main approaches to the one-dimensional system of coupled oscillators, whose complete description requires the knowledge of all the displacements ϕ_i with respect to the equilibrium positions $\phi_i = 0$. Two questions remain: (i) the Lagrangian (3.3) is not as yet completely specified, as the summation requires the knowledge of ϕ_0 and ϕ_{N+1}; (ii) the physical problem is not well characterized, as the boundary conditions are missing. Both problems are solved by taking periodic conditions: $\phi_{i+N} = \phi_i$. This corresponds to making the extremities join each other ($N = 0$). In fact, we had been cheating a bit when we wrote Eq.(3.3). As the summation runs from $i = 1$ to $i = N$, that expression only acquires a meaning *after* a periodic condition $\phi_0 = \phi_N$, $\phi_{N+1} = \phi_1$ is imposed.

§ 3.5 The degrees of freedom are coupled in Eqs. (3.1). To solve the equations, it is highly convenient to pass into the system of normal coordinates, in terms of which the degrees of freedom decouple from each other. Such coordinates $\tilde{\phi}_i$, and their conjugate momenta $\tilde{\pi}_i$, will be such that

$$\phi_j = \tfrac{1}{\sqrt{N}} \sum_{n=-N/2}^{N/2} e^{i\frac{2\pi n j}{N}}\, \tilde{\phi}_n \quad ; \quad \pi_j = \tfrac{1}{\sqrt{N}} \sum_{n=-N/2}^{N/2} e^{i\frac{2\pi n j}{N}}\, \tilde{\pi}_n \ . \tag{3.5}$$

These expressions can be inverted by using the Kronecker identity

$$\frac{1}{N}\sum_{n=1}^{N} e^{i\frac{2\pi n(j-j')}{N}} = \delta_{jj'} \tag{3.6}$$

to give

$$\tilde{\phi}_m = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} e^{-i\frac{2\pi nm}{N}}\,\phi_n \quad ; \quad \tilde{\pi}_m = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} e^{-i\frac{2\pi nm}{N}}\,\pi_n \ . \tag{3.7}$$

Though we have passed from the real ϕ_i, π_i to the complex variables $\tilde{\phi}_i$, $\tilde{\pi}_i$, the number of independent variables is the same because $\tilde{\phi}_{-j} = \tilde{\phi}_j^*$ and $\tilde{\pi}_{-j} = \tilde{\pi}_j^*$. We find easily that the only nonvanishing Poisson brackets can be summed up as $\{\tilde{\phi}_i, \tilde{\pi}_j^*\} = \delta_{ij}$. It is not difficult to show, with the help of (3.6), that

$$\sum_{j=1}^{N} {\pi_j}^2 = \sum_{j=-N/2}^{N/2} \tilde{\pi}_j^*\,\tilde{\pi}_j \ ,$$

and

$$\frac{K}{2}\sum_{j=1}^{N}(\phi_j - \phi_{j-1})^2 = \frac{m}{2}\sum_{j=-N/2}^{N/2} \tilde{\phi}_j\tilde{\phi}_j^* \left[4\omega^2 \sin^2\frac{\pi j}{N}\right] \ .$$

Consequently, in normal coordinates, the Hamiltonian function reads

$$H = \tfrac{1}{2}\textstyle\sum_{n=-N/2}^{N/2}\left[\frac{\tilde{\pi}_n^*\,\tilde{\pi}_n}{m} + m\left(2\omega\sin\frac{\pi n}{N}\right)^2 \tilde{\phi}_n^*\,\tilde{\phi}_n\right] \ . \tag{3.8}$$

We see that the new oscillators, with frequencies

$$w_k = 2\omega\sin\frac{\pi k}{N} \tag{3.9}$$

instead of ω, decouple entirely from each other. They are, of course, the normal modes of the system. The equations of motion become simply

$$\ddot{\tilde{\phi}}_j + {\omega_j}^2\,\tilde{\phi}_j = 0 \tag{3.10}$$

with solutions

$$\tilde{\phi}_n(t) = \tilde{\phi}_n(0)\cos\omega_n t + \frac{1}{\omega_n}\,\dot{\tilde{\phi}}_n(0)\sin\omega_n t \ . \tag{3.11}$$

§ 3.6 The normal modes of vibration of the system are thus the Fourier components of the degrees of freedom. They are "collective" degrees of freedom, in the sense that each mode contains information on all the original degrees of freedom. On the space of degrees of freedom, they are "global".

The oscillation frequency (3.9) is such that

$$\begin{aligned} -2\omega \le \omega_k \le 0 \ \text{ for } \ k \ \text{ in the interval } \ -N/2 \le k \le 0 \ ; \\ 2\omega \ge \omega_k \ge 0 \ \text{ for } \ k \ \text{ in the interval } \ N/2 \ge k \ge 0 \ . \end{aligned}$$

In consequence, the decomposition (3.5) can be separated into positive- and negative-frequency parts:

$$\begin{aligned}\phi_j &= \tfrac{1}{\sqrt{N}} \left[\sum_{k=-N/2}^{-1/2} + \sum_{k=1/2}^{N/2} \right] e^{i\ \frac{2\pi kj}{N}}\ \tilde{\phi}_k \\ &= \tfrac{1}{\sqrt{N}} \sum_{k=1/2}^{N/2} \left[e^{i\frac{2\pi kj}{N}}\ \tilde{\phi}_k + e^{-i\frac{2\pi kj}{N}}\ \tilde{\phi}_{-k} \right] \ . \end{aligned} \tag{3.12}$$

For each one of these oscillators, we can introduce new variables a_k and a_k^* as

$$a_k = \frac{1}{\sqrt{2m\omega_k}} \left[m\omega_k \tilde{\phi}_k + i\tilde{\pi}_k \right] \tag{3.13}$$

$$a_k^* = \frac{1}{\sqrt{2m\omega_k}} \left[m\omega_k \tilde{\phi}_k^* - i\tilde{\pi}_k^* \right] \ . \tag{3.14}$$

As $a_k^* \neq a_{-k}$, the total number of variables remains the same. The only nonvanishing Poisson brackets are now $\{a_i, a_j^*\} = -\,i\,\delta_{ij}$. The equations of motion become $\dot{a}_k = -\,i\,\omega_k\,a_k$, with solutions

$$a_k(t) = e^{-i\omega_k t} a_k(0). \tag{3.15}$$

Both a_k and a_{-k}^* will have frequencies with the same sign. It is convenient to redefine ω_k as the positive object $\omega_k = 2\omega|\sin(\pi k/N)|$ and take the mass $m = 1$. Once this is made,

$$\tilde{\phi}_k = \frac{1}{\sqrt{2\omega_k}} \left[a_k + a_{-k}^* \right] \ .$$

Thus, $\tilde{\phi}_k$ has only contributions of frequencies with the same sign. If we establish by convention that these frequencies are to be called positive, it is easy to see that $\tilde{\phi}_{-k} = \tilde{\phi}_k^*$ will only have contributions of negative frequencies. Let us substitute $\tilde{\phi}_k$ in (3.5):

$$\phi_j(t) = \tfrac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2} \frac{1}{\sqrt{2\omega_k}} \left[e^{i2\pi kj/N}\ a_k + e^{-i2\pi kj/N}\ a_k^* \right] \ . \tag{3.16}$$

§ 3.7 We now intend to change into the continuum case. There is a well- known recipe to do it, inspired in the trick to go from finite to continuum Fourier transforms. We know where we want to arrive at, and we find a procedure to accomplish it. But it should be clear that it is only that — a practical recipe.

The prescription goes as follows. First, we take the limits $a \to 0$ and $N \to \infty$ simultaneously, but in such a way that the length value $L = Na$ remains finite. On the same token, each intermediate label value k tends to infinity in a way such that the product ka retains a finite value (the "distance to the origin"). We call x

this value, $x = ka$. Combining dimensional reasons and the necessity to keep finite the kinetic energy, the summation $\sum_{n=1}^{N}$ and the degree of freedom ϕ must behave like

$$\sum_{n=1}^{N} \to \frac{1}{a}\int_0^L dx \quad ; \quad \phi_j \to \sqrt{a}\,\phi(x)\;. \tag{3.17}$$

We have then that

$$\frac{1}{a}\left(\phi_{j+1}-\phi_j\right) = \sqrt{a}\,\frac{\phi(x+a)-\phi(x)}{a} \to \sqrt{a}\,\frac{\partial\phi}{\partial x}\;.$$

The equation of motion (3.1) will now be

$$\ddot{\phi} = \omega^2 a^2\,\frac{1}{a}\left[\left(\frac{\partial\phi}{\partial x}\right)_x - \left(\frac{\partial\phi}{\partial x}\right)_{x-a}\right]$$

and, in the limit, turns up as the wave equation

$$\frac{\partial^2\phi}{\partial t^2} = c^2\,\frac{\partial^2\phi}{\partial x^2}\;, \tag{3.18}$$

with the parameter $c = \omega a$ as the velocity of wave propagation. The finiteness of c requires that the frequency ω become infinite. The Lagrangian density becomes

$$\mathcal{L}[\phi] = \frac{1}{2}\int_0^L dx\; m\left[\dot{\phi}^2 - c^2\left(\frac{\partial\phi}{\partial x}\right)^2\right]\;, \tag{3.19}$$

corresponding to the action

$$S[\phi] = \int_0^T dt\mathcal{L}[\phi] = \frac{1}{2}\int_0 dtdx\; m\left[\dot{\phi}^2 - c^2\left(\frac{\partial\phi}{\partial x}\right)^2\right]\;. \tag{3.20}$$

To obtain a continuum–infinite version of equations (3.5), it is enough to take

$$k = \frac{2\pi l}{L}\;; \quad \frac{2\pi}{L}\sum_{l=-N/2}^{N/2} \to \int dk\;; \quad \frac{L}{2\pi}\,\tilde{\phi}_k \to \tilde{\phi}_k\;, \tag{3.21}$$

arriving at

$$\phi(x) = \frac{1}{\sqrt{L}}\int dk\; e^{ikx}\,\tilde{\phi}_k \quad ; \quad \pi(x) = \frac{1}{\sqrt{L}}\int dk\; e^{ikx}\,\tilde{\pi}_k\;. \tag{3.22}$$

There are two different kinds of summation limits, and differences in the absorption of factors between ϕ and $\tilde{\phi}$, as seen in (3.17) and (3.21). The degrees, and their Fourier components, acquire dimensions in the process.[†]

[†] The function $\exp[ikx]$ is typically a wave which repeats itself when $x = 2\pi n/k$, with $n = \pm 1, 2, 3, \ldots$. As the number of times the wave repeats itself in a cycle of length 2π, k is called the wave-number. As an exponent can have no dimension, k must have dimension inverse to x.

The continuum limit of (3.16), with $\omega_k \to k\,\omega\,a \to kc \doteq \omega_k$ (the last dot–equality indicating a redefinition of the symbol ω_k), is

$$\phi(x,t) = \int \frac{dk}{\sqrt{2L\omega_k}} \left[e^{i(kx-\omega_k t)}\, a_k(0) + e^{-i(kx-\omega_k t)}\, a_k^*(0) \right] \ . \tag{3.23}$$

§ 3.8 This example may seem a parenthesis a bit too long. We are here using the vibrating line only as a suggestive illustration,[‡] which anticipates many points of interest. It gives due emphasis to the meaning of the position coordinate x, transmutation of the old label i: it should not be mistaken by a generalized coordinate. It is a parameter, appearing in the argument of the field $\phi(x,t)$ on equal footing with the time parameter. This is manifest in the fact that the equation of motion (3.18) does not come from (3.19) as the usual Lagrange equation,

$$\frac{\delta \mathcal{L}}{\delta \phi} = \frac{\partial \mathcal{L}}{\partial \phi} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\phi}} = 0 \ , \tag{3.24}$$

but through its covariant form,

$$\frac{\delta \mathcal{L}}{\delta \phi_i} = \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi_i} = \frac{\partial \mathcal{L}}{\partial \phi_i} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{\phi}_i} + \frac{\partial}{\partial \vec{x}} \cdot \frac{\partial \mathcal{L}}{\partial \vec{\partial} \phi_i} \ . \tag{3.25}$$

Comment 3.1 The expression in (3.24) is the Lagrange derivative of Classical Mechanics when the degree of freedom and its first time derivative are enough to fix the problem. When higher-order derivatives are necessary, the derivative is

$$\frac{\delta}{\delta q^k} = \frac{\partial}{\partial q^k} - \frac{d}{dt}\frac{\partial}{\partial \dot{q}^k} + \frac{d^2}{dt^2}\frac{\partial}{\partial \ddot{q}^k} - \dots \ , \tag{3.26}$$

with alternating successive signs. This derivative takes into account the effect of a coordinate transformation on the time derivatives. It is the first example of a covariant derivative: $\frac{\delta}{\delta q^k} F(q,\dot{q},\ddot{q})$ transforms just as $F(q,\dot{q},\ddot{q})$ under coordinate transformations.

We could say further that the equation of motion endures a "covariantization", a process by which the time variable t and the "space" variable x acquire a similar status. As in Special Relativity, they become coordinates on one same space, "spacetime". One aspect of this effect is the change undergone by the action, from (3.4) to (3.20). In the latter, space and time parameters are equally integrated over. And the Lagrange derivative changes accordingly. There is actually more: the above wave equation is not invariant under transformations analogous to those of non- relativistic Physics, the Galilei transforms $x' = x - vt$ and $t' = t$. It is invariant

[‡] It has, of course, an interest by itself. Deeper developments to can be found in A. Askar, *Lattice Dynamical Foundations of Continuum Theories*, World Scientific, Singapore, 1985.

under transformations analogous to those of relativistic Physics, alike to Lorentz transformations: $x' = \gamma(x - vt)$ and $t' = \gamma(t - vx/c^2)$ with $\gamma = (1 - v^2/c^2)^{-1/2}$.

Let us insist: we have defined a classical field as a continuous infinity (labeled by x) of degrees of freedom, each one described by a function. The generalized coordinates are the very fields $\phi(x)$, one for each value of x. Thus, the parameter x spans the space of degrees of freedom. The equation of motion is of the kind to be found later, governing relativistic fields. It is usual to call it "field equation" in the continuum case.

One further remark: the decomposition into Fourier components is extensively used in canonical field quantization. In that case, it is the normal modes which are quantized as oscillators, leading to the quantization rules for the fields themselves. There is actually a further, usually overlooked, proviso: in order to qualify as a field, the infinite variables describing the system must be really at work. It may happen that many, or most of them, are quiescent. In the above use of Fourier analysis, this would show up if most of the modes were not actually active ($\tilde{\phi}_n = 0$ for many values of n). We might, in that case, talk of a quiescent field. An example is the field of fluid velocities in a laminar flow, for which most of the Fourier components are not active. They become active at the onset of turbulence.[§]

3.2 Non-Material Fields

§ 3.9 The vibrating line involves a *material* field, a field of displacements describing the mechanics of a would-be 1-dimensional elastic medium. Material fields abound, for example, in Fluid Mechanics of inhomogeneous media: the density, the concentration of each type of constituent, the velocity, the local pressure and temperature in the inhomogeneous case, and so on. The concept of field extends to non-mechanical systems (examples: the electric field, the gravitational field, ...) exhibiting a continuous infinity of degrees of freedom. In that case, the notion of field is actually inevitable for the description of interactions, as the alternative — the description of interactions by action at a distance — has never been given a simple, satisfactory formulation. Though implicit in the work of Galilei and Newton, the notion has been explicitly and systematically used by Faraday and has led to the complete description of the classical electromagnetic phenomena synthesized in Maxwell's equations.

In Wave Mechanics, the state of a system is characterized by a wavefunction $\psi(\vec{x}, t)$ (or better, by the ray to which $\psi(\vec{x}, t)$ belongs in Hilbert space), and its time

[§] Arnold Sommerfeld, *Mechanics of Deformable Bodies*, Academic Press, New York, 1967 (mainly section 38).

evolution is ruled by the prototype of nonrelativistic wave equation, the Schrödinger equation. The wavefunction must be known at each point of spacetime, so that the system requires a continuous infinity of values to be described. Function $\psi(\vec{x},t)$ has, consequently, the role of a field. It is usual to obtain it from Classical Mechanics by using the so-called quantization rules, by which classical quantities become operators acting on the wavefunction. Depending on the "representation", some quantities become differential operators and other are given by a simple product. Thus, the above $\psi(\vec{x},t)$ corresponds to the configuration–space representation, in which the Hamiltonian and the 3-momenta are given by

$$H \to i\,\hbar\,\frac{\partial}{\partial t}\;; \tag{3.27}$$

$$\vec{p} \to \frac{\hbar}{i}\,\vec{\nabla}\;, \tag{3.28}$$

and $\vec{x}$ is the operator acting on $\psi(\vec{x},t)$ according to $\psi(\vec{x},t) \to \vec{x}\;\psi(\vec{x},t)$.[¶] In the case of a free particle, in which $H = p^2/2m$, these rules lead to the free Schrödinger equation

$$i\hbar\,\frac{\partial}{\partial t}\,\psi(\vec{x},t) = -\,\frac{\hbar^2}{2m}\,\vec{\nabla}^2\,\psi(\vec{x},t)\;. \tag{3.29}$$

We shall from now on avoid the use of arrows, indicating 3–vectors by bold–faced characters. For example, the 3–momentum will be written $\mathbf{p} = -\,i\,\hbar\nabla$.

3.2.1 Optional reading: the Quantum Line

§ 3.10 Another version[‖] of the rules is summarized in the prescription relating Poisson brackets to commutators: $i\hbar\{\ ,\ \} \to [\ ,\ ]$. Thus, the oscillators of section 3.1 can be quantized by taking the variables a_i, a_j^* into operators a_i, $a_j^\dagger$ annihilating and creating quanta, and going from $\{a_i, a_j^*\} = -i\;\delta_{ij}$ into $[a_i, a_j^\dagger] = \;\delta_{ij}$. Let us recall a few elementary facts on the quantum harmonic oscillator. It is ruled by the Hamiltonian operator

$$\hat{H} = \tfrac{1}{2}\left(m\omega^2\hat{q}^2 + \frac{1}{m}\;\hat{p}^2\right).$$

[¶] These rules have been of great help in guessing most of the basic facts of Quantum Mechanics. But it should be clear that they are only guides, not immune to ambiguities. For example, they must be changed in the presence of spin.

[‖] This paragraph is, as the title above announces, optional. It supposes some knowledge of Quantum Mechanics and introduces some notions of the so-called "second quantization" formalism, which lies outside the scope of the present course.

We change into simpler dimensionless variables $\hat{Q} = \sqrt{\frac{m\omega}{\hbar}}\ \hat{q}$ and $\hat{P} = \frac{1}{\sqrt{m\omega\hbar}}\ \hat{p}$, such that $[\hat{Q}, \hat{P}] = i$. In the configuration space representation, $\hat{H}$ takes the form

$$\hat{H} = \tfrac{1}{2}\ \hbar\omega \left(Q^2 - \frac{d^2}{dQ^2} \right)$$

and has eigenfunctions $\psi_n(Q)$ with eigenvalues $E_n = (n + \frac{1}{2})\hbar\omega$. One passes into the occupation-number representation by defining annihilation and creation operators, respectively

$$\hat{a} = \tfrac{1}{\sqrt{2}}\ \left(\hat{Q} + i\hat{P} \right) = \tfrac{1}{\sqrt{2}} \left(Q + \ \frac{d}{dQ} \right)$$

and

$$\hat{a}^\dagger = \tfrac{1}{\sqrt{2}}\ \left(\hat{Q} - i\hat{P} \right) = \tfrac{1}{\sqrt{2}} \left(Q - \frac{d}{dQ} \right) ,$$

as well as the occupation-number operator $\hat{N} = \hat{a}^\dagger \hat{a}$. We find immediately that $[\hat{a}, \hat{a}^\dagger] = 1$ and $\hat{H} = \hbar\omega(\hat{N} + \frac{1}{2})$.

One next introduces a Fock space, generated by the set $\{|n>\}$ of eigenkets $|n>$ of $\hat{N}$. These kets are normalized in such a way that $\hat{a}|n> = \sqrt{n}\,|n-1>$; $\hat{a}^\dagger|n> = \sqrt{n+1}\,|n+1>$; $\hat{N}|n> = n\,|n>$. The number n is interpreted as the number of quanta. The state with zero quanta $|0>$, such that $\hat{a}|0> = 0$ or $\hat{N}|0> = 0$, is that with the minimum energy and is called the "vacuum". Its energy, by the way, is not zero — it is $\hat{H}|0> = (1/2)\hbar\omega|0>$. Each state $|n>$ can be obtained from the vacuum by creating n quanta: $|n> = (1/\sqrt{n!})\ (\hat{a}^\dagger)^n|0>$. Thus, this occupation-number representation describes the oscillator in terms of excitation quanta. When there is only one oscillator, all the quanta are identical — they are characterized by the energy, which is the same for all. From these states one can pass to other representations: usual wavefunctions in configuration space are $\psi_n(x) = <\ x\,|\,n>$, wavefunctions in momentum space are $\psi_n(p) = <p\,|\,n>$, etc.

Comment 3.2 Turning again to the problem of coordinates on phase space: in Quantum Mechanics, $\{q^k\}$ and $\{p_k\}$ become operators, represented by matrices. To specify a matrix one needs *all* its entries, which are, in the case, infinite.

All this can be generalized to a system with an arbitrary number of independent oscillators. It is enough to add a label to distinguish them. The kets, in that case, will be multiple. Annihilation and creation operators are defined for each oscillator, and a general state will be

$$|n_1, n_2, n_3, \cdots> = \frac{1}{\sqrt{n_1!\, n_2!\, n_3! \ldots}}\ (\hat{a}_1^\dagger)^{n_1}\ (\hat{a}_2^\dagger)^{n_2}\ (\hat{a}_3^\dagger)^{n_3} \ldots |0, 0, 0, \cdots> \ .$$

The total number of quanta in a given state will be the eigenvalue of the operator $\hat{N} = \sum_i \hat{a}_i^\dagger\ \hat{a}_i$.

The vibrating line has a continuous infinity of oscillators, each one characterized by the momentum k. The Fock space will consist of a continuous infinity of kets, collectively indicated, for example, by $|\{n_k\}>$. The field $\phi(x,t)$ in (3.23) becomes consequently also an operator, which is the quantized field of the material line:

$$\begin{aligned}\phi(x,t) &= \int \frac{dk}{\sqrt{2L\omega_k}} \left[e^{i(kx-\omega_k t)}\, a_k(0) + e^{-i(kx-\omega_k t)}\, a^\dagger{}_k(0)\right] \\ &= \int \frac{dk}{\sqrt{2L\omega_k}} \left[e^{ikx}\, a_k(t) + e^{-ikx}\, a^\dagger{}_k(t)\right] \ . \end{aligned} \tag{3.30}$$

In the last step use has been made of (3.15). Applied to the vacuum $|\{n_k = 0\}>$, this field spans the one-quantum states:

$$\phi(x,t)|0> = \int \frac{dk}{\sqrt{2L\omega_k}} \exp\left[-i(kx-\omega_k t)\right] |0,0,\ldots,1_k,\cdots> \ ,$$

where the ket in the integrand indicates the state with one quantum of momentum k. Well, the vibrating line stands for a system of material mechanic oscillators, necessarily quantal. This "quantum line" is the simplest example of a quantum field.

A simple direct computation gives the complete, collective quantization rules, including all the degrees of freedom:

$$\begin{aligned}[\phi(x,t), \phi(x',t)] &= 0 \ ; \\ [\pi(x,t), \pi(x',t)] &= 0 \ ; \\ [\phi(x,t), \pi(x',t)] &= i\,\hbar\,\delta(x-x') \ . \end{aligned} \tag{3.31}$$

3.3 Wavefields

§ 3.11 Wavefunctions are precisely the kind of fields we shall be most concerned with. The word "classical" acquires here a more precise meaning: the field $\psi(\mathbf{x},t)$ will be classical as long as it is an usual function, whose values are classical (real or complex) numbers (for short, "c–numbers"). In that case, it will belong to spaces on which hold the same algebraic properties of the complex numbers. Field $\psi(x)$ $= \psi(\mathbf{x},t)$ will no more be classical when it belongs to function spaces with more involved algebras (for example, when not all fields commute with each other) and can be no more treated as an ordinary function. It is the case of quantized fields, which are certain functionals or distributions, inhabiting spaces with non-trivial (though well-defined) internal algebras. Thus, as it stands, "classical" here means merely "non-quantum". The standard procedure begins with classical fields and proceeds to quantize them by changing their algebras. And the general structure,

be it Lagrangian or Hamiltonian, is transferred to the quantum stage, so that the preliminary study of classical fields is inescapable. It would be highly desirable to have a means of getting at the quantum description of a system without the previous knowledge of its classical description, not the least reason being the possible existence of purely quantum systems with no classical counterpart. Or with many of them, as there is no reason to believe that the classical limit of Quantum Mechanics be unique. This would avoid the intermediate procedure of "quantization". For the time being, no such course to a direct quantum description is in sight.

The fields appearing in relativistic field theory are not, in general, of the material type seen above. For that reason we shall not have the same phenomenological, immediate intuition which has conducted us to conceive, in the example of the vibrating line, the harmonic oscillator as a first reasonable trial. In other words, the access to dynamics is far more difficult. Phenomenology gives information of a general nature on the system, basically its symmetries. Symmetries, though important also for non-relativistic systems, become the one basic tool when the energies involved are high enough to impose the use of a relativistic approach.

A very important fact is that *every symmetry of the Lagrangian is also a symmetry of the equations of motion*. The procedure for high energies is rather inverse to that used for a mechanical system. For the latter, in general, the equations of motion are obtained phenomenologically and the Lagrangian leading to them (if existent) is found afterwards. In relativistic Field Theory, most commonly we start from a Lagrangian which is invariant under the symmetries suggested by the phenomenological results, because then the field equations will have the same symmetries. In particular, the Lagrangian involving relativistic fields will be invariant under the transformations of the Poincaré group. This is imposed by Special Relativity: the behavior of the system does not depend on the inertial frame used to observe it. That is where the adjective "relativistic" comes from. Furthermore, depending on the system, other symmetries can be present, some of them "external" (parity, conformal symmetry, ...), other "internal" (isospin, flavor, color, ...). The Lagrangian approach is specially convenient to account for symmetries, and is more largely used for that reason. It will be dominating in this text.

3.4 Internal Transformations

§ 3.12 Let us repeat ourselves. Besides those related to transformations taking place on spacetime, Nature exhibits other kinds of symmetries. As spacetime is taken to be our *external* space, such symmetries are said to be *internal*. They are

taken into account by supposing the existence of other spaces to which fields belong and on which transformations are represented in a way as analogous as possible to those on spacetime.

The general arguments on symmetries and invariance of the Lagrangian extend to these symmetries. If the theory is invariant under such internal transformations (say, symmetries related to the conservation of isospin, flavor, color, etc), the Lagrangian will be invariant and the fields will necessarily belong to representations of the corresponding groups. Fields invariant under any transformation of G, so that $\phi'(x) = \phi(x)$, are supposed to stay in a singlet (0-dimensional) representation.

Consider then those transformations changing only the functional form of the fields. We shall consider only a very particular kind amongst all the possibilities of such changes: the fields will be supposed to have components in some "interior" space, and changes will be only combinations of these components. This means that such "interior" transformations will be supposed to be represented by linear representations of the symmetry groups. The fields will present indices related to such internal representations. A scalar field (meaning: a *Lorentz* scalar field) ϕ will, for example, appear as ϕ_a, the a indicating a direction in an internal carrier vector space. A gauge potential is a Lorentz 4-vector belonging also to a representation of the gauge group and will appear in the form $A^a{}_\mu$. There is here, of course, a physicists' bias. Physicists are used to calling "fields" the *components* of certain mathematical objects, and we shall not fight this long–established attitude. The main ideas are best introduced through an example. The simplest non-trivial example of internal symmetry is provided by isospin.

Isospin has been introduced by Heisenberg in the fifties to account for what was then called the "charge independence" of the strong interactions. The proton and the neutron had been observed to have identical strong interactions, and almost the same mass. The mass difference was supposed to be of electromagnetic origin. As long as we could consider strong interactions alone, and forget about electromagnetic interactions, they were one and the same particle. Or better: they were seen as components of a double wavefunction, a doublet like (2.25), called the "nucleon":

$$N = \begin{pmatrix} p \\ n \end{pmatrix} .$$

A pure proton would be the analogous to the spin up state; the neutron, the spin down. As all that had nothing to do with spin, the name (isobaric spin, later) isospin was coined. The Lagrangian describing strong interactions would be invariant under "internal" rotations, formally identical with the above described, but in another, "internal" space. All this talk about "internal" things (space, components,

wavefunctions) is only to help intuition. It refers only to behavior under changes in the functional form of the fields, changes independent of their arguments (that is, of spacetime). Coming back to isospin: we have said that the proton-neutron system, or the nucleon, was attributed isospin 1/2. The pions appear with 3 possible charges, but their strong interactions ignore that difference. Thus, they also exhibit charge independence and, by an analogy with the rotation group, were accommodated in an isospin = 1 representation. In this way an internal symmetry was revealed: particles were cased in carrier vector spaces, the symmetry says that transformations in those spaces were irrelevant to Physics. In the isospin case the group was supposed to be just $SU(2)$, and two representations were immediately known. A field without isospin would be cased in the scalar representation. This was the starting point of a very powerful method. Once particles (quanta of fields) were found experimentally, people tried to accommodate them into representations ("multiplets") of some group. One same multiplet for particles of close masses, different components for different charges. The strong interactions would not "see" the components, only each multiplet as a whole. It would in this way be "independent of charge". The isospin rotations are formally the same as given above, though without any spacetime realization. The indices refer to internal space. For Pauli matrices, another notation became usual: τ_1, τ_2 and τ_3 instead of σ_1, σ_2 and σ_3. A rotation like (2.24) in an isospin = 1/2 spinor representation, as that of the nucleon, will consequently be written

$$N'(x) = e^{\frac{i}{2}\omega^k \tau_k}\ N(x)\ . \tag{3.32}$$

§ 3.13 Now: it may happen that the internal transformation parameters be different at different points of spacetime. When the transformation parameters ω^a are independent of spacetime points, the above transformation is called a *global gauge transformation* (old name: gauge transformations of the first kind). When ω^a depend on the point, the transformation is a *local gauge transformation* (old name: gauge transformations of the second kind). The formalism would remain much the same, except for the derivatives. It is clear that a derivative ∂_μ, when applied to

$$N'(x) = e^{\frac{i}{2}\ \omega^k(x)\tau_k}\ N(x)$$

will have an extra contribution:

$$\partial_\mu\ N'(x) = e^{\frac{i}{2}\omega^k(x)\tau_k}\ \partial_\mu N(x) + [\tfrac{i}{2}\ \partial_\mu\omega^k(x)\tau_k]\ e^{\frac{i}{2}\omega^k(x)\tau_k}\ N(x)\ .$$

Thus, usual derivatives will not be covariant. To find invariant Lagrangians it will be necessary to define a modified, *covariant derivative* including an extra field,

$$D_\mu = \partial_\mu + \tfrac{i}{2}A^k{}_\mu(x)\,\tau_k\ , \tag{3.33}$$

such that

$$[\partial_\mu + \tfrac{i}{2} A^{k\prime}{}_\mu(x)\tau_k]\; N'(x) = e^{\frac{i}{2}\omega^k(x)\tau_k}\; [\partial_\mu + \tfrac{i}{2} A^j{}_\mu(x)\tau_j]\; N(x)\;. \tag{3.34}$$

$A^k{}_\mu(x)$ is a *gauge potential*, with a peculiar behavior under transformations. This behavior will be just peculiar enough to compensate the misbehavior of the derivative:

$$A^{k\prime}{}_\mu(x)\frac{\tau_k}{2} = e^{\frac{i}{2}\;\omega^k(x)\tau_k}\;\; \left[A^k{}_\mu(x) - \partial_\mu\omega^k(x)\right]\frac{\tau_k}{2}\, e^{-\frac{i}{2}\;\omega^k(x)\tau_k}\;.$$

With $U(x) = e^{\frac{i}{2}\;\omega^k(x)\tau_k}$, this is the same as

$$\tfrac{i}{2} A^{k\prime}{}_\mu(x)\tau_k = U(x)\left[\partial_\mu + \tfrac{i}{2} A^k{}_\mu(x)\tau_k\right] U^{-1}(x)\;. \tag{3.35}$$

Exercise 3.1 Try to show that, if $A^k{}_\mu(x)$ transforms according to (3.35), then the derivative defined in (3.33) is indeed a covariant derivative, that is, Eq.(3.34) holds. This is why the derivative is called "covariant" derivative: applied to a field, it transforms just in the same way as the field itself. ■

The physical consequences are overwhelming. The presence of a point-dependent symmetry imposes the presence of new fields. These fields have been found to mediate most of the fundamental interactions of Nature. When related to external symmetries, they turn up as the Christoffel symbols in gravitation theory. For internal symmetries, they appear in two distinct families. As the gluon fields (with gauge group $SU(3)$), they are the mediating fields of chromodynamics, supposed to describe the strong interactions between the quarks. In the electroweak theory describing electrodynamics and weak interactions, they turn up, after a certain symmetry–breaking process, as the fields describing the photon, the Z^0 and the pair $W^\pm$. Mathematically, they are related to connections.

Chapter 4

General Formalism

Before we start with a general description of the Lagrangian formalism, it would be wise to assess what we can and what we cannot expect from it. For example, we shall be supposing a complete equivalence between Lagrange and Hamilton approaches and shall use variational principles without too much restraint. As it would not be practical to stop at every step to inquire on its validity, let us rather list once for all its main qualities and shortcomings. We shall use the language of Classical Mechanics, because it is simpler and more intuitive.

§ 4.1 Positive points of the Lagrangian formalism:

1. with respect to Newtonian mechanics, that of Lagrange is simpler: instead of vectors (forces), it works with scalars, the kinetic energy T and the potential V; forces, even those related to constraints, are no more fundamental;

2. the procedure is standardized: we write T and V in terms of generalized coordinates $\{q^j\}$, we form the Lagrangian $L = T - V$ and the equations of motion are those of Lagrange:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^j}\right) - \frac{\partial \mathcal{L}}{\partial q^j} = 0 \; ;$$

3. these equations hold in any system of generalized coordinates, while Newton's have different forms in each coordinate system, leading to the famous "fictitious" forces in the non–cartesian case; notice that, for systems with two or more degrees of freedom, it is the whole set of equations which is invariant under a change of system of generalized coordinates, and not each one;

4. it provides an extremely elegant approach to symmetries and to the dynamical invariants of a theory;

5. from the point of view of manipulations, we frequently gain from the fact that L is defined up to a temporal total derivative of a function of coordinates and time: $\mathcal{L}$ and
$$\mathcal{L}' = \mathcal{L} + \frac{df(q,t)}{dt}$$
lead to the same equations; two physical systems are equivalent when there is a coordinate transformation making their Lagrangian differ only by such a total derivative; in the relativistic case, systems described by Lagrangians differing by 4-divergences are equivalent;

6. the formalism is easy to extend; to the continuous case as illustrated in chapter (3); to the relativistic case as we shall see in the following; it applies to an unbelievable variety of systems: elastic fields, electromagnetism, elementary particles, electric circuits, solids, liquids, etc; as announced, we are using the language of Classical Mechanics, but these qualitative comments hold in general;

7. the structural analogy resultant from this unicity in the variety is extremely useful: once a particular procedure is found to be fruitful in a particular Lagrangian theory, its application to other cases suggests itself immediately;

8. the approach is globalized directly: the action functional contains information on the system as a whole, including boundary conditions; it lends itself directly to quantization via the integral functional methods.

§ 4.2 Negative points of the Lagrangian formalism:

1. going from the Newtonian to the Lagrangian formulation is only possible if

   the system is holonomic (that is, if all its constraints are integrable), or almost; *

   the forces are derived from a potential (also a bit generalizable);

   the constraints do not produce any work;

2. not every equation of motion can be obtained as the Euler–Lagrange equation of some Lagrangian function (famous counter–example: the Navier–Stokes equation) — in order to be, it must satisfy the conditions of a theorem due to Vainberg;

3. when the Lagrangian function does exist, it is not necessarily unique: the equation for a free particle, $\ddot{q} = 0$, comes from the Lagrangian $\mathcal{L}' = \dot{q} \ln \dot{q}$, besides the usual $\mathcal{L} = \dot{q}^2/2$; actually, it comes from any Lagrangian of the form $\mathcal{L} = f(\dot{q})$, provided $f'' \neq 0$;

4. the Lagrangian function has, as a rule, a lower degree of symmetry than the equations of motion; for example, $\ddot{q} = 0$ is invariant under the scaling transformation $q \to kq$, where k is a constant, but the unusual Lagrangian $\mathcal{L}'$ above is not;

5. qualifying item (8) above, the Lagrangian formulation does not lead to a unique quantization procedure; actually, even the equations of motion, which the present discussion may seem to suggest to be more fundamental, fail to determine a unique quantization procedure.

§ 4.3 From the Hamiltonian point of view, two physical systems are equivalent when there is a canonical transformation taking coordinates and momenta of one system into the coordinates and momenta of the other. The superiority of the Hamiltonian formulation rests precisely in its invariance under canonical transformations, more general than the generalized coordinate transformations. The Lagrangian and Hamiltonian formulations are not always equivalent — for them to be, in Classical Mechanics, it is necessary that the condition

$$\det \left[\frac{\partial^2 \mathcal{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right] \neq 0$$

* Some anholonomic cases are amenable to a Lagrangian treatment, but they are exceptional. Because the usual quantization procedure uses, directly or indirectly, the Lagrangian formulation, there are many constrained systems which we do not know how to quantize.

hold,[†] which forebodes difficulties in the zero–mass cases.

§ 4.4 Classical Mechanics is a very difficult science, with many questions as yet without answer. For instance, is there one formulation which is more fundamental? When they are not equivalent, which is the correct one? Classical Mechanics has even become a kind of *chasse gardée* for mathematical sophistication. Even seemingly simple systems, as that of two independent oscillators with incommensurate frequencies, can suggest matter for further research. Well, we cannot expect to solve every problem. The method of science — trial and error — leads anyhow to surprisingly many good answers. For example, we have a beautiful quantum description of the Helium atom, despite the problematic stability of the corresponding 3–body classical system. These words, to justify the rather unworried presentation which follows.

§ 4.5 The Hamiltonian formalism has another peculiarity, not mentioned in the previous discussion: it can be generalized to the case of fields while keeping a complete analogy with Classical Mechanics. This is frequently an advantage, but can become a hindrance in relativistic theories, because of the distinct role played by time with respect to space coordinates. The Lagrangian formalism, on the other hand, allows a explicitly covariant approach to the field equations: the four variables representing points of spacetime have an equal status. This simplifies life considerably and is the main reason for its wide use in relativistic field theories.

4.1 Lagrangian Approach

4.1.1 Relativistic Lagrangians

§ 4.6 Let us start thinking about the form a Lagrangian function should have. In a relativistic theory it should, to begin with, be invariant under the transformations of the Poincaré group. Invariance of a Lagrangian under a transformation ensures the covariance of the Euler–Lagrange equations under the transformation. This is the reason for which Lagrangians, in relativistic Field Theory, are supposed to be Poincaré invariant. The fields are the degrees of freedom, supposed to provide with their gradients (first–order derivatives with respect to space and time coordinates, corresponding to the "velocities") — a complete characterization of the system. A second–order derivative would lead to a third–order equation, and in

[†] This is Donkin's theorem.

that case the fields and velocities would be insufficient to describe the system. We take for Lagrangians, therefore, invariant functions of the fields and their gradients.[‡] Furthermore, our Lagrangians will have no explicit dependence on space or time coordinates, which are only parameters. These comments summarize the underlying spirit of Field Theory: *the state of the system is characterized by the fields and their first derivatives.* And, though we are heating up to a long discussion of special–relativistic fields, these considerations would hold for a non–relativistic (or Galilei–relativistic) theory, with the Galilei group taking the place of the Poincaré group. Another condition which we shall impose is that *the Lagrangian be real.* The reason is that no classical system has ever been found to suggest any kind of complex energy. In the quantum case, the Lagrangian should be self–adjoint, or simply hermitian. A non–hermitian Lagrangian would break probability conservation, thereby violating the scattering matrix unitarity. But it can be used in some cases (under the form of "optical potentials") to describe non–isolated systems. Next, we take a *Lagrangian density as simple as possible.* This is a rather loose condition. Of course, given an invariant $\mathcal{L}$, also $\mathcal{L}^{17}$, $\arctan(\mathcal{L})$ or any function of $\mathcal{L}$, are invariant. We take the simplest possible invariant functional of the fields and their derivatives leading to results confirmed by experiment. The theories of polynomial type, with the Lagrangian density a polynomial in the fields and their derivatives, are always the first trial. Models with non–polynomial terms are, however, increasingly studied. And a last condition, this one also dictated by simplicity. *The Lagrangian will be supposed to depend, at each point of spacetime, only on the values of the fields and their derivatives in an infinitesimal neighborhood of that point.* In this case, the Lagrangian — and the theory — is said to be "local". Non–local theories are in principle conceivable, but they are extremely complex and of a rather uncontrollable diversity. We shall ignore them.

