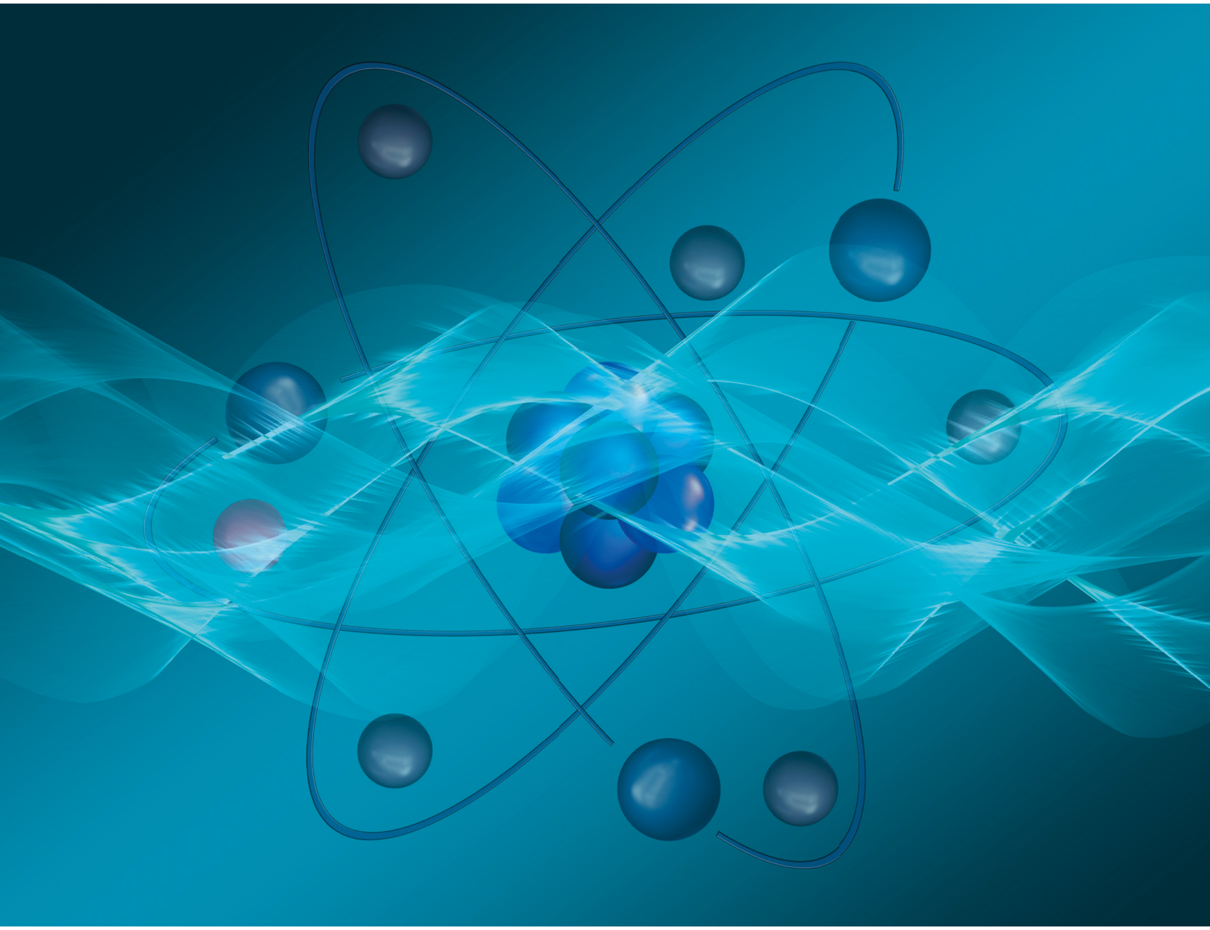
JIŘÍ HOŘEJŠÍ

Lectures on QUANTUM FIELD THEORY



MatfyzPress
Publishing House

Lectures on Quantum Field Theory

Jiří Hořejší

Published by Karolinum Press and MatfyzPress Publishing House

KAROLINUM PRESS
Karolinum Press is a publishing department of Charles University
Ovocný trh 560/5, 116 36 Prague 1, Czech Republic
www.karolinum.cz

MATFYZPRESS PUBLISHING HOUSE
MatfyzPress Publishing House, Faculty of Mathematics and Physics of Charles University
Sokolovská 83, 186 75 Prague 8, Czech Republic
www.matfyzpress.cz

First edition

A catalogue record for this book is available from the National Library of the Czech Republic.