§ 4.7 Thus, given a set of fields $\{\phi_i(x)\}$, the Lagrangian density will have the form

$$\mathcal{L}(x) = \mathcal{L}[\boldsymbol{\phi}(x), \partial_\mu \boldsymbol{\phi}(x)], \tag{4.1}$$

where we use the notations $\boldsymbol{\phi}(x) = \{\phi_1(x), \phi_2(x), \phi_3(x), \dots\}$ and $\partial_\mu \boldsymbol{\phi}(x) = \{\partial_\mu \phi_1(x), \partial_\mu \phi_2(x), \partial_\mu \phi_3(x), \dots\}$. Actually, as it is defined up to a gradient, the density has no need to be a complete invariant. Under a Poincaré transformation, it can acquire a gradient. In what follows, only its integral will be taken as invariant. From the density $\mathcal{L}(x)$, one could obtain directly the field equations, simply by generalizing the procedure of Classical Mechanics so as to abrogate the special status of the time

[‡] Function of fields mean function of the *forms* of the fields, of the way they depend on their arguments. Thus, what we have are actually functionals.

parameter $t = x^0/c$ with respect to the space coordinates. We shall, however, follow a more instructive method, arriving at the field equations through the mediation of a variational principle, Hamilton's principle. Here an adaptation of that of particle classical mechanics to the continuum case, it states that the action, defined by

$$A[\boldsymbol{\phi}] = \int d^4x \mathcal{L}(x) \ , \tag{4.2}$$

is minimal for all states actually occurring, that is, for all solutions of the equations of motion.

Notice that the action contains more on the system than the Lagrangian density or the equations. The integration in (4.2) covers the whole spacetime region occupied by the physical system to be described by the fields. Thus, the action includes information on this region. In this sense, it is a "global" characteristic of the system. In the case of particle mechanics, what is defined is the action of a trajectory γ,

$$A[\gamma] = \int_\gamma dt \, \mathcal{L}(\boldsymbol{x}, t) \ , \tag{4.3}$$

the integral being performed along γ. The action is thus a function (better: a functional) of the trajectory. It is global, it depends on the whole of γ. In the relativistic case, it becomes a functional of the "field configuration" $\boldsymbol{\phi}$ and depends on the domain of spacetime occupied by the system, whose boundaries are 3-dimensional surfaces assuming the role of γ's end-points. In terms of fields, this region is delimited by the boundary conditions, which are consequently incorporated in the action. In principle, the action contains all the conceivable information on the system.

4.1.2 Simplified Treatment

§ 4.8 Hamilton's principle states that, for an arbitrary variation of the degrees of freedom all over the system, the action stands fixed. The physical states (the solutions of the equations of motion) are characterized by minima of the action as a functional of the fields:

$$\delta A[\boldsymbol{\phi}] = 0 \ . \tag{4.4}$$

Let us be more clear: the action (4.2) does not depend on x, as it is an integral over x. $A[\boldsymbol{\phi}]$ depends actually on the functional form of $\boldsymbol{\phi}$. It is as a functional, on the space of which the $\boldsymbol{\phi}_k$'s are coordinates, that A is minimal. It is in the $\boldsymbol{\phi}$-space that A is to be differentiated, and the states actually realized correspond to the values which minimize A.

Let us start with the standard treatment leading to the Euler-Lagrange equations. The differential of A will be

$$\delta A = \delta \int d^4x \mathcal{L}(\boldsymbol{\phi}, \partial_\mu \boldsymbol{\phi}) = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi_i} \delta\phi_i + \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \delta\left(\partial_\mu \phi_i\right) \right] .$$

Notice that spacetime is unaffected by differentiation in functional space: the integration region is kept fixed. This is the covariant analogous to the classical mechanical procedure of taking variations of the trajectories at a fixed value of the time parameter "t".[§] The measure d^4x is also kept fixed. We shall see later that its variation, though important in other aspects, does not contribute to the Euler-Lagrange equations [see below, Eq.(4.29) and the discussion leading to (4.33)]. We shall also use, for that reason,

$$\delta(\partial_\mu \phi_i) = \partial_\mu(\delta\phi_i) \tag{4.5}$$

(we shall also elaborate on this point later – see comment below equation (4.24)). The last term in the right–hand side is

$$\int d^4x \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \delta\left(\partial_\mu \phi_i\right) = \int d^4x \left\{ \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \delta\phi_i \right] - \delta\phi_i \, \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \right] \right\} .$$

With the help of the 4-dimensional Gauss theorem, we have

$$\int d^4x \left\{ \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \delta\phi_i \right] \right\} = \int d\sigma^\alpha \frac{\partial \mathcal{L}}{\partial \left(\partial_\alpha \phi_i\right)} \delta\phi_i \, .$$

This term reduces consequently to an integration over a 3-dimensional hypersurface, the boundary of the integration region — or the boundary of the system. The variation $\delta\phi_i$ is arbitrary over all the interior of the system, but on the boundary we take $\delta\phi_i = 0$ (this is the covariant analogous to the classical mechanical procedure of taking null variations at the trajectories end–points). Then,

$$\delta A = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \right] \delta\phi_i \, . \tag{4.6}$$

As the $\delta\phi_i$ are arbitrary, $\delta A = 0$ implies the Euler–Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} = 0 \, . \tag{4.7}$$

§ 4.9 It is worth noticing that, as (4.6) holds always under the supposed conditions, the Euler–Lagrange equations imply an extremal of the functional (4.2). In functional analysis, expression (4.6) provides the definition of the *functional derivative*

[§] See for instance H. Goldstein, *Classical Mechanics*, Addison–Wesley, Reading, Mass., 1982.

of $A[\boldsymbol{\phi}]$ with respect to ϕ_i: it is the term between brackets in the integrand. We write

$$\frac{\delta A[\boldsymbol{\phi}]}{\delta \phi_i} = \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} \; . \tag{4.8}$$

In the particular case of variational calculus, this functional derivative is called *Lagrange derivative*. If $\mathcal{L}$ depends on higher–order derivatives of $\boldsymbol{\phi}$, the Lagrange derivative takes the form

$$\frac{\delta A[\boldsymbol{\phi}]}{\delta \phi_i} = \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi_i\right)} + \partial_\mu \partial_\nu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \partial_\nu \phi_i\right)} - \partial_\mu \partial_\nu \partial_\lambda \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \partial_\nu \partial_\lambda \phi_i\right)} + \cdots \tag{4.9}$$

with alternating successive signs. This is the functional version of (3.26). As this is one of the conditions imposed above, at least for relativistic fields we shall stop at the second term in the right–hand side.

Exercise 4.1 Scalar fields satisfy the Klein-Gordon equation:

$$\left(\Box + m^2\right) \, \phi(x) \; = \; 0 \; .$$

This equation comes from Eq.(1.101),

$$p^2 = m^2 c^2$$

by using the quantization rules (3.27) and (3.28). The d'Alembertian operator $\Box$ is simply the Laplace operator in 4-dimensional Minkowski space. In Cartesian coordinates $\{x^\alpha\}$, in terms of which the Lorentz metric is $\eta = diag(1, -1, -1, -1)$,

$$\Box = \eta^{\alpha\beta} \partial_\alpha \partial_\beta = \partial^\alpha \partial_\alpha = \partial_0 \partial_0 - \partial_1 \partial_1 - \partial_2 \partial_2 - \partial_3 \partial_3 \; .$$

Show that the Klein-Gordon equation comes out as the Euler-Lagrange equation of the Lagrangian

$$\mathcal{L} = \tfrac{1}{2} \, \left[\partial_\mu \phi \partial^\mu \phi \; - \; m^2 \phi^2\right] \; .$$

Show that it comes also from

$$\mathcal{L} = \tfrac{1}{2} \, \phi [\Box \; + \; m^2] \phi \; ,$$

taking into account terms up to second order in the Lagrangian derivative. ■

4.1.3 Rules of Functional Calculus

We intend now to introduce some formal devices, which both ease the manipulations and provide better understanding. They are — though presented here as expedients of a practical nature, without any pretense to rigor — the stepping stones of variational calculus. To do it, we shall show an alternative road to the Euler–Lagrange equation (4.7).

§ 4.10 Consider a point y *interior* to the domain on which the field and the system it describes is present, and the variation $\delta\phi(y)$ at that point. Instead of the functional *differential* of A previously used, we shall take the *derivative* of

$$A[\phi] = \int d^4x \mathcal{L}[\phi(x), \partial_\mu \phi(x)] \ .$$

We write

$$\frac{\delta A[\phi]}{\delta\phi(y)} = \int d^4x \left[\frac{\partial\mathcal{L}(x)}{\partial\phi(y)} + \frac{\partial\mathcal{L}(x)}{\partial(\partial_\mu\phi(x))}\frac{\delta(\partial_\mu\phi(x))}{\delta\phi(y)}\right] \ . \tag{4.10}$$

As the variation is well–defined, unique at each point, we shall agree to put

$$\frac{\delta\phi(x)}{\delta\phi(y)} = \delta^4(x-y) \ . \tag{4.11}$$

This can also be written

$$\delta\phi(x) = \int d^4x\, \delta^4(x-y)\delta\phi(y) \ .$$

This is to say that $\delta^4(x-y)$ is the functional derivative of ϕ with respect to itself. Recall the case of the vibrating line, where x takes the place of the index i and the differential $d\phi_i$ is replaced by the variation $\delta\phi(x)$. There is a clear analogy between (4.6), written in terms of the Lagrange derivative

$$\delta A[\phi] = \int d^4x \frac{\delta A[\phi]}{\delta\phi(x)}\,\delta\phi(x) \ , \tag{4.12}$$

and the usual differential of a several–variables function f,

$$df = \sum_i \frac{\partial f}{\partial x^i} dx^i \ .$$

Expression (4.11) is the analogue of the well–known relation

$$\frac{dx^i}{dx^j} = \delta^i_j \ ,$$

which holds for Cartesian coordinates. Notice also that the variation at point y has nothing to do with the variation at point x — they are variations of distinct degrees of freedom, so that

$$\frac{\delta}{\delta\phi(y)}\frac{\partial}{\partial x^\mu}\phi(x) = \frac{\partial}{\partial x^\mu}\frac{\delta}{\delta\phi(y)}\phi(x) = \frac{\partial}{\partial x^\mu}\delta^4(x-y) \ . \tag{4.13}$$

At this point, (4.10) can be written

$$\frac{\delta A[\phi]}{\delta\phi(y)} = \int d^4x \left[\frac{\partial\mathcal{L}(x)}{\partial\phi(x)}\delta^4(x-y) + \frac{\partial\mathcal{L}(x)}{\partial(\partial_\mu\phi(x))}\frac{\partial}{\partial x_\mu}\delta^4(x-y)\right] \ .$$

Let us take separately the last term: it is

$$\int d^4x\, \partial_\mu \left[\frac{\partial \mathcal{L}(x)}{\partial(\partial_\mu \phi(x))}\, \delta^4(x-y) \right] - \int d^4x \left[\partial_\mu \frac{\partial \mathcal{L}(x)}{\partial(\partial_\mu \phi(x))} \right] \delta^4(x-y) \; .$$

Using again the four–dimensional Gauss theorem, we see that the term with the total derivative is an integral on the boundary of the system, and — as y is in its interior — it vanishes. The delta factors out in all the remaining terms and

$$\begin{aligned} \frac{\delta A[\phi]}{\delta \phi(y)} &= \int d^4x \left[\frac{\partial \mathcal{L}(x)}{\partial \phi(x)} - \partial_\mu \frac{\partial \mathcal{L}(x)}{\partial(\partial_\mu \phi(x))} \right] \delta^4(x-y) \\ &= \frac{\partial \mathcal{L}(y)}{\partial \phi(y)} - \partial_\mu \frac{\partial \mathcal{L}(y)}{\partial(\partial_\mu \phi(y))} \; , \end{aligned} \tag{4.14}$$

just the expression turning up in the Euler–Lagrange equation (4.7). There is more than simple analogy in the steps taken above. They are justified by Functional Analysis and the whole derivation is essentially correct. Advanced calculations in Field Theory can be very involved and the functional methods, which simplify them, are more and more used. It is not always possible to prove the results in a quite rigorous way and they should, whenever feasible, be checked in some independent manner. Here, these procedures will be regarded as a kind of stenographic language. Rigorous proofs involve defining suitable topology and integration measure, besides verifying the convergence at each step.

4.1.4 Variations

§ 4.11 A physical system will thus be characterized as a whole by some symmetry-invariant action functional like (4.2),

$$A[\boldsymbol{\phi}] = \int d^4x \mathcal{L}[\boldsymbol{\phi}] \; , \tag{4.15}$$

where $\boldsymbol{\phi}$ represents collectively all the involved fields. Let us examine in some more detail the total variation[¶] of the action functional under the simultaneous change of the coordinates according to

$$x'^\mu = x^\mu + \delta x^\mu \; , \tag{4.16}$$

and of the fields according to

$$\phi'_i(x') = \phi_i(x) + \delta\phi_i(x) \; . \tag{4.17}$$

[¶] When said like that, without further specification, "variation" means the first–variation, or the first–order variation.

Let us first define the change *in the functional form* of $\phi_i(x)$,

$$\bar{\delta}\phi_i(x) = \phi'_i(x) - \phi_i(x) \; . \tag{4.18}$$

Consider

$$\phi'_i(x') = \phi'_i(x + \delta x) \approx \phi'_i(x) + \partial_\mu \phi'_i(x)\, \delta x^\mu \; .$$

The last term is

$$\partial_\mu \phi'_i(x)\, \delta x^\mu = \partial_\mu \phi_i(x)\, \delta x^\mu + \partial_\mu \bar{\delta}\phi_i(x)\, \delta x^\mu \approx \partial_\mu \phi_i(x)\, \delta x^\mu$$

to first order, so that

$$\phi'_i(x') = \phi'_i(x) + \partial_\mu \phi_i\, \delta x^\mu \; .$$

We can use (4.18) to write

$$\phi'_i(x') = \phi_i(x) + \bar{\delta}\phi_i(x) + \delta x^\lambda \, \frac{\partial \phi_i(x)}{\partial x^\lambda} \; . \tag{4.19}$$

or

$$\delta\phi_i(x) = \bar{\delta}\phi_i(x) + \delta x^\mu \, \partial_\mu \phi_i(x) \; . \tag{4.20}$$

The total variation of a field ϕ_i is, in this way, separated into two parts: the first term represent changes in the functional forms of the fields themselves, the second term the variation due to changes $\delta x^\mu = x'^\mu - x^\mu$ in the coordinates, that is, in the argument. By an iteration procedure we find, always retaining only terms to first order,

$$\frac{\partial x^\sigma}{\partial x'^\mu} = \delta^\sigma{}_\mu - \frac{\partial \delta x^\sigma}{\partial x'^\mu} = \delta^\sigma{}_\mu - \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial \delta x^\sigma}{\partial x^\rho} = \delta^\sigma{}_\mu - \left(\delta^\rho{}_\mu - \frac{\partial \delta x^\rho}{\partial x^\mu}\right)\frac{\partial \delta x^\sigma}{\partial x^\rho}$$

$$= \delta^\sigma{}_\mu - \frac{\partial \delta x^\sigma}{\partial x^\mu} \quad ,$$

and consequently

$$\frac{\partial}{\partial x'^\mu} = \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial}{\partial x^\sigma} = \left(\delta^\sigma{}_\mu - \frac{\partial \delta x^\sigma}{\partial x'^\mu}\right)\frac{\partial}{\partial x^\sigma} = \left(\delta^\sigma{}_\mu - \frac{\partial \delta x^\sigma}{\partial x^\mu}\right)\frac{\partial}{\partial x^\sigma} \; ,$$

which is

$$\frac{\partial}{\partial x'^\mu} = \frac{\partial}{\partial x^\mu} - \frac{\partial \delta x^\sigma}{\partial x^\mu}\frac{\partial}{\partial x^\sigma} \; . \tag{4.21}$$

The derivative $\partial/\partial x^\mu$ of expression (4.20) is

$$\partial_\mu \left[\delta\phi_i(x)\right] = \partial_\mu \left[\bar{\delta}\phi_i(x)\right] + \partial_\mu \left(\delta x^\lambda\right)\, \partial_\lambda \phi_i(x) + \delta x^\lambda \, \partial_\mu \partial_\lambda \phi_i(x) \; . \tag{4.22}$$

On the other hand, if we apply (4.21) to (4.19) we find, always retaining only the first order terms,

$$\delta\left[\partial_\mu \phi_i(x)\right] = \partial_\mu \, \bar{\delta}\phi_i(x) + \delta x^\lambda \, \partial_\mu \, \partial_\lambda \phi_i(x) \; . \tag{4.23}$$

Comparing this with (4.22), we arrive at the "commutator"

$$[\partial_\mu, \delta]\, \phi_i(x) = \left[\partial_\mu(\delta x^\lambda)\right]\, \partial_\lambda \phi_i(x)\ . \tag{4.24}$$

Thus, we can commute ∂_μ and δ only if the spacetime variation δx^λ is point–independent. This is the case of equation (4.5), because there the spacetime variable was kept fixed.

Concerning the purely functional variations,
(i) take the derivative of (4.20):

$$\partial_\mu \delta\phi_i(x) = \partial_\mu \bar{\delta}\phi_i(x) + \partial_\sigma \phi_i(x)\, \partial_\mu \delta x^\sigma + \delta x^\sigma\, \partial_\mu \partial_\sigma \phi_i(x)\ ;$$

(ii) apply (4.20) to the derivative function $\partial_\mu \phi_i(x)$:

$$\delta\partial_\mu \phi_i(x) = \bar{\delta}\partial_\mu \phi_i(x) + \delta x^\sigma\, \partial_\sigma \partial_\mu \phi_i(x)\ ;$$

(iii) take the difference of both expressions, using (4.24) to obtain the expected result

$$\bar{\delta}\partial_\mu \phi_i(x) = \partial_\mu \bar{\delta}\phi_i(x)\ .$$

The change in the functional form of the derivative of $\phi_i(x)$ is the derivative of the change in the functional form of $\phi_i(x)$. We can spell it in commutator form,

$$\left[\partial_\mu, \bar{\delta}\right]\, \phi_i(x) = 0\ . \tag{4.25}$$

Let us go back to the action functional (4.15). It does not depend on x. Its is a functional of the fields, depending on the integration domain. Variations (4.16) and (4.17) will have effects of two kinds: changes in the integration volume and in the Lagrangian density. We shall indicate this by writing

$$\delta A[\phi] = \int \left[\delta(d^4x)\mathcal{L} + d^4x\, \delta\mathcal{L}\right]\ . \tag{4.26}$$

To calculate the first term it is enough to recall that, given a coordinate transformation as (4.16), the change in the volume element is fixed by the Jacobian,

$$d^4x' = \left|\frac{\partial x'}{\partial x}\right|\, d^4x\ . \tag{4.27}$$

The Jacobian determinant $\left|\frac{\partial x'}{\partial x}\right|$ is

$$\begin{vmatrix} \frac{\partial x'^0}{\partial x^0} & \frac{\partial x'^0}{\partial x^1} & \frac{\partial x'^0}{\partial x^2} & \frac{\partial x'^0}{\partial x^3} \\ \frac{\partial x'^1}{\partial x^0} & \frac{\partial x'^1}{\partial x^1} & \frac{\partial x'^1}{\partial x^2} & \frac{\partial x'^1}{\partial x^3} \\ \frac{\partial x'^2}{\partial x^0} & \frac{\partial x'^2}{\partial x^1} & \frac{\partial x'^2}{\partial x^2} & \frac{\partial x'^2}{\partial x^3} \\ \frac{\partial x'^3}{\partial x^0} & \frac{\partial x'^3}{\partial x^1} & \frac{\partial x'^3}{\partial x^2} & \frac{\partial x'^3}{\partial x^3} \end{vmatrix} \approx \begin{vmatrix} \frac{\partial(x^0+\delta x^0)}{\partial x^0} & \frac{\partial(x^0+\delta x^0)}{\partial x^1} & \frac{\partial(x^0+\delta x^0)}{\partial x^2} & \frac{\partial(x^0+\delta x^0)}{\partial x^3} \\ \frac{\partial(x^1+\delta x^1)}{\partial x^0} & \frac{\partial(x^1+\delta x^1)}{\partial x^1} & \frac{\partial(x^1+\delta x^1)}{\partial x^2} & \frac{\partial(x^1+\delta x^1)}{\partial x^3} \\ \frac{\partial(x^2+\delta x^2)}{\partial x^0} & \frac{\partial(x^2+\delta x^2)}{\partial x^1} & \frac{\partial(x^2+\delta x^2)}{\partial x^2} & \frac{\partial(x^2+\delta x^2)}{\partial x^3} \\ \frac{\partial(x^3+\delta x^3)}{\partial x^0} & \frac{\partial(x^3+\delta x^3)}{\partial x^1} & \frac{\partial(x^3+\delta x^3)}{\partial x^2} & \frac{\partial(x^3+\delta x^3)}{\partial x^3} \end{vmatrix}$$

$$= \begin{vmatrix} 1+\frac{\partial\delta x^0}{\partial x^0} & \frac{\partial\delta x^0}{\partial x^1} & \frac{\partial\delta x^0}{\partial x^2} & \frac{\partial\delta x^0}{\partial x^3} \\ \frac{\partial\delta x^1}{\partial x^0} & 1+\frac{\partial\delta x^1}{\partial x^1} & \frac{\partial\delta x^1}{\partial x^2} & \frac{\partial\delta x^1}{\partial x^3} \\ \frac{\partial\delta x^2}{\partial x^0} & \frac{\partial\delta x^2}{\partial x^1} & 1+\frac{\partial\delta x^2}{\partial x^2} & \frac{\partial\delta x^2}{\partial x^3} \\ \frac{\partial\delta x^3}{\partial x^0} & \frac{\partial\delta x^3}{\partial x^1} & \frac{\partial\delta x^3}{\partial x^2} & 1+\frac{\partial\delta x^3}{\partial x^3} \end{vmatrix} .$$

Thus, to first order,

$$\left|\frac{\partial x'}{\partial x}\right| \approx 1 + \partial_\mu\left(\delta x^\mu\right) \ , \tag{4.28}$$

so that

$$\delta(d^4x) = \partial_\mu\left(\delta x^\mu\right)\, d^4x \ . \tag{4.29}$$

We have thus the first contribution to (4.26). The variation of the Lagrangian density will have two contributions, one coming from the coordinate variations and another coming from the variations in the functional form of the fields:

$$\delta\mathcal{L} = \left(\partial_\mu\mathcal{L}\right)\delta x^\mu + \bar{\delta}\mathcal{L} \ , \tag{4.30}$$

where

$$\bar{\delta}\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi_i}\,\bar{\delta}\phi_i + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\bar{\delta}(\partial_\mu\phi_i) \ . \tag{4.31}$$

The dependence of $\mathcal{L}$ on x comes exclusively through the fields — a basic hypothesis of Field Theory. Equation (4.25) authorizes commuting $\bar{\delta}$ and ∂_μ, leading to

$$\bar{\delta}\mathcal{L} = \frac{\delta\mathcal{L}}{\delta\phi_i}\,\bar{\delta}\phi_i + \partial_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\bar{\delta}\phi_i\right] \ . \tag{4.32}$$

with $\delta\mathcal{L}/\delta\phi_i$ the Lagrangian derivative.

Attention should be paid to the different notations $\partial\mathcal{L}/\partial\phi_i$ and $\delta\mathcal{L}/\delta\phi_i$. Notice that it would not be clear that we are allowed, while manipulating the second term in (4.31), to perform an integration by parts, as here also the integration boundaries are varying.

Finally, putting together (4.29), (4.30) and (4.32), we arrive at the expression of the action first–variation:

$$\delta A[\phi] = \int d^4x\left\{\frac{\delta\mathcal{L}}{\delta\phi_i}\,\bar{\delta}\phi_i + \partial_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\bar{\delta}\phi_i + \mathcal{L}\,\delta x^\mu\right]\right\} \ . \tag{4.33}$$

This is a most important formula, instrumental in our future derivations of the Noether theorems relating Lagrangian symmetries to conserved charges.

4.2 The First Noether Theorem

§ 4.12 Symmetries, we recall, are transformations leaving formally invariant the equations of motion. This means that, if applied to a solution of the equations,

they lead to another solution, compatible with the same boundary conditions. A sufficient condition for the equations to be invariant is, as we have said many times, that the Lagrangian be invariant.

When the equations of motion are obtained from a variational principle, like Hamilton's, it is possible to establish a relationship between the symmetries and the integrals of motion. This is the content of Noether's first theorem. In condensed form, this theorem says the following: to each transformation which leaves the action invariant corresponds a combination of the fields and their derivatives which is also invariant. The transformation can include changes in the coordinates and alterations in the fields, the latter being related to the former or not. Since their publication in 1918, the theorem has been subjected to many extensions and adaptations. We shall here present a version specially adapted to Field Theory.[‖]

The transformations will be supposed to be continuous and connected to the identity. For all that will concern us here, it will be sufficient to consider the first-order infinitesimal case. Equation (4.33) gives the response of the action functional to variations both in the fields and in the spacetime coordinates:

$$\delta A[\phi] = \int d^4x \left[\frac{\delta \mathcal{L}}{\delta \phi_i} \, \bar{\delta}\phi_i + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \, \bar{\delta}\phi_i + \mathcal{L} \, \delta x^\mu \right) \right]$$

From this expression will come the two Noether theorems we intend to study. The first will be concerned with global transformations, that is, with transformations which are the same at all points of the system. In other words, the transformation parameters (the group parameters) will be point-independent. The second theorem is concerned with transformations which change from point to point.

As we have seen in page 51, a well-defined behavior under transformations require that each field belong to some representation of the corresponding group. Suppose that T_a, for $a = 1, 2, \ldots, N =$ *group dimension*, are the generators in some matrix representation to which ϕ belongs, and that ϕ_i are the components. Then,

$$\phi'_i(x') = \left[\mathrm{e}^{\omega^a T_a} \right]_{ij} \phi_j(x) \tag{4.34}$$

will be the general form of the transformation undergone by ϕ. In the case of a global transformation the parameters ω^a will be constants; in the case of a local transformation, they will be point-dependent, $\omega^a = \omega^a(x)$. In the infinitesimal case, with very small parameters $\delta\omega^a$,

$$\phi'_i(x') \approx \left[\delta_{ij} + \delta\omega^a (T_a)_{ij} \right] \phi_j(x) \; ,$$

[‖] A very complete treatment can be found in N. P. Konopleva and V. N. Popov, *Gauge Fields*, Harwood Academic Plub., Chur, 1981.

$$= \phi_i(x) + \delta\omega^a (T_a)_{ij} \phi_j(x) \ . \tag{4.35}$$

In terms of the transformation parameters, variations (4.16) and (4.17) will be, to first order,

$$\delta x^\mu = \frac{\delta x^\mu}{\delta \omega^a} \delta\omega^a \ ; \tag{4.36}$$

$$\delta\phi_i(x) = \frac{\delta\phi_i(x)}{\delta\omega^a} \delta\omega^a \ . \tag{4.37}$$

The field variation at a fixed point x will be

$$\bar{\delta}\phi_i(x) = \left[\frac{\delta\phi_i(x)}{\delta\omega^a} - \frac{\partial\phi_i(x)}{\partial x^\mu} \frac{\delta x^\mu}{\delta\omega^a} \right] \delta\omega^a \ . \tag{4.38}$$

It is clear that, from (4.35),

$$\frac{\delta\phi_i(x)}{\delta\omega^a} = (T_a)_{ij}\phi_j(x) \ . \tag{4.39}$$

To get an idea on how to obtain $\frac{\delta x^\mu}{\delta\omega^a}$, consider again the particular case of a rotation around the third axis: in the infinitesimal case,

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} + \delta\omega^3 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \ ,$$

that is, $\frac{\delta x^\mu}{\delta\omega^3} = \mu$-th component of $(0, -x^2, x^1, 0)^T$. For example, $\frac{\delta x^2}{\delta\omega^3} = x^1$. This is, of course, consistent with the kinematic representation of equations (2.40, 2.41), in which

$$x'^\mu = x^\mu + \delta\omega^3 \left(x^1\partial_2 - x^2\partial_1 \right) x^\mu = x^\mu + \delta\omega^3 \left(x^1\delta^\mu_2 - x^2\delta^\mu_1 \right) \ .$$

4.2.1 Symmetries and Conserved Charges

§ 4.13 In search of the first Noether theorem, consider constant parameters $\delta\omega^a$ and take (4.36), (4.37) and (4.38) into (4.33). The action variation becomes

$$\delta A[\phi] = \int d^4x \left\{ \frac{\delta\mathcal{L}}{\delta\phi_i} \frac{\bar{\delta}\phi_i(x)}{\delta\omega^a} + \partial_\mu \left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)} \frac{\bar{\delta}\phi_i}{\delta\omega^a} + \mathcal{L} \frac{\delta x^\mu}{\delta\omega^a} \right] \right\} \delta\omega^a \ . \tag{4.40}$$

When the variation comes from a symmetry transformation, the action must remain invariant: $\delta A = 0$. The $\delta\omega^a$'s appear factorized. They are small, but totally arbitrary. Thus, in order that the integral vanish for any $\delta\omega^a$ it is necessary that

the integrand vanishes. In other words, action invariance under a global transformation imposes the vanishing of the derivative of A with respect to the corresponding constant (but otherwise arbitrary) parameter $\delta\omega^a$. The condition for that is

$$\frac{\delta\mathcal{L}}{\delta\phi_i}\,\frac{\bar{\delta}\phi_i(x)}{\delta\omega^a} = -\,\partial_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\frac{\bar{\delta}\phi_i}{\delta\omega^a} + \mathcal{L}\,\frac{\delta x^\mu}{\delta\omega^a}\right]\,. \tag{4.41}$$

This is the content of the historical Noether theorem: if the action is invariant under the transformations of an N-dimensional group, then there are N linear combinations of the Lagrangian derivatives which reduce to divergences.

§ 4.14 When the field ϕ_i is a solution of the Euler-Lagrange equation $\frac{\delta\mathcal{L}}{\delta\phi_i} = 0$, the current

$$J_a{}^\mu = -\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\frac{\bar{\delta}\phi_i}{\delta\omega^a} + \mathcal{L}\,\frac{\delta x^\mu}{\delta\omega^a}\right]\,. \tag{4.42}$$

will have vanishing divergence:

$$\partial_\mu J_a{}^\mu = 0\,. \tag{4.43}$$

There will be one conserved current for each group generator. Each will result in a conserved charge (that is, an integral of motion). To see this, take in spacetime a volume unbounded in the space-like directions, but limited in time by two space-like surfaces w_1 and w_2. Integrating (4.43) over this volume, we get an integral over the boundary surface, composed of w_1, w_2 and the time-like boundaries supposed to be at infinity. If we now suppose the current to be zero at infinity on these boundaries, we remain with

$$\int_{w_1} d\sigma_\mu\, J_a{}^\mu = \int_{w_2} d\sigma_\mu\, J_a{}^\mu\,. \tag{4.44}$$

This means that the integral $Q_a = \int_{w_n} d\sigma_\mu\, J_a{}^\mu$, taken over a space-like (hyper)surface w_n, is independent of which w_n one takes, provided the current vanishes at space infinity. In a more prosaic way: take an axis $x^0 = ct$ and as spaces the planes given by t = constant; then the integral will be the same on any such plane — will be time-independent. In effect, integrating (4.43) in d^3x,

$$\frac{d}{dx^0}\int d^3x J_a{}^0(x) = -\int_{space} d^3x\partial_i J_a{}^i(x) = -\int_{bound} d\sigma_i J_a{}^i(x) = 0\,. \tag{4.45}$$

Thus, to each group generator will correspond a "charge"

$$Q_a = \int d^3x J_a{}^0(\mathbf{x},t)\,, \tag{4.46}$$

which is conserved,

$$\frac{d}{dt}\,Q_a = 0\,. \tag{4.47}$$

§ 4.15 We shall see below, when we study the main Lagrangians, applications of all that. A few comments:

(i) the ("Noether") current is not unique; addition of the divergence of any antisymmetric tensor, $J_a{}^{\mu} \to J_a{}^{\mu} + \partial_\lambda A_a{}^{\mu\lambda}$, with $A_a{}^{\mu\lambda} = -A_a{}^{\lambda\mu}$, gives another conserved current (as $\partial_\mu \partial_\lambda A_a{}^{\mu\lambda} = 0$) and the charge will not change if the tensor $A_a{}^{\mu\lambda}$ vanishes at the space infinity.

(ii) the theorem is frequently presented in the physical literature as just (4.47): to each transformation leaving indifferent the action (and consequently the field equations) corresponds an invariant, a constant of motion. In the mathematical literature, the theorem is (4.41). In this last, historical form, it is possible to show an inverse theorem: if there are N linearly independent combinations of the Lagrangian derivatives reducing to divergences, then the action is invariant under the transformations of some N dimensional group. Whether there is or not some kind of inverse for the "physical" version is not clear. If the question is whether there is a symmetry corresponding to any integral of motion, the answer is no. The so-called "topological invariants" are not related to symmetries — we shall see an example below.

(iii) Equation (4.41) holds always, provided there is a symmetry of the action functional, for fields satisfying or not the equation of motion; it provides consequently information on the "space of states" of the system presenting the symmetry. Relations of this kind, independent of the field equations, are called "strong relations". We shall say a little more about that in the discussion of the second Noether theorem.

4.2.2 The Basic Spacetime Symmetries

The last term in (4.42) takes into account precisely transformations to which spacetime coordinates are sensitive. We have seen that physical fields *must* have a well–defined behavior under spacetime translations and Lorentz transformations. These can be realized as coordinate transformations on Minkowski spacetime and will consequently lead to currents, the densities of energy-momentum and 4-dimensional angular-momentum, engendered by every physical field. They are the most important field characteristics. Other spacetime symmetries may be eventually present, and will be left for later examination.

Translations and Energy-Momentum

§ 4.16 Consider an infinitesimal translation ($\alpha = 0, 1, 2, 3$)

$$x'^{\mu} = x^{\mu} + \delta x^{\mu} = x^{\mu} + \frac{\delta x^{\mu}}{\delta a^{\alpha}}\, \delta a^{\alpha} \ .$$

We can, in this case, take the x^{μ} themselves as parameters,

$$\frac{\delta x^{\mu}}{\delta a^{\alpha}} = \delta^{\mu}_{\alpha} \ . \tag{4.48}$$

Fields are Lorentz tensors and spinors, and as such unaffected by translations: $\delta\phi_i/\delta a^{\alpha} = 0$. Consequently, from (4.38),

$$\bar{\delta}\phi_i = -\ (\partial_{\mu}\phi_i)\, \delta^{\mu}_{\alpha}\, \delta a^{\alpha} = -\ (\partial_{\alpha}\phi_i)\, \delta a^{\alpha}.$$

The Noether current (4.42) related to invariance under translations will be

$$\Theta_{\alpha}{}^{\mu} := \frac{\partial \mathcal{L}}{\partial \partial_{\mu}\phi_i}\, \partial_{\alpha}\phi_i - \delta^{\mu}_{\alpha}\mathcal{L} \ . \tag{4.49}$$

Exercise 4.2 Use Eq.(4.20) to rewrite the general Noether current (4.42) as

$$J_a{}^{\mu} = -\ \frac{\partial \mathcal{L}}{\partial(\partial_{\mu}\phi_i)}\, \frac{\delta\phi_i}{\delta\omega^a} + \Theta_{\alpha}{}^{\mu}\, \frac{\delta x^{\alpha}}{\delta\omega^a} \ . \tag{4.50}$$

■

Due to the choice (4.48), the translation–algebra index α appears as a spacetime index and $\Theta_{\alpha}{}^{\mu}$ is a second-order tensor. With all indices raised,

$$\Theta^{\lambda\mu} = \eta^{\lambda\alpha}\, \Theta_{\alpha}{}^{\mu} = \frac{\partial \mathcal{L}}{\partial \partial_{\mu}\phi_i}\, \partial^{\lambda}\phi_i - \eta^{\lambda\mu}\mathcal{L} \ . \tag{4.51}$$

If we look for the analogous in Classical Mechanics, we find that this quantity corresponds to the "stress–energy tensor". In Field Theory it is usual to call it the *canonical energy–momentum tensor density*. The corresponding conserved charges will be

$$P^{\lambda} = \int d^3x \Theta^{\lambda 0}(x). \tag{4.52}$$

In particular, the charge P^0 corresponds to the Hamiltonian, with Θ^{00} the energy density. For covariance reasons, P^{λ} must then be the 4-momentum and Θ^{k0} its space density.

Summing up: the invariants related to the translations are the momentum components and the Noether currents constitute the energy–momentum tensor density.

Exercise 4.3 Again the real scalar field. From the first Lagrangian in Exercise 4.1,

$$\mathcal{L} = \tfrac{1}{2}\,\left[\partial_\mu\phi\partial^\mu\phi \;-\; m^2\phi^2\right]\,,$$

find the energy–momentum tensor density

$$\Theta^{\lambda\mu} = \partial^\lambda\phi\partial^\mu\phi \;-\; \eta^{\lambda\mu}\mathcal{L}\,.$$

Show that the energy density can be put into the positive form

$$\Theta^{00} = \tfrac{1}{2}\,\left[\partial_0\phi\;\partial_0\phi + \partial_i\phi\;\partial_i\phi \;+ m^2\;\phi^2\right]\,.$$

■

Lorentz Transformations and Angular Momentum

§ 4.17 Consider now infinitesimal rotations in spacetime (Lorentz transformations):

$$x'^\mu = x^\mu + \delta x^\mu = x^\mu + \tfrac{1}{2}\,\frac{\delta x^\mu}{\delta\omega^{\alpha\beta}}\,\delta\omega^{\alpha\beta}\;.$$

Recall that Cartesian spacetime coordinates transform according to the matrix vector representation (see § 2.28), whose matrices have entries given in Eq.(2.51):

$$[J_{\alpha\beta}]^\mu{}_\nu = i\left(\eta_{\alpha\nu}\delta^\mu_\beta - \eta_{\beta\nu}\delta^\mu_\alpha\right)\;.$$

Consequently,

$$x'^\mu = \left[\exp\left(\tfrac{i}{2}\delta\omega^{\alpha\beta}\,J_{\alpha\beta}\right)\right]^\mu{}_\nu\,x^\nu \approx \left[\delta^\mu_\nu + \tfrac{i}{2}\delta\omega^{\alpha\beta}\;(J_{\alpha\beta})^\mu{}_\nu\right]\;x^\nu =$$

$$= x^\mu + \tfrac{1}{2}\;\left(\delta\omega^{\mu\nu} - \delta\omega^{\nu\mu}\right)\;x_\nu\;. \tag{4.53}$$

We use then $\delta\omega^{\alpha\beta} = -\;\delta\omega^{\beta\alpha}$ to obtain

$$x'^\mu = x^\mu + \delta\omega^{\mu\nu}\,x_\nu\;. \tag{4.54}$$

Therefore,

$$\frac{\delta x^\mu}{\delta\omega^{\alpha\beta}} = \left(\delta^\mu_\alpha\,x_\beta - \delta^\mu_\beta\,x_\alpha\right)\;. \tag{4.55}$$

Exercise 4.4 Check this formula. ■

The Noether current corresponding to Lorentz transformations is known as the total angular momentum current density:

$$M^\mu{}_{\alpha\beta} = -\;\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}\,\frac{\bar\delta\phi_i}{\delta\omega^{\alpha\beta}} - \mathcal{L}\,\frac{\delta x^\mu}{\delta\omega^{\alpha\beta}}. \tag{4.56}$$

Using (4.38) or (4.50), it can be rewritten in the form

$$M^{\mu}{}_{\alpha\beta} = \Theta_{\alpha}{}^{\mu}\, x_{\beta} - \Theta_{\beta}{}^{\mu}\, x_{\alpha} - \frac{\partial \mathcal{L}}{\partial \partial_{\mu}\phi_i}\, \frac{\delta \phi_i}{\delta \omega^{\alpha\beta}} \ . \tag{4.57}$$

The first part,

$$L^{\mu}{}_{\alpha\beta} = \Theta_{\alpha}{}^{\mu}\, x_{\beta} - \Theta_{\beta}{}^{\mu}\, x_{\alpha} \ , \tag{4.58}$$

is present for all fields and is called the *orbital angular–momentum density* tensor. The last term is the *spin current density*,

$$S^{\mu}{}_{\alpha\beta} = - \frac{\partial \mathcal{L}}{\partial \partial_{\mu}\phi_i}\, \frac{\delta \phi_i}{\delta \omega^{\alpha\beta}} \ , \tag{4.59}$$

which appears only when the field is not a Lorentz singlet. In effect, a scalar field is defined as a field such that

$$\delta\phi(x) = \phi'(x') - \phi(x) = 0,$$

which is to say that

$$\frac{\delta \phi(x)}{\delta \omega^{\alpha\beta}} = 0 \ . \tag{4.60}$$

The spin density clearly vanishes. We find also

$$\frac{\bar{\delta} \phi(x)}{\delta \omega^{\alpha\beta}} = x_{\alpha}\partial_{\beta}\phi - x_{\beta}\partial_{\alpha}\phi \ ,$$

which helps to understand what happens to the scalar field: the field changes its functional form so as to just compensate the change in the argument. This compensation does not happen for other fields. A (Lorentzian) vector field, for example, is a field transforming like x^{μ}:

$$\phi'^{\mu}(x') = \left[\exp\left(\tfrac{i}{2}\ \delta\omega^{\alpha\beta}\, J_{\alpha\beta}\right)\right]^{\mu}{}_{\nu}\, \phi^{\nu}(x)$$

$$\approx \phi^{\mu}(x) + \tfrac{i}{2}\, \delta\omega^{\alpha\beta}\ (J_{\alpha\beta})^{\mu}{}_{\nu}\, \phi^{\nu}(x) = \phi^{\mu}(x) + \ \delta\omega^{\mu}{}_{\nu}\, \phi^{\nu}(x) \ . \tag{4.61}$$

We find then [Cf. eq.(4.55)]

$$\frac{\delta \phi^{\mu}(x)}{\delta \omega^{\alpha\beta}} = \delta^{\mu}_{\alpha}\phi_{\beta} \ - \delta^{\mu}_{\beta}\phi_{\alpha} \ ; \tag{4.62}$$

$$\frac{\bar{\delta} \phi^{\mu}(x)}{\delta \omega^{\alpha\beta}} = \delta^{\mu}_{\alpha}\phi_{\beta} \ - \delta^{\mu}_{\beta}\phi_{\alpha} \ + \ [x_{\alpha}\partial_{\beta} - x_{\beta}\partial_{\alpha}]\, \phi^{\mu} \ . \tag{4.63}$$

The last term is analogous to that of the scalar case and compensates the argument change, but there is a non-vanishing net variation $\delta\phi^{\mu}(x)$ as a response to

the Lorentz transformation. The spin density is exactly the contribution to $M^{\mu}{}_{\alpha\beta}$ coming from this "intrinsic" response.

From $\partial_\mu M^{\mu}{}_{\alpha\beta} = 0$ and $\partial_\mu \Theta_\alpha{}^\mu = 0$ follows

$$\partial_\mu S^{\mu}{}_{\alpha\beta} = \Theta_{\beta\alpha} - \Theta_{\alpha\beta} \ . \tag{4.64}$$

The antisymmetric part of the canonical energy–momentum density tensor measures the breaking of pure–spin conservation. Of course, there is no *a priori* reason for the spin to be conserved separately, but this happens when the canonical energy–momentum density tensor is symmetric. From the conservation of the orbital angular momentum for scalar fields, it comes that the energy–momentum is symmetric for those fields.

§ 4.18 The energy–momentum density tensor is the source for Einstein's equations

$$R_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu} R - g_{\mu\nu}\Lambda = \tfrac{8\pi G}{c^4}\, T_{\mu\nu}$$

for the gravitational field. Nevertheless, the left-hand side is symmetric in the two indices. The source current $T_{\mu\nu}$ representing energy and momentum must therefore be a symmetric tensor, which is not in general the case of the above canonical tensor $\Theta_{\lambda\mu}$. The solution to this conundrum comes from the fact that it is possible to obtain, from the canonical tensor, a symmetric tensor which differs from it by the total divergence of an antisymmetric tensor. An example is the Belinfante tensor

$$\Theta_B{}^{\lambda\mu} = \Theta^{\lambda\mu} + \tfrac{1}{2}\, \partial_\nu \left(S^{\nu\lambda\mu} - S^{\mu\lambda\nu} - S^{\lambda\mu\nu} \right) \ , \tag{4.65}$$

where $S^{\mu\rho\sigma} = \eta^{\rho\alpha}\eta^{\sigma\beta} S^{\mu}{}_{\alpha\beta}$.

Exercise 4.5 (Facultative) Show that $\Theta_B{}^{\lambda\mu}$ is indeed symmetric, by calculating $\Theta_B{}^{\lambda\mu}$ - $\Theta_B{}^{\mu\lambda}$ and using Eq.(4.64). ■

Examples of these conserved quantities will be seen in greater detail in the study of specific fields. We have said that, besides the ever–present translation and Lorentz invariances, other spacetime symmetries can manifest themselves. The most important is the conformal symmetry, which shows up when no scale–fixing parameter (such as a mass) is present in the theory. We shall not consider it here.

4.2.3 Internal Symmetries

§ 4.19 Any field has necessarily a well–defined behavior under Poincaré transformations. It can further belong to a representation $U(G)$ of an internal transformation

group G, as in the example of § 3.12. Under a transformation given by the element $g \in G$, its behavior is generically represented by

$$\phi_i'(x) = [U(g)]_i^j \ \phi_j(x) = \left[e^{i\omega^a T_a}\right]_i^j \ \phi_j(x) \ . \tag{4.66}$$

The T_a are the G generators in the U representation. In the presence of a local gauge invariance (see section 3.4), fields like the above $\phi_i(x)$ appear in physical Lagrangians in two ways. First, as free fields. Second, combined into certain currents which couple to gauge potentials. In consequence, those currents appear as sources in the right-hand side of the equations of motion for the gauge fields. For this reason such $\phi_i(x)$ are called *source fields*. Gauge potentials, on the other hand, mediate the interactions between the source fields. They are written as $A_\mu = J_a A_\mu^a$, with J_a the generators in the adjoint representation of G. Under a transformation $g = e^{i\omega^a J_a}$, they change according to (3.35),

$$A_\mu'(x) = g A_\mu(x) g^{-1} + g\partial_\mu g^{-1} \ . \tag{4.67}$$

A covariant derivative depends on the field ϕ_i on which it applies: A_μ will be written with the generators in the representation of ϕ_i. Field strengths $F_{\mu\nu} = J_a F_{\mu\nu}^a$ transform according to

$$F_{\mu\nu}'(x) = g F_{\mu\nu}(x) g^{-1} \ .$$

When G is an abelian group, the transformation is a product of phase transformations. As they are quite independent, it is enough to consider the one-dimensional case:

$$\phi'(x) = e^{-iq\alpha} \ \phi(x) \ . \tag{4.68}$$

The quantity q is a constant, playing the role of a 1-dimensional generator. Notice that, if $\phi(x)$ were supposed to be real, the transformation would cause a duplication of the number of degrees of freedom, which is incompatible with the idea of a symmetry. Thus, this phase transformation only makes sense if $\phi(x)$ is complex. In that case, (4.68) should be paired with

$$\phi^{*\prime}(x) = e^{iq\alpha} \ \phi^*(x) \ . \tag{4.69}$$

The argument above could be retaken at this point: $\phi = \phi^*$ would imply $q\alpha = 0$. The infinitesimal versions of the two equations are $\bar{\delta}\phi(x) = -iq\delta\alpha\phi(x)$ and $\bar{\delta}\phi^*(x)$ $= iq\delta\alpha\phi^*(x)$. The Noether current (4.42) will then be given by

$$J^\mu = -\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\frac{\bar{\delta}\phi}{\delta\alpha} + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^*)}\frac{\bar{\delta}\phi^*}{\delta\alpha}\right] = iq\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\phi - \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^*)}\phi^*\right] \ . \tag{4.70}$$

4.3 The Second Noether Theorem

§ 4.20 Let us once again go back to equation (4.33): it gives the variation of the global action of the system in terms of the variations of the coordinates (4.16) and the fields (4.17). One point is of basic importance: in no moment, during its derivation, have we used (as we did while obtaining the equations of motion) the boundary conditions. In other words, we have made no integration by parts. This means that, if we consider some sub–domain of the system, equation (4.33) will give the action variation in that region, provided of course that the integration take place only on it.

Let us make now a stronger supposition: that, under symmetry transformations, the action be an extremal in each sub–region of the system. The action remains unmoved under transformations in a small region around each point in the system. We shall have, in that case, a *local symmetry* and $\delta A = 0$ on an arbitrary volume inside the system. That requires the vanishing of the integrand everywhere, leading to the *Lie equation*:

$$\frac{\delta \mathcal{L}}{\delta \phi_i}\,\bar{\delta}\phi_i + \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)}\,\bar{\delta}\phi_i + \mathcal{L}\,\delta x^\mu \right] = 0. \tag{4.71}$$

From this expression we can get again (4.41) for constant parameter transformations and the consequent Noether theorem. But we can obtain something more: instead of using the functional differentiation, we can use functional derivations. We could proceed in the spirit of equation (4.10), getting directly the functional derivative of equation (4.33):

$$\frac{\delta A[\phi]}{\delta \omega^a(y)} = \frac{\delta \mathcal{L}(x)}{\delta \phi_i(x)}\,\frac{\bar{\delta}\phi_i(x)}{\delta \omega^a(y)} + \partial_\mu \left[\frac{\partial \mathcal{L}(x)}{\partial(\partial_\mu \phi_i(x))}\,\frac{\bar{\delta}\phi_i(x)}{\delta \omega^a(y)} + \mathcal{L}\,\frac{\delta x^\mu}{\delta \omega^a(y)} \right]. \tag{4.72}$$

Point-dependent transformations are characteristic of gauge theories. Here, we shall only call attention to some consequences of one fact in those theories. There are two kinds of fields in those theories, "source" fields transforming according to

$$\bar{\delta}\phi(x) = \delta\omega^a(x) T_a \phi(x), \tag{4.73}$$

and gauge potentials transforming according to

$$\bar{\delta} A^a{}_\mu(x) = f^a{}_{bc}\,\delta\omega^b(x) A^c{}_\mu(x) - \partial_\mu \delta\omega^a(x). \tag{4.74}$$

In both cases, fields and parameters are at the same point x. Thus, in the simplified approach we are adopting, we use

$$\frac{\bar{\delta}\phi(x)}{\delta \omega^a(y)} = \delta^4(x-y)\; T_a \phi(x);$$

$$\frac{\bar{\delta}A^a{}_\mu(x)}{\delta\omega^b(y)} = f^a{}_{bc}A^c{}_\mu(x)\ \delta^4(x-y) - \delta^a_b\partial_\mu\delta^4(x-y).$$

The δ's ensure locality. For any y interior to the system, these expressions lead to the vanishing of the divergence term. What remains is

$$\frac{\delta A[\phi,A]}{\delta\omega^a(y)} = \frac{\delta\mathcal{L}(y)}{\delta\phi(y)}\,T_a\phi(x) + \frac{\delta\mathcal{L}(y)}{\delta A^b{}_\mu(y)}f^a{}_{bc}A^c{}_\mu(y) + \partial_\mu\frac{\delta\mathcal{L}(y)}{\delta A^a{}_\mu(y)}\ . \tag{4.75}$$

The second Noether theorem says that, in the presence of a local symmetry related to a group with N generators, that is when

$$\frac{\delta A[\phi]}{\delta\omega^a(y)} = 0\ , \tag{4.76}$$

there are N independent relations between the Lagrange derivatives and their derivatives. This is what we obtain from the above equation:

$$\partial_\mu\frac{\delta\mathcal{L}(y)}{\delta A^a{}_\mu(y)} + \frac{\delta\mathcal{L}(y)}{\delta A^b{}_\mu(y)}f^a{}_{bc}A^c{}_\mu(y) = -\ \frac{\delta\mathcal{L}(y)}{\delta\phi_i(y)}\,T_a\phi(x). \tag{4.77}$$

Notice that the equations of motion have not been used. These relations are "strong", they hold independently of the solutions, reflecting the symmetries of the very space of possible states of the system.

More detail will be given in the section on gauge theories, but a few general comments can be made here. Define the object

$$J_a{}^\mu(x) = -\ \frac{\delta\mathcal{L}(x)}{\delta A^a{}_\mu(x)}\ . \tag{4.78}$$

The last expression above takes the form

$$\partial_\mu J_a{}^\mu(x) - f_{abc}A^b{}_\mu(x)J^{c\mu}(x) = -\ \frac{\delta\mathcal{L}(x)}{\delta\phi(x)}\,T_a\ \phi(x). \tag{4.79}$$

The total operator acting on J in the left-hand side will be the covariant derivative, actually a covariant divergence. In the right-hand side appears a factor resembling the Euler-Lagrange form, which is zero for solutions of the field equations. Then, a weak result would be: the "current" J has vanishing covariant divergence. In gauge theories, J is the current produced by the sources. The second Noether theorem does not lead to conserved quantities, but establishes constraints on the possible sources.

There will be conserved charges under the additional proviso that the local transformations become constant transformations outside the system. We shall come back to these points presently.

4.4 Topological Conservation Laws

§ 4.21 As we have said, not every conserved quantity is related to a symmetry. Let us see a simple example in two–dimensional spacetime, with coordinates $x^0 = vt$ and $x^1 = x$. Let a field be given by a scalar function $\phi(x,t)$ and consider the totally antisymmetric symbol in two dimensions, $\epsilon_{\mu\nu}$, $\epsilon_{01} = -\epsilon_{10} = 1$; $\epsilon_{00} = \epsilon_{11} = 0$. The current defined by $j_\mu(x,t) = \epsilon_{\mu\nu}\partial^\nu\phi$ will be automatically conserved:

$$\partial^\mu j_\mu(x,t) = 0.$$

This is an identity, valid independently of any Lagrangian we may have chosen — that is, of any dynamics. This kind of conservation *precedes* dynamics, it comes from something still more fundamental. This would be a simple curiosity if it did not apply to well–defined cases. But it does. A well-known example is that of a field obeying the so–called sine-Gordon equation

$$\frac{1}{v^2}\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} = -\ \sin\phi(x,t). \tag{4.80}$$

A particular solution is the solitary wave of Figure 4.1, given by

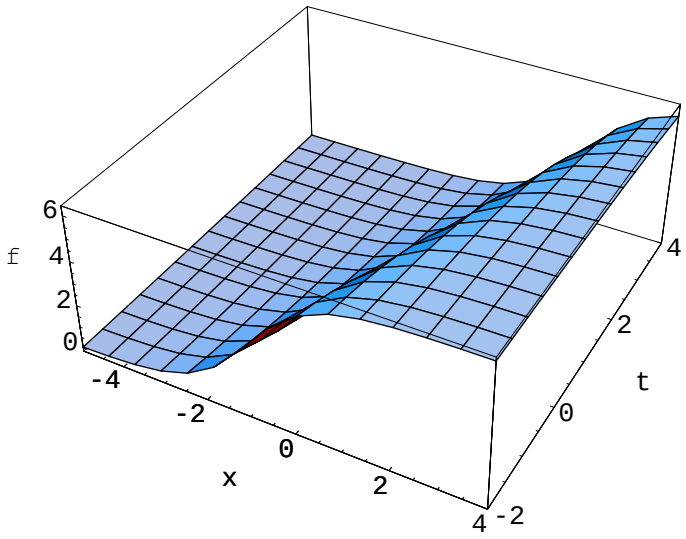


Figure 4.1: *Sine-Gordon wave (4.81) with $v = 0.5$.*

$$\phi(x,t) = 4\arctan\ \exp\left[\frac{x-vt}{\sqrt{1-v^2}}\right]. \tag{4.81}$$

Let us use it to unravel the meaning of the current conservation above. For this solution,

$$J_1 = \epsilon_{01}\partial^0\phi = \frac{v}{\sqrt{1-v^2}}\ \frac{4e^{\gamma(x-vt)}}{1+e^{2\gamma(x-vt)}} = 2\gamma v\ \frac{1}{\cosh[\gamma(x-vt)]}\quad .$$

Thus, $J_1 \to 0$ for $x \to \pm\infty$. But $\partial^0 J_0 + \partial^1 J_1 = 0$ implies

$$\frac{\partial}{\partial(vt)} \int_{-\infty}^{+\infty} dx J_0(x,t) = \int_{-\infty}^{+\infty} dx \frac{\partial J_1(x,t)}{\partial x} = J_1(+\infty,t) - J_1(-\infty,t) = 0.$$

We have thus a conserved "charge". We can normalize things conveniently and calculate

$$n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dx J_0(x,t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dx \epsilon_{01} \partial^1 \phi(x,t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dx \frac{\partial \phi(x,t)}{\partial x} = \frac{1}{2\pi} [\phi(+\infty,t) - \phi(-\infty,t)]. \tag{4.82}$$

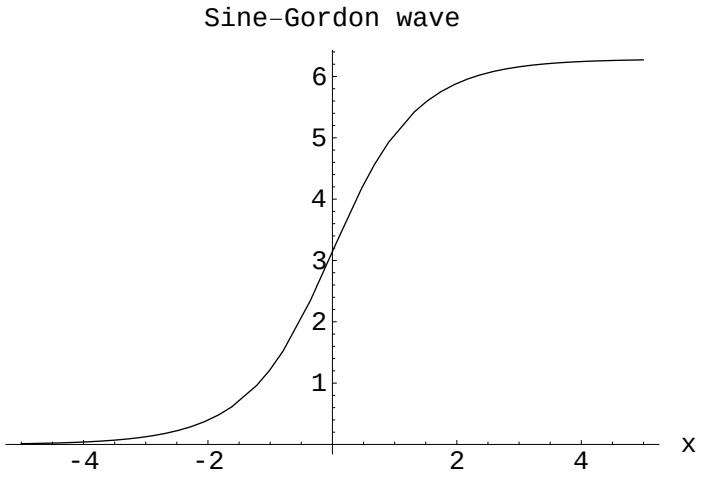


Figure 4.2: *Space cross-section of the above sine-Gordon wave at* $t = 0$.

The aspect of solution (4.81) at a fixed time is given in Figure 4.2: it has $\phi(+\infty,t) = 2\pi$ and $\phi(-\infty,t) = 0$. It follows that $n = 1$. Waves like that are called "solitons" (or solitary waves) and turn up as solutions of many non–linear equations. The sine-Gordon equation has also solutions with many solitons (which "grow" many times 2 π), inverted solutions with the wave decreasing ("anti–solitons") and still solutions combining r solitons and s anti–solitons. For such solutions, the above number n is $n = r - s =$ "soliton number". We have, at the beginning of the discussion, carefully avoided saying anything on boundary conditions, which are different for each kind of solution. For fixed boundary conditions, however, there will be all the solutions with a fixed n. This number n is an example of invariant related to the topology of the fields (or of the space of solutions). This kind of invariant is, for that reason, called "a topological number". It is not related to any symmetry and cannot be obtained through the Noether theorem. Other non-linear equations exhibit solitonic solutions with non–Noetherian conserved charges. The best known

are mostly in two-dimensional space [(1+1)-space, one dimension for space, one for time], like the Korteweg-de Vries (KdV) equation.

There are thus two kinds of conserved quantities in Physics: those coming from Noether's theorem – conserved *along* solutions of the equations of motion – and the topological invariants, which come from the global, topological properties of the space of states.

§ 4.22 A Note on the Hamiltonian Approach To pass into the Hamiltonian formalism, we must first define the momentum conjugate to each ϕ_i. A majority of authors follow the classic analogy, putting

$$\pi_i \doteq \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi_i)} \ ,$$

which gives the time parameter a favored role from the start. A few others introduce a 4-vector momentum for each ϕ_i,

$$\pi_i{}^{\mu} \doteq \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \ , \tag{4.83}$$

in terms of which the field equations become compact indeed:

$$\partial_\mu \pi_i{}^{\mu} = \frac{\partial \mathcal{L}}{\partial \phi_i} \ . \tag{4.84}$$

Comment 4.1 We have been talking about a "minimum" of the action but it is enough to have an extremum to arrive at the field equations. To go into the details of the principle of minimal action, we should study also the second variation.

Chapter 5

Vector Fields

We now proceed to a detailed discussion of the main fields which have been found to describe elementary particles — and, consequently, the fundamental interactions — in Nature. They are classified, as repeatedly announced, by their behavior under transformations of the Poincaré group. It is fortunate that, at least for the time being, only particles and fields belonging to the lowest representations — those of small dimensions — seem to play a basic role. They also have different characters according to their spins being integers or half-integers. Integer-spin particles are called *bosons*, and their fields are *bosonic fields*. Half-integer-spin particles are called *fermions*, and their fields are *fermionic fields*. Bosonic and fermionic particles have quite distinct statistical behaviour.

An electric charge creates a field, to which responds any other electrically charged object. This field — the electromagnetic field — is a vector field. The charged objects interact with each other electromagnetically and we say that the electromagnetic field "mediates" that interaction. The same happens with other interactions. With the remarkable exception of gravitation, all the known fundamental interactions of Nature are mediated by vector fields. Because of its fundamental nature, we start our study with bosons of spin 1 — that is, with vector fields.

As the name indicates, a vector field is a set of four fields transforming as the components of a Lorentz vector. It can be indicated as

$$\phi(x) = \begin{pmatrix} \phi^0(x) \\ \phi^1(x) \\ \phi^2(x) \\ \phi^3(x) \end{pmatrix} .$$

This means that the set transforms according to the vector representation of the

Lorentz group, that of the cartesian coordinates x^μ:

$$\phi'^\mu(x') = \Lambda^\mu{}_\nu \phi^\nu(x)\,. \tag{5.1}$$

The vector field is real or complex if each one of its components is real or complex. The complex field is formally richer, and includes the real field as a particular case. In Nature, we find both kinds among the mediating fields of the electroweak interactions, whose quanta have been found in 1983: the bosons $W^\pm_\mu$ are described by a pair of complex conjugate massive vector fields and the boson Z^0_μ by a real massive vector field. The photon field A_μ is a real vector field with vanishing mass. It is the mediating field of electromagnetism and will be discussed in an independent chapter.

5.1 Real Vector Fields

§ 5.1 The simplest, most natural Lagrangian for a real vector field ϕ_μ of mass M would be

$$\mathcal{L} = -\,\tfrac{1}{2}\left[(\partial_\mu\phi_\nu)(\partial^\mu\phi^\nu) - M^2\phi^\nu\phi_\nu)\right], \tag{5.2}$$

which is, actually, the covariant sum of four Lagrangians of the scalar type. But here, unlike the internal components of the complex scalar field, the components are truly those of a vector and the way they are contracted in the Lagrangian is a requirement of Lorentz invariance. This Lagrangian leads naturally to four independent Klein-Gordon equations,

$$(\Box + M^2)\phi_\mu(x) = 0. \tag{5.3}$$

The energy-momentum density tensor is symmetric,

$$\Theta^{\mu\nu} = -\,\partial^\mu\phi_\lambda\partial^\nu\phi^\lambda - \eta^{\mu\nu}\mathcal{L}. \tag{5.4}$$

The spin density obtained from $\mathcal{L}$ is the first we have the opportunity to write down:

$$S^\mu_{(\alpha\beta)}(x) = -\,\phi_\alpha(x)\partial^\mu\phi_\beta(x) + \phi_\beta(x)\partial^\mu\phi_\alpha(x). \tag{5.5}$$

It has clearly vanishing divergence for the solutions, which is to be expected given the symmetry of the canonical energy-momentum density. The spin $S^\mu_{(\alpha\beta)} = \int d^3x\; S^0_{(\alpha\beta)}(x)$ and the orbital angular momentum are, consequently, separately conserved.

There is, however, a difficulty. The terms containing the component $\phi^0(x)$ contribute to $\mathcal{L}$ with opposite sign with respect to the terms of the other components,

so that $\mathcal{L}$ has not a well-defined sign. This leads to an energy density which is not positive-definite:

$$\begin{aligned}\Theta^{00} = &- \tfrac{1}{2}(\partial^0\phi^0)^2 - \tfrac{1}{2}(\vec{\nabla}\phi^0)^2 - \tfrac{M^2}{2}(\phi^0)^2 \\ &+ \tfrac{1}{2}\sum_{j=1}^{3}\left[(\partial_0\phi_j)^2 + (\vec{\nabla}\phi_j)^2 + M^2\phi_j^2\right]. \end{aligned} \tag{5.6}$$

The three positive contributions are alike those of the scalar case, but there are negative terms. A Hamiltonian which is not positive–definite is a serious defect in a theory proposed to describe a free, non-interacting system. A supplementary condition must be introduced in order to correct it. The only condition which is invariant and linear in the fields is

$$\partial_\mu\phi^\mu = 0, \tag{5.7}$$

which should hold at each point of the system. We could put also a constant in the right-hand side, but this would add an arbitrary constant to the theory. The above condition reduces to three the number of independent degrees of freedom and, when used to eliminate the miscreant ϕ^0, does lead to a positive-definite energy. We shall not prove it here, as it requires the use of Fourier analysis in detail.

The physical system is described by the Lagrangian *plus* the supplementary condition. This is a novel situation: the Lagrangian alone does not determine, via the minimal action principle, the acceptable conditions. There is something amiss with the Lagrangian (5.2). The problem can be circumvented by using another, the Wentzel–Pauli Lagrangian

$$\mathcal{L}' = \mathcal{L} + \tfrac{1}{2}(\partial_\mu\phi_\nu)(\partial^\nu\phi^\mu) \tag{5.8}$$

$$= -\tfrac{1}{4}\left[(\partial_\mu\phi_\nu - \partial_\nu\phi_\mu)(\partial^\mu\phi^\nu - \partial^\nu\phi^\mu)\right] + \tfrac{M^2}{2}\phi^\nu\phi_\nu. \tag{5.9}$$

The Euler-Lagrange equation coming from this Lagrangian is the Proca equation

$$\Box\,\phi^\nu - \partial^\nu(\partial_\mu\phi^\mu) + M^2\phi^\nu = 0, \tag{5.10}$$

whose solutions satisfy automatically the supplementary condition if $M \neq 0$. In effect, taking ∂_ν of the equation we obtain $M^2\partial_\nu\phi^\nu = 0$. Thus, the Lagragian $\mathcal{L}'$ automatically implements the supplementary condition. It is interesting to introduce the variable $F_{\mu\nu} = \partial_\mu\phi_\nu - \partial_\nu\phi_\mu$, in terms of which the Lagrangian takes the form

$$\mathcal{L}' = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \tfrac{M^2}{2}\phi^\nu\phi_\nu \tag{5.11}$$

and the Proca equation becomes

$$\partial_\mu F^{\mu\nu} + M^2\phi^\nu = 0. \tag{5.12}$$

Exercise 5.1 Obtain (5.12) from (5.11). ■

The Wentzel–Pauli Lagrangian is used by most modern authors. It leads to the energy-momentum density

$$\Theta'^{\alpha}{}_{\mu} = - F_{\mu\nu}\partial^{\alpha}\phi^{\nu} - \delta^{\alpha}_{\mu}\mathcal{L}' \tag{5.13}$$

and the spin density

$$S'^{\mu}{}_{\alpha\beta} = S^{\mu}{}_{\alpha\beta} + \phi_{\alpha}\partial_{\beta}\phi^{\mu} - \phi_{\beta}\partial_{\alpha}\phi^{\mu} = \phi_{\beta}F^{\mu}{}_{\alpha} - \phi_{\alpha}F^{\mu}{}_{\beta} \; . \tag{5.14}$$

In the case $M = 0$, which includes the electromagnetic field, $\mathcal{L}'$ does not implement the supplementary condition. In that case, however, $\mathcal{L}'$ has a special, extra symmetry: it does not change if the field changes as

$$\phi^{\mu}(x) \rightarrow \phi'^{\mu}(x) = \phi^{\mu}(x) + \partial^{\mu}f(x),$$

for any differentiable function $f(x)$. This is a gauge invariance and allows one a lot of freedom in choosing the field. Each choice of the field is called "a gauge". In particular, if $\partial_{\mu}\phi^{\mu} = g(x)$, it is possible to implement the supplementary condition whenever a solution f can be found for the Poisson equation $\Box f(x) = \text{-}\ g(x)$, which would lead immediately to $\partial_{\mu}\phi'^{\mu} = 0$. This choice is called the *Lorenz gauge* (not Lorentz!). As the Lagrangian is invariant under any "change of gauge", the physical results found in that particular gauge hold true in general. As already announced, the electromagnetic field will deserve a special chapter.

5.2 Complex Vector Fields

§ 5.2 Let us examine the case of the complex vector field, which includes the real case. As said above, the electroweak bosons $W^{\pm}_{\mu}$ are described by a pair of complex conjugate massive vector fields. The Lagrangian is

$$\mathcal{L} = - \left[\partial_{\mu}\phi^{*\nu}\partial^{\mu}\phi_{\nu} - M^2\phi^{*\nu}\phi_{\nu}\right] \; , \tag{5.15}$$

and the subsidiary conditions,

$$\partial_{\mu}\phi^{\mu} = 0 \;\; ; \;\; \partial_{\mu}\phi^{*\mu} = 0 \, . \tag{5.16}$$

The equations of motion come out as

$$\left(\Box + M^2 \right) \phi^{\nu} = 0 \; ; \; \left(\Box + M^2 \right) \phi^{*\nu} = 0 \, . \tag{5.17}$$

The energy-momentum density tensor is symmetric:

$$\Theta^{\alpha\mu} = -\,\partial^\mu\phi^*_\nu\partial^\alpha\phi^\nu - \partial^\mu\phi_\nu\partial^\alpha\phi^{*\nu} - \eta^{\alpha\mu}\mathcal{L}, \tag{5.18}$$

and in special

$$\Theta^{00} = -\,\left(\partial^0\phi^*_\nu\partial^0\phi^\nu + \partial_i\phi^{*\nu}\partial_i\phi_\nu\right) - M^2\phi^\nu\phi^*_\nu, \tag{5.19}$$

$$\Theta^{i0} = -\left(\partial^i\phi^*_\nu\partial^0\phi^\nu + \partial^i\phi_\nu\partial_0\phi^{*\nu}\right). \tag{5.20}$$

The spin density will be

$$S^\mu_{(\alpha\beta)} = -\,\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_\nu)}\frac{\delta\phi_\nu}{\delta\omega^{\alpha\beta}} \;-\; \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^*_\nu)}\frac{\delta\phi^*_\nu}{\delta\omega^{\alpha\beta}}\;. \tag{5.21}$$

Using (4.62) we find

$$S^\mu_{(\alpha\beta)} = \phi^*_\beta\overset{\leftrightarrow}{\partial^\mu}\phi_\alpha - \phi^*_\alpha\overset{\leftrightarrow}{\partial^\mu}\phi_\beta\,. \tag{5.22}$$

Comment 5.1 It is a useful notation to put all antisymmetrized indices inside square brackets. Thus, the above expression is

$$S^\mu_{(\alpha\beta)} = \phi^*_{[\beta}\overset{\leftrightarrow}{\partial^\mu}\phi_{\alpha]}.$$

Spin itself will be

$$S_{(\alpha\beta)} = \int d^3x S^0_{(\alpha\beta)}. \tag{5.23}$$

Thanks to the symmetry of $\Theta^{\alpha\mu}$, $S_{(\alpha\beta)}$ is conserved. We can separate the purely spatial components and define the (tri–)vector spin:

$$S_i = \tfrac{1}{2}\epsilon_{ijk}S_{(jk)} = \tfrac{1}{2}\epsilon_{ijk}\int d^3x\; S^0_{(jk)}. \tag{5.24}$$

In the quantized theory, this conserved vector $\vec{S}$ will be the spin of the field quanta.

§ 5.3 The Wentzel-Pauli Lagrangian would come as follows: first, introduce the complex vector field as the combination of two real vector fields

$$\phi_\mu = \phi^{(1)}_\mu + i\phi^{(2)}_\mu;$$

$$\phi^{*\mu} = \phi^{(1)}_\mu - i\phi^{(2)}_\mu.$$

The Lagrangian is then the sum of two Lagrangians like (5.11):

$$\mathcal{L}' = -\,\tfrac{1}{4}\,F^{(1)}_{\mu\nu}F^{\mu\nu}_{(1)} - \;\tfrac{1}{4}\,F^{(2)}_{\mu\nu}F^{\mu\nu}_{(2)} + \;\tfrac{M^2}{2}\,\phi^\nu_{(1)}\phi^{(1)}_\nu + \;\tfrac{M^2}{2}\,\phi^\nu_{(2)}\phi^{(2)}_\nu. \tag{5.25}$$

This is the same as

$$\mathcal{L}' = -\,\tfrac{1}{4}\,F^*_{\mu\nu}F^{\mu\nu} + \;\tfrac{M^2}{2}\,\phi^\nu\phi^*_\nu\,. \tag{5.26}$$

Two Proca equations come out in consequence:

$$\partial_\mu F^{\mu\nu} + M^2 \phi^\nu = 0. \tag{5.27}$$

$$\partial_\mu F^{*\mu\nu} + M^2 \phi^{*\nu} = 0. \tag{5.28}$$

The ensuing energy-momentum density tensor is:

$$\Theta^{\alpha\mu} = -\,\tfrac{1}{2}\left(F^{*\mu\nu}\partial^\alpha \phi_\nu + F^{\mu\nu}\partial^\alpha \phi^*_\nu\right) - \eta^{\alpha\mu}\mathcal{L}, \tag{5.29}$$

with in particular

$$\Theta^{00} = \tfrac{1}{2}\left[(\partial_0\phi_k)(\partial_0\phi^*_k) - (\partial_i\phi_0\partial_i\phi^*_0\right] + \tfrac{1}{4}F^*_{ij}F^{ij} - \tfrac{M^2}{2}\phi^\nu\phi^*_\nu. \tag{5.30}$$

The spin density will be

$$S^\mu_{(\alpha\beta)} = \tfrac{1}{2}\left[F^{*\mu}{}_\alpha\phi_\beta - F^{*\mu}{}_\beta\phi_\alpha + F^\mu{}_\alpha\phi^*_\beta - F^\mu{}_\beta\phi^*_\alpha\right]\ . \tag{5.31}$$

As in the complex scalar case, there will be a global $U(1)$ symmetry. The corresponding Noether current will be

$$J^\mu = -\,i\left[\phi^{*\nu}\partial^\mu\phi_\nu - (\partial^\mu\phi^{*\nu})\phi_\nu\right], \tag{5.32}$$

with the charge

$$Q = -\,i\int d^3x\left[\phi^{*\nu}\overset{\leftrightarrow}{\partial^0}\phi_\nu\right]. \tag{5.33}$$

Theories with massive vector fields as above interacting with other fields are, in general (perturbatively) "unrenormalizable". This means that, once quantized, they lead to infinite values for certain finite quantities. The exceptions are the gauge fields, of which the simplest example is the electromagnetic field.

Chapter 6

Electromagnetic Field

This is the best known of all fields. In what follows we shall only outline the general aspects of the electromagnetic field, emphasizing some features which are more specific to it while exhibiting some properties it shares with fields in general.

6.1 Maxwell's Equations

§ 6.1 Classical form. The electric field $\vec{E}$ and the magnetic field $\vec{H}$ created by a charge density ρ and a current density $\vec{j}$ satisfy Maxwell's equations:

$$\text{rot } \vec{E} + \frac{\partial \vec{H}}{\partial ct} = 0 \tag{6.1}$$

$$\text{div } \vec{H} = 0 \tag{6.2}$$

$$\text{div } \vec{E} = 4\pi \rho \tag{6.3}$$

$$\text{rot } \vec{H} - \frac{\partial \vec{E}}{\partial ct} = \frac{4\pi}{c} \vec{j}. \tag{6.4}$$

Fields $\vec{E}$ and $\vec{H}$ can exist (that is, it is possible to have $\vec{E} \neq 0$ and/or $\vec{H} \neq 0$) in domains of space where $\rho = 0$ and $\vec{j} = 0$. This is the case, for example, of radio waves traveling far away from their sources. In such a sourceless case, Maxwell's equations present a special symmetry, called *duality*: they remain invariant if we exchange

$$\vec{E} \to \vec{H} \quad \text{and} \quad \vec{H} \to -\vec{E}. \tag{6.5}$$

If we wished to extend this symmetry to the case with nonvanishing sources, it would be necessary to postulate the existence of a magnetic charge density, to be introduced in the right hand side of (6.2) and a magnetic current density to be introduced in the right hand side of (6.1). A magnetic density would be carried

by magnetic monopoles. Unlike electric monopoles which turn up at every corner, magnetic monopoles have never been observed, despite many efforts.

From equations (6.3) and (6.4) follows the local conservation of charge, or continuity equation:

$$\frac{\partial \rho}{\partial t} + \text{div } \vec{j} = 0. \tag{6.6}$$

§ 6.2 Integral version. Let us recall the meaning of Maxwell's equations, which is better understood in their integral versions, which are laws going after illustrious names:

1. the Gauss law:

$$\int_{S=\partial V} d\vec{S} \cdot \vec{E} = 4\pi \int_V \rho \, d^3x \, ; \tag{6.7}$$

   the flux of $\vec{E}$ through a closed surface S is equal to the charge contained in the volume V of which S is the border;

2. the Faraday induction law:

$$\int_{l=\partial S} d\vec{l} \cdot \vec{E} = - \int_S d\vec{S} \cdot \frac{\partial \vec{H}}{\partial ct} \, ; \tag{6.8}$$

   the time-variation of the magnetic flux crossing a surface S creates a circulation of the electric field along the line l which in the border of S;

3. there are no magnetic monopoles:

$$\int_{S=\partial V} d\vec{S} \cdot \vec{H} = 0; \tag{6.9}$$

   the flux of $\vec{H}$ through any closed surface S vanishes; or, no surface can contain a magnetic charge;

4. Maxwell's displacement current:

$$\int_{l=\partial S} d\vec{l} \cdot \vec{H} = \int_S d\vec{S} \cdot \left(\frac{4\pi}{c} \, \vec{j} + \frac{\partial \vec{E}}{\partial ct} \right) ; \tag{6.10}$$

   the circulation of $\vec{H}$ along a closed line bordering a surface S is due not only to the current through S, but also to the flux of the time variation of $\vec{E}$. Recall that $\frac{1}{4\pi} \frac{\partial \vec{E}}{\partial ct}$ is the "displacement current". In its absence, the above equation spells *Ampère's law.*

§ 6.3 Maxwell tensor. Though established long before the advent of Special Relativity, Maxwell's equations were already relativistic-covariant. This is not explicit in the expressions given above. It is convenient to pass into a covariant notation. Let us define the antisymmetric tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{6.11}$$

where A_μ is the electromagnetic potential. Historically, the electromagnetic potential has been suggested by Eq.(6.2): if $\vec{H}$ has vanishing divergence (in a simply-connected region), then it is the curl of some vector:

$$\vec{H} = \text{rot } \vec{A}. \tag{6.12}$$

By covariance, there must exist a time-component A^0 satisfying, because of Eq.(6.1),

$$\vec{E} = -\vec{\nabla} A^0 - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}\ . \tag{6.13}$$

Expression (6.11) is the explicitly covariant expression: it gives $E^i = -\partial_i A_0 - \partial_0 A^i$ (the usual gradient is a covector, with lower indices). Notice that A_α is not an observable, as it is defined up to a *gauge transformation*

$$A^\alpha(x) \Rightarrow A'^\alpha(x) = A^\alpha(x) + \partial^\alpha \phi(x) \tag{6.14}$$

with ϕ an arbitrary function. This transformation leaves $F_{\alpha\beta}$ in (6.11) invariant.

From these relations we see that

$$F^{i0} = E^i \quad \text{and} \quad F^{ij} = \epsilon^{ijk} H^k. \tag{6.15}$$

matrix, whose entries we shall denote $F^{\alpha\beta} = -F^{\beta\alpha}$:

$$(F^{\alpha\beta}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -H^3 & H^2 \\ E^2 & H^3 & 0 & -H^1 \\ E^3 & -H^2 & H^1 & 0 \end{pmatrix}. \tag{6.16}$$

This is the electromagnetic tensor, or Maxwell tensor, or still "field strength". Let us define also its dual,

$$(\tilde{F}^{\alpha\beta}) = (\tfrac{1}{2}\,\epsilon^{\alpha\beta\gamma\delta} F_{\gamma\delta}) = \begin{pmatrix} 0 & -H^1 & -H^2 & -H^3 \\ H^1 & 0 & E^3 & -E^2 \\ H^2 & -E^3 & 0 & E^1 \\ H^3 & E^2 & -E^1 & 0 \end{pmatrix}, \tag{6.17}$$

which is obtained from the first by performing the duality transformation (6.5) for each entry.

6.2 Transformations of $\vec{E}$ and $\vec{H}$

§ 6.4 We shall here be interested in the behavior of the electric and magnetic fields under boosts. This means, in view of Eq.(6.16), the behavior of $F^{\alpha\beta}$. It will be necessary, consequently, to examine the behavior of tensors.