ISBN 978-80-246-5791-2 (Karolinum Press)
ISBN 978-80-7378-500-0 (MatfyzPress Publishing House)
ISBN 978-80-246-5792-9 (Karolinum Press, pdf)
ISBN 978-80-7378-508-6 (MatfyzPress Publishing House, pdf)

Charles University
Karolinum Press

www.karolinum.cz
ebooks@karolinum.cz

MatfyzPress
Publishing House

MatfyzPress

www.matfyzpress.cz
info@matfyzpress.cz

Contents

Preface

This work covers the material of a two-semester course of quantum field theory (QFT) that I taught for more than 20 years at the Charles University and Czech Technical University in Prague. For years, I was reluctant to write up such a set of lecture notes, since the current literature in this area is quite rich and there are dozens of books on the subject. However, eventually I was forced to do it, because of the pandemy of the infamous coronavirus that has broken out in spring 2020. I comment on this in more detail below. Conceptually, my approach is traditional, starting with several introductory chapters on the relativistic quantum mechanics. Then, after a brief interlude on the classical field theory, one proceeds to the quantization of free fields and to some elementary examples of field interactions, the basic tool being the Dyson perturbation expansion of the S-matrix in the interaction representation. The pragmatic aim of the first half of the text (chapters 1–25) is to arrive at the basic techniques for calculations of Feynman diagrams in the lowest perturbative order, as well as for the computation of the particle decay rates and scattering cross sections. This is just the matter that should be ideally explained during the first (winter) semester, since a part of the curriculum in the second (summer) semester, at least for some students, is a course on the standard model of particle physics, where a Feynman diagram calculation is an everyday occurrence. The second half (chapters 26–50) represents topics to be explained during the second semester and the main theme here is quantum electrodynamics at the level of one-loop diagrams, including techniques of regularization of ultraviolet divergences and renormalization. In this way, the whole material of the present lecture notes is divided into 50 chapters and each of them corresponds, roughly, to a 90 min. lecture (the total number of QFT lectures in a given academic year is about fifty). I would like to stress that the text is really intended to have the character of lecture notes, which means that, among other things, some explicit calculations are shown here in greater detail than in most of the representative monographs and textbooks, so as to make the life of a QFT beginner easier. Throughout the text one also encounters numerous hints to possible independent calculations, addressed to interested diligent readers; some of the problems in question may also serve as appropriate topics for tutorials. Admittedly, readers that are not quite fond of performing independent calculations may find the repeated offers of problems left to them as “instructive exercises” somewhat disturbing (or even annoying); anyway, there are just about three dozen of such hints in the whole text, i.e. less than one per chapter on average.

As I have indicated above, these lecture notes have been written under rather special circumstances, during the protracted coronavirus (COVID-19) crisis in 2020 and 2021. It was a situation that people of my generation have experienced never before, so let me add some personal recollection (which is, admittedly, somewhat emotional). The outbreak of the pandemy was officially announced in March 2020. Thus, on Wednesday, March 11, the personal attendance of students in the lecture rooms was banned “until further notice” and I decided to write immediately the text of a lecture scheduled for Thursday, to be able to send it to students via e-mail. Such a procedure seemed to me more efficient than a system of videoconferences or so, and I hoped also that the students' opinion would coincide with that of the aspiring student in Goethe's Faust, expressed in a dialogue with Mephistopheles, namely, "You won't need to tell me twice! I think, myself, it's very helpful, too, that one can take back home, and use, what someone's penned in black and white".[1] In any case, it is obvious that a carefully written text is more durable than lectures presented on a blackboard and erased immediately after the classes. Thus I went on in this manner, sticking to the maxim "nulla dies sine linea", till the end of May when the semester terminates. When the summer semester and the students' exams were over, I returned to the material of the envisaged next winter semester and continued writing down the relevant lectures so as to have a complete set (in musical terms, "da capo al fine"). In the meantime, I had to put together a collection of lectures for another course, aimed at a more advanced audience (25 chapters as well). In this way, the whole work has been basically completed in May 2021, with the nasty virus still around. Then there followed a period of transforming the manuscript full of handwritten formulae into a user-friendly electronic file, as well as gradual detailed proofreading of the text, mostly during the academic year 2021/2022. This was largely finished in autumn 2022, when the pandemy was fading away, but was overshadowed by even more tragic events — of course, I have in mind the absurd criminal war that Russia started against Ukraine.