We have in § 1.29 defined a Lorentz vector as any set $V = (V^0, V^1, V^2, V^3)$ of four quantities transforming like (ct, x, y, z). In the case of a boost like that given by Eqs. (1.50) and (1.51), with a velocity v along the axis Ox ($v = v_x$), they will have the behavior

$$\begin{aligned} A^1 &= \gamma(A'^1 + \frac{v}{c}\, A'^0) \\ A^2 &= A'^2 \; ; \; A^3 = A'^3 \\ A^0 &= \gamma(A'^0 + \frac{v}{c}\, A'^1) \; . \end{aligned} \tag{6.18}$$

We have in § 1.32 introduced the four-vector current (1.73)

$$j^\alpha = e\; u^\alpha = e\; \gamma\; \left(1, \frac{v^x}{c}, \frac{v^y}{c}, \frac{v^z}{c}\right)$$

and the four-vector potential (1.74)

$$A = (A^\alpha) = (\phi, \vec{A}) = (\phi, A^x, A^y, A^z) \; .$$

They will both transform in the above way. For instance,

$$\phi = \gamma(\phi' + \frac{v}{c}\, A'^x) \; .$$

This is the case for upper-indexed vectors (contravariant vectors). Lower-indexed vectors (covariant vectors, or covectors as A_α, for example), once contracted with upper-indexed vectors give invariants, scalars like $j^\alpha A_\alpha$. It follows that they transform just in the inverse way. The inverse transformation, in the case of the boost above, corresponds to exchange the roles of the reference frames or, more simply, to change the sign of the boost velocity. Thus,

$$\begin{aligned} A_1 &= \gamma(A'_1 - \frac{v}{c}\, A'_0) \\ A_2 &= A'_2 \; ; \; A_3 = A'_3 \\ A_0 &= \gamma(A'_0 - \frac{v}{c}\, A'_1) \; . \end{aligned} \tag{6.19}$$

§ 6.5 To see how this appears in matrix language, recall that the usual notation for a general Lorentz transformation is $x^{\alpha'} = \Lambda^{\alpha'}{}_\beta x^\beta$. We have in Eqs. (6.18) an

inverse transformation, corresponding to $x^\alpha = (\Lambda^{-1})^\alpha{}_{\beta'} x^{\beta'}$. Precisely, we have a transformation of type $A^\alpha = (\Lambda^{-1})^\alpha{}_{\beta'} A^{\beta'}$. In detail, Eqs. (6.18) are

$$\begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A^{0'} \\ A^{1'} \\ A^{2'} \\ A^{3'} \end{pmatrix} \tag{6.20}$$

and Eqs. (6.19) are

$$\begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_{0'} \\ A_{1'} \\ A_{2'} \\ A_{3'} \end{pmatrix} . \tag{6.21}$$

Exercise 6.1 Verify that the matrices appearing in Eqs. (6.20) and Eqs. (6.21)

1. have determinants = 1 and

2. are inverse to each other.

■

§ 6.6 The cases above are, of course, particular boosts. Under general Lorentz transformations (see § 1.34) vectors and covectors change as given by Eqs. (1.79) and (1.81). Second order tensors follow (1.80). That is to say that they transform like objects whose components are the products of vectors and/or covector components. We are anyhow interested only in the behavior of the antisymmetric tensor $F^{\alpha\beta}$ under boosts.

§ 6.7 Let us go back to the case of § 6.5. To get the behavior of a tensor $T^{\alpha\beta}$, the simplest procedure is to consider A as in (6.18) together with another vector

$$\begin{aligned} B^1 &= \gamma(B'^1 + \frac{v}{c} B'^0) \\ B^2 &= B'^2 \; ; \; B^3 = B'^3 \\ B^0 &= \gamma(B'^0 + \frac{v}{c} B'^1) \; . \end{aligned} \tag{6.22}$$

$T^{\alpha\beta}$ will transform like $A^\alpha B^\beta$. Some examples are:

$$\begin{aligned}
A^1B^1 &= \gamma^2(A'^1B'^1 + \frac{v}{c}\ A'^1B'^0 + \frac{v}{c}\ B'^1A'^0 + \frac{v^2}{c^2}\ B'^0A'^0) \\
\therefore\ T^{11} &= \gamma^2(T'^{11} + \frac{v}{c}\ T'^{10} + \frac{v}{c}\ T'^{10} + \frac{v^2}{c^2}\ T'^{00}) \qquad (6.23) \\
A^1B^0 &= \gamma^2(A'^1B'^0 + \frac{v}{c}\ A'^1B'^1 + \frac{v}{c}\ A'^0B'^0 + \frac{v^2}{c^2}\ A'^0B'^1) \\
\therefore\ T^{10} &= \gamma^2(T'^{10} + \frac{v}{c}\ T'^{11} + \frac{v}{c}\ T'^{00} + \frac{v^2}{c^2}\ T'^{01}) \qquad (6.24) \\
A^1B^2 &= \gamma(A'^1B'^2 + \frac{v}{c}\ A'^0B'^2) \ \text{ and } \ A^1B^3 = \gamma(A'^1B'^3 + \frac{v}{c}\ A'^0B'^3) \\
\therefore\ T^{12} &= \gamma(T'^{12} + \frac{v}{c}\ T'^{02}) \ \text{ and } \ T^{13} = \gamma(T'^{13} + \frac{v}{c}\ T'^{03}) \qquad (6.25) \\
A^2B^2 &= A'^2B'^2 \ ; \ A^2B^3 = A'^2B'^3 \ ; \ A^3B^3 = A'^3B'^3 \\
\therefore\ T^{22} &= T'^{22}\ ; \ \text{and also } \ T^{33} = T'^{33}\ ,\ T^{23} = T'^{23} \qquad (6.26)
\end{aligned}$$

$$\begin{aligned}
A^0B^0 &= \gamma^2(A'^0B'^0 + \frac{v}{c}\ A'^0B'^1 + \frac{v}{c}\ A'^1B'^0 + \frac{v^2}{c^2}\ A'^1B'^1) \\
\therefore\ T^{00} &= \gamma^2(T'^{00} + \frac{v}{c}\ T'^{01} + \frac{v}{c}\ T'^{10} + \frac{v^2}{c^2}\ T'^{11}) \qquad (6.27) \\
A^0B^1 &= \gamma^2(A'^0B'^1 + \frac{v}{c}\ A'^0B'^0 + A'^1B'^1 + \frac{v^2}{c^2}\ A'^1B'^0) \\
\therefore\ T^{01} &= \gamma^2(T'^{01} + \frac{v}{c}\ T'^{00} + T'^{11} + \frac{v^2}{c^2}\ T'^{10}) \qquad (6.28) \\
A^0B^2 &= \gamma(A'^0B'^2 + \frac{v}{c}\ A'^1B'^2)\ ; \ \ A^0B^3 = \gamma(A'^0B'^3 + \frac{v}{c}\ A'^1B'^3) \\
\therefore\ T^{02} &= \gamma(T'^{02} + \frac{v}{c}\ T'^{12}) \ \text{ and } \ T^{03} = \gamma(T'^{03} + \frac{v}{c}\ T'^{13}). \qquad (6.29)
\end{aligned}$$

A few possibilities are missing, but these are more than enough to tackle the question of tensors with well-defined symmetries in the indices. Take for example $E_x = F^{01}$, antisymmetric: F^{01} will transform like $A^0B^1 - A^1B^0 = \gamma^2(1-\frac{v^2}{c^2})(A'^0B'^1 - A'^1B'^0) = F'^{01} = E'_x$. Another case: $H_z = F^{12}$ transforms like $A^1B^2 - A^2B^1 = \gamma(F'^{12} + \frac{v}{c}\ F'^{02})$ and $\therefore\ H_z = \gamma[H'_z + \frac{v}{c}\ E'_y]$. In this way we find the cases:

$$\begin{aligned}
E_x &= E'_x \\
E_y &= \gamma(E'_y + \frac{v}{c}\ H'_z) \\
E_z &= \gamma(E'_z - \frac{v}{c}\ H'_y) \\
H_x &= H'_x \\
H_y &= \gamma(H'_y - \frac{v}{c}\ E'_z) \\
H_z &= \gamma(H'_z + \frac{v}{c}\ E'_y). \qquad (6.30)
\end{aligned}$$

The left-hand side gives the fields as seen from a frame K, the right-hand side as seen from a frame K'. Electric and magnetic fields depend, as we see, on the reference frame from which they are looked at.

§ 6.8 All this is actually much simpler in the matrix version of § 6.5. The transformation above is

$$F^{\alpha\beta} = (\Lambda^{-1})^{\alpha}{}_{\alpha'}(\Lambda^{-1})^{\beta}{}_{\beta'}F^{\alpha'\beta'} = (\Lambda^{-1})^{\alpha}{}_{\alpha'}F^{\alpha'\beta'}(\Lambda^{-1})^{\beta}{}_{\beta'},$$

or $F = (\Lambda^{-1})F'(\Lambda^{-1})^T$. As the matrix Λ^{-1} in (6.20) is symmetric, $(\Lambda^{-1})^T = \Lambda^{-1}$,

$$F = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -H^3 & H^2 \\ E^2 & H^3 & 0 & -H^1 \\ E^3 & -H^2 & H^1 & 0 \end{pmatrix} = (\Lambda^{-1})F'(\Lambda^{-1})^T =$$

$$\begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -E^{1'} & -E^{2'} & -E^{3'} \\ E^{1'} & 0 & -H^{3'} & H^{2'} \\ E^{2'} & H^{3'} & 0 & -H^{1'} \\ E^{3'} & -H^{2'} & H^{1'} & 0 \end{pmatrix} \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The result of the matrix products is an equality which is equivalent to the set of equations (6.30):

$$F = \begin{pmatrix} 0 & -E^{1'} & -\gamma(E^{2'}+\beta H^{3'}) & -\gamma(E^{3'}-\beta H^{2'}) \\ E^{1'} & 0 & -\gamma(H^{3'}+\beta E^{2'}) & \gamma(H^{2'}-\beta E^{3'}) \\ \gamma(E^{2'}+\beta H^{3'}) & \gamma(H^{3'}+\beta E^{2'}) & 0 & -H^{1'} \\ \gamma(E^{3'}-\beta H^{2'}) & -\gamma(H^{2'}-\beta E^{3'}) & H^{1'} & 0 \end{pmatrix}.$$

§ 6.9 Suppose that the magnetic field vanishes in frame K': $\vec{H}' = 0$. In frame K,

$$\begin{aligned} E_x &= E'_x \\ E_y &= \gamma E'_y \\ E_z &= \gamma E'_z \\ H_x &= H'_x = 0 \\ H_y &= -\gamma\frac{v}{c}E'_z = -\frac{v}{c}E_z \\ H_z &= \gamma\frac{v}{c}E'_y = \frac{v}{c}E_y. \end{aligned} \tag{6.31}$$

Thus, in a way, an electric field in frame K' turns up as a magnetic field in frame K. A magnetic field which is zero in one frame appears very effectively in another. The relationship between $\vec{E}$ and $\vec{H}$ can, in frame K, be summed up as

$$c\vec{H} = \vec{v} \times \vec{E}. \tag{6.32}$$

$\vec{E}$ and $\vec{H}$ are clearly orthogonal to each other.

Suppose now that it is the electric field which vanishes in frame K': $\vec{E}' = 0$. This time, in frame K,

$$\begin{aligned} H_x &= H'_x \\ H_y &= \gamma H'_y \\ H_z &= \gamma H'_z \\ E_x &= E'_x = 0 \\ E_y &= \gamma \frac{v}{c} H'_z = \frac{v}{c} H_z \\ E_z &= -\gamma \frac{v}{c} H'_y = -\frac{v}{c} H_y. \end{aligned} \tag{6.33}$$

The relationship between $\vec{E}$ and $\vec{H}$ in frame K is now encapsulated in

$$c\vec{E} = -\vec{v} \times \vec{H}. \tag{6.34}$$

Also in this case $\vec{E}$ and $\vec{H}$ are orthogonal to each other. It is possible to show in general that, whenever a frame exists in which either $\vec{E}$ or $\vec{H}$ vanish, there is another frame in which they are orthogonal. And vice versa: if a frame exists in which they are orthogonal, there exists another frame in which one of them is zero.

Electric and magnetic fields cannot, of course, simply convert into each other by a change of frames. The electric field is a true vector, while the magnetic field is a pseudo-vector. Some other vector must be present to avoid parity violation. This is done by the velocity v in Eqs. (6.32) and (6.34).

6.3 Covariant Form of Maxwell's Equations

§ 6.10 Equations (6.1-6.4) are covariant, but this is not evident at first sight. They can be easily put into an explicitly covariant form in terms of the Maxwell tensor. Indeed, if we introduce the four-vector $j^\alpha = (\rho c, \vec{j})$, Maxwell's equations take the form

$$\partial_\alpha \tilde{F}^{\alpha\beta} = 0; \tag{6.35}$$

$$\partial_\alpha F^{\alpha\beta} = j^\beta. \tag{6.36}$$

Equation (6.35) includes the first pair of Maxwell's equations, Eqs. (6.1) and (6.2). Equation (6.36) includes the second pair, Eqs.(6.3) and (6.4). Notice that now the duality symmetry (6.5) in the absence of source becomes simply the exchange of $F^{\alpha\beta}$ by its dual $\tilde{F}^{\alpha\beta}$.

Exercise 6.2 Verify that the equations (6.35) and (6.36) are equivalent to the set (6.1)-(6.4). ■

Allied to the antisymmetry of $F^{\alpha\beta}$, Eq. (6.36) implies

$$\partial_\alpha j^\alpha = 0, \tag{6.37}$$

which is just the continuity equation (6.6).

There is a certain confusion in the standard language used in this subject. What is usually called "electromagnetic field" is the pair $(\vec{E}, \vec{H})$ or, if we prefer, the tensor $F^{\alpha\beta}$. These are the observables of the theory. Here, however, the role analogous to the basic fields of previous sections will be played by the electromagnetic potential A_α. Classical objects only see the field strength $F_{\alpha\beta}$.

Though it is in principle possible to work with variables $\vec{E}$ and $\vec{H}$, the covariant formulation turns out to be very complicated. Classically, this would nevertheless be justified, on the basis of $\vec{E}$ and $\vec{H}$ being observables and A_α not. In the quantum case, however, it is known that $\vec{E}$ and $\vec{H}$ are not able to describe all the facets of the electromagnetic field. The circulation $\oint_\gamma \vec{A} \cdot d\vec{l}$ of $\vec{A}$ along a closed curve γ can be measured as a wavefunction phase (the Bohm-Aharonov effect). In consequence, it will be through the vector field A_α that electromagnetism will de described, though we shall keep for it the name "potential". The reason behind that is a special character of A_α: it is a *gauge field* and deeply different from standard vector fields. This will be discussed in the chapter on gauge fields.

§ 6.11 The wave equation. We arrive at the field equation for A_α by taking (6.11) into (6.36):

$$\Box A^\alpha - \partial^\alpha(\partial_\beta A^\beta) = j^\alpha. \tag{6.38}$$

At the same time, (6.11) makes of (6.35) an identity. Actually, (6.35) is a 4-dimensional version of div $\vec{H} = 0$, more precisely it is div $F = 0$. And (6.11) expresses $F = \text{rot } A$ in the 4-dimensional case. Finally, taken together, definition (6.11) and Eq.(6.38) have the same content as Maxwell's equations.

Let us compare the sourceless case ($j^\alpha = 0$) with the Proca equation for a general vector field, Eq.(5.10). We see that A^α is a real vector field with zero mass. The difficulties with the non-positive energy turn up here again, with an additional problem: the subsidiary condition

$$\partial_\alpha A^\alpha = 0 \tag{6.39}$$

cannot be included automatically into any Lagrangean. In compensation, we have here an additional freedom of choice: function ϕ in (6.14) can be chosen at will. Given any A^α, choosing ϕ obeying

$$\Box\, \phi + \partial_\alpha A^\alpha = 0 \tag{6.40}$$

will lead to $\partial_\alpha A'^\alpha = 0$. The left-hand side of Eq. (6.38) is invariant under transformation (6.14): this is the "gauge invariance" of the theory. When we choose a particular A'^α in (6.14), we say that we are "fixing a gauge". In particular, a potential satisfying Eq.(6.39) is said to be "in the Lorenz gauge". In that gauge, the wave equation reduces to the *d'Alembert equation*

$$\Box A^\alpha = j^\alpha. \tag{6.41}$$

Notice that the Lorenz condition (6.39) does not fix A^α completely. To begin with, we can pass into the Lorenz gauge from any A^α by choosing a ϕ obeying (6.40). We have $\partial_\alpha A'^\alpha = 0$. But then a new transformation $A''^\alpha = A'^\alpha + \partial^\alpha \phi'$ with $\Box\phi' = 0$ (ϕ' is "harmonic") will take into the same condition for A''^α, $\partial_\alpha A''^\alpha = 0$. Thus, a potential in the Lorenz gauge is determined up to a gradient of a harmonic scalar. This additional freedom can be used to eliminate one of the components of A^α, for example A^0: choose ϕ' such that $\partial_0\phi' = -\ A'_0$; then we shall have $A''_0 = 0$ at any point $x = (\vec{x}, t)$. In that case, $\partial_0 A''_0 = 0$ and the Lorenz condition takes the form

$$\vec{\nabla} \cdot \vec{A} = 0 \quad ; \quad A^0 = 0. \tag{6.42}$$

This gauge is known as the radiation gauge, or Coulomb gauge. The choice is not explicitly covariant, but can be made in each inertial frame. If we pass into the momentum representation through the Fourier transform

$$A_\mu(x) = \frac{1}{(2\pi)^{3/2}} \int d^4k\ \delta(k^2) e^{ikx} A_\mu(k), \tag{6.43}$$

Equations (6.42) become

$$A^0(k) = 0, \tag{6.44}$$

$$\vec{k} \cdot \vec{A}(k) = 0. \tag{6.45}$$

The latter is a *transversality condition*. In this gauge, of the initial four components of the potential, only two remain independent. Notice that the massive vector field, once submitted to the transversality condition, remained with three independent components. The zero mass field has the extra condition (6.44), which reduces the number of components to two. As seen above, this is due to the gauge invariance. The relation between vanishing mass and gauge invariance is simple to see: in the Lagrangean (5.11) for the vector field, only the mass term is not gauge invariant. And let us repeat; although (6.44) and (6.45) are not covariant, it is possible to choose the ϕ's in each frame so as to realize them. Therefore, the electromagnetic field has only two independent components, and they can be taken as orthogonal to the propagation direction $\frac{\vec{k}}{|\vec{k}|}$. As said at the beginning of section 7.2, it usual to

forget about the infinity of degrees of freedom represented by each component and talk about each components as one degree. In this language, the electromagnetic field has two degrees of freedom.

§ 6.12 Field invariants. We verify the relationships

$$F_{\alpha\beta}F^{\alpha\beta} = -\ \tilde{F}_{\alpha\beta}\tilde{F}^{\alpha\beta} = -\ 2\ (\vec{E}^2 - \vec{H}^2). \tag{6.46}$$

$$F_{\alpha\beta}\tilde{F}^{\alpha\beta} = -\ 4\ \vec{E}\cdot\vec{H}. \tag{6.47}$$

These two expressions are contractions of Lorentz tensors and, consequently, Lorentz invariants. Contraction (6.47) is not, however, invariant under parity transformation: it is a pseudoscalar. It is, furthermore, a total derivative.

6.4 Lagrangian, Spin, Energy

§ 6.13 Let us go back to the Lagrangians (5.2) and (5.11) for the vector field. In the case of vanishing mass they become respectively

$$\mathcal{L} = -\ \tfrac{1}{2}(\partial_\mu\phi_\nu)(\partial^\mu\phi^\nu), \tag{6.48}$$

and

$$\mathcal{L}' = -\ \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{6.49}$$

Notice that only Eqs.(6.36), corresponding to the second pair, follow from these Lagrangians. Equations (6.35), the first pair, are not really field equations — they are identities holding automatically for any $F_{\alpha\beta}$ of the form (6.11). Notice further that, as announced, it is the potential A^α which plays the role of fundamental field: it is by taking variations with respect to it that the field equations are got from the Lagrangians.

As a Lagrangian for the sourceless field, (6.49) has many advantages:

1. its is explicitly invariant under gauge transformations;

2. by Eq.(6.46), it is explicitly invariant under the exchange of $F^{\alpha\beta}$ with its dual $\tilde{F}^{\alpha\beta}$ and is, consequently, duality invariant;

3. by the second equality in (6.46), it is written in terms of classical observables; notice that invariant (6.47) cannot be used as long as experiment support strict parity conservation in electromagnetism; furthermore, as a total derivative, it would give no local equation through the minimal action procedure; by the pure Lagrangian method, it actually gives identity (6.35), the first pair of Maxwell's equations;

4. $F^{\alpha\beta}$ has a clear mathematical meaning, as the four-dimensional rotational of the vector field.

Expression (6.48) is not explicitly gauge invariant. It seems simpler in the massive case, and with it calculations are indeed simpler. For gauge transformations with $\Box\phi = 0$, it is clearly invariant up to total divergences. As to (6.49), it is explicitly invariant, and differs from (6.48) by a term

$$\tfrac{1}{2}\ \partial_\mu(A_\nu\partial^\nu A^\mu) - \tfrac{1}{2}\ A_\nu\partial^\nu\partial_\mu A^\mu,$$

which only reduces to a pure divergence when the Lorenz condition is satisfied. We shall use (6.48) only for exercise (see Comment 6.1 below).

Maxwell's equations with sources are obtained by adding to (6.48) or (6.49) the coupling term

$$\mathcal{L}_I = -\ jA = -\ j_\mu A^\mu. \tag{6.50}$$

Notice that, under a gauge transformation, $\mathcal{L}_I$ acquires an extra term equal to $[-\ j_\mu\partial^\mu\phi]$, which is a total divergence $[-\ \partial^\mu(j_\mu\phi)]$ due to current conservation. This conservation is, consequently, related to gauge invariance.

Exercise 6.3 Find Eq.(6.36) from Eqs.(6.49) and (6.50). ■

Comment 6.1 Using (6.48), it is immediate to arrive at

$$\Theta^{\alpha\beta} = -\ (\partial^\alpha A_\gamma\partial^\beta A^\gamma) - \eta^{\alpha\beta}\mathcal{L}, \tag{6.51}$$

that is,

$$\Theta^{00} = -\ \tfrac{1}{2}\ (\partial^0 A_\gamma\partial^0 A^\gamma) + \tfrac{1}{2}\ (\partial_j A_\gamma\partial^j A^\gamma) \tag{6.52}$$

and

$$\Theta^{i0} = -\ (\partial^i A_\gamma\partial^0 A^\gamma). \tag{6.53}$$

We also obtain

$$S^\mu{}_{\alpha\beta} = A_\beta\partial^\mu A_\alpha - A_\alpha\partial^\mu A_\beta. \tag{6.54}$$

The spin density will be then given by

$$S^0{}_{ij} = -\ A_i\overset{\leftrightarrow}{\partial^0}A_j, \tag{6.55}$$

and the spin vector by

$$S_i = \tfrac{1}{2}\ \epsilon_{ijk}\int d^3x A_j\overset{\leftrightarrow}{\partial^0}A_k. \tag{6.56}$$

These expressions differ from the more usual ones, but allow us to use what has been said on the vector fields.

The more usual treatment starts from Lagrangian (6.49). The canonical energy–momentum density turns out to be

$$\Theta^{\alpha\beta} = F^{\nu\beta}\partial^{\alpha}A_{\nu} + \tfrac{1}{4}\,\eta^{\alpha\beta}F_{\gamma\delta}F^{\gamma\delta}\,. \tag{6.57}$$

The angular momentum density tensor is

$$M^{\mu\alpha\beta} = \Theta^{\mu\alpha}x^{\beta} - \Theta^{\mu\beta}x^{\alpha} + F^{\mu\alpha}A^{\beta} - F^{\mu\beta}A^{\alpha}, \tag{6.58}$$

expression in which we recognize the spin density tensor

$$S^{\mu\alpha\beta} = F^{\mu\alpha}A^{\beta} - F^{\mu\beta}A^{\alpha}. \tag{6.59}$$

The spin itself will be given by

$$S^{ij} = \int d^3x\; S^{0ij} = \int d^3x\; [F^{0i}A^{j} - F^{0j}A^{i}],$$

or, with $S_k = \frac{1}{2}\,\epsilon_{kij}S^{ij}$,

$$\vec{S} = \int d^3x\; \mathbf{E}\times\mathbf{A}\;. \tag{6.60}$$

Notice that, once (6.59) is known, it is possible to obtain the symmetrized energy–momentum (4.65):

$$\Theta_B^{\alpha\beta} = F^{\alpha\gamma}F_{\gamma}{}^{\beta} + \tfrac{1}{4}\,\eta^{\alpha\beta}F_{\gamma\delta}F^{\gamma\delta}. \tag{6.61}$$

The energy density takes on, up to an acceptable divergence, the value $\Theta^{00} = \frac{1}{2}\;(\mathbf{E}^2 + \mathbf{H}^2)$, usually presented in the standard form

$$\Theta^{00} = \frac{1}{8\pi}\;\left(\mathbf{E}^2 + \mathbf{H}^2\right), \tag{6.62}$$

(for which the notation W is frequently used) and so does the momentum density

$$\Theta^{i0} = (\mathbf{E}\times\mathbf{H})^{i}\,. \tag{6.63}$$

The vector

$$\mathbf{S} = \frac{c}{4\pi}\;\mathbf{E}\times\mathbf{H} \tag{6.64}$$

is called the *Poynting vector*, and measures the energy flux of the electromagnetic field. This has the dimension (energy × c)/volume. Consequently, the flux of momentum density is actually $\mathbf{S}/c^2$.

Comment 6.2 It is a historical misfortune that the same notation be currently used for spin and for the Poynting vector.

6.5 Motion of a Charged Particle

§ 6.14 We have examined the case of a free particle in §1.36, where the action

$$S = - mc \int ds$$

has been used. Let us see, through an example, what happens when a force is present. Consider the case of a charged test particle. The coupling of a particle of charge e to an electromagnetic potential A is given by $A_\alpha j^\alpha = e\ A_a u^\alpha$, as said in §1.32. The action along a curve is, consequently,

$$S_{em}[\alpha] = - \frac{e}{c} \int_\alpha A_\alpha u^\alpha ds = - \frac{e}{c} \int_\alpha A_\alpha dx^\alpha .$$

with a factor to give the correct dimension. The variation is

$$\delta S_{em}[\alpha] = - \frac{e}{c} \int_\alpha \delta A_\alpha dx^\alpha - \frac{e}{c} \int_\alpha A_\alpha d\delta x^\alpha = - \frac{e}{c} \int_\alpha \delta A_\alpha dx^\alpha + \frac{e}{c} \int_\alpha dA_\beta \delta x^\beta$$

$$= - \frac{e}{c} \int_\alpha \partial_\beta A_\alpha \delta x^\beta dx^\alpha + \frac{e}{c} \int_\alpha \partial_\alpha A_\beta \delta x^\beta dx^\alpha = - \frac{e}{c} \int_\alpha [\partial_\beta A_\alpha - \partial_\alpha A_\beta] \delta x^\beta \frac{dx^\alpha}{ds} ds$$

$$= - \frac{e}{c} \int_\alpha F_{\beta\alpha} u^\alpha \delta x^\beta ds \ ,$$

where

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \ \partial_\beta A_\alpha \ . \tag{6.65}$$

Combining the two pieces, the variation of the total action

$$S = -mc \int_P^Q ds - \frac{e}{c} \int_P^Q A_\alpha dx^\alpha \tag{6.66}$$

is

$$\delta S = \int_P^Q \left[\eta_{\alpha\beta}\ mc\ \frac{d}{ds}\ u^\alpha - \frac{e}{c}\ F_{\beta\alpha} u^\alpha \right]\ \delta x^\beta ds.$$

The extremal satisfies

$$mc\ \frac{du^\alpha}{ds} = \frac{e}{c}\ F^\alpha{}_\beta\ u^\beta. \tag{6.67}$$

This is the relativistic version of the Lorentz force law. It has the general form (1.89).

§ 6.15 Let us rewrite the action (6.66) in the form

$$S = - mc \int \left(1 + \frac{e}{mc^2}\ A_\alpha U^\alpha \right) ds. \tag{6.68}$$

We can use the same trick which has led to Eq.(1.91) to isolate the time variable — use Eq.(1.65), $ds = cdt/\gamma$ — and the Lagrangian. It is necessary to take into account also Eqs.(1.68) and (1.74),

$$U = \gamma \ (1, \vec{v}/c) \quad \text{and} \quad A = \left(\phi, \vec{A}\right),$$

to arrive at

$$L = -\, mc^2 \sqrt{1 - \frac{v^2}{c^2}} + \frac{e}{c}\vec{A} \cdot \vec{v} - e\phi. \tag{6.69}$$

Exercise 6.4 Fill in the details to get Eq.(6.69) from Eq.(6.68). ■

The momentum canonically conjugate to the position variable $\vec{x}$ is now

$$\vec{P} = \frac{\partial L}{\partial \vec{v}} = \gamma m \vec{v} + \frac{e}{c}\vec{A} = \vec{p} + \frac{e}{c}\vec{A}, \tag{6.70}$$

where $\vec{p}$ is the free-particle momentum (1.93). This modification of the free momentum to the real conjugate momentum in the presence of interaction,

$$\vec{p} \ \Rightarrow \ \vec{p} + \tfrac{e}{c}\vec{A} \tag{6.71}$$

seems to have been, historically, the first version of the so-called *minimal coupling rule*.

§ 6.16 The Hamiltonian is

$$H = \vec{v} \cdot \frac{\partial L}{\partial \vec{v}} \ - \ L = \gamma mc^2 + e\phi. \tag{6.72}$$

Exercise 6.5 Again, fill in the details to get this expression. ■

In consequence, the complete minimal coupling rule says: in order to take the electromagnetic interaction of a particle into account, simply add $\frac{e}{c}\vec{A}$ to its free momentum and $e\phi$ to its free energy.

§ 6.17 The covariant version (6.67) of the Lorentz force law

$$\frac{dp^\alpha}{ds} = \frac{e}{c}\, F^\alpha{}_\beta \, U^\beta$$

can be decomposed into three space- and one time-components. Recall (i) the definition (1.100) of the four-momentum, $p = mcU = \gamma(mc, m\mathbf{v})$; (ii) that $F^{i0} = E^i$

and $F^{ij} = \epsilon^{ij}{}_k H^k$. Then, again using Eqs.(1.68) and (1.74), as well as Eq.(1.65) under the form $\frac{d}{ds} = \frac{\gamma}{c}\frac{d}{dt}$, we find

the Lorentz force law proper,

$$\dot{\vec{p}} = \frac{d}{dt}\,[m\gamma\vec{v}] = e\;\vec{E} + \frac{e}{c}\;\vec{v}\times\vec{H} \tag{6.73}$$

and the energy time variation,

$$\frac{d\mathcal{E}}{dt} = \frac{d}{dt}\;\gamma mc^2 = e\;\vec{E}\cdot\vec{v}. \tag{6.74}$$

Notice that the particle energy remains constant in time if the field is purely magnetic.

Exercise 6.6 Verify that $\frac{d\mathcal{E}}{dt} = \vec{v}\cdot\frac{d\vec{p}}{dt}$. This is always valid. Here, it turns up trivially. ■

§ 6.18 Constant and uniform fields An electromagnetic field is said to be *constant* when it does not change with time. In that case, the potentials ϕ and $\vec{A}$ depend only on the space position $\vec{x}$. Or, in wider generality: some gauge can be chosen in which that happens – this is enough. In that case, the second of Maxwell's equations, (6.2), keeps the form $\mathrm{div}\vec{H} = 0$ but the first (6.1) reduces to $\mathrm{rot}\;\vec{E} = 0$. This means that we still have $\vec{H} = \mathrm{rot}\;\vec{A}$, but now $\vec{E}$ is a pure gradient. Equation (6.13) gives

$$\vec{E} = -\;\mathrm{grad}\;\phi. \tag{6.75}$$

Here the only possible gauge addition to ϕ is a constant, in principle arbitrary. In most cases we fix the value of ϕ at some point (frequently, $\phi(\vec{x}) \to 0$ when $\vec{x} \to \infty$). Once this is done, the scalar potential is uniquely defined. The same is *not* true of the vector potential: it remains defined only up to a gradient of an arbitrary function of $\vec{x}$.

There is a specially simple particular case: when $\vec{E}$ and $\vec{H}$ have also the same values at every point $\vec{x}$. In that case the electromagnetic field is said to be *uniform*. When $\vec{E}$ is a constant vector, then its relation to ϕ can be reversed:

$$\phi = -\;\vec{E}\cdot\vec{x}. \tag{6.76}$$

Exercise 6.7 Find (6.75) from (6.76), recalling that the i-th component of the gradient of a scalar product is given by

$$[\mathrm{grad}\;(\vec{a}\cdot\vec{b})]_i = \partial_i(a_j b_j).$$

■

And if $\vec{H}$ is a constant vector, then its relation to $\vec{A}$ can be reversed:

$$\vec{A} = \tfrac{1}{2}\ \vec{H} \times \vec{x}. \tag{6.77}$$

Exercise 6.8 Verify that (6.77) leads indeed to $\vec{H} = \text{rot}\ \vec{A}$, if H is a constant vector. Use

$$[\text{rot}\ (\vec{a} \times \vec{b})]_k = \epsilon_{kij}\partial_i(\vec{a} \times \vec{b})_j = \epsilon_{kij}\partial_i(\epsilon_{jrs}a_r b_s) = \epsilon_{kij}\epsilon_{jrs}\partial_i(a_r b_s),$$

and then one of the contractions of Exercise 1.7. ■

The next two paragraphs show two rather unrealistic exercises, intended to fix some ideas about the probing of fields by particles. They examine the motions of a charged particle in uniform constant electric and magnetic fields, forgetting the current produced by those very motions.

§ 6.19 Motion in a uniform constant electric field In a uniform constant electric field, the Lorentz force law (6.73) reduces to

$$\dot{\vec{p}} = e\ \vec{E}. \tag{6.78}$$

As $\vec{E}$ is fixed both in time and space, it defines a preferred direction. Let us choose the cartesian coordinates axes in such a way that $\vec{E}$ stands along the Ox direction. A charged particle will have a motion governed by

$$\dot{p}_x = eE \quad ; \quad \dot{p}_y = 0 \quad ; \quad \dot{p}_z = 0$$

$$\therefore \quad p_x(t) = p_x(0) + eEt \quad ; \quad p_y(t) = p_y(0) \quad ; \quad p_z(t) = p_z(0).$$

Suppose the charged particle has initially only one momentum component, along Oy. In that case $p_x(0) = 0$, $p_y(0) = p_0$, $p_z(0) = 0$, and

$$p_x(t) = eEt \quad ; \quad p_y(t) = p_0 \quad ; \quad p_z(t) = 0.$$

To find the velocity it is better to use Eq.(1.96),

$$\vec{v} = \frac{\vec{p}c^2}{\mathcal{E}}\ ,$$

with the energy $\mathcal{E}$ given in the present case by

$$\mathcal{E} = \sqrt{m^2c^4 + \vec{p}^2c^2} = \sqrt{m^2c^4 + p_0^2c^2 + (eEct)^2} = \sqrt{\mathcal{E}_0^2 + (eEt)^2c^2}\ ,$$

where

$$\mathcal{E}_0 = \sqrt{m^2c^4 + p_0^2c^2}$$

is the particle energy at start. Then,

$$v_x = \frac{dx}{dt} = \frac{p_x c^2}{\mathcal{E}} = \frac{eEtc^2}{\sqrt{\mathcal{E}_0^2 + (eEct)^2}} \ ;$$

$$v_y = \frac{dy}{dt} = \frac{p_y c^2}{\mathcal{E}} = \frac{p_0 c^2}{\sqrt{\mathcal{E}_0^2 + (eEct)^2}} \ .$$

As $p_z(t) = 0$, the motion will take place on the plane xy. The integrations give

$$x(t) = \sqrt{\left(\frac{\mathcal{E}_0}{eE}\right)^2 + c^2 t^2} \ \ ; \ \ y(t) = \frac{p_0 c}{eE} \ \text{arcsinh} \ \frac{eEct}{\mathcal{E}_0}. \tag{6.79}$$

The choice of initial conditions ($x(0) = \frac{\mathcal{E}_0}{eE}$, $y(0) = 0$) has been made so as to make simpler to get the equation of the trajectory, which is found by eliminating t:

$$x = \frac{\mathcal{E}_0}{eE} \cosh \frac{eEy}{p_0 c} \ . \tag{6.80}$$

As cosh z $\approx 1 + \frac{1}{2}z^2$, the non-relativistic limit gives a parabola.

§ 6.20 Motion in a uniform constant magnetic field In this case, the Lorentz force law (6.73) reduces to

$$\dot{\vec{p}} = \frac{e}{c} \ \vec{v} \times \vec{H}. \tag{6.81}$$

We have called attention below Eq.(6.74) to the fact that the energy remains constant in time if the field is purely magnetic. It will be, consequently, very convenient to use also here Eq.(1.96),

$$\vec{p} = \frac{\mathcal{E}\vec{v}}{c^2} \ .$$

The equation becomes

$$\frac{\mathcal{E}}{c^2}\frac{d\vec{v}}{dt} = \frac{e}{c} \ \vec{v} \times \vec{H},$$

or

$$\frac{d\vec{v}}{dt} = \frac{ec}{\mathcal{E}} \ \vec{v} \times \vec{H}. \tag{6.82}$$

It will be also convenient to introduce the notation

$$\omega = \frac{ecH}{\mathcal{E}} \ . \tag{6.83}$$

If we now choose the axes so that $\vec{H}$ lies along Oz, the equations take the forms

$$\dot{v}_x = \omega v_y \ \ ; \ \ \dot{v}_y = -\ \omega v_x \ \ ; \ \ \dot{v}_z = 0. \tag{6.84}$$

A first integration gives, with convenient integration constants,

$$v_x = v_0 \cos \omega t \ \ ; \ \ v_y = -\ v_0 \sin \omega t \ \ ; \ \ v_z = v_{0z} = \text{constant}.$$

The absolute value of the velocity on plane xy is time-independent, as $v_0^2 = v_x^2 + v_y^2$. The final solution is

$$x = x_0 + \frac{v_0}{\omega}\sin\omega t \quad ; \quad y = y_0 + \frac{v_0}{\omega}\cos\omega t \quad ; \quad z = z_0 + v_{0z}t. \tag{6.85}$$

On the plane xy the particle performs a rotation of frequency ω and radius $\frac{v_0}{\omega}$ around the point (x_0, y_0). Combined with the uniform motion along the axis Oz, the trajectory is a helix whose axis lies along the direction of $\vec{H}$. If $v_{0z} = 0$, of course, the motion will remain on plane xy. Notice further that the motion only takes place when the particle is initially "injected" with some velocity $v_0 \neq 0$.

6.6 Electrostatics and Magnetostatics

§ 6.21 Electrostatics. In a particular frame, we may have non-uniform but time-independent fields $\vec{E}$ or $\vec{H}$. Let us begin with the case $\vec{H} = 0$. Maxwell's equations reduce to

$$\text{rot } \vec{E} = 0 \tag{6.86}$$

$$\text{div } \vec{E} = 4\pi\,\rho. \tag{6.87}$$

The first equation again implies (6.75), $\vec{E} = -\text{ grad } \phi$. Together with (6.87), we have that necessarily

$$\Delta\phi = -\,4\pi\,\rho. \tag{6.88}$$

This is the *Poisson equation*. Recall that, in cartesian coordinates,

$$\Delta\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}\ . \tag{6.89}$$

The sourceless case ($\rho = 0$) leads to the *Laplace equation*

$$\Delta\phi = 0. \tag{6.90}$$

This is probably the most important equation of Mathematical Physics. It turns up in a wide variety of domains, from Music (a function ϕ satisfying it is called a *harmonic*) to Topology. Given a set of boundary conditions, it has a large number of elementary solutions (in the spherically symmetrical case, "spherical harmonics"). As it is a linear equation, the general solution is a superposition of those. This means that the space of solutions is a vector space, for which the elementary solutions constitute a base.

The equation imposes some strict conditions on the solutions. For example, the signs of the second derivatives cannot be all the same and, consequently, they can have neither maxima nor minima.

Let us consider a point-like particle of charge e. Such a system has spherical symmetry around the particle. We can fix the particle at the origin of a spherical coordinate system. The field $\vec{E}$ created at a point (r, θ, φ) will have an absolute value depending only on the distance to the origin, which is r. Notice that the origin itself, the very site of the particle, is excluded. The flux through a surface of fixed radius r will be

$$\int_{S=\partial V} d\vec{S} \cdot \vec{E} = r^2 E \int d\varphi d(\cos\theta) = 4\pi\ r^2 E.$$

On the other hand, from the Gauss law (6.7), this must be $4\pi\ \int_V \rho d^3x = 4\pi\ e$, so that

$$E = \frac{e}{r^2}.$$

As a vector field with no rotational [by Eq.(6.86)], $\vec{E}$ must then be given by *Coulomb's law* :

$$\vec{E} = \frac{e}{r^3}\ \vec{r}\ . \tag{6.91}$$

The potential will be

$$\phi = \frac{e}{r}\ . \tag{6.92}$$

The origin has been excluded. The point-like charge at $\vec{r} = 0$ can be introduced as $e\delta^3(\vec{r})$. In fact, a detailed Fourier analysis shows that

$$\Delta\ \frac{e}{r} = -\ 4\pi e\delta^3(\vec{r})\ . \tag{6.93}$$

Because the equation is linear, the field created by a set of charges $(e_1, e_2, e_3, ...)$ will be the sum of the fields created by the individual charges e_k. If the point at which the field is to be measured stands at a distance r_1 of the charge e_1, at a distance r_2 of the charge e_2, in short at a distance r_i of the charge e_i, then

$$\phi = \sum_i \frac{e_i}{r_i}\ . \tag{6.94}$$

In the case of a continuum distribution of charges with density ρ, the solution for the Poisson equation will be

$$\phi = \int d^3x\ \frac{\rho}{r(x)}\ . \tag{6.95}$$

The charge in a volume element d^3x will be ρd^3x, and the variable $r(x)$ represents the distance between that volume element and the point at which the field is to be measured.

We have discussed fields created by electrically-charged pointwise particles. We could have said: by electric monopoles. There is no magnetic analogous to such fields, as there are no magnetic monopoles.

§ 6.22 Magnetostatics. We may have $\vec{E} \equiv 0$ in a particular frame. Maxwell's equations reduce to

$$\frac{\partial \vec{H}}{\partial t} = 0 \tag{6.96}$$

$$\text{div } \vec{H} = 0 \tag{6.97}$$

$$c \text{ rot } \vec{H} = 4\pi \vec{j}. \tag{6.98}$$

We learn from the first equation above that $\vec{E} \equiv 0$ implies that $\vec{H}$ is constant in time. As there are no magnetic charges, the only possible source of a pure magnetic field is a current $\vec{j}$ produced by electric charges in motion. The system is not really static, but can be made to be stationary by a device: take the time average of all fields over a large lapse. We shall use the "hat" notation $\widehat{\mathbf{H}}$, $\widehat{\mathbf{E}}$, etc for these averages. For example,

$$\widehat{\mathbf{H}} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{H}(t)dt\,.$$

Notice that the time average of a time derivative vanishes for any quantity which remains finite. Thus,

$$\widehat{\frac{dE^k}{dt}} = \frac{1}{T} \int_0^T \frac{dE^k}{dt} dt = \frac{E^k(T) - E^k(0)}{T} \to 0$$

for T large enough.

It is better to take back the complete Maxwell's equations. Once the averages are taken, only remain

$$\text{div } \widehat{\mathbf{H}} = 0 \tag{6.99}$$

$$c \text{ rot } \widehat{\mathbf{H}} = 4\pi \widehat{\mathbf{j}}. \tag{6.100}$$

We can then introduce an average potential vector such that

$$\text{rot } \widehat{\mathbf{A}} = \widehat{\mathbf{H}}, \tag{6.101}$$

which will consequently obey

$$c \text{ rot rot } \widehat{\mathbf{A}} = c \text{ grad div } \widehat{\mathbf{A}} - c\, \Delta\widehat{\mathbf{A}} = 4\pi\, \widehat{\mathbf{j}}.$$

Exercise 6.9 Show that

$$\text{rot rot } \vec{V} = \text{grad div } \vec{V} - \Delta \vec{V}.$$

■

Now we choose the Coulomb gauge (6.42), div $\widehat{\mathbf{A}} = 0$ and remain with

$$\Delta \widehat{\mathbf{A}} = -\frac{4\pi}{c}\,\widehat{\mathbf{j}}. \tag{6.102}$$

This is now the Poisson equation for each component of $\widehat{\mathbf{A}}$, and the solution can be obtained by analogy with the electrostatic case. Equation (6.95) will lead then to

$$\widehat{\mathbf{A}} = \frac{4\pi}{c} \int d^3x\, \frac{\widehat{\mathbf{j}}}{r}. \tag{6.103}$$

The magnetic field will be

$$\widehat{\mathbf{H}} = \text{rot } \widehat{\mathbf{A}} = \frac{4\pi}{c}\ \text{rot} \int d^3x\, \frac{\widehat{\mathbf{j}}}{r(x)}\ . \tag{6.104}$$

Exercise 6.10 Show that, if f is a function,

$$\text{rot } (f\vec{V}) = f\ \text{rot } \vec{V} + (\text{grad } f) \times \vec{V}.$$

■

The rot operator acts only on r, which represents the point at which the field is to be measured. It can consequently be introduced into the integral. The average current is integrated, and can be taken as constant. Thus, rot $\widehat{\mathbf{j}} = 0$ and rot $\left(\frac{\widehat{\mathbf{j}}}{r}\right) =$ grad $\left(\frac{1}{r}\right) \times \widehat{\mathbf{j}} = \widehat{\mathbf{j}} \times \frac{\mathbf{r}}{r^3}$. We thus arrive at

$$\widehat{\mathbf{H}} = \frac{4\pi}{c} \int d^3x\, \widehat{\mathbf{j}} \times \frac{\mathbf{r}}{r^3}. \tag{6.105}$$

This is the *Biot-Savart law*.

6.7 Electromagnetic Waves

§ 6.23 The wave equation Let us go back to the wave equation (6.41) in the Lorenz gauge. We shall actually consider only the sourceless case,

$$\Box A^\alpha = 0\ , \tag{6.106}$$

and have in mind (for nomenclature, for example) the most important of electromagnetic waves, light waves. The Lorenz condition does not entirely fix the gauge. If we use the radiation gauge (6.42)

$$\text{div } \vec{A} = 0 \;\; ; \;\; A^0 = \phi = 0, \tag{6.107}$$

only the vector potential remains, and for it Eq.(6.106) can be written

$$\Delta \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = 0 \; . \tag{6.108}$$

What about the fields $\vec{E}$ and $\vec{H}$? Besides the inevitable $\vec{H} = \text{rot } \vec{A}$, we have from (6.13) that

$$\vec{E} = - \frac{1}{c} \frac{\partial \vec{A}}{\partial t} \; .$$

It is easy to see that both $\vec{E}$ and $\vec{H}$ satisfy that same equation. Thus, the electromagnetic potential, the electric field and the magnetic field all obey the same equation.

Exercise 6.11 Verify the statement above, by applying the operators rot and $\frac{\partial}{\partial t}$ to (6.108). ■

Furthermore, that equation is the same for each component of $\vec{A}$, $\vec{E}$ and $\vec{H}$. It is consequently enough to examine the equation as holding for one component, as for a function $f(\vec{x}, t)$:

$$\frac{1}{c^2} \frac{\partial^2 f(\vec{x}, t)}{\partial t^2} - \Delta f(\vec{x}, t) = 0 \; . \tag{6.109}$$

§ 6.24 Plane waves A solution of the electromagnetic wave equation is said to be a *plane wave* when the fields ($\vec{A}$, $\vec{E}$, $\vec{H}$) depends only on one of the space coordinates: for example, when the function above is $f(\vec{x}, t) = f(x, t)$. The wave equation becomes

$$\frac{\partial^2 f(x,t)}{\partial t^2} - c^2 \frac{\partial^2 f(x,t)}{\partial x^2} = 0 \; , \tag{6.110}$$

which is the same as

$$\left[\frac{\partial}{\partial t} - c \frac{\partial}{\partial x} \right] \left[\frac{\partial}{\partial t} + c \frac{\partial}{\partial x} \right] f(x,t) = 0 \; . \tag{6.111}$$

It is convenient to make a change of coordinates so that each of the above bracketed expression becomes a simple derivative. Such coordinates are

$$\xi = t - \frac{x}{c} \;\; ; \;\; \zeta = t + \frac{x}{c},$$

or

$$t = \tfrac{1}{2} \, (\zeta + \xi) \;\; ; \;\; \tfrac{x}{c} = \tfrac{1}{2} \, (\zeta - \xi).$$

In that case,

$$\frac{\partial}{\partial \xi} = \tfrac{1}{2}\ \left(\tfrac{\partial}{\partial t} - c\ \tfrac{\partial}{\partial x}\right)\ \ ;\ \ \tfrac{\partial}{\partial \zeta} = \tfrac{1}{2}\ \left(\tfrac{\partial}{\partial t} + c\ \tfrac{\partial}{\partial x}\right).$$

The wave equation acquires the aspect

$$\frac{\partial^2 f(\xi,\zeta)}{\partial \xi \partial \zeta} = 0.$$

The solution has, consequently, the form

$$f(x,t) = f(\xi,\zeta) = f_1(\xi) + f_2(\zeta) = f_1\left(t - \frac{x}{c}\right) + f_2\left(t + \frac{x}{c}\right), \tag{6.112}$$

where the single-argument functions f_1 and f_1 are arbitrary. Let us examine the meaning of this solution. Suppose first $f_2(\zeta) = 0$, so that $f(x,t) = f_1(\xi) = f_1\left(t - \frac{x}{c}\right)$. Fix the plane $x = \textit{constant}$: on that plane, the field changes with time at each point. On the other hand, at fixed t, the field is different for different values of x. Nevertheless, the field will have the same value every time the variables t and $\frac{x}{c}$ satisfy the relation $t - \frac{x}{c} =$ a constant, that is, when

$$x = K + ct.$$

If the field has a certain value at $t = 0$ at the point x, it will have that same value after a time t at a point situated at a distance ct from x. Take that value of the field: we can say that *that value* "propagates" along the axis Ox with velocity c. We say, more simply, that the *field* propagates along the axis Ox with the velocity of light. The solution $f_1\left(t - \frac{x}{c}\right)$ represents a plane wave propagating with the velocity of light in the positive sense of Ox. The same analysis leads to the conclusion that the solution $f_2\left(t + \frac{x}{c}\right)$ represents a plane wave propagating in the negative sense of Ox with the velocity of light. The general solution is therefore a superposition of two plane waves, one advancing along Ox, the other traveling in the inverse sense.

The condition for the radiation gauge (6.107) is here simply

$$\frac{\partial A_x}{\partial x} = 0.$$

In this gauge, the component A_x is constant in space. The wave equation says then that $\frac{\partial^2 A_x}{\partial t^2} = 0$, or $\frac{\partial A_x}{\partial t} =$ constant. This would say that the electric field $E_x =$ constant, not a wave. $\vec{E}$ can only have components orthogonal to the direction of propagation. Furthermore, $A_x = 0$ if we are looking for wave solutions. It follows that the electromagnetic potential is always perpendicular to the axis Ox, that is, to the direction of the plane wave propagation.

Consider a plane wave progressing along Ox. All the field variables depend only on $\left(t - \frac{x}{c}\right)$. From the first of relations

$$\vec{E} = -\frac{\partial \vec{A}}{\partial ct} \quad \text{and} \quad \vec{H} = \text{rot } \vec{A}$$

we obtain

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial \xi} \ . \tag{6.113}$$

The second is

$$H_k = \epsilon_k{}^{ij}\frac{\partial}{\partial x^i}A_j(t - x/c) = \epsilon_k{}^{1j}\frac{\partial}{\partial x}A_j(t - x/c) = -\frac{1}{c}\,\epsilon_k{}^{1j}\frac{\partial}{\partial \xi}A_j(\xi)$$

$$\therefore\ H_k = \epsilon_k{}^{1j}E_j(\xi) \quad \therefore \quad H_z = E_y \ \ ; \ \ H_y = -\,E_z \ \ ; \ \ H_x = 0\ . \tag{6.114}$$

These relations can be put together by using the unit vector along the direction of propagation, $\vec{n} = (1, 0, 0)$. Then we verify that

$$\vec{H} = \vec{n} \times \vec{E}. \tag{6.115}$$

In a plane wave, the electric and the magnetic fields are orthogonal to each other and to the direction of propagation. These are *transversal waves*. We see from (6.114) that $\vec{E}$ and $\vec{H}$ have the same absolute values.

The energy flux of a plane wave field will be given by the Poynting vector (6.64)

$$\vec{S} = \frac{c}{4\pi}\,\vec{E} \times \left(\vec{n} \times \vec{E}\right) = \frac{c}{4\pi}\,E^2\vec{n} = \frac{c}{4\pi}\,H^2\vec{n}. \tag{6.116}$$

The energy flux is carried by a plane wave along its direction of propagation. The energy density (6.62) will be

$$W = \frac{1}{8\pi}\,\left(\vec{E}^2 + \vec{H}^2\right) = \frac{1}{4\pi}\,\vec{E}^2 = \frac{1}{4\pi}\,\vec{H}^2$$

so that

$$\vec{S} = c\,W\vec{n}. \tag{6.117}$$

As said below Eq.(6.64) the momentum density is $\vec{S}/c^2$. For a plane wave, this is $W\vec{n}/c$. Thus, for an electromagnetic plane wave, the relation between the energy W and the momentum W/c is the same as that for particles traveling at the velocity of light, Eq.(1.97).

§ 6.25 Monochromatic plane waves Let us now consider solutions closer to our simplest intuitive idea of wave: suppose the above field is a periodic function of time. This means that all the quantities characterizing the field — the components

of $\vec{A}$, $\vec{E}$ and $\vec{H}$ — involve time in the forms $\cos(\omega t + \alpha)$ and/or $\sin(\omega t + \alpha)$. In that case, much can be said in a purely qualitative discussion. The unique time multiplier ω will be the wave *frequency*. In applications to Optics it appears as the light frequency. Now, light with a single frequency means light with a single color, and for this reason such waves are called *monochromatic*. For expressions of the form

$$f = a\cos\omega t + b\sin\omega t \tag{6.118}$$

the second time derivative will always satisfy $\frac{\partial^2 f}{\partial t^2} = -\,\omega^2 f$. The wave equation gives then, for the space part,

$$\Delta f + \frac{\omega^2}{c^2}\, f = 0\ . \tag{6.119}$$

Take now a wave traveling along the Ox axis in the positive direction. We have seen that in that case the wave will depend only on the variable $\xi = t - \frac{x}{c}$. To fix the ideas, lets us consider the quantity $\vec{A}$. Instead of (6.118), we can use

$$\vec{A} = \vec{A}_0 e^{-i\omega\xi} = \vec{A}_0 e^{-i\omega(t-x/c)}, \tag{6.120}$$

with $\vec{A}_0$ a constant complex vector.

Exercise 6.12 Verify that (6.120) satisfies Eq.(6.119). ■

Fields $\vec{E}$ and $\vec{H}$ will have analogous expressions. At fixed time and given an initial value of the field, the wave comes back to that value at a distance $x = \lambda$ given by

$$\lambda = \frac{2\pi c}{\omega}. \tag{6.121}$$

This "length of one wave" is the *wavelength*. And, given the above-defined unit vector $\vec{n}$ along the propagation direction, the *wave vector* $\vec{k}$ is defined as

$$\vec{k} = \frac{\omega}{c}\ \vec{n}. \tag{6.122}$$

Representation (6.120) can then be rewritten as

$$\vec{A} = \vec{A}_0\ e^{i(\vec{k}\cdot\vec{r}-\omega t)}. \tag{6.123}$$

The expression $(\vec{k}\cdot\vec{r}-\omega t)$ is the wave *phase*. We can actually introduce a four-vector

$$k = \left(\frac{\omega}{c},\ \vec{k}\right) = \frac{\omega}{c}\ (1,\ \vec{n}) \tag{6.124}$$

which will be such that $k_\alpha x^\alpha = (\omega t - \vec{k}\cdot\vec{r})$. We see that k_α is a null vector, or a light-like vector: $k_\alpha k^\alpha = 0$. This comes also from the fact that

$$\vec{A} = \vec{A}_0\ e^{-ik_\alpha x^\alpha} \tag{6.125}$$

must be a solution of the wave equation. Taking all this into Eqs.(6.113, 6.115), we obtain

$$\vec{E} = ik\vec{A} \quad ; \quad \vec{H} = i\vec{k} \times \vec{A}. \tag{6.126}$$

§ 6.26 Doppler effect The behavior of the 4-vector (6.124), when seen from different reference frames, leads to an important effect. Suppose a wave (such as a light beam) is emitted from a (source) frame K_S with a 4-vector $k_{(S)} = \frac{\omega_S}{c}\ (1,\ \vec{n})$ towards another (receptor) frame K_R in the direction of $\vec{n}$. Suppose $\vec{v}$ is the velocity of the source, or of K_S, which moves towards K_R along the axis $0x$. The latter will see $k_{(R)} = \frac{\omega_R}{c}\ (1, 1, 0, 0)$. As the distance separating K_S and K_R, as well as the unit vector $\vec{n}$, lie along $0x$, $k_{(S)} = \frac{\omega_S}{c}\ (1, 1, 0, 0)$.

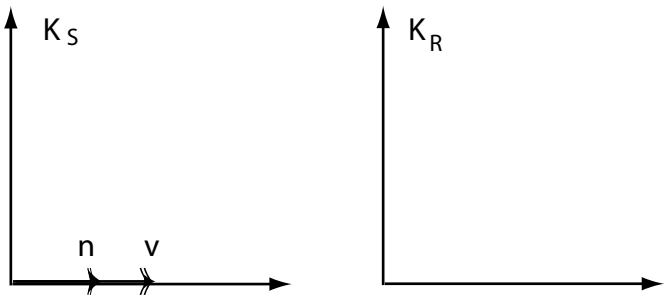


Figure 6.1: *Scheme for the Doppler effect.*

By the transformation laws for four-vectors,

$$k^0_{(S)} = \gamma \left[k^0_{(R)} - \frac{v}{c}\ k^1_{(R)} \right].$$

$$\frac{\omega_S}{c} = \gamma \left[\frac{\omega_R}{c} - \frac{v}{c}\ \frac{\omega_R}{c} \right] \quad \therefore \quad \omega_S = \gamma \omega_R \left[1 - \frac{v}{c} \right]$$

$$\therefore \quad \omega_R = \omega_S\ \frac{\sqrt{1 - v^2/c^2}}{1 - \frac{v}{c}} = \omega_S \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}$$

To consider the case in which the source, instead of moving towards the receptor, moves away from it (always along $0x$), it is enough to invert the sign of v above. We have thus the two opposite cases:

$$\omega_R = \omega_S \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}} \ > \ \omega_S \quad \text{(blue shift)} \ .$$

and

$$\omega_R = \omega_S \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \ < \ \omega_S \quad \text{(red shift)} \ .$$

The terminology, as usual, takes from Optics: a light ray emitted by an approaching source is seen with its frequency displaced towards higher values (bluer for visible light).

Chapter 7

Scalar Fields

7.1 Real Scalar Fields

§ 7.1 These are the simplest relativistic fields, defined in page 93 as those which are, on the whole, Lorentz invariant. They satisfy the Klein-Gordon equation:

$$(\Box + m^2)\ \phi(x)\ =\ 0\ . \tag{7.1}$$

We are using units with $c = 1$ and $\hbar = 1$. We shall see below that real scalar fields describe neutral mesons.

Comment 7.1 Schrödinger has found this wave equation even before he found his famous nonrelativistic equation – but discarded it because it led to negative-energy solutions for free states – for which at that time there was no interpretation.

The D'Alembertian operator $\Box$ is simply the Laplace operator in 4-dimensional Minkowski space. In Cartesian coordinates $\{x^\alpha\}$, in terms of which the Lorentz metric is $\eta = \text{diag}\ (1, -1, -1, -1)$ and thus coordinate–independent, it is

$$\Box = \eta^{\alpha\beta}\partial_\alpha\partial_\beta = \partial^\alpha\partial_\alpha = \partial_0\partial_0 - \partial_1\partial_1 - \partial_2\partial_2 - \partial_3\partial_3\ . \tag{7.2}$$

The Klein-Gordon equation describes the field in absence of any source, that is, in absence of any interaction. It is the simplest relativistic adaptation of the Schrödinger equation, actually the Poincaré invariant $P_\mu P^\mu = m^2$ of § 2.27, with $P^\mu = i\,\partial^\mu$, applied to the field $\phi(x)$. This means that, as a matter of fact,

every relativistic field satisfies the Klein-Gordon equation.

We introduce it as *the* equation for scalar fields because it is the only equation they satisfy. Other fields either obey to equations implying the Klein-Gordon equation

(as the Dirac fields) or obey the Klein-Gordon equation plus some supplementary conditions (as the higher-spin fields).

Source terms can be added in the right-hand side, but for the time being we examine the sourceless equation. As seen in Exercise 4.1, the equation comes out as the Euler-Lagrange equation of the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\ [\partial_\mu\phi\partial^\mu\phi\ -\ m^2\phi^2]\ , \tag{7.3}$$

or, if we prefer, from

$$\mathcal{L} = \tfrac{1}{2}\ \phi\ [\ \Box\ +\ m^2]\ \phi\ , \tag{7.4}$$

which differs from the previous one only by a total divergence. When using the second version, we should not forget to consider terms up to second order in the Lagrangian derivative. The energy-momentum tensor density, whose general expression is (4.49), comes easily from (7.3):

$$\Theta^{\lambda\mu} = \partial^\lambda\phi\partial^\mu\phi\ -\ \eta^{\lambda\mu}\mathcal{L}\ . \tag{7.5}$$

In particular, the densities of momenta

$$\Theta^{0i} = \partial^0\phi\ \partial^i\phi \tag{7.6}$$

and of energy

$$\Theta^{00} = \tfrac{1}{2}\ [\partial_0\phi\ \partial_0\phi + \partial_i\phi\ \partial_i\phi\ + m^2\ \phi^2]\ . \tag{7.7}$$

Notice that, as a summation of real squares, this expression is always positive. It is of course to be expected that the energy density be positive in the absence of interactions.

The fact that θ^{00} is positive leads, by the way, to a criterion for the presence of a field. We might ask on which region of spacetime is some field $\phi(x)$ really present. As any contribution of the field adds up a positive quantity, we can say that the field is present on every point x at which $\theta^{00} > 0$ and absent wherever $\theta^{00} = 0$.

As to the angular momentum, we have seen that it reduces to the orbital part in this case.

Of course, a field can interact with other fields, or with itself. The study of free fields is of interest because it allows the introduction of notions and methods, but in itself a free field is rather empty of physical content: the real characteristics of the system it supposedly describes can only be assessed, measured, through interactions with other systems, described by other fields. These characteristics are described precisely by the responses of the system to exterior influences. The Lagrangian of a theory is the sum of free Lagrangians, fixing the fields which are at work, and

of "interaction Lagrangians", which try to describe the interplay between them. In the case of an isolated scalar field one tries, for reasons of simplicity, to describe the self-interaction by monomial terms like $\lambda\phi^3$,$\lambda\phi^4$,$\lambda\phi^8$,$\cdots$ $\lambda\phi^n$. More involved, non-polynomial interaction Lagrangians (such as $\cos(\alpha\phi)$ and $\exp(\alpha\phi)$, which lead to the Sine-Gordon and the Liouville equations) can be of great interest. In 4-dimensional spacetime, only the Lagrangian $\mathcal{L}_{int} = \lambda\phi^4$ seems able to produce a coherent theory in the quantum case. The other lead to uncontrollable infinities.

Exercise 7.1 Find the Euler-Lagrange equation of the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\ \left[\partial_\mu\phi\partial^\mu\phi\ -\ m^2\phi^2\right] + \frac{\lambda}{4}\ \phi^4\ . \tag{7.8}$$

Find the energy-momentum density tensor. ■

7.2 Complex Scalar Fields

§ 7.2 A complex field is equivalent to two real fields ϕ_1 and ϕ_2, put together as

$$\phi(x) = \phi_1(x)\ + i\ \phi_2(x)\ . \tag{7.9}$$

In the jargon of field theory it usual to forget about the infinity of degrees of freedom represented by each component and talk about each components as if it were "one" degree. There are then "two" independent degrees of freedom, and we can use either $\phi_1(x), \phi_2(x)$ or the pair $\phi(x), \phi^*(x)$. The Lagrangian, which leads to two independent Klein-Gordon equations, is

$$\mathcal{L} = \partial_\mu\phi^*\partial^\mu\phi - m^2\phi^*\phi = \partial_\mu\phi_1\partial^\mu\phi_1 - m^2\phi_1^2 + \partial_\mu\phi_2\partial^\mu\phi_2 - m^2\phi_2^2\ . \tag{7.10}$$

The energy-momentum density tensor is now

$$\Theta^{\mu\lambda} = \partial^\mu\phi^*\partial^\lambda\phi + \partial^\mu\phi\partial^\lambda\phi^* - \eta^{\mu\lambda}\mathcal{L}. \tag{7.11}$$

The density of energy (notice: always positive) is

$$\Theta^{00} = \partial_0\phi^*\partial_0\phi + (\vec{\nabla}\phi^*)\cdot(\vec{\nabla}\phi) + m^2\phi^*\phi\ , \tag{7.12}$$

(where $\vec{\nabla}$ is the 3–dimensional, space gradient) and that of momentum,

$$\Theta^{0i} = \partial_0\phi^*\partial^i\phi + \partial^i\phi^*\partial_0\phi. \tag{7.13}$$

Exercise 7.2 Verify formulas (7.11) and (7.12). ■

Like in the case of the real field, the spin density is zero also here

$$S^{\mu}{}_{\alpha\beta} = 0. \tag{7.14}$$

§ 7.3 A question comes up naturally at the sight of (7.10): $\phi(x)$ has two "components", but it is a Lorentz scalar. What is the meaning of these components? In order to answer this question, let us begin by noting that the Lagrangian has a supplementary invariance, absent in the real scalar case: it does not change under the transformations

$$\begin{aligned} \phi(x) \to \phi'(x) &= e^{i\alpha}\phi(x) \ ; \\ \phi^*(x) \to \phi'^*(x) &= e^{-i\alpha}\phi^*(x) \ , \end{aligned} \tag{7.15}$$

where α is an arbitrary constant. This transformation takes place only in the space of fields, leaving spacetime untouched. It is a rotation in the complex field plane, the same as

$$\begin{aligned} \phi'_1 &= \phi_1 \cos\alpha - \phi_2 \sin\alpha \ ; \\ \phi'_2 &= \phi_1 \sin\alpha + \phi_2 \cos\alpha \ . \end{aligned} \tag{7.16}$$

Let us see what Ms. Noether would have to say about the invariance under these transformations (called gauge transformations of first kind, or global gauge transformations). Looking at the terms in the current, we shall have

$$\frac{\delta x^\mu}{\delta\omega^a} = 0; \quad \frac{\bar\delta\phi}{\delta\omega^a} = \frac{\bar\delta\phi}{\delta\alpha} = i\phi; \quad \frac{\bar\delta\phi^*}{\delta\omega^a} = \frac{\bar\delta\phi^*}{\bar\delta\alpha} = -i\phi^*. \tag{7.17}$$

Consequently, the Noether current will be

$$J^\mu(x) = i\ [\phi^*(\partial^\mu\phi) - (\partial_\mu\phi^*)\phi] = i\ \phi^*\ \overleftrightarrow{\partial^\mu}\ \phi \ . \tag{7.18}$$

Even had we guessed this expression without Noether's help, we would know that its divergence vanishes: from the very field equations

$$\Box\phi(x) + m^2\phi(x) = 0 \quad \text{and} \quad \Box\phi^*(x) + m^2\phi^*(x) = 0, \tag{7.19}$$

we see that

$$\begin{aligned} \partial_\mu J^\mu = i[\partial_\mu\phi^*\partial^\mu\phi + \phi^*(\Box\phi) - (\Box\phi^*)\phi - \partial_\mu\phi^*\partial^\mu\phi] = \\ i[\phi^*(\Box\phi) - (\Box\phi^*)\phi] = im^2[\phi^*\phi - \phi^*\phi] = 0. \end{aligned} \tag{7.20}$$

Because it is a conservation law quite independent of spacetime (related to an "internal" symmetry, with the complex plane in the role of internal space), it is highly tempting to attribute to the corresponding charge,

$$Q = \int d^3x J^0(x) \ , \tag{7.21}$$

the role of electric charge. This would be wrong.

§ 7.4 To understand this point, which is fundamental, let us begin by remarking that there are — besides the "topological charges" mentioned in section 4.4 — two kinds of charge in Physics. As far as is known nowadays, the barionic number and the flavor hypercharge, for example, are not associated to any field, while the electric charge is the source of the electromagnetic field. There are thus charges related to certain fields, and others which, at least apparently, are not. A Lagrangian like (7.10) cannot really describe a field with electric charge: as soon as such a charge is present, another field is automatically created, which must appear also in the Lagrangian. In other words, electrically charged fields cannot be described by a free Lagrangian.

Let us see what happens if, in (7.15), the angle becomes dependent on the space-time position:

$$\begin{aligned} \phi(x) \to \phi'(x) &= e^{i\alpha(x)}\phi(x) \; ; \\ \phi^*(x) \to \phi'^*(x) &= e^{-i\alpha(x)}\phi^*(x) \; . \end{aligned} \tag{7.22}$$

A function like $e^{i\alpha}$, or $e^{i\alpha(x)}$, can be seen as a complex matrix with a single entry. It will be, of course, a unitary matrix. The set of such unitary 1–dimensional matrices form a group, denoted $U(1)$ or $SO(2)$. Equations (7.15) describe a transformation belonging to $U(1)$ which is the same at all points of spacetime. We try to represent this case in Figure 7.1, in which the x-axis represents spacetime. On the other hand, Eqs.(7.22) describe a transformation of $U(1)$ which is different at each point

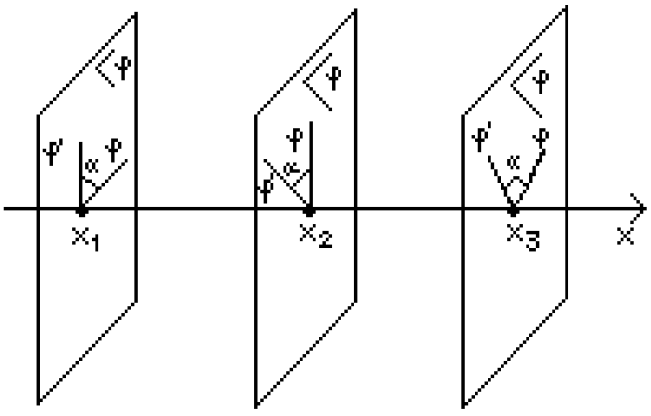


Figure 7.1: *A global gauge transformation is the same at every point.*

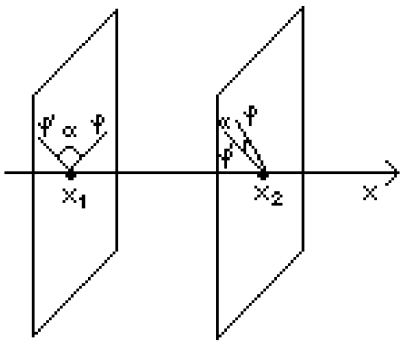


Figure 7.2: *Local gauge transformations: different at each point.*

of spacetime, as indicated in Figure 7.2. In the first case, the degrees of freedom are everywhere rotated of the same angle α: the transformation is "global". In the second case, called gauge transformations of second kind, that particular degree of freedom which is indexed by x is rotated by the angle $\alpha(x)$. At different events, the angle can be different. It is a different element of the same group $U(1)$ which is at work.

Comment 7.2 Nomenclature has wavered a little. Nowadays, when people say "gauge transformations", they mean usually gauge transformations of second kind. Because these expressions are so telling, it would perhaps be better to call them "global" and "local" gauge transformations.

§ 7.5 For local gauge transformations a new problem arises: the Lagrangian (7.10) is no more invariant:

$$\mathcal{L} \to \mathcal{L}' = [\partial_\mu - i\alpha_\mu]\phi^*[\partial^\mu + i\alpha^\mu]\phi - m^2\phi^*\phi, \tag{7.23}$$

where $\alpha_\mu = \partial_\mu \alpha(x)$. It all happens as if the derivatives were changed, and only them. The mass term remains invariant. Is there another case in which derivatives get changed? Let us now recall two properties of the electromagnetic 4–vector potential A_μ:

1. it is defined only up to a 4-divergence, the same field (the same value of the observable $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ being given by A_μ and by

$$A'_\mu = A_\mu - \partial_\mu \alpha \ , \tag{7.24}$$

for any function $\alpha(x)$.

2. its coupling to other fields is of the so-called minimal type: it changes the canonical momenta p^μ of charged particles (or other fields) to $p^\mu - A^\mu$ and, consequently, the derivatives ∂^μ to $(\partial^\mu + eA^\mu)$ (classically, the electric charge "e" can be absorbed into A^μ).

Thus, if we want that the Lagrangian describe a charged field, we must modify the derivatives in (7.10) to allow the presence of an electromagnetic field. It becomes

$$\mathcal{L} = [(\partial_\mu - iA_\mu)\phi^*]\,[(\partial^\mu + iA^\mu)\phi] - m^2\phi^*\phi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{7.25}$$

the last term being the Lagrangian of the free electromagnetic field. And now a very beautiful thing happens: (7.25) is invariant under the gauge transformations described by (7.22) and (7.24) together. The potential A_μ compensates, through its indetermination (7.24), the variance of ϕ. It is the simplest known *gauge potential*. Its simplicity is due to the related group $U(1)$, which has one single generator and is, consequently, abelian.

§ 7.6 The above procedure can be generalized. Transformations (7.22) are not, of course, the most general admissible unitarity transformations. Given any "internal" group with generators $\{J_a\}$, transformations like $\phi(x) \to \phi'(x) = e^{i\alpha^a(x)J_a}\phi(x)$ are admissible. What has been done can be adapted to provide an invariant Lagrangian. Such a "gaugefication" or "localization" of a symmetry has been first performed by Yang and Mills around 1953 for $SU(2)$, which is a simple but non-abelian group.* It is the starting point of gauge theories.

Comment 7.3 If we proceed to quantize the theory, particles come up as quanta of each field: neutral particles as neutral fields, charged particles as quanta of charged fields. Thus, an approach like the above one is necessary to describe charged particles.

Back to the discussion about charges: those related to a global conservation do not require (or: are not sources of) additional fields. Those related to fields, like the electric charge, are related to a local invariance, or to a gauge invariance. There will be a great difference between the abelian and the non–abelian cases.

Well, let us try to sum up what we have learned here. First, charge-carrying fields must be complex. In this case, the field and its complex conjugate are described together. If the charge creates another field, it is necessary to adapt the Lagrangian. In their quantum versions, the field and its complex conjugate describe jointly particles and antiparticles. Thus, the field ϕ either creates an antiparticle or annihilates a particle; ϕ^* either creates a particle or annihilates an antiparticle. Furthermore,

* C. N. Yang and R. L . Mills, *Phys. Rev.* **96** (1954) 191.

ϕ and ϕ^* couple with opposite signs with the electromagnetic field: it is enough to check the signs in (7.25). A last comment: if $\phi = \phi^*$, case of the real field, the current (7.18) vanishes identically, and all additive charges are zero. Consequently, real fields will describe particles which are equal to their antiparticles, which do not carry any additive charge and are unable to interact electromagnetically.

Chapter 8

Dirac Fields

8.1 Dirac Equation

§ 8.1 The first attempts to extend Quantum Mechanics to the relativistic case made use of the same conversion rules (3.27) and (3.28) through which the Schrödinger equation had been found. In the relativistic case the momentum is given by Eq.(1.100) as a fourvector $p = (\mathcal{E}/c, \mathbf{p})$ such that $p_\mu p^\mu = \frac{\mathcal{E}^2}{c^2} - \mathbf{p}\cdot\mathbf{p} = m^2c^2$. The two rules can be put together into a covariant set as

$$p_\mu \Rightarrow i\hbar \frac{\partial}{\partial x^\mu} \quad .$$

Notice the signs:

$$p_0 = \frac{H}{c} \Rightarrow i\hbar\frac{\partial}{\partial x^0} \quad ; \quad x^0 = x_0 = ct \quad ; \quad p_j \Rightarrow i\hbar\frac{\partial}{\partial x^j} = -i\,\hbar\frac{\partial}{\partial x_j} = \frac{\hbar}{i}\nabla^j \quad .$$

There is, however, a problem: the relativistic Hamiltonian does not lend itself to such a simple "translation", since

$$H = \sqrt{\vec{p}^{\,2}c^2 + m^2c^4} \Rightarrow \sqrt{m^2c^4 - \hbar^2c^2\vec{\nabla}^2} \; .$$

This is a non-local operator: once expanded, it requires the knowledge of arbitrarily high–order derivatives of the wavefunction. One can think of using H^2 instead of H, and the result is the Klein–Gordon equation:

$$-\hbar^2\frac{\partial^2}{\partial t^2}\,\psi(\vec{x},t) = -\hbar^2c^2\vec{\nabla}^2\psi(\vec{x},t) + m^2c^4\psi(\vec{x},t)\,. \tag{8.1}$$

As previously said, this equation is, in a certain sense, "compulsory", as it states that the field is an eigenstate of the Poincaré group invariant operator $P_\mu P^\mu$ with eigenvalue m^2c^2. Every field corresponding to a particle of mass m must satisfy

this condition. Of course, once we use H^2, we shall be introducing negative energy solutions for a free system: there is no reason to exclude $H = -\sqrt{\mathbf{p}^2c^2 + m^2c^4}$. We have above (in our toy model of Section 3.1) separated the fields into components of positive and negative frequencies, ready to interpret the latter as related to antiparticles. This, of course, because we now know the solution of the problem. At that time, negative-energy solutions caused great discomfort and led Dirac to a quest which led him, in the long run, to quite unexpected results.

Comment 8.1 Summing up, this problem led to

(i) the conversion of a non–local problem into a local one,

(ii) the discovery of antimatter, and

(iii) the uncovering of the wealth of statistics in Physics.

§ 8.2 He started by seeking a new way to "extract the square root" of the operator H^2. He looked for an equation in which squaring $\frac{\partial}{\partial t}$ and $\vec{\nabla}$ were not necessary. In other words, he looked for a linear, first-order equation both in t and $\vec{x}$. He began with an equation for the square root

$$\sqrt{\vec{p}^2c^2 + m^2c^4} = c\ \vec{\alpha} \cdot \vec{p}\ + \beta mc^2, \tag{8.2}$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and β are constants to be found. Taking the square, one arrives at the conditions

$$\begin{aligned} &(a) \quad {\alpha_1}^2 = {\alpha_2}^2 = {\alpha_3}^2 = \beta^2 = 1; \\ &(b) \quad \alpha_k\ \beta + \beta\ \alpha_k = 0 \quad \text{for } k = 1, 2, 3; \\ &(c) \quad \alpha_i\alpha_j + \alpha_j\alpha_i = 0 \quad \text{for } i, j = 1, 2, 3, \text{ but } i \neq j. \end{aligned} \tag{8.3}$$

These conditions cannot be met if α_k and β are real or complex numbers. They can be satisfied, however, if they are matrices, with the number "1" replaced by the identity matrix I wherever it appears. In that case, as the equation corresponding to (8.2) is the matrix equation (the "Hamiltonian form" of the Dirac equation)

$$H\psi(\vec{x}, t) = i\hbar\ \frac{\partial}{\partial t}\psi(\vec{x}, t) = \ \frac{\hbar}{i}\ c\ \ \vec{\alpha} \cdot \vec{\nabla}\ \psi(\vec{x}, t)\ +\ \beta mc^2\psi(\vec{x}, t), \tag{8.4}$$

the wavefunction will be necessarily a column-vector, on which the matrices act. Notice that, once conditions (8.3) are satisfied, ψ will also obey the Klein-Gordon equation which is, as said, mandatory. We must thus look at (8.4) as an equation involving four matrices (complex, $n \times n$ for the time being) and the n-vector ψ. As H should be hermitian, so should α_k and β be: $\alpha_k^\dagger = \alpha_k$, $\beta^\dagger = \beta$. Take one of them (the reasoning which follows holds for each one). Being hermitian, it has real

eigenvalues and there exists a similarity transformation which diagonalizes it. By condition (a) in (8.3), these eigenvalues can be either + 1 or − 1. Furthermore, conditions (a) and (b) say that

$$\begin{aligned} \mathrm{tr}\,\alpha_k = \mathrm{tr}\,(\beta^2\alpha_k) = \mathrm{tr}\,(\beta\alpha_k\beta) = -\mathrm{tr}\,\alpha_k \rightarrow \mathrm{tr}\,\alpha_k = 0; \\ \mathrm{tr}\,\beta = \mathrm{tr}\,(\beta\alpha_k^2) = \mathrm{tr}\,(\alpha_k\beta\alpha_k) = -\mathrm{tr}\,\beta \rightarrow \mathrm{tr}\,\beta = 0. \end{aligned}$$

The sum of the eigenvalues vanishes, so that there must be an equal number of eigenvalues +1 and − 1. Consequently, the number of eigenvalues is even: n is even. The first possibility would be $n = 2$, but there are not four 2×2 matrices which are hermitian, independent and distinct from the identity. There are only three [for example, the Pauli matrices (2.23)].

§ 8.3 The minimal possible value of n for which the α_k's and β can be realized is 4. We shall make the choice

$$\alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}; \quad \beta = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \tag{8.5}$$

where the σ_i's are the Pauli matrices, $\sigma_0 = I$ is the 2×2 identity matrix and "0" is the 2×2 matrix with all entries equal to zero.

Comment 8.2 Notice that the argument holds as long as four matrices are needed. If $m = 0$, β disappears and three 2×2 matrices (say, again the Pauli matrices) suffice. In this case the particle is described by a Weyl spinor, or Pauli spinor.

§ 8.4 It is good to keep in mind that any other set of matrices obtained from that one by similarity will also satisfy (8.3) and can be used equivalently. Each such a set of matrices is called a "representation". The above choice will be called the "Dirac representation". Equation (8.4) is now

$$i\hbar\,\frac{\partial}{\partial t}\psi(\vec{x},t) = \begin{pmatrix} mc^2\,I & \frac{\hbar}{i}\,c\,\vec{\sigma}\cdot\vec{\nabla} \\ \frac{\hbar}{i}\,c\,\vec{\sigma}\cdot\vec{\nabla} & -mc^2\,I \end{pmatrix}\psi(\vec{x},t), \tag{8.6}$$

where $\psi(\vec{x},t) = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$. The hermitian conjugate $\psi^\dagger = \begin{pmatrix} \psi_1^* & \psi_2^* & \psi_3^* & \psi_4^* \end{pmatrix}$ will obey the hermitian conjugate of the above equation,

$$-\,i\hbar\,\frac{\partial}{\partial t}\,\psi^\dagger(\vec{x},t) = \psi^\dagger(\vec{x},t)\begin{pmatrix} mc^2\,I & -\frac{\hbar}{i}\,c\,\vec{\sigma}\cdot\overleftarrow{\nabla} \\ -\frac{\hbar}{i}\,c\,\vec{\sigma}\cdot\overleftarrow{\nabla} & -mc^2\,I \end{pmatrix}. \tag{8.7}$$

Notice that, due to the order inversion inbuilt in hermitian conjugation, $(AB)^\dagger = B^\dagger A^\dagger$, the gradient $\overleftarrow{\nabla}$ now "attacks" $\psi^\dagger$ from its right side.

Let us now multiply the equation (8.6) for ψ on the left by $\psi^\dagger$, multiply the equation (8.7) for $\psi^\dagger$ on the right by ψ, and subtract the results. The result is

$$i\hbar \frac{\partial}{\partial t} \left[\psi^\dagger(\vec{x},t)\psi(\vec{x},t)\right] = \\ = \frac{\hbar}{i}\, c\, \psi^\dagger(\vec{x},t) \begin{pmatrix} 0 & \vec{\sigma}\cdot(\overleftarrow{\nabla}+\vec{\nabla}) \\ \vec{\sigma}\cdot(\overleftarrow{\nabla}+\vec{\nabla}) & 0 \end{pmatrix} \psi(\vec{x},t), \quad (8.8)$$

that is,

$$i\hbar \frac{\partial}{\partial t}\left[\psi^\dagger\psi\right] = \frac{\hbar}{i}\, c\, \mathrm{div}\left[\psi^\dagger\, \vec{\alpha}\, \psi\right]. \quad (8.9)$$

This expression is reminiscent of the continuity equation which, in non-relativistic Quantum Mechanics, states the conservation of probability:

$$\frac{\partial \rho}{\partial t} + \mathrm{div}\ \vec{j} = 0, \quad (8.10)$$

where

$$\rho = \psi^\dagger\psi \quad (8.11)$$

is the density of probability and

$$j^k = c\, \psi^\dagger \alpha^k\, \psi \quad (8.12)$$

is the k-th component of the probability current. From this continuity equation (and Gauss theorem) we obtain

$$\frac{\partial}{\partial t}\int d^3x\ \psi^\dagger(\vec{x},t)\psi(\vec{x},t) = 0. \quad (8.13)$$

§ 8.5 Up to this point, all we have said is that the α_k's and β are hermitian matrices with vanishing trace. But the above continuity equation should be put into the covariant form $\partial_\mu\ j^\mu = 0$ and this would require something else of them: $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, up to this point only a notation, must actually be such that $\psi^\dagger\psi$ and c $\psi^\dagger\alpha_k\psi$ constitute a Lorentz four-vector: ρ must be the temporal component and j_k the space components. Furthermore, it suggests that $c\alpha_k$ be a velocity. This hint will be corroborated below.

§ 8.6 It is clear, above all, that the Dirac equation (8.4) must be covariant. Before going into that, let us try to grasp something of the physical meaning of the equation. We shall see later that it describes particles of spin $\frac{1}{2}$. For that reason, we shall frequently take the liberty of referring to ψ as the "electron wavefunction" and talk

of the electron as if it were the only particle in view. Let us examine the case of the electron at rest. As $\vec{p} = 0$, the equation reduces to

$$i\hbar \frac{\partial}{\partial t}\psi(\vec{x},t) = \beta mc^2 \psi(\vec{x},t). \tag{8.14}$$

The de Broglie wavelength $\lambda = \hbar/p$ is infinite and ψ must be uniform over all the space, as $\vec{p}\,\psi = 0 \Rightarrow \vec{\nabla}\,\psi(\vec{x},t) = 0$. This is also coherent with the interpretation of $c\,\vec{\alpha}$ as the velocity. Using the Dirac representation,

$$\begin{pmatrix} i\hbar\frac{\partial}{\partial t} - mc^2 & 0 & 0 & 0 \\ 0 & i\hbar\frac{\partial}{\partial t} - mc^2 & 0 & 0 \\ 0 & 0 & i\hbar\frac{\partial}{\partial t} + mc^2 & 0 \\ 0 & 0 & 0 & i\hbar\frac{\partial}{\partial t} + mc^2 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = 0. \tag{8.15}$$

There are four independent solutions,

$$\begin{aligned} \psi_1 &= e^{-\frac{i}{\hbar}mc^2 t}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}; \quad \psi_2 = e^{-\frac{i}{\hbar}mc^2 t}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}; \\ \psi_3 &= e^{+\frac{i}{\hbar}mc^2 t}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \quad \psi_4 = e^{+\frac{i}{\hbar}mc^2 t}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned} \tag{8.16}$$

We have thus a first drawback: ψ_3 and ψ_4 are solutions with negative energy. In quantum theory they are interpreted as wavefunctions describing antiparticles (positrons).

§ 8.7 The electron has an electric charge, and consequently couples to the electromagnetic field. We shall introduce in (8.4) an electromagnetic field through the minimal coupling prescription:

$$p_\mu \Rightarrow p_\mu - \frac{e}{c}A_\mu, \quad \text{with} \quad p_\mu = i\hbar\frac{\partial}{\partial x^\mu}.$$

The result is

$$i\hbar\,\frac{\partial}{\partial t}\,\psi(\vec{x},t) = \Big[\; c\vec{\alpha}\cdot\Big(\vec{p} - \;\frac{e}{c}\vec{A}\Big) \;+\; \beta mc^2 + e\phi\Big]\,\psi(\vec{x},t), \tag{8.17}$$

where $A_0 = \phi$. Recall that an electric point–charge in an electromagnetic field has the interaction energy $H_I = -\,\frac{e}{c}\vec{v}\cdot\vec{A} + \;e\phi$, which appears above under the form $H_I = -\,\frac{e}{c}\,c\vec{\alpha}\cdot\vec{A} + e\phi$. This validates the interpretation of matrix $c\,\vec{\alpha}$ as the velocity in this theory: $c\,\vec{\alpha}$ will be the velocity operator.

8.2 Non-Relativistic Limit: Pauli Equation

§ 8.8 In order to examine the non-relativistic limit, we suppose the mass term βmc^2 to be much larger than the other energies involved. As the positive-energy and negative-energy components will have different behavior, it will be convenient to introduce the two $2-$component columns

$$\tilde{L} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}; \quad \tilde{S} = \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix}. \tag{8.18}$$

Using $\psi = \begin{pmatrix} \tilde{L} \\ \tilde{S} \end{pmatrix}$ in (8.17),

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} \tilde{L} \\ \tilde{S} \end{pmatrix} = \begin{pmatrix} [mc^2 + e\phi]I & c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A}) \\ c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A}) & [-mc^2 + e\phi]I \end{pmatrix} \begin{pmatrix} \tilde{L} \\ \tilde{S} \end{pmatrix} \tag{8.19}$$

or

$$\begin{pmatrix} i\hbar\frac{\partial}{\partial t}\tilde{L} \\ i\hbar\frac{\partial}{\partial t}\tilde{S} \end{pmatrix} = \begin{pmatrix} [mc^2 + e\phi]\tilde{L} + c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})\tilde{S} \\ c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})\tilde{L} + [-mc^2 + e\phi]\tilde{S} \end{pmatrix} \tag{8.20}$$

Now: the larger part of the energy will be concentrated in mc^2; then, the strongest time–variation will be dominated by this term. This means that, if we look for solutions of the form

$$\begin{pmatrix} \tilde{L} \\ \tilde{S} \end{pmatrix} = e^{-\frac{i}{\hbar}mc^2 t} \begin{pmatrix} L \\ S \end{pmatrix}, \tag{8.21}$$

most of the time variation will be isolated in the exponential, and $\begin{pmatrix} L \\ S \end{pmatrix}$ will vary slowly with t. The equation then becomes

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} L \\ S \end{pmatrix} = \begin{pmatrix} e\phi L + c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})S \\ c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})L + [e\phi - 2mc^2]S \end{pmatrix}. \tag{8.22}$$

Consider the equation for S:

$$i\hbar \frac{\partial}{\partial t} S = c\,\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})L + [e\phi - 2mc^2]S. \tag{8.23}$$

Let us proceed to still another approximation: as S vary slowly, we neglect $i\hbar\frac{\partial}{\partial t}S$. As also $e\phi \ll 2mc^2$, we arrive at the expression

$$S \approx \frac{\vec{\sigma}\cdot(\vec{p} - \frac{e}{c}\vec{A})}{2mc}\, L. \tag{8.24}$$

We see that S ("small") is indeed very small in comparison to L ("large"): S/L is of the order v/c. The components ψ_3 and ψ_4 are for that reason called the "small components" of the Dirac wavefunction, ψ_1 and ψ_2 being the "large components". Because it associates in this way two Pauli spinors, one large and one small, the 4-component representation is called the "bispinor representation". Taking the above approximated S into the equation for $i\hbar\frac{\partial}{\partial t}L$, we find

$$i\hbar\frac{\partial}{\partial t}L = \left[\frac{\vec{\sigma}\cdot(\vec{p}-\frac{e}{c}\vec{A})\;\vec{\sigma}\cdot(\vec{p}-\frac{e}{c}\vec{A})}{2m} + e\phi\right] L. \tag{8.25}$$

This can be put into a more readable form by using the identity $\sigma^i\,\sigma^j = \delta^{ij} + i\,\epsilon^{ij}{}_k\,\sigma^k$, which leads to

$$(\vec{\sigma}\cdot\vec{a})\,(\vec{\sigma}\cdot\vec{b}) = \vec{a}\cdot\vec{b} + \;i\;\vec{\sigma}\cdot\left(\vec{a}\times\vec{b}\right)\;. \tag{8.26}$$

Let us first look at the vector–product term. As $\vec{p} = \frac{\hbar}{i}\vec{\nabla}$, then

$$\begin{aligned}\left[(\vec{p}-\frac{e}{c}\vec{A})\times(\vec{p}-\frac{e}{c}\vec{A})\right]_i L &= \epsilon_i{}^{jk}\left(\frac{\hbar}{i}\partial_j - \frac{e}{c}A_j\right)\left(\frac{\hbar}{i}\partial_k - \frac{e}{c}A_k\right)L \\ &= -\frac{\hbar}{i}\frac{e}{c}\,\epsilon_i{}^{jk}\,(\partial_j A_k)L = -\frac{\hbar}{i}\frac{e}{c}\,B_i\,L\;.\end{aligned}$$

The magnetic field $\vec{B}$ (in vacuum, = our previous $\vec{H}$) turns up. Adding now the scalar–product term,

$$\left[\vec{\sigma}\cdot(\vec{p}-\frac{e}{c}\vec{A})\right]\left[\vec{\sigma}\cdot(\vec{p}-\frac{e}{c}\vec{A})\right] = \left(\vec{p}-\frac{e}{c}\vec{A}\right)^2 - \frac{e\hbar}{c}\,\vec{\sigma}\cdot\vec{B}. \tag{8.27}$$

The equation becomes the 2−component *Pauli equation*

$$i\hbar\frac{\partial}{\partial t}L = \left[\frac{\left(\vec{p}-\frac{e}{c}\vec{A}\right)^2}{2m} - \frac{e\hbar}{2mc}\,\vec{\sigma}\cdot\vec{B}\;+e\phi\right] L, \tag{8.28}$$

which describes a spin–$\frac{1}{2}$ electron: $L = \exp[\frac{i}{\hbar}mc^2t]\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$. Under a rotation, as shown in Exercise 2.7, it transforms as a member of the representation $j=\frac{1}{2}$.

8.3 Covariance

In the Hamiltonian form (8.4) of the Dirac equation, time and space play distinct roles. To go into the so-called "covariant form", we first define new matrices, the celebrated Dirac's "gamma matrices", as

$$\gamma^0 = \beta\;;\;\gamma^i = \beta\alpha^i. \tag{8.29}$$

Multiplying the equation by $\frac{\beta}{c}$ on the left, we find $i\hbar\gamma^\mu \frac{\partial\psi}{\partial x^\mu} = mc\psi$, or

$$(i\hbar\gamma^\mu\partial_\mu - mc)\psi(x) = (\gamma^\mu p_\mu - mc)\,\psi = 0. \tag{8.30}$$

Of current use is Feynman's "slash" notation: we write $\not{p} = \gamma^\mu p_\mu$, $\not{\partial} = \gamma^\mu\partial_\mu$, and so on, and the above equation is written diversely as

$$(i\hbar\ \not{\partial} - mc)\psi = 0 \tag{8.31}$$

or

$$(\not{p} - mc)\psi = 0. \tag{8.32}$$

In the presence of an electromagnetic field,

$$(\not{p} - \frac{e}{c}\ \not{A} - mc)\psi = 0, \tag{8.33}$$

where, of course, $\not{A} = \gamma^\mu A_\mu$. In terms of the gamma matrices, conditions (8.3) acquire a compact form,

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = \{\gamma^\mu, \gamma^\nu\} = 2\ \eta^{\mu\nu}\ I. \tag{8.34}$$

Comment 8.3 Rewrite conditions (8.3) in the equivalent form

$$\begin{array}{ll} (e) & \beta^2 \ = \ I \leftrightarrow \gamma^0\gamma^0 + \gamma^0\gamma^0 = 2\,I \\ (f) & \alpha^k\ \gamma^0 + \gamma^0\ \alpha^k = 0 \quad \text{for } k = 1,2,3; \\ (g) & \alpha^i\alpha^j + \alpha^j\alpha^i = 2\,\delta^{ij}\,I \quad \text{for } i,j = 1,2,3. \end{array} \tag{8.35}$$

Multiply (g) by γ^0 on both sides and use (f) to commute the α^k's with γ^0 and obtain

$$\gamma^i\gamma^j + \gamma^j\gamma^i = -\,2\,\delta^{ij}\,I. \tag{8.36}$$

Together with (e), this is just (8.34). By the way, (8.34) is the operation table of a particular example of "Clifford algebra".

The right-hand side exhibits the Lorentz metric times the unit 4×4 matrix I. And, in the middle, there is a first: an anticommutator comes forth. For $i = 1,2,3$ the matrix γ^i is antihermitian [because $(\gamma^i)^\dagger = (\beta\alpha^i)^\dagger = (\alpha^i)^\dagger\beta^\dagger = \alpha^i\beta = -\,\beta\alpha^i = -\,\gamma^i$], whereas γ^0 $(= \beta)$ remains hermitian. Other properties are $(\gamma^i)^2 = -\,I$ and $(\gamma^0)^2 = I$. In the Dirac representation the γ's have the forms

$$\gamma_i = \begin{pmatrix} 0 & \sigma_i \\ -\,\sigma_i & 0 \end{pmatrix};\ \gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}. \tag{8.37}$$

This is also called the "Pauli—Dirac representation" of the gamma matrices. As we have emphasized, any other set of matrices γ^μ obtained from those by a similarity transformation is equally acceptable. As a rule, for each particular problem there is a special representation which is the best suited. We shall see later many other matrices of interest, such as those representing change of parity and charge conjugation. They are obtained from the γ's, and differ from one representation to the other. The details are of purely technical interest and the subject goes under the nickname "gammalogy". Instead of examining the many possible cases, we shall here concentrate in the Dirac representation.

Comment 8.4 The Dirac equation becomes real in the "Majorana representation". The solutions are then superpositions of real functions. The γ's are given by $\gamma^\mu_{Majorana} = U\gamma^\mu_{Dirac}U^{-1}$, with $U = U^{-1} = \frac{1}{\sqrt{2}}\begin{pmatrix} I & \sigma^2 \\ \sigma^2 & -I \end{pmatrix}$.

Comment 8.5 The "chiral representation" is of interest in approaching neutrinos and the violation of CP invariance: $\gamma^\mu_{chiral} = U\gamma^\mu_{Dirac}U^{-1}$, with $U = \frac{1}{\sqrt{2}}\begin{pmatrix} I & -I \\ I & I \end{pmatrix}$.

Notice $(\not{p})^2 = \gamma^\mu p_\mu \gamma^\nu p_\nu = \frac{1}{2}\{\gamma^\mu, \gamma^\nu\}p_\mu p_\nu = \eta^{\mu\nu}p_\mu p_\nu\, I = p^\mu p_\mu\, I$ (the identity matrix I is frequently left unwritten in standard notation). It follows that the Dirac equation can be seen as coming from a factorization of the Klein-Gordon equation:

$$(p^\mu p_\mu - m^2c^2)\psi = [\gamma^\mu p_\mu + mc][\gamma^\nu p_\nu - mc]\psi = 0. \tag{8.38}$$

This comes, of course, from the conditions (8.3) or (8.34), which have been imposed just to attain this objective. It follows also that a solution of the Dirac equation is a solution of the Klein-Gordon equation though, of course, the Klein-Gordon equation can have solutions which do not satisfy the Dirac equation.

Let us now examine the Lorentz covariance of (8.30). Recall that, in order to start talking about invariance, covariance, etc, the field $\psi(x) = \psi(\vec{x}, t)$ must belong to (the carrier space of) a linear representation of the Lorentz group, so that a matrix $U(\Lambda)$ must exist such that the field $\psi'(x')$, seen in another frame (that frame in which $x^{\alpha'} = \Lambda^{\alpha'}{}_\beta\, x^\beta$, $p^{\alpha'} = \Lambda^{\alpha'}{}_\beta\, p^\beta$, etc) is related to $\psi(x)$ by $\psi'(x') = U(\Lambda)\psi(x)$. What we shall do is to determine $U(\Lambda)$ so as to ensure the covariance of the equation. In other words, we shall find the representation to which $\psi(x)$ belongs.

Multiplying the equation on the left by $U(\Lambda)$ and substituting $\psi(x) = U(\Lambda)^{-1}$ $\psi'(x')$, we arrive at

$$\left[U(\Lambda)\ \gamma^\alpha p_\alpha\ U(\Lambda)^{-1} - mc\right]\psi'(x') = 0.$$

In order to identify this with $\left[\gamma^{\alpha'} p_{\alpha'} - mc\right] \psi'(x') = 0$, it will be necessary that

$$U(\Lambda)\ \gamma^{\alpha}\ U(\Lambda)^{-1} = \gamma^{\alpha'} \Lambda_{\alpha'}{}^{\alpha}, \tag{8.39}$$

or

$$\gamma^{\beta'} = \Lambda^{\beta'}{}_{\alpha}\ U(\Lambda)\ \gamma^{\alpha}\ U(\Lambda)^{-1}. \tag{8.40}$$

This means that the set $\{\gamma^{\alpha}\}$ must constitute a 4-vector of matrices, and that $U(\Lambda)$ acts in the space of such matrices, as the representative of Λ. Notice however that, as the Minkowski metric tensor is invariant under Lorentz transformations,

$$\eta_{ab} = \Lambda^{c}{}_{a}\, \Lambda^{d}{}_{b}\, \eta_{cd}, \tag{8.41}$$

if we take into account the relation (8.34) between the gamma matrices and the spacetime metric, we conclude that γ^{β} does not change under such transformation either. In other words, like Eq. (8.41) for η_{ab}, the transformation (8.40) must actually be written without the "primes":

$$\gamma^{\beta} = \Lambda^{\beta}{}_{\alpha}\ U(\Lambda)\ \gamma^{\alpha}\ U(\Lambda)^{-1}. \tag{8.42}$$

In fact, the meaning of Eq. (8.39) is that the Lorentz transformation in the spinor indices of γ^{α} is equivalent to the **inverse** Lorentz transformation in the 4-vector index of γ^{α}. The "total" Lorentz transformation of γ^{α}, therefore, as given by Eq. (8.42), implies that γ^{α} is invariant under Lorentz transformations.

For later convenience we shall write the transformation equation in the form

$$\psi'(x') = U(\Lambda)\psi(x) = e^{-\frac{i}{4}\omega^{\alpha\beta}\sigma_{\alpha\beta}}\psi(x), \tag{8.43}$$

and look for matrices $\sigma_{\alpha\beta}$ apt to do the job. The $\omega^{\alpha\beta}$'s are the Lorentz group parameters of the transformation Λ. We are here paying tribute to a notation which became standard for historical reasons. The generators will actually turn out to be $\frac{1}{2}\sigma_{\alpha\beta}$. And the double index leads to double counting, rendering necessary an extra $\frac{1}{2}$ factor. The factor $\frac{1}{4}$ in the exponent owes its origin to these two $\frac{1}{2}$ factors. Now, for the specific case of the vector representation,

$$U(\Lambda) = \exp\left[-\frac{i}{2}\,\omega^{\alpha\beta}\, M_{\alpha\beta}\right], \tag{8.44}$$

where the matrix representative of the Lorentz group generators in the vector representation is

$$(M_{\alpha\beta})^{\mu}{}_{\lambda} = -\,(M_{\beta\alpha})^{\mu}{}_{\lambda} = -\,(M_{\alpha\beta})_{\lambda}{}^{\mu} = i(\eta_{\alpha\lambda}\delta^{\mu}_{\beta} - \eta_{\beta\lambda}\delta^{\mu}_{\alpha}). \tag{8.45}$$

For an infinitesimal transformation, the group element is

$$\Lambda^{\mu}{}_{\lambda} \approx \left[I - \tfrac{i}{2}\delta\omega^{\alpha\beta}M_{\alpha\beta}\right]^{\mu}{}_{\lambda} = \delta^{\mu}_{\lambda} - \tfrac{i}{2}\delta\omega^{\alpha\beta}(M_{\alpha\beta})^{\mu}{}_{\lambda} = \delta^{\mu}_{\lambda} - \delta\omega^{\mu}{}_{\lambda}. \tag{8.46}$$

The right-hand side of (8.39) will then be

$$\gamma^{\beta'}\Lambda_{\beta'}{}^{\epsilon} \approx \gamma^{\epsilon} - \tfrac{i}{2}\delta\omega^{\alpha\beta}(M_{\alpha\beta})^{\epsilon}{}_{\beta'}\gamma^{\beta'}.$$

On the other hand, for the spinor representation,

$$U(\Lambda) \approx I - \tfrac{i}{4}\delta\omega^{\alpha\beta}\sigma_{\alpha\beta},$$

and, to the first order, we find for the left-hand side of (8.39):

$$U\gamma^{\mu}U^{-1} \approx \gamma^{\mu} - \tfrac{i}{4}\delta\omega^{\alpha\beta}[\sigma_{\alpha\beta}, \gamma^{\mu}].$$

It is consequently necessary that

$$[\tfrac{1}{2}\,\sigma_{\alpha\beta}, \gamma^{\epsilon}] = (M_{\alpha\beta})^{\epsilon}{}_{\delta}\;\gamma^{\delta}. \tag{8.47}$$

Clearly, $\sigma_{\alpha\beta}$ must be antisymmetric in the two indices, and the first idea coming to the mind does work: the matrices

$$\sigma_{\alpha\beta} = \tfrac{i}{2}\;(\gamma_{\alpha}\gamma_{\beta} - \gamma_{\beta}\gamma_{\alpha}) = \tfrac{i}{2}\;[\gamma_{\alpha}, \gamma_{\beta}] \tag{8.48}$$

satisfy condition (8.47). In effect:

1. Separating the product $\gamma_{\alpha}\gamma_{\beta}$ into its symmetric and antisymmetric parts, $\gamma_{\alpha}\gamma_{\beta} = \frac{1}{2}\{\gamma_{\alpha}, \gamma_{\beta}\} + \frac{1}{2}[\gamma_{\alpha}, \gamma_{\beta}]$, we arrive at the useful identity

$$\gamma_{\alpha}\gamma_{\beta} = \eta_{\alpha\beta} - i\;\sigma_{\alpha\beta}. \tag{8.49}$$

2. From this, $[\sigma_{\alpha\beta}, \gamma_{\epsilon}] = i[\gamma_{\alpha}\gamma_{\beta}, \gamma_{\epsilon}]$.

3. The result follows then from using (8.34) twice:

$$\begin{aligned}
[\sigma_{\alpha\beta}, \gamma_{\epsilon}] &= i(\gamma_{\alpha}\gamma_{\beta}\gamma_{\epsilon} - \gamma_{\epsilon}\gamma_{\alpha}\gamma_{\beta}) \\
&= i(-\gamma_{\alpha}\gamma_{\epsilon}\gamma_{\beta} + 2\gamma_{\alpha}\eta_{\beta\epsilon} + \gamma_{\alpha}\gamma_{\epsilon}\gamma_{\beta} - 2\gamma_{\beta}\eta_{\alpha\epsilon}) \\
&= 2i(\gamma_{\alpha}\eta_{\beta\epsilon} - \gamma_{\beta}\eta_{\alpha\epsilon}).
\end{aligned}$$

More than that, the $\sigma_{\alpha\beta}$'s are such that

$$[\tfrac{1}{2}\,\sigma_{\alpha\beta}, \tfrac{1}{2}\,\sigma_{\gamma\delta}] = i\;(\eta_{\beta\gamma}\tfrac{1}{2}\,\sigma_{\alpha\delta} - \eta_{\alpha\gamma}\tfrac{1}{2}\,\sigma_{\beta\delta} + \eta_{\alpha\delta}\tfrac{1}{2}\,\sigma_{\beta\gamma} - \eta_{\beta\delta}\tfrac{1}{2}\,\sigma_{\alpha\gamma}), \tag{8.50}$$

which shows that $\frac{1}{2}\,\sigma_{\alpha\beta}$ is a Lorentz generator. As each matrix $M_{\alpha\beta}$, each matrix $\frac{1}{2}\,\sigma_{\alpha\beta}$ is a generator of a representation of the Lie algebra of the Lorentz group. The $M_{\alpha\beta}$'s generate the vector representation, the $\frac{1}{2}\,\sigma_{\alpha\beta}$'s generate the bispinor representation. Expression (8.47) is the infinitesimal version of (8.39). It states again that the gamma's constitute a 4–vector. Thus, covariance of the Dirac equation requires that the Dirac field $\psi(x)$ belong to the bispinor representation $U(\Lambda)$ generated by the above $\frac{1}{2}\;\sigma_{\alpha\beta}$'s.

The form of the multiplication table (8.50) is general: any set of Lorentz generators will satisfy it. It characterizes the Lie algebra of the Lorentz group. It is clear, however, that the particular form of the matrices $\sigma_{\alpha\beta}$ depend on the "representation" we are using for the matrices γ. In the Pauli-Dirac "representation" we are using, the $\sigma_{\alpha\beta}$'s are particularly simple:

$$\sigma_{ij} = \begin{pmatrix} \epsilon_{ijk}\;\sigma_k & 0 \\ 0 & \epsilon_{ijk}\;\sigma_k \end{pmatrix}; \quad \sigma^{0i} = i\;\alpha^i = \begin{pmatrix} 0 & i\sigma^i \\ i\sigma^i & 0 \end{pmatrix}. \tag{8.51}$$

Notice that $U(\Lambda)$ is not, in general, unitary. From the hermiticity properties of the γ's,

$$(\gamma^i)^\dagger = -\;\gamma^i \quad \text{and} \quad (\gamma^0)^\dagger = \gamma^0, \tag{8.52}$$

we get

$$(\sigma_{ij})^\dagger = \sigma_{ij} \quad \text{and} \quad (\sigma_{0j})^\dagger = -\,\sigma_{0j}. \tag{8.53}$$

Consequently, in this "representation", $U(\Lambda)$ will be unitary for the rotations, but not for the boosts.

Comment 8.6 Actually, there can be no unitary representation for all the members of the Lorentz group with finite matrices. This comes from a general result from the theory of groups: a non-compact group has no finite unitary representations.

A useful property comes from (8.52): sandwiching a gamma matrix between two γ^0 yields the hermitian conjugate, a property that propagates to the $\sigma^{\alpha\beta}$:

$$\gamma^0\gamma^\alpha\gamma^0 = \gamma^{\alpha\dagger} \;\rightarrow\; \gamma^0\sigma_{\alpha\beta}\gamma^0 = \sigma^\dagger_{\alpha\beta}\;. \tag{8.54}$$

Applying the latter order by order in the expansion of U, a result of interest in future calculations comes out:

$$U^{-1}(\Lambda) = \gamma_0 U^\dagger(\Lambda)\gamma_0\,. \tag{8.55}$$

Before finishing this section, let us consider the specially important examples of Lorentz transformations which are rotations. Take the particular case of a rotation

of an angle ϕ around the axis Oz, generated by

$$J^z = J^3 = J^{12} = \tfrac{1}{2}\,\sigma^{12} = \tfrac{1}{2}\begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix} . \tag{8.56}$$

In this case,

$$\psi'(x') = \begin{pmatrix} \tilde{L}'(x') \\ \tilde{S}'(x') \end{pmatrix} = e^{-\frac{i}{2}\phi\,\sigma^{12}} \begin{pmatrix} \tilde{L}(x) \\ \tilde{S}(x) \end{pmatrix} . \tag{8.57}$$

Given the (diagonal) form of σ^{12}, the transformation will act separately on $\tilde{L}(x)$ and $\tilde{S}(x)$, and for each one we shall have the behavior of a Pauli spinor under rotations. The expression for the rotation of an angle $\vec{\alpha}$ acting on a Pauli spinor $\chi(x)$ is

$$\begin{aligned} \chi'(x') &= e^{-i\vec{\alpha}\cdot\vec{J}}\chi(x) = e^{-\frac{i}{2}\vec{\alpha}\cdot\vec{\sigma}}\chi(x) && (8.58) \\ &= \left[\cos\frac{|\,\vec{\alpha}\,|}{2} - i\,\frac{\vec{\alpha}\cdot\vec{\sigma}}{|\,\vec{\alpha}\,|}\,\sin\frac{|\,\vec{\alpha}\,|}{2}\right]\,\chi(x). && (8.59) \end{aligned}$$

For the case of an angle φ around the axis Oz,

$$\chi'(x') = \left[\cos\frac{\varphi}{2} - i\,\sigma^3\,\sin\frac{\varphi}{2}\right]\,\chi(x). \tag{8.60}$$

A complete rotation in configuration space, $\varphi = 2\pi$ leads to $\chi'(x') = -\chi(x)$. A double rotation ($\varphi = 4\pi$) will be necessary to bring the wavefunction back to its initial value. Both $\tilde{L}(x)$ and $\tilde{S}(x)$ in (8.57) are Pauli spinors. We shall see later that the presence of two spinors comes from the fact (mentioned previously) that, in a relativistic theory, particles and antiparticles come up in a joint venture.

The probability density is conserved and covariant. Indeed, from (8.11) and (8.12), the current is

$$j^\mu(x) = c\;\psi^\dagger(x)\gamma^0\gamma^\mu\psi(x). \tag{8.61}$$

In another frame, it will be (using successively (8.55), (8.39) and (8.61))

$$\begin{aligned} j'^\mu(x') &= c\;\psi'^\dagger(x')\gamma^0\gamma^\mu\psi'(x') = c\;\psi^\dagger(x)U^\dagger\gamma^0\gamma^\mu U\psi(x) \\ &= c\;\psi^\dagger(x)\gamma^0\gamma^0U^\dagger\gamma^0\gamma^\mu U\psi(x) = c\;\psi^\dagger(x)\gamma^0U^{-1}\gamma^\mu U\psi(x) \\ &= c\;\psi^\dagger(x)\gamma^0\Lambda^\mu{}_\nu\gamma^\nu\psi(x) = c\;\Lambda^\mu{}_\nu\psi^\dagger(x)\gamma^0\gamma^\nu\psi(x) = \Lambda^\mu{}_\nu j^\nu(x)\;. \end{aligned}$$

That is, the density current transforms as it should — as a Lorentz 4-vector. In consequence, the continuity equation

$$\partial_\mu j^\mu(x) \;=\; 0 \tag{8.62}$$

is invariant.

8.4 Lagrangian Formalism

The matrix γ^0 ($= \gamma_0$ in our convention) has, as it could be guessed from its origin, a role rather different from the other γ matrices. We have seen its relationship to hermitian conjugacy. It plays actually other special roles. For example, the hermitian conjugate function $\psi^\dagger(x)$ appears most of times in the combination $\psi^\dagger(x)\gamma_0$. This is so frequent that another definition of "conjugate function" becomes convenient. We call

$$\overline{\psi}(x) = \psi^\dagger(x)\gamma_0 \tag{8.63}$$

the "adjoint wavefunction". The current (8.61) is then written

$$j^\mu(x) = c\,\overline{\psi}(x)\gamma^\mu\psi(x). \tag{8.64}$$

The adjoint function changes under a Lorentz transformation according to

$$\overline{\psi}'(x') = \psi'^\dagger(x')\gamma_0 = [U\psi(x)]^\dagger\gamma_0 = \psi(x)^\dagger U^\dagger\gamma_0 = \psi(x)^\dagger\gamma_0\gamma_0 U^\dagger\gamma_0,$$

so that

$$\overline{\psi}'(x') = \overline{\psi}(x)U^{-1}, \tag{8.65}$$

where use has been made of Eq. (8.55).

Having introduced the adjoint function, it is easy to see now that the Dirac equation (8.30) comes from the Lagrangian density

$$\mathcal{L} = \tfrac{i}{2}\hbar c\left[\ \overline{\psi}\ \gamma^\mu\ \partial_\mu\psi - (\partial_\mu\overline{\psi})\ \gamma^\mu\ \psi\ \right] - m\,c^2\ \overline{\psi}\ \psi, \tag{8.66}$$

by variation with respect to $\overline{\psi}$. Variation with respect to ψ leads to the equation for the adjoint field,

$$i\hbar(\partial_\mu\overline{\psi})\ \gamma^\mu\ +\ m\,c\ \overline{\psi}\ =\ 0, \tag{8.67}$$

also usually written in the form

$$\overline{\psi}\left[i\hbar\ \overleftarrow{\partial\!\!\!/}\ +\ m\,c\ \right]\ =\ 0, \tag{8.68}$$

a parody of the equation for ψ, which is

$$\left[i\hbar\ \overrightarrow{\partial\!\!\!/}\ -\ m\,c\ \right]\psi = 0. \tag{8.69}$$

From the Lagrangian above it follows the canonical energy-momentum tensor (4.51)

$$\theta^{\mu\nu} =\ \tfrac{i}{2}\hbar c\ \left[\ \overline{\psi}\ \gamma^\mu\ \partial^\nu\psi - (\partial^\nu\overline{\psi})\ \gamma^\mu\ \psi\ \right], \tag{8.70}$$

and the current [for any charge, as given in (4.70)]

$$J^\mu = c\,\overline{\psi}\ \gamma^\mu\ \psi\ . \tag{8.71}$$

To obtain the spin density (4.59), we first get the Lorentz transformations

$$\delta\psi(x) = -\tfrac{i}{4}\ \sigma_{\alpha\beta}\ \psi(x)\delta\omega^{\alpha\beta};\quad \delta\overline{\psi}(x) = \tfrac{i}{4}\ \overline{\psi}(x)\sigma_{\alpha\beta}\ \delta\omega^{\alpha\beta}. \tag{8.72}$$

Then, we get

$$S^{\mu}{}_{\alpha\beta} = -\tfrac{1}{4}\hbar c\ \overline{\psi}(x)(\gamma^{\mu}\sigma_{\alpha\beta} + \sigma_{\alpha\beta}\gamma^{\mu})\ \psi(x) = -\tfrac{1}{4}\hbar c\ \overline{\psi}(x)\{\gamma^{\mu},\ \sigma_{\alpha\beta}\}\ \psi(x)\ . \tag{8.73}$$

As the canonical energy-momentum tensor $\theta^{\mu\nu}$ is not symmetric, the spin tensor is not separately conserved. What is conserved in this case is the total angular momentum density

$$J^{\mu}{}_{\alpha\beta} = S^{\mu}{}_{\alpha\beta} + L^{\mu}{}_{\alpha\beta},$$

with

$$L^{\mu}{}_{\alpha\beta} = x_{\alpha}\,\theta^{\mu}{}_{\beta} - x_{\beta}\,\theta^{\mu}{}_{\alpha}$$

representing the orbital angular momentum density.

8.5 Parity

Another special role reserved to γ^0 is related to the parity transformation. To keep on with a notation which became usual, we shall use

$$\Lambda^{(P)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{8.74}$$

for the matrix representing the parity transformation in cartesian coordinates on spacetime, and $P = U(\Lambda^{(P)})$ for its representative acting on bispinors. Repeating what has been done to impose Lorentz covariance on the Dirac equation, if we define

$$\psi'(x') = P\psi(x),$$

we arrive at the same condition (8.39) for $U(\Lambda^{(P)})$:

$$P\ \gamma^{\alpha}\ P^{-1} = \gamma^{\beta'}\Lambda^{(P)\alpha}_{\beta'}. \tag{8.75}$$

Comment 8.7 From the signs in (8.74), $\gamma^{\beta'}\Lambda^{(P)\alpha}_{\beta'} = \gamma^0\delta^{\alpha}_0 - \gamma^i\delta^{\alpha}_i = (\gamma^0)^{\dagger}\delta^{\alpha}_0 + (\gamma^i)^{\dagger}\delta^{\alpha}_i = (\gamma^{\alpha})^{\dagger}$. Use then the first equality in (8.54) to find the solution below.

The solution is any matrix of the form

$$P \ = e^{i\varphi} \ \gamma^0, \tag{8.76}$$

with φ an arbitrary phase. Notice that P is unitary, and satisfies an equation analogous to (8.55), that is

$$P^{-1} = \gamma_0 \ P^\dagger \gamma_0. \tag{8.77}$$

Thus, in spinor space, the parity transformation is represented by

$$\psi'(x') = P \ \psi(x) = \psi'(-\ \vec{x}, t) = e^{i\varphi} \ \gamma^0 \psi(x). \tag{8.78}$$

This is a first example of a property which turns up frequently: operators acting on the Dirac bispinors are represented by complex 4×4 matrices (we shall see some exceptions later). These will be linear combinations of any set of complex 4×4 matrices forming a basis for their algebra (recall: an algebra is a vector space on which a binary internal operation — here the matrix product — is defined; and a basis for a vector space of dimension d is any set of d members which are linearly independent).

Actually, any 4×4 matrix may be expanded on a basis of 16 matrices. The algebra generated by the γ matrices — which is an example of the so called Clifford algebra — is the complete algebra of these 4×4 matrices. To see that, we need to introduce the notation

$$\gamma_5 \equiv \gamma^5 = i \, \gamma^0 \gamma^1 \gamma^2 \gamma^3 = \frac{i}{4!} \, \epsilon_{\alpha\beta\gamma\delta} \ \gamma^\alpha \gamma^\beta \gamma^\gamma \gamma^\delta. \tag{8.79}$$

In the Pauli-Dirac representation, γ^5 is given by

$$\gamma^5 = \gamma_5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \ . \tag{8.80}$$

Some of its properties are

$$\{\gamma^5, \gamma^\alpha\} = 0, \tag{8.81}$$

and

$$[\gamma^5, \sigma_{\alpha\beta}] = 0. \tag{8.82}$$

This last property propagates from the spinor generators to the whole representation $U(\Lambda)$:

$$[\gamma^5, U(\Lambda)] = 0. \tag{8.83}$$

Notice that (8.81) says that γ^5 inverts parity:

$$P \ \gamma^5 = -\ \gamma^5 \ P \ . \tag{8.84}$$

This means that, if ψ is a state with definite parity, say $P\psi = +\psi$, then $\gamma^5\psi$ will have opposite eigenvalue: $P\gamma^5\psi = -\gamma^5\psi$.

It can then be shown that the following 16 matrices form a basis for the 4×4 matrices:

$$\begin{aligned}
\Gamma^S &= I \\
\Gamma^V_\mu &= \gamma_\mu \\
\Gamma^T_{\mu\nu} &= \sigma_{\mu\nu} \\
\Gamma^A_\mu &= \gamma_5\,\gamma_\mu \\
\Gamma^P_\mu &= \gamma_5.
\end{aligned}$$

8.6 Charge Conjugation

As we have said and repeated, the description of the electron by the wavefunction (or field) $\psi(x)$ cannot be dissociated from its antiparticle, the positron. If we take a certain positive-energy spinor to describe the electron, the small components will describe a negative-energy "piece" of it. This leads to a difficulty in the sign of the total field energy, which is solved by the adoption of the anticommutative quantization rules. These lead to well-defined expressions for the energy and the charge. The positrons appear as particles with positive energy, but charges opposite to those of the electron. Thus, in the presence of an electromagnetic field, the electron satisfies

$$[i\,\hbar\,\gamma^\mu\partial_\mu - \frac{e}{c}\,\gamma^\mu A_\mu - mc]\psi(x) = 0, \tag{8.85}$$

whereas the positron will satisfy

$$[i\,\hbar\,\gamma^\mu\partial_\mu + \frac{e}{c}\,\gamma^\mu A_\mu - mc]\psi_c(x) = 0. \tag{8.86}$$

Notice that in everything we have done up to now the charge sign has played no role. We could exchange the above equation, ascribing the first to the positron and the second to the electron. The sign of the charge is conventional — only the relative sign is meaningful. What we are going to show is the existence of a correspondence which, to each solution of one of them, provides a solution of the other. To each electron corresponds a particle which differs from it only by the sign of the charge. The operation $\mathcal{C}$ describing this correspondence is called *charge conjugation*. It gives the positron wavefunction $\psi_c(x)$ from the electron wavefunction $\psi(x)$, and vice-versa:

$$\psi_c(x) = \mathcal{C}\,\psi(x).$$

It is, like parity, an involution: $\mathcal{C}^2 = I$. This operation is not given by a simple action of a matrix on $\psi(x)$: there is no 4×4 matrix leading one into the other solution of the two equations above. Notice that we want only to change the relative sign between the kinetic term and the charge term. The complex conjugate of (8.85) is

$$\left[i\, \gamma^{\mu *} (\hbar\, \partial_\mu - \; i\, \frac{e}{c} A_\mu) + mc \right] \psi^*(x) = 0\ . \tag{8.87}$$

To arrive at a solution of (8.86), we should find a matrix taking $\gamma^{\mu *}$ into $(-\,\gamma^\mu)$. It is traditional to write such a matrix in the form $C\gamma^0$:

$$(C\gamma^0)(\gamma^{\mu *})(C\gamma^0)^{-1} = -\,\gamma^\mu\ .$$

If this matrix exists, then

$$\psi_c \equiv \mathcal{C}\,\psi = C\gamma^0\ \psi^*,$$

which would be the desired solution. Now, it so happens that the matrix does exist, and its explicit form depends on the γ-representation used. Let us proceed in the Pauli-Dirac representation, in which $\psi_c \equiv \mathcal{C}\,\psi = C\gamma^0\ \psi^* = C\gamma^0\ (\gamma^0)^T\ \overline{\psi}^{\,T} = C\,\overline{\psi}^{\,T}$. Furthermore, as $\gamma^0\gamma^{\mu\dagger}\gamma^0 = \gamma^\mu$, or $\gamma^0\gamma^{\mu *}\gamma^0 = \gamma^{\mu T}$, C must have the effect $C^{-1}\gamma^\mu C = -\,\gamma^{\mu T}$. As γ^1 and γ^3 are already equal to minus their transposes, C must simply commute with them. As γ^0 and γ^2 are equal to their transposes, C must anticommute with them. Then, up to a phase which will not interest us,

$$C = \; i\, \gamma^2\gamma^0.$$

Notice that $C = -\,C^{-1} = -\,C^\dagger = -\,C^T$. Given now a solution $\psi(x)$ of (8.85), its charge conjugate is

$$\psi_c = i\,\gamma^2\,\psi^* = i\,\gamma^2\,\gamma^0\,\overline{\psi}^{\,T}.$$

8.7 Time Reversal and $\mathcal{CPT}$

The Klein-Gordon equation, being quadratic in the time variable, is automatically invariant under time reversal. Such an invariance reflects our intuitive notion by which, if we look backwards at the motion picture of the evolution of a particle without any energy dissipation, we would see it retrace, though in inverse order, all the points prescribed by the same equation of motion, with inverse initial velocity. In other words, the equation of motion itself must be invariant under time reversal.

Let us take the Dirac equation in its Hamiltonian form (8.17) (with $\hbar = c = 1$), in the presence of an external electromagnetic field:

$$i\,\frac{\partial}{\partial t}\ \psi(\vec{x},t) = \; \left[\vec{\alpha}\cdot(-i\,\vec{\nabla} - e\,\vec{A}) + \;\beta m + e\phi \right] \psi(\vec{x},t). \tag{8.88}$$

A look at the wave equations $\Box \vec{A} = \vec{j}$ and $\Box \phi = \rho$, will tell us that

$$\vec{A}'(\vec{x}, -t) = -\vec{A}(\vec{x}, t) \quad \text{and} \quad \phi'(\vec{x}, -t) = \phi(\vec{x}, t).$$

We shall look for an operation $\mathcal{T}$ implementing time reversal:

$$\psi'(\vec{x}, -t) = \mathcal{T}\psi(\vec{x}, t) \ .$$

One can verify that a simple matrix operation will not do the work. The first member of the above Dirac equation, as well as the usual treatment of the Schrödinger equation, suggest the use of the complex-conjugate equation. In effect, what we shall look for (and find) is a 4×4 matrix T such that

$$\psi'(\vec{x}, t' = -t) = \mathcal{T}\psi(\vec{x}, t) = T\psi^*(\vec{x}, t) \ .$$

Taking the inverse of this expression into the complex conjugate of (8.88), we get

$$i \ \frac{\partial}{\partial t'} \ \psi'(\vec{x}, t') = \ \left[T \ \vec{\alpha}^* \ T^{-1} \cdot (i\vec{\nabla}' + e \ \vec{A}') + T\beta^* T^{-1} m + e\phi' \right] \psi'(\vec{x}, t') \ .$$

To obtain the Dirac equation with reversed time, T must commute with β and α_2, while anticommuting with α_1 and α_3. Up to another phase which we shall not discuss,

$$T = -i\alpha_1\alpha_3 = i\gamma^1\gamma^3.$$

Thus,

$$\mathcal{T}\psi(\vec{x}, t) = \psi'(\vec{x}, -t) = i\gamma^1\gamma^3\psi^*(\vec{x}, t).$$

The time-reversal operation is anti-unitary, and was introduced by Wigner ("Wigner time reversal").

Let us now examine the successive application of the operations $\mathcal{T}$, $\mathcal{C}$ and P:

$$P\mathcal{C}\mathcal{T} \ \psi(\vec{x}, t) = P\mathcal{C}i\gamma^1\gamma^3\psi^*(\vec{x}, t) = Pi\gamma^2[i\gamma^1\gamma^3\psi^*(\vec{x}, t)]^*,$$

or

$$\psi_{P\mathcal{C}\mathcal{T}}(x') \equiv P\mathcal{C}\mathcal{T} \ \psi(\vec{x}, t) = ie^{i\phi}\gamma^5\psi(\vec{x}, t) \ .$$

What we have just seen is a particular case of a very general theorem of the theory of relativistic fields, which says that every possible state for a system of particles is also possible for a system with antiparticles, though with reversed space and time. This $\mathcal{C}P\mathcal{T}$ *theorem* states that $\mathcal{C}P\mathcal{T}$ is an invariance of any Lorentz covariant system which is causal and local.

Chapter 9

Gauge Fields

9.1 Introduction

The study of free fields is essential to introduce the basic notions and methods but has, by itself, small physical content. The attributes of a system can only be discovered by studying its responses to exterior influences. The characteristics of the system supposedly described by the field only can be found and measured via interactions with other systems. In the spirit of field theory, according to which everything must be ultimately described through the mediation of fields, that would mean interactions with other fields. Furthermore, a free field can, due to the symmetries imposed, require the presence of another. A complex scalar field, for instance, has a charge that, if interpreted as the electric charge, calls for (or is the cause of, or still is the source of) another field, the electromagnetic field.

The problem of how to introduce interactions in a relativistic theory has been the object of long discussions. The old notion of potential presented great difficulties. There are still problems in the classical theory (with a finite number of degrees of freedom!). We shall not be concerned with those questions. In field theory, the simplest, straightest way to introduce interaction in a coherent way is provided by the Lagrangian formalism. What is done in practice is to write a total Lagrangian formed by two pieces. The first — the kinematical part — is the sum of the free Lagrangians of all the fields involved. The second has terms representing the interactions supposed to be at work. This takes the general form

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{int}}. \tag{9.1}$$

It is then necessary to calculate the consequences and compare with experiment. In a nutshell: trial and error! For the above referred to complex scalar plus electro-

magnetism case, the total Lagrangian is (7.25):

$$\mathcal{L} = [\partial_\mu - iA_\mu]\phi^* \; [\partial^\mu + iA^\mu]\phi - m^2\phi^*\phi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \; . \tag{9.2}$$

This Lagrangian is the starting point of the electrodynamics of charged scalar mesons (such as $\pi^\pm$ and $K^\pm$). Actually, experiment provides an *a priori* guideline, by establishing conservation laws for the system to be described. Such laws reflect symmetries of the candidate Lagrangian. The above Lagrangian has, besides the Poincaré invariance, an invariance under the gauge transformations

$$\phi'(x) = e^{i\alpha(x)}\phi(x) \; ; \tag{9.3}$$

$$\phi'^*(x) = e^{-i\alpha(x)}\phi^*(x) \; ; \tag{9.4}$$

$$A'_\mu = A_\mu - \partial_\mu\alpha(x) \; . \tag{9.5}$$

Basically, two criteria are used when looking for a Lagrangian: symmetry and (if not redundant) simplicity. We select the simplest combination of fields respecting the symmetries related to the conservation laws. The fields are previously chosen as members of linear representations of the Lorentz group and, to get $\mathcal{L}_{int}$ as a scalar, only their contractions are allowed. Quantum theory adds other requirements, because not every $\mathcal{L}_{int}$ leads to well-defined values for the ensuing calculated quantities. Many lead to infinite values for quantities known to be finite. Actually all of them lead to infinities, but there is a well-established procedure to make them finite (to "renormalize" them). When this procedure fails, $\mathcal{L}_{int}$ is said to be "non-renormalizable" and discarded. This requirement is extremely severe, and eliminates all but a few field combinations. For example, amongst the many Lagrangians conceivable to represent the interaction of a real scalar field with itself, such as

$$\lambda\phi^3, \lambda\phi^4, \lambda\phi^n, \cos[\alpha\phi], e^{\alpha\phi}, \text{etc},$$

only $\lambda\phi^4$ is entirely acceptable in 4−dimensional spacetime (λ and α are "coupling constants"), as leading to a consistent quantum theory. All the other lead to incurable infinities. Thus, the Lagrangian for a self–interacting real scalar field must be

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu\phi \; \partial^\mu\phi - \frac{m^2}{2}\phi^2 - \tfrac{\lambda}{4!}\phi^4 \tag{9.6}$$

(the numerical factor 4! is merely conventional, but convenient in calculations). For other fields the same happens: almost all thinkable Lagrangians lead to unacceptable quantum theories.

A Dirac field in the presence of an electromagnetic field can be obtained from the free Lagrangians by the minimal coupling prescription:

$$\mathcal{L} = \tfrac{i}{2}\left[\overline{\psi}\gamma^\mu\{\partial_\mu - i\frac{e}{c}A_\mu\}\psi - (\{\partial_\mu + i\frac{e}{c}A_\mu\}\overline{\psi})\gamma^\mu\psi\right] - m\overline{\psi}\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \; , \tag{9.7}$$

or

$$\mathcal{L} = \tfrac{i}{2}\left[\overline{\psi}\gamma^\mu\partial_\mu\psi - (\partial_\mu\overline{\psi})\gamma^\mu\psi\right] - m\overline{\psi}\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{e}{c}\overline{\psi}\gamma^\mu A_\mu\psi \ . \tag{9.8}$$

This is the starting point of electrodynamics proper. The interaction Lagrangian is

$$\mathcal{L}_{int} = \frac{e}{c}\overline{\psi}\gamma^\mu\psi A_\mu = j^\mu A_\mu \ . \tag{9.9}$$

For charged self–interacting scalar mesons, the minimal coupling prescription yields the Lagrangian

$$\mathcal{L} = [\partial_\mu - iA_\mu]\phi^* \ [\partial^\mu + iA^\mu]\phi - m^2\phi^*\phi - \tfrac{\lambda}{4!}|\phi^*\phi|^2 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \ , \tag{9.10}$$

which is the same as

$$\mathcal{L} = \partial_\mu\phi^* \ \partial^\mu\phi - m^2\phi^*\phi - \tfrac{\lambda}{4!}|\phi^*\phi|^2 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + A^\mu A_\mu\phi^*\phi - J^\mu A_\mu \ . \tag{9.11}$$

The pion-nucleon coupling is the celebrated Yukawa interaction

$$\mathcal{L}_{int} = g\overline{\psi}\gamma^5\psi\phi \ . \tag{9.12}$$

The γ^5 is necessary if we want that parity be preserved — recall that the pion field is not a scalar, but a pseudo-scalar (at the time of its first proposal, as not even the existence of the pion was known, the γ^5 was "overlooked").

The basic question is: supposing we know the symmetries of a given system, is there a systematic procedure to obtain a Lagrangian? A positive answer is given by gauge theories. Given a free Lagrangian and a symmetry group, they teach us how to obtain the total Lagrangian, which is symmetric (and renormalizable!). The procedure actually generalizes that used to introduce the electromagnetic field in (9.8) and (9.10), which are Lagrangians of a gauge theory. In both, A_μ represents the gauge potential. The difference, in the case of more involved groups like $SU(2), SU(3), SU(5)$, etc, comes essentially from the non-commutativity of their transformations, which engenders self-interactions of the gauge field. Given a symmetry group, we have beforehand to decide (by experimental observation) whether or not the charges involved create a field. If they do not, the Lagrangian is the simplest scalar formed from all the multiplets necessary to classify the particles involved.

9.2 The Notion of Gauge Symmetry

Gauge theories involve a symmetry group (the gauge group) and a prescription (the minimal coupling prescription) to introduce coupling (that is, interactions)

between fields in such a way that the symmetry is preserved around each point of spacetime. They account for three of the four known fundamental interactions of Nature (gravitation, at least for the time being, stands apart). Namely:

Electrodynamics. Describes the electromagnetic interaction of all physical particles; the gauge group is $U(1)$, and the gauge potentials represent the photons; it is the theory showing the best agreement with experimental data.

Weinberg–Salam Theory. It is a gauge theory for electromagnetic and weak interactions; the group is $SU(2) \otimes U(1)$, and the gauge potentials are given by A^μ, corresponding to the photon, and three massive fields W^+_μ, W^-_μ and Z^0_μ, corresponding to the experimentally detected homonym particles; the symmetry is broken, in a way such that the bosons W^+_μ, W^-_μ and Z^0_μ acquire masses; it is not really a unified theory, as it keeps two distinct, independent coupling constants; it has an impressive experimental record.

Chromodynamics. It is a gauge theory for the 8-dimensional SU(3)-color group, and supposed to describe the interactions between the quarks. The gauge potentials are related to the gluons; favored by a good phenomenological evidence, though not completely established. Neither quarks nor gluons have been observed in free state. Despite great efforts, nobody has as yet been able to explain this "color confinement" within the theory.

Take, to fix the ideas, a scalar field endowed with supplementary degrees of freedom (internal, alien to spacetime). If these degrees of freedom assume N values, the field will actually be a set ϕ of N fields, $\phi = \{\phi_i\}$. The behavior must be well-defined, that is, ϕ must belong to some representation of the group, called the "gauge group". For simplicity, one supposes that only linear representations are at work. This means that each group element will be represented by an $N \times N$ matrix U, and the corresponding transformation will be given by

$$\phi_i(x) \to \phi'_i(x) = U_{ij}\ \phi_j(x). \tag{9.13}$$

Notice that the gauge transformation is a transformation at a fixed spacetime point x. The number N depends on the representation, and $i, j = 1, 2, \ldots, N$. The group element U will have the form

$$U(\alpha) = \exp[\alpha^a T_a], \tag{9.14}$$

where α^a, with $a = 1, 2, \ldots, \mathrm{d} = \text{group dimension} = \text{number of generators}\}$ is the set of group parameters, and each T_a is the matrix representing the generator of

transformations along α^a in the representation to which ϕ belongs. In any representation, the representatives of the generators will satisfy the same commutation rules

$$[T_a, T_b] = f^c{}_{ab}\, T_c. \tag{9.15}$$

The $f^c{}_{ab}$'s are the structure constants of the group.

For an infinitesimal transformation, the parameters $\delta\alpha_a$ are small enough so that higher orders are negligible with respect to the first. In this case,

$$\phi'_i(x) = [\exp(\delta\alpha_a\, T^a)]_{ij}\, \phi_j(x) \approx (I + \delta\alpha_a\, T^a)_{ij}\, \phi_j(x), \tag{9.16}$$

that is,

$$\bar{\delta}\phi_i(x) \equiv \phi'_i(x) - \phi_i(x) = (\delta\alpha_a\, T^a)_{ij}\, \phi_j(x). \tag{9.17}$$

In matrix language,

$$\bar{\delta}\phi(x) = \delta\alpha_a\, T^a\, \phi(x). \tag{9.18}$$

9.3 Global Transformations

As long as we suppose constant $\delta\alpha_a$, things are simple. Denoting by $\mathcal{L} = \mathcal{L}[\phi]$ the lagrangian of the field ϕ, its invariance will be written as

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi_i(x)}\, \frac{\bar{\delta}\phi_i(x)}{\delta\alpha_a} + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i(x)}\, \frac{\bar{\delta}(\partial_\mu\phi_i(x))}{\delta\alpha_a} = 0. \tag{9.19}$$

Taking the derivative of (9.17), we find

$$\bar{\delta}\partial_\mu\phi_i = \delta\alpha^c\, (T_c)_{ij}\, \partial_\mu\phi_j, \tag{9.20}$$

so that

$$\frac{\partial\mathcal{L}}{\partial\phi_i}\, (T_c)_{ij}\, \phi_j + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}\, (T_c)_{ij}\, \partial_\mu\phi_j = 0 \tag{9.21}$$

for each generator T_c. Using the Lagrange derivative

$$\frac{\delta\mathcal{L}}{\delta\phi} = \frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i},$$

this is the same as

$$\frac{\delta\mathcal{L}}{\delta\phi}\, (T_c)_{ij}\, \phi_j + \left(\partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}\right)\, (T_c)_{ij}\, \phi_j + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}\, (T_c)_{ij}\, \partial_\mu\phi_j = 0,$$

or

$$\frac{\delta\mathcal{L}}{\delta\phi}\, T_c\phi + \partial_\mu \left[\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} T_c\phi\right] = 0. \tag{9.22}$$

The Noether current will be just

$$J^\mu_c = -\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}\frac{\delta\phi_i}{\delta\alpha^c} = -\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi_i}(T_c)_{ij}\phi_j = -\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}T_c\phi, \tag{9.23}$$

so that

$$\frac{\delta\mathcal{L}}{\delta\phi}T_c\phi = \partial_\mu J^\mu_c. \tag{9.24}$$

The conservation of current comes then directly from the equations of motion $\frac{\delta\mathcal{L}}{\delta\phi} = 0$. Such transformations, with spacetime-independent parameters, will be the same for all events and are consequently called **global transformations**. In the representation of ϕ it is always possible (*almost* always: the symmetry group must be semi–simple) to define an internal scalar product $\phi_i\phi^i$ which is invariant under the group transformations. The invariant Lagrangian will then be

$$\mathcal{L}[\phi] = \tfrac{1}{2}\left[(\partial_\mu\phi_i)^\dagger(\partial^\mu\phi^i) - m^2\ \phi^\dagger_i\phi^i\ \right]. \tag{9.25}$$

9.4 Local Transformations

Suppose now that the parameters in (9.14) are event-dependent, that is, functions of the point in spacetime. Under such **local transformations** the Lagrangian is no more invariant. There is now a new term in $\bar\delta(\partial_\mu\phi_i)$: taking again the derivative of (9.17), we obtain, instead of (9.20),

$$\bar\delta\partial_\mu\phi_i(x) = \delta\alpha^c(x)\ (T_c)_{ij}\ \partial_\mu\phi_j(x) + \partial_\mu\delta\alpha^c(x)(T_c)_{ij}\ \phi_j(x). \tag{9.26}$$

The Lagrangian variation is now

$$\delta\mathcal{L} = \delta\alpha^c\left[\frac{\partial\mathcal{L}}{\partial\phi}\ T_c\phi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\ T_c\partial_\mu\phi\right] + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\ T_c\phi(\partial_\mu\delta\alpha^c). \tag{9.27}$$

The term inside the brackets is the variation by global transformations (9.21), under which the Lagrangian *is* invariant. Thus, by (9.23),

$$\delta\mathcal{L} = -\ J^\mu_c(\partial_\mu\delta\alpha^c). \tag{9.28}$$

All this is reminiscent of what we have seen in Section 7.2, when discussing the complex scalar fields. Also there we had found a breaking in the Lagrangian invariance when the parameters became point–dependent. Terms in the derivatives of the parameters broke the invariance. How did we fare in that case? We have recalled the gauge indeterminacy of the electromagnetic potential, and found that it was possible to use that freedom to compensate the parameter derivative by

a gauge transformation. This, of course, *provided the electromagnetic potential* A_μ *were present*. We were forced to introduce A_μ if we wanted to restore the invariance. There, we have done it through the minimal coupling prescription, by which the derivative is modified. That is what we shall do here: we shall define a new covariant derivative,

$$\mathcal{D}_\mu\phi = \partial_\mu\phi + A_\mu\phi, \tag{9.29}$$

with the obvious difference that A_μ must, now, be a matrix

$$A_\mu = A^a{}_\mu\ T_a. \tag{9.30}$$

Consequently,

$$(\mathcal{D}_\mu\phi)_i = \partial_\mu\phi_i + A^a{}_\mu\ (T_a)_{ij}\phi_j =: (\mathcal{D}_\mu)_{ij}\phi_j =: \mathcal{D}_\mu\phi_i, \tag{9.31}$$

where we have profited to exhibit some usual notations. Let us calculate the variation of this covariant derivative:

$$\begin{aligned}\bar\delta[\mathcal{D}_\mu\phi_i] &= \bar\delta(\partial_\mu\phi_i) + \bar\delta A^a{}_\mu\ (T_a)_{ij}\phi_j + A^a{}_\mu\ (T_a)_{ij}\bar\delta\phi_j = \delta\alpha^c(T_c)_{ij}\partial_\mu\phi_j \\ &+ (\partial_\mu\delta\alpha^c)(T_c)_{ij}\phi_j + \bar\delta A^a{}_\mu(T_a)_{ij}\phi_j + \delta\alpha^c A^a{}_\mu(T_a)_{ij}(T_c)_{jk}\phi_k.\end{aligned}$$

Introducing again $\mathcal{D}_\mu\phi$ instead of $\partial_\mu\phi$, subtracting the added term, using the commutation rules and rearranging the terms, we arrive at

$$\bar\delta[\mathcal{D}_\mu\phi_i] = \delta\alpha^c(T_c)_{ij}\mathcal{D}_\mu\phi_j + [\bar\delta A^c{}_\mu + \partial_\mu\delta\alpha^c - \delta\alpha^d f^c{}_{de}A^e{}_\mu](T_c)_{ik}\phi_k.$$

This variation will come back to the form (9.20),

$$\bar\delta(\mathcal{D}_\mu\phi_i) = \delta\alpha^c(T_c)_{ij}\mathcal{D}_\mu\phi_j, \tag{9.32}$$

if the second term vanishes, that is, provided A_μ transforms according to

$$\bar\delta A^c{}_\mu = -(\partial_\mu\delta\alpha^c + f^c{}_{ed}A^e{}_\mu\delta\alpha^d) =: -\ \mathcal{D}_\mu(\delta\alpha^c). \tag{9.33}$$

Notice that (9.32) attributes to the covariant derivative (as (9.20) gave to the usual derivative) the same behavior the fields have under transformations. That is where the name *covariant derivative* comes from. Under a global transformation, the usual derivative is already automatically covariant.

9.5 Local Noether Theorem

According to the minimal coupling prescription, the original Lagrangian (9.25) has to be modified by the change

$$\partial_\mu\phi_i \to \mathcal{D}_\mu\phi_i$$

It then becomes

$$\mathcal{L}' \equiv \mathcal{L}'[\phi] = \tfrac{1}{2}\ [(\mathcal{D}_\mu\phi_i)^\dagger(\mathcal{D}^\mu\phi^i) - m^2\ \phi_i^\dagger\phi^i\ ]. \tag{9.34}$$

We are supposing a real ϕ_i. As we want (9.13) to be a unitary transformation, the generator matrices must be anti-hermitian, $T_a^\dagger = -\ T_a$ (if we want to use hermitian matrices for the generators T_a, it is necessary to add a factor i in the exponent of (9.14)). In consequence,

$$(\mathcal{D}_\mu\phi)_i^\dagger = \partial_\mu\phi_i^\dagger - A^a{}_\mu\ \phi_j^\dagger\ (T_a)_{ji}. \tag{9.35}$$

Imposing $(\delta\mathcal{L}'/\delta\phi_i^\dagger) = 0$, the equation of motion comes out as

$$(\mathcal{D}_\mu\mathcal{D}^\mu\phi)_i + m^2\phi_i = 0. \tag{9.36}$$

The Lagrangian variation will be, now,

$$\delta\mathcal{L}' = \frac{\partial\mathcal{L}'}{\partial\phi_i}\delta\alpha^c(T_c)_{ij}\phi_j + \frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi_i}\delta\alpha^c(T_c)_{ij}(\mathcal{D}_\mu\phi)_j + \text{hc}\,,$$

with "hc" meaning the "hermitian conjugate". Equivalently,

$$\begin{aligned}\delta\mathcal{L}' = \delta\alpha^c\Bigg[&\frac{\delta\mathcal{L}'}{\delta\phi_i}(T_c)_{ij}\phi_j + \left(\mathcal{D}_\mu^\dagger\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right)_i(T_c)_{ij}\phi_j + \\ &\left(\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right)_i(T_c)_{ij}(\mathcal{D}_\mu\phi)_j\Bigg] + \text{hc},\end{aligned} \tag{9.37}$$

where we have added the second term and subtracted it by absorption into the Lagrange derivative $(\delta\mathcal{L}'/\delta\phi_i)$. Writing now explicitly the covariant derivatives, the last two terms give

$$\begin{aligned}&\partial_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi_i}(T_c)_{ij}\phi_j\right] - A^a_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right]_i(T_a)_{ij}(T_c)_{jk}\phi_k \\ &\qquad + A^a_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right]_i(T_c)_{ij}(T_a)_{jk}\phi_k \\ =\ &\partial_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}T_c\phi\right] - A^a_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right]\ [T_a, T_c]\phi \\ =\ &\partial_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}(T_c)\phi\right] - A^a_\mu\left[\frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\right]\ f^b{}_{ac}T_b\phi \\ =\ &-\left(\partial_\mu J^\mu{}_c - f^b{}_{ac}A^a{}_\mu J^\mu{}_b\right) \equiv -\left(\partial_\mu J^\mu{}_c + f^b{}_{ca}A^a{}_\mu J^\mu{}_b\right) \\ =\ &-D_\mu J^\mu{}_c,\end{aligned}$$

where

$$J^\mu{}_c = -\ \frac{\partial\mathcal{L}'}{\partial\mathcal{D}_\mu\phi}\ T_c\phi \tag{9.38}$$

(compare with (9.23)). The variation of the Lagrangian is then

$$\delta\mathcal{L}' = \delta\alpha^c \left[\frac{\delta\mathcal{L}'}{\delta\phi} T_c\phi - \mathcal{D}_\mu J^\mu{}_c\right]. \tag{9.39}$$

For the solutions of the field equation, that is, for $\delta\mathcal{L}'/\delta\phi = 0$, the invariance of the lagrangian gives, for each component in the algebra,

$$\mathcal{D}_\mu J^\mu{}_c = \partial_\mu J^\mu{}_c + f^b{}_{ca} A^a_\mu J^\mu{}_b = 0. \tag{9.40}$$

In terms of the matrices $J^\mu := J^\mu{}_c T^c$ and $A_\mu := A^a_\mu T_a$, the covariant divergence becomes

$$\mathcal{D}_\mu J^\mu = \partial_\mu J^\mu + [A_\mu, J^\mu]. \tag{9.41}$$

This is the same expression found above, if we can use the cyclic property ($f^b{}_{ac} = f^c{}_{ba} = f^a{}_{cb}$) of the structure constants, valid if the group is semi–simple. In this case, the conservation law assumes the form

$$\mathcal{D}_\mu J^\mu = \partial_\mu J^\mu + [A_\mu, J^\mu] = 0. \tag{9.42}$$

This covariant derivative differs from that of (9.29). The covariant derivative of a quantity depends on how the quantity is represented: ϕ is a column vector and there $\mathcal{D}_\mu$ acts as a matrix on a column. The current J_μ is a matrix, and $\mathcal{D}_\mu$ acts on it through a commutator. We shall see that the same happens to A_μ. The covariant derivative depends also on the spacetime indices, in a way quite analogous to the usual differentials. In the case above we have a divergence.

Equation (9.42) is not a real conservation law. The current is not conserved ($\partial_\mu J^\mu \neq 0$), it has only vanishing *covariant* divergence. This is exactly the concern of the second Noether theorem. It does not lead directly to a conserved quantity. It is a constraint imposed on the current to ensure the invariance of the modified Lagrangian.

The modified Lagrangian,

$$\mathcal{L}'[\phi] = \mathcal{L}'[\phi_i, \mathcal{D}_\mu\phi_j] \tag{9.43}$$

will depend on A^a_μ only through $\mathcal{D}_\mu\phi$. The current (9.38) can then take a simpler form. From (9.31) for fixed values of a and μ,

$$(T_a)_{ij}\phi_j = \frac{\partial(\mathcal{D}_\mu\phi)_i}{\partial A^a{}_\mu}, \tag{9.44}$$

so that

$$J^\mu{}_a = -\frac{\partial\mathcal{L}'}{\partial A^a{}_\mu}. \tag{9.45}$$

9.6 Field Strength and Bianchi Identity

Unlike usual derivatives, covariant derivatives do not commute: it is easy to check that

$$[\mathcal{D}_\mu, \mathcal{D}_\nu]\phi = F^a{}_{\mu\nu}\, T_a\, \phi, \tag{9.46}$$

where

$$F^a{}_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\mu + f^a{}_{bc} A^b{}_\mu A^c{}_\nu. \tag{9.47}$$

The matrix

$$F_{\mu\nu} = T_a F^a{}_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \tag{9.48}$$

is the field strength. If the group is abelian, $f^a{}_{bc} = 0$, and the last term does not exist. If, furthermore, the group has only one generator, the expression above reduces to that holding for the electromagnetic field. Matrix (9.48), thus, generalizes the electromagnetic field strength to the non-abelian case.

We can write (9.33) in matrix form:

$$\bar{\delta} A_\mu = -\partial_\mu \delta\alpha + [\delta\alpha, A_\mu], \tag{9.49}$$

with $\delta\alpha = T_a \delta\alpha^a$. It is then easy to find that

$$\bar{\delta} F_{\mu\nu} = [\delta\alpha, F_{\mu\nu}], \tag{9.50}$$

which is the same as

$$\bar{\delta} F^c{}_{\mu\nu} = \delta\alpha^a f^c{}_{ab} F^b{}_{\mu\nu}. \tag{9.51}$$

This is the infinitesimal form of

$$F'_{\mu\nu} = e^{\alpha^a(x) T_a} F_{\mu\nu} e^{-\alpha^b(x) T_b}. \tag{9.52}$$

This means that F is covariant. We see that, in the abelian case, $F_{\mu\nu}$ is simply invariant. This is the case, in particular, of electromagnetism. In the general case, $F_{\mu\nu}$ behaves as a matrix — it is not invariant, but covariant. On the other hand, (9.49) is the infinitesimal version of

$$A'_\mu = e^{\alpha^a T_a} A_\mu e^{-\alpha^b T_b} + e^{\alpha^a T_a} \partial_\mu e^{-\alpha^b T_b}, \tag{9.53}$$

which shows that A_μ is not strictly covariant. This is the expression for the gauge transformation of A_μ in the general non-abelian case.

We have seen that the expression of the covariant derivative changes in each case. Acting on a Lorentz scalar, which is furthermore a column vector in internal space, it has the form (9.29). Acting on a Lorentz vector which is furthermore a

matrix in internal space, it can assume two forms, corresponding to the divergence and the rotational in usual vector analysis. That corresponding to the divergence we have seen in (9.41). As to the rotational, it has exactly the form given in (9.48): the field strength is the covariant derivative of the potential. This generalizes the relation $\mathbf{B} = \mathrm{rot}\mathbf{A}$ of electromagnetism. It is frequent to write symbolically

$$F_{\mu\nu} = \mathcal{D}_\mu A_\nu, \tag{9.54}$$

meaning by that just (9.48). The kind of derivative (divergence or rotational) depends on the resultant indices (contracted or not), and the name covariant derivative is used for both. This is a physicists' practice, which actually mixes up two quite distinct mathematical notions, that of exterior derivative and that of coderivative (the derivative of the dual).

Gauge theories are very near to differential geometry, and the most appropriate language to treat them is that of differential forms. Well, also a tensor like $F_{\mu\nu}$ has its covariant derivatives, with and without contraction. One of them is

$$\mathcal{D}_\rho F_{\mu\nu} = T_a \left[\partial_\rho F^a{}_{\mu\nu} + f^a{}_{bc} A^b{}_\rho F^c{}_{\mu\nu}\right] = \partial_\rho F_{\mu\nu} + [A_\rho, F_{\mu\nu}]. \tag{9.55}$$

From the very definition of $F_{\mu\nu}$ we obtain, by using the Jacobi identity

$$[T_a, [T_b, T_c]] + [T_c, [T_a, T_b]] + [T_b, [T_c, T_a]] = 0, \tag{9.56}$$

the following identity:

$$\mathcal{D}_\rho F_{\mu\nu} + \mathcal{D}_\nu F_{\rho\mu} + \mathcal{D}_\mu F_{\nu\rho} = 0. \tag{9.57}$$

The indices are exchanged cyclically from term to term. This **Bianchi identity** generalizes to the non-abelian case the so-called first pair of Maxwell's equations. Recall that those equations do not follow from the electromagnetic Lagrangian, and in this sense are not dynamical.

9.7 Gauge Lagrangian and Field Equation

Concerning the Lagrangian for the gauge field itself, it is possible to show that, in order to be gauge–invariant, it can depend on the potential A_μ only through the field strength $F_{\mu\nu}$. Its simplest expression is (formally) the same as that of the electromagnetism:

$$\mathcal{L}_G = -\tfrac{1}{4}\, F^a{}_{\mu\nu} F_a{}^{\mu\nu}. \tag{9.58}$$

We could think of using the dual of $F_{\mu\nu}$, defined as

$$\tilde{F}_a{}^{\mu\nu} = \tfrac{1}{2}\, \epsilon^{\mu\nu\rho\sigma} F_{a\rho\sigma}. \tag{9.59}$$

It so happens that $\tilde{F}_a{}^{\mu\nu}\tilde{F}^a{}_{\mu\nu}$ is proportional to the above $\mathcal{L}_G$, so that it adds nothing to dynamics. And also that $F_a{}^{\mu\nu}\tilde{F}^a{}_{\mu\nu}$ is an exact differential (the divergence of a certain current), which does not contribute to the equations of motion. Actually,

$$C = \int d^4x F_a{}^{\mu\nu}\tilde{F}^a{}_{\mu\nu} \tag{9.60}$$

contains information on the topology involved. It is an invariant number, which allows to classify the gauge fields into families. Its values are topological numbers, analogous to that seen for the sine-Gordon field of section 4.4.

The total Lagrangian of the system will then be (9.58) plus (9.43):

$$\mathcal{L} = \mathcal{L}'[\phi_i, \mathcal{D}_\mu\phi_j] - \tfrac{1}{4}\ F^a{}_{\mu\nu}F_a{}^{\mu\nu}. \tag{9.61}$$

The corresponding Euler-Lagrange equation is

$$\partial_\mu F^{a\mu\nu} + f^a{}_{bc}A^b{}_\mu F^{c\mu\nu} = J^{a\nu}, \tag{9.62}$$

where use has been made of (9.45), and of the cyclic property already mentioned,

$$f_{abc} = f_{cab} = f_{bca}, \tag{9.63}$$

valid for semisimple groups. On such groups (which, by definition, have no invariant abelian subgroup), there exists a metric, the Killing-Cartan metric

$$\gamma_{ab} = f^c{}_{ad}f^d{}_{bc}\ , \tag{9.64}$$

which can be used to rise and lower internal indices. We have used it implicitly every time some lower internal index appeared as in (9.58). A Lie group is a differential manifold, and γ_{ab} is a metric on that group manifold, which has the special property of being invariant under the group transformations. In this sense, all those expressions are scalar products in internal space, invariant under the group transformations. The Lagrangian, an invariant, can only have indices contracted in this way.

The field equations (9.62) are called the Yang-Mills equations. They govern the field mediating all known interaction–mediating fields, if we exclude the case of gravitation. They can be written in matrix form as

$$\partial_\mu F^{\mu\nu} + [A_\mu, F^{\mu\nu}] = J^\nu. \tag{9.65}$$

The left-hand side is the second form of the covariant derivative of $F^{\mu\nu}$, to which we have alluded above. This equation generalizes (the second pair of) Maxwell's

equations to the non-abelian case with several internal degrees of freedom. If, by analogy with (9.45), we define

$$j^{\nu}{}_{a} = -\frac{\partial \mathcal{L}_G}{\partial A^{a}{}_{\nu}} \tag{9.66}$$

as the current of the gauge field itself, we shall have

$$j^{\nu}{}_{a} = A^{b}{}_{\mu} f_{abc} F^{c\nu\mu}. \tag{9.67}$$

In matrix form, it is

$$j^{\nu} = T^{a} j^{\nu}{}_{a} = [A_\mu, F^{\nu\mu}] = -[A_\mu, F^{\mu\nu}]. \tag{9.68}$$

The Yang-Mills equations become

$$\partial_\mu F^{\mu\nu} = j^{\nu} + J^{\nu}, \tag{9.69}$$

from which

$$\partial_\nu (j^{\nu} + J^{\nu}) = 0. \tag{9.70}$$

We see here what happens concerning current conservation. It is not only J^μ (the external source current) which is to be considered, but the total current, including the gauge field current j^μ itself. The meaning of the "self-current" j^μ is important: the gauge field can be its own source. This effect comes from the non-linear character of the theory, which is a consequence of the non-abelian character of the gauge group. Non-abelian gauge fields, even in the absence of external sources, are highly non-trivial, because they are self-producing. They are never actually "free", as they are always, at least, in interaction with each other. In the quantum case, the quanta of the gauge fields carry themselves the charges of the theory (as if the photons carried electric charges).

This interpretation, though satisfactory from the point of view of the conservation law, is not without difficulties. The problem is that the total current $j^\nu + J^\nu$ is not gauge-covariant. This is reflected in the charges themselves, which are given by

$$Q = \int_V d^3x \partial_\mu F^{\mu 0} = \int_V d^3x \partial_i F^{i0} = \int_{\partial V} d^2\sigma^i F_{i0}. \tag{9.71}$$

As F is covariant (see (9.52)), the charge will change, under a gauge transformation, as

$$Q \Rightarrow Q' = \int_{\partial V} d^2\sigma^i \; U(x) F_{i0} U^{-1}(x). \tag{9.72}$$

Thus, only if we suppose that $U(x) = e^{\alpha^a(x) T_a}$ becomes constant on a far enough spacelike surface ∂V, can we extract U from inside the integral and get covariant charges, that is, charges satisfying

$$Q' = U \; Q \; U^{-1}. \tag{9.73}$$

This imposes a limitation on the local gauge invariance. The charges only make sense if the transformations become global (that is, constant) at ∂V. The latter can be placed, if we like, at space infinity. This problem is quite analogous to that of General Relativity, in which the total energy-momentum (which plays there the role of the above current) is not covariant and, as a consequence, the energy (one of the corresponding charges) can only be defined for asymptotically flat spaces.

The energy-momentum tensor of a gauge field will have the same form of that of the electromagnetic field:

$$\Theta^{\mu\nu} = F_a{}^{\mu\rho} F^{a\nu}{}_{\rho} - \tfrac{1}{4}\, \eta^{\mu\nu} F_a{}^{\rho\sigma} F^a{}_{\rho\sigma}. \tag{9.74}$$

9.8 Final Remarks

Let us repeat ourselves a bit. The field ϕ above can belong to any linear representation of the gauge group G. For each representation a covariant derivative is defined: take the generators T^a in some representation, and use (9.31). In a singlet representation, the second term in (9.31) vanishes and the covariant derivative reduces to the usual derivative. A singlet field does not "feel" the gauge potential. A representation of special significance is the adjoint representation, a matrix $d \times d$ (d = dim G) representation in which the generators are matrices whose entries are the structure constants [see Eq.(2.16) for a variant],

$$(T_a)^c{}_b = f^c{}_{ab}\,. \tag{9.75}$$

In this case, the indices $i, j, k, \ldots$ used above, vary with the same range as the indices $a, b, c, \ldots$. If ϕ belongs to the adjoint representation, it will be a matrix $\phi = T_a \phi^a$ instead of a column, and the covariant derivative has the form

$$\mathcal{D}_\mu \phi = \partial_\mu \phi + [A_\mu, \phi]. \tag{9.76}$$

The fields A_μ and $F_{\mu\nu}$, in particular, belong to the adjoint representation. The expressions (9.47) and (9.48) give the covariant derivative of a vector field in the adjoint representation. The left-hand side of (9.57) is the expression of the covariant derivative of $F_{\mu\nu}$ (there, $\mathcal{D}_\mu$ is given by (9.55)). Thus, the true covariant derivative of an antisymmetric second-order tensor in the adjoint representation is the cyclic sum in (9.57). Besides those covariant derivatives, there are the covariant coderivatives or, roughly speaking, the derivatives of the dual fields. The Yang-Mills equations state that the covariant coderivative of the field strength equals the source current.

The reader may think that we are exchanging the roles. After all, it is the dual who seems to be derived in (9.57), as there a cyclic sum is at work; whereas only the

field, and not its dual, appears in (9.62). The reason for this apparent contradiction is a certain opposition between the nomenclatures used by physicists and mathematicians. Physicists are used to write everything in terms of components, while mathematicians view ϕ, $A = A_\mu dx^\mu$ and $F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu$ as differential forms and write things in invariant language, with no components in sight. Differential forms inhabit integrands. In the integration sign, also the measure is written in invariant form, so as to have the same expression in any coordinate system. Instead of $\int d^4x$, mathematicians write

$$\int \sqrt{|g|}\; d^4x = \int J\; d^4x.$$

This is symbolic, but the jacobian determinant J, whose presence is compulsory in invariant language, contains a Levi-Civita symbol $\epsilon^{\mu\nu\rho\sigma}$. There is, thus, a difference of a Levi-Civita symbol between physicists and mathematicians (and many other, of which we shall say nothing). For example, in mathematical language the action (9.58) is written (with the measure always omitted)

$$\mathcal{L}_G = -\frac{1}{2}\int \mathrm{tr}(F \wedge \tilde{F}), \tag{9.77}$$

with the dual in the scene. This is the reason to say that the Bianchi identity states the vanishing of the covariant derivative of $F_{\mu\nu}$, whereas the sourceless Yang–Mills equation states the vanishing of its covariant coderivative.

The best experimentally verified of all gauge theories is the Weinberg-Salam model, a theory whose gauge symmetry is broken. In the process of symmetry breaking, the Lagrangian remains invariant, but the lowest-energy state (the fundamental state) is not symmetric under gauge transformations (this is called "spontaneous breakdown" of a symmetry). Thus, even electrodynamics is a subtheory of a symmetry-broken theory. Notice that in (9.58) there is no mass term: the quanta of the gauge field have zero masses. This changes in a broken theory and the particles (the "gauge bosons") emerging in the Weinberg-Salam theory have, with the exception of the photon, masses between 90 and 100 Gev. They are the W^+, W^- and Z^0, experimentally discovered in the eighties. And all that is perhaps related to what is nowadays the central question of Elementary Particle Physics, the confinement problem. Chromodynamics is a gauge theory for the group SU(3), supposed to account for the interactions between quarks. It has a very good phenomenological record, but nobody knows why free quarks are not found in Nature. They are supposed to be confined to the interior of the particles by the very gauge field which mediates their interaction. There is a large evidence for that, but it has been impossible up to now to demonstrate quark confinement using the theory. Also the gauge quanta ("gluons") are never seen in free state (a phenomenon called

"shielding"), another property which should be deduced from the theory. Actually, the calculations are very, very complicated, and it is not known whether the theory explains these properties or not.

The structure of gauge theories is fairly geometric. Only the dynamic part (Lagrangian and Yang-Mills equations) is actually Physics. All the characters above have their mathematical counterparts, sometimes with different names. As an interchange of names become more and more frequent in the physical literature, we give a short glossary:

Physics Name	**Mathematics Name**
gauge potential	connection
field strength	curvature
gauge group	structure group
internal space	fiber
external space (spacetime)	base manifold
spacetime + internal space	fiber bundle

Locally, around each point of spacetime, the complete space (base + fiber) is a direct product of both spaces. But fiber bundles, globally, are not necessarily the direct product of the base manifold and the fiber. The simplest examples: a torus is the (global) direct product of two circles; a cylinder is a (global) direct product of a circle and a straight line; the Möbius band is a direct product locally, but not globally.

Chapter 10

Gravitational Field

10.1 General Concepts

All elementary particles feel gravitation the same. More specifically, particles with different masses experience a different gravitational force, but in such a way that all of them acquire the same acceleration and, given the same initial conditions, follow the same path. Such *universality* of response is the most fundamental characteristic of the gravitational interaction. It is a unique property, peculiar to gravitation: no other basic interaction of Nature has it.

Due to universality, the gravitational interaction admits a description which makes no use of the concept of *force*. In this description, instead of acting through a force, the presence of a gravitational field is represented by a deformation of the spacetime structure. This deformation, according to General relativity, preserves the pseudo-riemannian character of the *flat* Minkowski spacetime, the non-deformed spacetime that represents absence of gravitation. In other words, the presence of a gravitational field is supposed to produce *curvature*, but no other kind of spacetime deformation.

A free particle in flat space follows a straight line, that is, a curve keeping a constant direction. A geodesic is a curve keeping a constant direction on a curved space. As the only effect of the gravitational interaction is to bend spacetime so as to endow it with curvature, a particle submitted exclusively to gravity will follow a geodesic of the deformed spacetime.

This is the approach of Einstein's General Relativity, according to which the gravitational interaction is described by a *geometrization*. It is important to remark that only an interaction presenting the property of universality can be described by such a geometrization.

10.2 The Equivalence Principle

Equivalence is a guiding principle, which inspired Einstein in his construction of General Relativity. It is firmly rooted on experience.* In its most usual form, the Principle includes three sub–principles: the weak, the strong, and that which is called "Einstein's equivalence principle". Let us shortly list them with a few comments.

- The **weak equivalence principle** states the universality of free fall, or the equality *inertial mass* = *gravitational mass*. It can be stated in the form:

  In a gravitational field, all pointlike structureless particles follow one same path. That path is fixed once given (i) *an initial position* $x(t_0)$ *and* (ii) *the correspondent velocity* $\dot{x}(t_0)$.

  This leads to an equation of motion which is a second-order ordinary differential equation. No characteristic of any special particle, no particular property appears in the equation. Gravitation is consequently universal. Being universal, it can be seen as a property of space itself. It determines geometrical properties which are common to all particles. The weak equivalence principle goes back to Galileo. It raises to the status of fundamental principle a deep experimental fact: the equality of inertial (m_i) and gravitational (m_g) masses of all bodies. If these masses were not equal, Newton's second law would be written as

  $$\vec{F} = m_i \, \vec{a},$$

  whereas the law of gravitation would be

  $$\vec{F} = m_g \, \vec{g},$$

  with $\vec{g}$ the acceleration produced by a gravitational field. The acceleration at a given point would then be

  $$\vec{a} = \frac{m_g}{m_i} \, \vec{g}.$$

  and would be different for different bodies. Along the history, many different experiments have been performed to test for this difference, all of them yielding a negative result.

* Those interested in the experimental status will find a recent appraisal in C. M. Will, *The Confrontation between General Relativity and Experiment*, arXiv:gr-qc/0510072 16 Out 2005. Theoretical issues are discussed by B. Mashhoon, *Measurement Theory and General Relativity*, gr-qc/0003014, and *Relativity and Nonlocality*, gr-qc/0011013 v2.

- The **strong equivalence principle** (Einstein's lift) says that

  gravitation can be made to vanish locally through an appropriate choice of frame.

  It requires that, for any and every particle, and at each point x_0, there exists a frame in which $\ddot{x}^\mu = 0$.[†]

- **Einstein's equivalence principle** requires, besides the weak principle, the local validity of Poincaré invariance — that is, of Special Relativity. It can be stated in the form:

  Every law of physics reduces locally to that of Special Relativity through an appropriate choice of frame.

  This invariance is, in Minkowski space, summed up in the Lorentz metric. The requirement suggests that the above deformation caused by gravitation is a change in that metric.

Forces equally felt by all bodies were known since long. They are the inertial forces, whose name comes from their turning up in non-inertial frames. Examples on Earth (not an inertial system!) are the centrifugal force and the Coriolis force. Universality of inertial forces has been the first hint towards General Relativity. A second ingredient is the notion of field. The concept allows the best approach to interactions coherent with Special Relativity. All known forces are mediated by fields on spacetime. Now, if gravitation is to be represented by a field, it should, by the considerations above, be a universal field, equally felt by every particle. It should change spacetime itself. And, of all the fields present in a space, the metric — the first fundamental form, as it is also called — seemed to be the basic one. The simplest way to change spacetime would be to change its metric. Furthermore, the metric does change when looked at from a non-inertial frame, where the inertial forces are present. The gravitational field, therefore, is represented by the spacetime metric. In the absence of gravitation, the spacetime metric reduces to the Minkowski metric.

[†] A precise, mathematically sound formulation of the strong principle can be found in R. Aldrovandi, P. B. Barros & J. G. Pereira: *The equivalence principle revisited*, Foundations of Physics **33** (2003) 545-575 — arXiv:gr-qc/0212034.

10.3 Pseudo-Riemannian Metric

Each spacetime is a 4-dimensional pseudo–riemannian manifold. Its main character is the fundamental form, or metric. For example, the spacetime of special relativity is the flat Minkowski spacetime. Minkowski space is the simplest, standard spacetime, and its metric, called the Lorentz metric, is denoted

$$\eta(x) = \eta_{ab} dx^a dx^b. \tag{10.1}$$

It is a rather trivial metric. Up to the signature, the Minkowski space is an Euclidean space, and as such can be covered by a single, global coordinate system. This system — the cartesian system — is the father of all coordinate systems, and just puts η in the diagonal form

$$\eta = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{10.2}$$

The Minkowski line element, therefore, is

$$ds^2 = \eta_{ab} dx^a dx^b = dx^0 dx^0 - dx^1 dx^1 - dx^2 dx^2 - dx^3 dx^3$$

or

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2. \tag{10.3}$$

On the other hand, the metric of a general 4-dimensional pseudo-riemannian spacetime will be denoted by

$$g(x) = g_{\mu\nu} dx^\mu dx^\nu. \tag{10.4}$$

We are using indices $\mu, \nu, \lambda, \ldots$ for the pseudo-riemannian spacetime, and $a, b, c, \ldots$ for the Minkowski spacetime. Like the Minkowski metric, it has signature 2. Being symmetric, the matrix $g(x) = (g_{\mu\nu})$ can be diagonalized. Signature concerns the signs of the eigenvalues: it is the number of eigenvalues with one sign minus the number of eigenvalues with the opposite sign. It is important because it is an invariant under changes of coordinates and vector bases. In the convention we shall adopt this means that, at any selected point P, it is possible to choose coordinates $\{x^\mu\}$, in terms of which $g_{\mu\nu}$ takes the form

$$g(P) = \begin{pmatrix} +|g_{00}| & 0 & 0 & 0 \\ 0 & -|g_{11}| & 0 & 0 \\ & 0 & -|g_{22}| & 0 \\ 0 & 0 & 0 & -|g_{33}| \end{pmatrix}. \tag{10.5}$$

10.4 The Notion of Connection

In a general pseudo-riemannian spacetime, the ordinary derivative of a tensor is not covariant under a general coordinate transformation $x^\mu \to x'^\mu$. In order to define a covariant derivative, it is necessary to introduce a "compensating field", that is, a *connection* which we will denote by Γ. The covariant derivative of a function ϕ (tensor of zero degree) is the usual derivative,

$$D_\mu \phi = \partial_\mu \phi,$$

which is automatically covariant.

Now, take a first order "contravariant" tensor ϕ^ν. Its covariant derivative will be given by

$$D_\mu \phi^\nu = \partial_\mu \phi^\nu + \Gamma^\nu{}_{\lambda\mu} \phi^\lambda. \tag{10.6}$$

The covariant derivative of a "covariant" vector ϕ_ν, on the other hand, is

$$D_\mu \phi_\nu = \partial_\mu \phi_\nu - \Gamma^\lambda{}_{\nu\mu} \phi_\lambda. \tag{10.7}$$

The metric tensor, in particular, will have the covariant derivative

$$D_\mu g_{\rho\sigma} = \partial_\mu g_{\rho\sigma} - \Gamma^\lambda{}_{\rho\mu} g_{\lambda\sigma} - \Gamma^\lambda{}_{\sigma\mu} g_{\rho\lambda}. \tag{10.8}$$

Let us take now a third order mixed tensor $\phi^\nu{}_{\rho\sigma}$. Its covariant derivative will be given by

$$D_\mu \phi^\nu{}_{\rho\sigma} = \partial_\mu \phi^\nu{}_{\rho\sigma} + \Gamma^\nu{}_{\lambda\mu} \phi^\lambda{}_{\rho\sigma} - \Gamma^\lambda{}_{\rho\mu} \phi^\nu{}_{\lambda\sigma} - \Gamma^\lambda{}_{\sigma\mu} \phi^\nu{}_{\rho\lambda}. \tag{10.9}$$

The rules to writing the covariant derivative are fairly illustrated in these examples. Notice the signs: positive for upper indices, negative for lower indices.

Under a general coordinate transformation $x^\mu \to x'^\mu$, in order to yield an appropriate behavior to the covariant derivative, the connection Γ must transform according to

$$\Gamma^\mu{}_{\nu\lambda} = \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x'^\gamma}{\partial x^\nu} \frac{\partial x'^\beta}{\partial x^\lambda} \Gamma'^\alpha{}_{\gamma\beta} + \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial^2 x'^\alpha}{\partial x^\nu \partial x^\lambda}. \tag{10.10}$$

This non-covariant behavior of the *connection* Γ makes of the covariant derivative a well-behaved, a really *covariant* object.

When the covariant derivative of a tensor is zero on a domain, this tensor is said to be "self–parallel" on the domain, or *parallel–transported*. An intuitive view of this notion can be get by noting that it exactly translates to curved space the idea of a straight line as a curve with maintains its direction along all its length. If the metric is parallel–transported, $D_\mu g_{\rho\sigma} = 0$, and the equation above gives the *metricity condition*

$$\partial_\mu g_{\rho\sigma} = \Gamma^\lambda{}_{\rho\mu} g_{\lambda\sigma} + \Gamma^\lambda{}_{\sigma\mu} g_{\rho\lambda} = \Gamma_{\sigma\rho\mu} + \Gamma_{\rho\sigma\mu} = 2\, \Gamma_{(\rho\sigma)\mu}, \tag{10.11}$$

where the symbol with lowered index is defined by $\Gamma_{\rho\sigma\mu} = g_{\rho\lambda}\Gamma^{\lambda}{}_{\sigma\mu}$ and the compact notation for the symmetrized part

$$\Gamma_{(\rho\sigma)\mu} = \tfrac{1}{2}\ \{\Gamma_{\rho\sigma\mu} + \Gamma_{\sigma\rho\mu}\}, \tag{10.12}$$

has been introduced. The analogous notation for the antisymmetrized part

$$\Gamma_{[\rho\sigma]\mu} = \tfrac{1}{2}\ \{\Gamma_{\rho\sigma\mu} - \Gamma_{\sigma\rho\mu}\} \tag{10.13}$$

is also very useful.

10.5 Curvature and Torsion

A connection defines covariant derivatives of general tensorial objects. It goes actually a little beyond tensors. A connection Γ defines a covariant derivative of itself. This gives, rather surprisingly, a tensor, the Riemann curvature tensor of the connection:

$$R^{\kappa}{}_{\lambda\rho\sigma} = \partial_\rho \Gamma^{\kappa}{}_{\lambda\sigma} - \partial_\sigma \Gamma^{\kappa}{}_{\lambda\rho} + \Gamma^{\kappa}{}_{\nu\rho}\Gamma^{\nu}{}_{\lambda\sigma} - \Gamma^{\kappa}{}_{\nu\sigma}\Gamma^{\nu}{}_{\lambda\rho}. \tag{10.14}$$

It is important to notice the position of the indices in this definition. Authors differ in that point, and these differences can lead to differences in the signs (for example, in the scalar curvature defined below). We are using all along notations consistent with the differential forms. There is a clear antisymmetry in the last two indices,

$$R^{\kappa}{}_{\lambda\rho\sigma} = -\ R^{\kappa}{}_{\lambda\sigma\rho}.$$

Other tensors can be obtained from the Riemann curvature tensor by contraction. The most important is the Ricci tensor

$$R_{\lambda\sigma} = R^{\rho}{}_{\lambda\rho\sigma} = \partial_\rho \Gamma^{\rho}{}_{\lambda\sigma} - \partial_\sigma \Gamma^{\rho}{}_{\lambda\rho} + \Gamma^{\rho}{}_{\nu\rho}\Gamma^{\nu}{}_{\lambda\sigma} - \Gamma^{\rho}{}_{\nu\sigma}\Gamma^{\nu}{}_{\lambda\rho}\ . \tag{10.15}$$

A further contraction with the metric tensor gives the scalar curvature

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{10.16}$$

Another important property of connections is their torsion tensor, which is defined by

$$T^{\lambda}{}_{\mu\nu} = \Gamma^{\lambda}{}_{\nu\mu} - \Gamma^{\lambda}{}_{\mu\nu} = 2\Gamma^{\lambda}{}_{[\nu\mu]}. \tag{10.17}$$

Notice that what exists is the curvature and torsion of a connection. Many connections are defined on a given space, each one with its curvature and torsion. It is common language to speak of "the curvature of space" and "torsion of space", but this only makes sense if a certain connection is assumed to be included in the very definition of that space.

10.6 The Levi-Civita Connection

There are, actually, infinite connections on a manifold, infinite objects behaving according to (10.10). And, given a metric, there are infinite connections satisfying the metricity condition. One of them, however, is special. It is given by

$$\overset{\circ}{\Gamma}{}^{\lambda}{}_{\rho\sigma} = \tfrac{1}{2}\, g^{\lambda\mu}[\partial_\rho g_{\sigma\mu} + \partial_\sigma g_{\rho\mu} - \partial_\mu g_{\rho\sigma}]. \tag{10.18}$$

It is the single connection satisfying the metricity condition

$$\overset{\circ}{D}_\mu g_{\rho\sigma} \equiv \partial_\mu g_{\rho\sigma} - \overset{\circ}{\Gamma}{}^{\lambda}{}_{\rho\mu} g_{\lambda\sigma} - \overset{\circ}{\Gamma}{}^{\lambda}{}_{\sigma\mu} g_{\rho\lambda} = 0, \tag{10.19}$$

and which is symmetric in the last two indices. This symmetry has a deep meaning as it means that torsion is vanishing. Connection (10.18) is called the *Levi–Civita connection*. Its components are also called the *Christoffel symbols*. It has, as said, a special relationship to the metric and is the only metric–preserving connection with zero torsion. Standard General Relativity works only with such a connection.

The curvature of a Levi-Civita connection,

$$\overset{\circ}{R}{}^{\kappa}{}_{\lambda\rho\sigma} = \partial_\rho \overset{\circ}{\Gamma}{}^{\kappa}{}_{\lambda\sigma} - \partial_\sigma \overset{\circ}{\Gamma}{}^{\kappa}{}_{\lambda\rho} + \overset{\circ}{\Gamma}{}^{\kappa}{}_{\nu\rho} \overset{\circ}{\Gamma}{}^{\nu}{}_{\lambda\sigma} - \overset{\circ}{\Gamma}{}^{\kappa}{}_{\nu\sigma} \overset{\circ}{\Gamma}{}^{\nu}{}_{\lambda\rho}\ , \tag{10.20}$$

has some special symmetries in the indices, which can be obtained from the detailed expression in terms of the metric:

$$\overset{\circ}{R}_{\kappa\lambda\rho\sigma} = -\overset{\circ}{R}_{\kappa\lambda\sigma\rho} = \overset{\circ}{R}_{\lambda\kappa\sigma\rho}, \tag{10.21}$$

as well as

$$\overset{\circ}{R}_{\kappa\lambda\rho\sigma} = \overset{\circ}{R}_{\rho\sigma\kappa\lambda}. \tag{10.22}$$

In these expressions,

$$\overset{\circ}{R}_{\kappa\lambda\rho\sigma} = g_{\kappa\mu} \overset{\circ}{R}{}^{\mu}{}_{\kappa\lambda\rho\sigma} = \partial_\rho \overset{\circ}{\Gamma}_{\kappa\lambda\sigma} - \partial_\sigma \overset{\circ}{\Gamma}_{\kappa\lambda\rho} + \overset{\circ}{\Gamma}_{\kappa\nu\rho} \overset{\circ}{\Gamma}{}^{\nu}{}_{\lambda\sigma} - \overset{\circ}{\Gamma}_{\kappa\nu\sigma} \overset{\circ}{\Gamma}{}^{\nu}{}_{\lambda\rho}\ , \tag{10.23}$$

where $\overset{\circ}{\Gamma}_{\mu\rho\sigma} = g_{\mu\nu} \overset{\circ}{\Gamma}{}^{\nu}{}_{\rho\sigma}$. As a consequence, the Ricci tensor is also symmetric:

$$\overset{\circ}{R}_{\mu\nu} = \overset{\circ}{R}_{\nu\mu}. \tag{10.24}$$

In consequence of these symmetries, the Ricci tensor (10.15) is essentially the only contracted second-order tensor obtained from the Riemann tensor. The scalar curvature will be now

$$\overset{\circ}{R} = g^{\mu\nu} \overset{\circ}{R}_{\mu\nu}. \tag{10.25}$$

10.7 Geodesics

As we have already seen, the action describing a free particle of mass m in the Minkowski spacetime is

$$S = -mc \int_a^b ds, \tag{10.26}$$

where

$$ds = (\eta_{ab} dx^a dx^b)^{1/2}. \tag{10.27}$$

In the presence of gravitation, that is, in a pseudo-riemannian spacetime, the action describing a particle of mass m is still that given by Eq. (10.26), but now with

$$ds = (g_{\mu\nu} dx^\mu dx^\nu)^{1/2}. \tag{10.28}$$

We see from this expression that the metric tensor modifies the line element. Taking the variation of S, the condition $\delta S = 0$ yields the equation of motion

$$\frac{du^\rho}{ds} + \overset{\circ}{\Gamma}{}^\rho{}_{\mu\nu}\, u^\mu\, u^\nu = 0, \tag{10.29}$$

where $u^\rho = dx^\rho/ds$ is the particle four-velocity. The solution of this equation of motion, called *geodesic equation*, gives the trajectory of the particle in the presence of gravitation.

An important property of the geodesic equation is that it does not involve the mass of the particle, a natural consequence of universality. Another important property is that it represents the vanishing of the covariant derivative of the four-velocity u^ρ along the trajectory of the particle:

$$u^\lambda \overset{\circ}{D}_\lambda u^\rho \equiv \frac{\overset{\circ}{D} u^\rho}{ds} = 0. \tag{10.30}$$

This is a consequence of the General Relativity approach to gravitation, in which the gravitational interaction is geometrized, and in which the concept of force is absent. According to this approach, gravitation produces a curvature in spacetime, and the gravitational interaction is achieved by letting (spinless) particles to follow the geodesics of this spacetime.

10.8 Bianchi Identities

A detailed calculation gives the simplest way to exhibit curvature. Consider a vector field U, with components U^α, and take twice the covariant derivative, getting $\overset{\circ}{D}_\gamma \overset{\circ}{D}_\beta U^\alpha$. Reverse then the order to obtain $\overset{\circ}{D}_\beta \overset{\circ}{D}_\gamma U^\alpha$ and compare. The result is

$$\overset{\circ}{D}_\gamma \overset{\circ}{D}_\beta U^\alpha - \overset{\circ}{D}_\beta \overset{\circ}{D}_\gamma U^\alpha = -\overset{\circ}{R}{}^\alpha{}_{\epsilon\beta\gamma} U^\epsilon. \tag{10.31}$$

Curvature turns up in the commutator of two covariant derivatives:

$$[\overset{\circ}{D}_\gamma, \overset{\circ}{D}_\beta]U^\alpha = -\overset{\circ}{R}{}^\alpha{}_{\epsilon\beta\gamma}U^\epsilon. \tag{10.32}$$

A detailed calculation leads also to some identities. One of them is

$$\overset{\circ}{R}{}^\kappa{}_{\lambda\rho\sigma} + \overset{\circ}{R}{}^\kappa{}_{\sigma\lambda\rho} + \overset{\circ}{R}{}^\kappa{}_{\rho\sigma\lambda} = 0\ . \tag{10.33}$$

Another one is

$$\overset{\circ}{D}_\mu \overset{\circ}{R}_{\kappa\lambda\rho\sigma} + \overset{\circ}{D}_\sigma \overset{\circ}{R}_{\kappa\lambda\mu\rho} + \overset{\circ}{D}_\rho \overset{\circ}{R}_{\kappa\lambda\sigma\mu} = 0 \tag{10.34}$$

Notice, in both cases, the cyclic rotation of three of the indices. These expressions are called respectively the first and the second Bianchi identities.

Now, as the metric has zero covariant derivative, it can be inserted in the second identity to contract indices in a convenient way. Contracting with $g^{\kappa\rho}$, it comes out

$$\overset{\circ}{D}_\mu \overset{\circ}{R}_{\lambda\sigma} - \overset{\circ}{D}_\sigma \overset{\circ}{R}_{\lambda\mu} + \overset{\circ}{D}_\rho \overset{\circ}{R}{}^\rho{}_{\lambda\sigma\mu} = 0.$$

A further contraction with $g^{\lambda\sigma}$ yields

$$\overset{\circ}{D}_\mu \overset{\circ}{R} - \overset{\circ}{D}_\sigma \overset{\circ}{R}{}^\sigma{}_\mu - \overset{\circ}{D}_\rho \overset{\circ}{R}{}^\rho{}_\mu = 0,$$

which is the same as

$$\overset{\circ}{D}_\mu \overset{\circ}{R} - 2\,\overset{\circ}{D}_\sigma \overset{\circ}{R}{}^\sigma{}_\mu = 0,$$

or

$$\overset{\circ}{D}_\mu \left[\overset{\circ}{R}{}^\mu{}_\nu - \tfrac{1}{2}\ \delta^\mu_\nu\ \overset{\circ}{R}\right] = 0. \tag{10.35}$$

This expression is the "contracted Bianchi identity". The tensor thus "covariantly conserved" will have an important role. Its totally covariant form,

$$G_{\mu\nu} = \overset{\circ}{R}_{\mu\nu} - \tfrac{1}{2}\ g_{\mu\nu}\ \overset{\circ}{R}, \tag{10.36}$$

is called the Einstein tensor. Its contraction with the metric gives the scalar curvature (up to a sign).

$$g^{\mu\nu}G_{\mu\nu} = -\overset{\circ}{R}. \tag{10.37}$$

When the Ricci tensor is related to the metric tensor by

$$\overset{\circ}{R}_{\mu\nu} = \lambda\ g_{\mu\nu}\ , \tag{10.38}$$

where λ is a constant, it is usual to say that we have an *Einstein space*. In that case, $\overset{\circ}{R} = 4\lambda$ and $G_{\mu\nu} = -\lambda g_{\mu\nu}$. Spaces in which $\overset{\circ}{R}$ is a constant are said to be *spaces of constant curvature*. This is the standard language. We insist that there is no such a thing as the curvature of space. Curvature is a characteristic of a connection, and many connections are defined on a given space.

10.9 Einstein's Field Equations

The Einstein tensor (10.36) is a purely geometrical second-order tensor which has vanishing covariant derivative. It is actually possible to prove that it is the only one. The energy-momentum tensor is a physical object with the same property. The next stroke of genius comes here. Einstein was convinced that some physical characteristic of the sources of a gravitational field should engender the deformation in spacetime, that is, in its geometry. He looked for a dynamical equation which gave, in the non-relativistic, classical limit, the newtonian theory. This means that he had to generalize the Poisson equation

$$\Delta V = 4\pi G\rho \tag{10.39}$$

within riemannian geometry. The $G_{\mu\nu}$ has second derivatives of the metric, and the energy-momentum tensor contains, as one of its components, the energy density.

He took then the bold step of equating them to each other, obtaining what we know nowadays to be the simplest possible generalization of the Poisson equation in a riemannian context:

$$R_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu}R = \tfrac{8\pi G}{c^4}\, T_{\mu\nu}. \tag{10.40}$$

This is the *Einstein equation*, which fixes the dynamics of a gravitational field. The constant in the right-hand side was at first unknown, but he fixed it when he obtained, in the due limit, the Poisson equation of the newtonian theory. The tensor in the right-hand side is the *symmetric* (or Belinfante-Rosenfeld) energy-momentum tensor (4.65) of a matter field, which is the source of the gravitational field.

Contracting (10.40) with $g^{\mu\nu}$, we find

$$R = -\,\tfrac{8\pi G}{c^4}\, T\ , \tag{10.41}$$

where $T = g^{\mu\nu}T_{\mu\nu}$. This result can be inserted back into the Einstein equation, to give it the form

$$R_{\mu\nu} = \tfrac{8\pi G}{c^4}\ \left[T_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu}T\right]. \tag{10.42}$$

Consider the sourceless case, in which $T_{\mu\nu} = 0$. It follows from the above equation that $R_{\mu\nu} = 0$ and, therefore, that $R = 0$. Notice that this does *not* imply $R^{\rho}{}_{\sigma\mu\nu} = 0$. The Riemann tensor can be nonvanishing even in the absence of source. Einstein's equations are non-linear and, in consequence, the gravitational field can engender itself. Absence of gravitation is signalled by $R^{\rho}{}_{\sigma\mu\nu} = 0$, which means a flat spacetime. This case — Minkowski spacetime — is a particular solution of the

sourceless equations. Beautiful examples of solutions without any source at all are the de Sitter spaces.

In reality, the Einstein tensor (10.36) is not the most general parallel-transported purely geometrical second-order tensor which has vanishing covariant derivative. The metric has the same property. Consequently, it is in principle possible to add a term $\Lambda g_{\mu\nu}$ to $G_{\mu\nu}$, with Λ a constant. Equation (10.40) becomes

$$R_{\mu\nu} - (\tfrac{1}{2}\ R + \Lambda) g_{\mu\nu} = \tfrac{8\pi G}{c^4}\ T_{\mu\nu}. \tag{10.43}$$

From the point of view of covariantly preserved objects, this equation is as valid as (10.40). In his first trial to apply his theory to cosmology, Einstein looked for a static solution. He found it, but it was unstable. He then added the term $\Lambda g_{\mu\nu}$ to make it stable, and gave to Λ the name *cosmological constant*. Later, when evidence for an expanding universe became clear, he called this "the biggest blunder in his life", and dropped the term. This is the same as putting $\Lambda = 0$. It was not a blunder: recent cosmological evidence claims for $\Lambda \neq 0$. Equation (10.43) is the *Einstein's equation with a cosmological term*. With this extra term, Eq. (10.42) becomes

$$R_{\mu\nu} = \frac{8\pi G}{c^4}\ \left[T_{\mu\nu} - \tfrac{1}{2}\ g_{\mu\nu} T\right] - \Lambda g_{\mu\nu}. \tag{10.44}$$

Finally, it is important to mention that Einstein's equations can be derived from an action functional, the so called Hilbert-Einstein action,

$$S[g] = \int \sqrt{-g}\ R\ d^4x, \tag{10.45}$$

where $g = \det(g_{\mu\nu})$.

10.10 The Schwarzschild Solution

Suppose we look for a solution of the Einstein equations which has spherical symmetry in the space section. This would correspond to central potentials in Classical Mechanics. It is better, in that case, to use spherical coordinates $(x^0, x^1, x^2, x^3) = (ct, r, \theta, \phi)$. This is one of the most studied of all solutions, and there is a standard notation for it. The interval is written in the form

$$ds^2 = \left(1 - \frac{2GM}{c^2\ r}\right) c^2 dt^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2) - \frac{dr^2}{1 - \frac{2GM}{c^2\ r}}. \tag{10.46}$$

Introducing the *Schwarzschild radius*

$$R_S = \frac{2GM}{c^2}, \tag{10.47}$$

the interval acquires the form

$$ds^2 = \left(1 - \frac{R_S}{r}\right) c^2 dt^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2) - \frac{dr^2}{1 - \frac{R_S}{r}}. \tag{10.48}$$

This is the solution found by K. Schwarzschild in 1916, soon after Einstein had presented his final version of General Relativity. It describes the field caused, outside it, by a symmetrically spherical source. We see that there is a singularity in the metric components at the value $r = R_S$. Its value for a body with the mass of the Sun would be $R_S \approx 3$ km. For a body with Earth's mass, $R_S \approx 0.9$ cm. For such objects, of course, there exists to real Schwarzschild radius. It would be well inside their matter distribution, where $T_{\mu\nu} \neq 0$ and the solution is not valid.

Index

aberration, 23
action, 90
    for a particle, 29
    for the vibrating line, 69
action-at-a-distance, 37
adjoint representation, 48, 197
algebra
    defined, 44
angular–momentum
    density tensor, 106

Belinfante
    energy-momentum, 108, 210
Bianchi identity, 194
Biot-Savart
    law, 155
blue shift, 161
boost, 20
Bose field
    introduced, 117
boson, 117

canonical basis
    for the linear algebra, 44
causality, 17
charge conjugation
    for Dirac fields, 179
charge independence, 81
chiral
    representation, 171
chromodynamics, 187
Classical Mechanics
    formulations, 67
Clifford algebra, 170
complex scalar field, 120
confinement, 187
contraction, 21
contravariant
    tensor, 27
coordinates
    normal, 70
Coulomb
    gauge, 143, 155
    law, 153
covariance, 7
covariant
    derivative, 83
    tensor, 27
    uses of the word, 27
covariant derivative, 190
covariantization, 74
CPT transformation
    for Dirac fields, 181

d'Alembert equation, 142
d'Alembertian, 93, 118
delta
    Kronecker, 70
Dirac
    gamma matrices, 170
    representation, 165
Dirac equation, 162
    covariant form, 170
    Hamiltonian form, 164

distance, 14
Doppler effect, 160

effect
  Doppler, 160
Einstein
  convention, 6
electrodynamics, 187
electromagnetic
  wave, 156
electromagnetic field, 132
  constant, 148
  uniform, 149
electromagnetism
  of a charged particle, 34
electrostatics, 152
electroweak interactions, 187
energy
  of a particle, 31
energy–momentum
  and Einstein equation, 108
  density tensor
    Belinfante, 108
    canonical, 105
energy-momentum
  Belinfante, 210
  of a gauge field, 197
equation
  d'Alembert, 142
  Dirac, 162
  Euler–Lagrange, 92
  field, 75
  Klein-Gordon, 93, 117, 171
  Korteweg-de Vries, 115
  Laplace, 152
  of motion, 75
  Pauli, 167
  Poisson, 152
  Proca, 127, 142
  Schrödinger, 76
  sine-Gordon, 113
  Yang-Mills, 195
equivalence principle, 202
event, 15

Fermat
  principle, 29
fermion, 117
field
  classical, 79
  criterion for the presence, 119
  equation, 75
  gauge, 184
  gravitational, 201
  introduced, 75
  mediating, 126
  non-material, 75
  quantum, 79
  relativistic
    scalar, 58
    tensor, 58
    vector, 58
  scalar, 107, 190
    complex, 120
    real, 117
  vector, 107, 125
    complex, 129
    real, 126
field strength, 193
  electromagnetic, 34
force law
  Lorentz, 148
  Newton, 3
four-momentum
  of a particle, 32
four-vector, 24

Fourier analysis, 71
frames
  accelerated, 1
  inertial, 1, 8
  non-inertial, 1, 36
functional calculus, 93
fundamental interactions, 187

Galilei group, 6
gamma factor, 22
gammalogy, 171
gauge, 128
  Coulomb, or radiation, 143, 155
  field, 184
  fixing, 142
  invariance, 142
  Lorenz, 128, 142
  potential, 124
  theory, 110
  transformation, 82, 141
    global, 121, 188
    local, 122, 189
generators
  of the Lorentz group, 64
gl(N, R), 46
gravitational field, 201
group
  Galilei, 6
  Lorentz, 55
  matrix, 45
  of internal transformations, 80
  of transformations, 38
  Poincaré, 20, 55
  rank, 51
  rotation, 48

Hamilton
  principle, 29, 90
Hamilton equations, 69
Hamiltonian
  for a particle, 32
  for the vibrating line, 69
  formulation of Mechanics, 67
harmonic, 152
Heisenberg
  and isospin, 81

identity
  Bianchi, 194
  Jacobi, 194
inertial frames, 1, 8
interval
  notion of, 15
  null, 17
  spacelike, 17
  timelike, 17
invariant
  Casimir, 50
involution, 180
ISO(r,s), 47
isospin, 81

Jacobi identity, 194

KdV equation, 115
Killing-Cartan metric, 51, 195
Klein-Gordon equation, 93, 117
  every field obeys, 118, 163
Korteweg-de Vries equation, 115
Kronecker
  symbol, 33, 48, 62
Kronecker delta, 70

Lagrange
  derivative
    and Noether, 111
    in Classical Mechanics, 74

in Field Theory, 93
Lagrangian
for a particle, 31
for the vibrating line, 69
formulation of Mechanics, 67
free and interaction, 119
Wentzel–Pauli, 127
Lagrangian approach
and Euler–Lagrange equations, 91
and Hamilton principle, 90
and symmetry, 89
general aspects, 85
Lagrangian properties
invariance, 89
locality, 90
low order, 89
reality, 89
simplicity, 89
Laplace equation, 152
law
Biot-Savart, 155
Coulomb, 153
Newton, 3
Lie
algebra
of a group, 41
equation, 110
light cone, 16
future and past, 17
linear
representation, 58
Liouville equation, 69
Lorentz
contraction, 21
force law, 33, 148
group, 19
generators, 64
metric, 15
tensor, 27
transformation, 26, 55, 65, 75
pure, or boost, 20
vector, 24, 27
Lorentz group
representation
bispinor, 174
nonunitary if finite, 174
vector, 174
Lorenz
gauge, 142
Lorenz gauge, 128

magnetostatics, 154
Majorana
representation, 171
Maxwell's tensor, 34
mediating field, 126
metric, 13
Killing-Cartan, 51, 195
Lorentz, 15
notion of, 13
minimal coupling
prescription, 148
Minkowski space, 15, 55
momentum
of a particle, 31

Noether
current, 102, 189
first theorem, 99
second theorem, 110
norm, 13
normal modes
and Fourier analysis, 71
normal coordinates, 70
null

interval, 17
vector, 24

orthogonality, 13

parity transformation, 38
for Dirac fields, 177
particle dynamics, 29
Pauli equation, 167
Pauli matrices, 52
Pauli—Dirac
representation, 171
Pauli–Lubanski operator, 62
Poincaré
group, 20, 26, 55, 63
transformation, 26, 55
Poisson
brackets, 69
equation, 152
Poynting vector, 146
principle
equivalence, 202
Fermat, 29
Hamilton, 29, 90
of causality, 36
of determinism, 8
of inertia, 8, 36
of relativity, 8, 36
Proca equation, 127, 142
proper
length, 21
time, 18

quantization
vibrating line, 77
quantization rules
for fields, 79

radiation
gauge, 143
rank, 51, 61
red shift, 161
relativity
galilean, 36
general, 37
special, 36
representation
adjoint, 48, 51, 197
fundamental, 48
linear, 45, 58
Lorentz vector, 125
spinor, 53
tensor, 54
vector, 50
representation of gammas
chiral, 171
Majorana, 171
Pauli—Dirac, 171
rotations, 48

scalar field
complex, 120
real, 117
scalar product, 13
Schrödinger, 117
Schrödinger equation, 76
sine-Gordon equation, 113
SO(3), 46
and SU(2), 48
soliton, 114
space
Fock, 77
Minkowski, 55
spacelike
interval, 17
vector, 24
spacetime

classical, 7
Minkowski, 15
special relativistic, 12
spin
density tensor, 107
of a particle, 61
operator, 63
spinor
Pauli,Weyl, 53
state
in Classical Mechanics, 67
in Quantum Mechanics, 76
SU(2)
and SO(3), 48
symmetry
and conserved charge, 103
gauge, 108, 109
internal, 108
Lorentz, 106
on spacetime, 104
translation, 104

tensor, 6
Lorentz, 27
theorem
Noether
first, 99
second, 110
time dilation, 19
time reversa
for Dirac fields, 181
time reversal, 39
Wigner, 182
timelike
nterval, 17
vector, 24
topological
conservation law, 113, 195
number, 115, 195
transformation
groups of, 38
transversality condition, 143
twin paradox, 19

U(N), 46

variation, 94, 95
action, 99
of action, 100
vector
contravariant, 135
covariant, or covector, 135
fields, 28
Lorentz, 24, 27
null, 24
representation, 125
spacelike, 24
timelike, 24
under rotations, 6
vector field, 125
vibrating line, 74
classical continuum, 72
quantum, 77

wave
electromagnetic, 156
equation, 156
monochromatic plane, 159
phase, 160
plane, 156
propagation, 157
transversal, 158
tvector, 160
wavefields, 79
wavelength, 160
Weinberg–Salam theory, 187
Wentzel–Pauli Lagrangian, 127

Wigner time reversal, 182
world line, 17

Yang-Mills equation, 195