When I started writing the lecture notes, in the gloomy atmosphere of the covid calamity on the rise, it came to my mind that there is a famous work of the world literature that was created under similar circumstances and survived over centuries. Yes, you guessed right; it is the Decameron by Giovanni Boccaccio. Its origin is widely known. It represents a collection of one hundred tales told by a group of ten young people who escaped from Florence, where the epidemic of plague broke out in 1348, and stayed in a hideout in the countryside to avoid the dangerous infection. Concerning my text, I have also written the lecture notes partly in a hideout (the "home office"). These consist of only fifty tales told by myself (not young anymore), concerning topics not so easily accessible to a general public and I certainly do not expect that my opus will become so famous as the Boccaccio's Decameron, or that it could survive through centuries. Nevertheless, I believe that it may have an appropriate (though inevitably limited) lifetime and may be useful for at least some students and other potentially interested scientifically minded readers. My primary aim has been to make it a comprehensible and digestible introduction to the rather difficult subject of quantum field theory, which, among others, forms a basis of the contemporary particle physics.

One last remark is perhaps in order here. In view of the above-mentioned origin of these lecture notes, it is to be expected that most of the potential readers will be university students fluent in Czech. Thus, I could not resist the temptation to include, occasionally, some notes concerning the Czech equivalents of the international English terminology, or even some elements of a common literary folklore. Hopefully, this might add some cheering moments to the serious scholarly style of the whole opus.

Acknowledgements

From what I have written above it might seem that I should thank the malicious coronavirus in the first place, for stimulating me to write up these lecture notes. But I will not, taking into account that, apart from the positive impact mentioned above, this dangerous invisible bug did also so much harm to so many people all over the world. Needless to say, my acknowledgements are aimed in a completely different, genuinely positive, direction. In particular, I recognize the work of my younger colleagues who conducted and supervised, during the previous years, the

[1] A translation into English by A. S. Kline, 2003. In Czech (in the classic translation by O. Fischer) it reads: "Tot' praktické, i heled'me se! To tělem duší při tom jsem. Neb co je černé na bílém, to vesele se domů nese."

tutorials related to my lectures. They are, in alphabetical order: Karol Kampf, Karel Kolář, Jiří Novotný and Martin Zdráhal. Further, I appreciate questions and comments that the students made throughout the years; this certainly led to many improvements of the style and contents of the lectures. Actually, I have also received some useful remarks from other colleagues; for instance, Walter Grimus from Vienna University has drawn my attention to the fact that the frequently cited "Lorentz condition" in electromagnetism is in fact "Lorenz condition". Finally, my great thanks are due to Tomáš Husek and Tomáš Kadavý, who recast my manuscript in LaTeX and thus made it ready for publication; the whole work matured to its present form in spring 2023.

Prague, May 2023 J. Hořejší

Conventions, notations and units

Unless stated otherwise, we use the natural system of units, in which $\hbar = c = 1$ (note that Peskin and Schroeder call it "God-given" units in their book [14]). Obviously, within such a system, the time and length have the same dimension, the energy, momentum and mass have the same dimension, inverse length has the dimension of a mass, etc. The passage from the economical natural system to ordinary units is quite straightforward. To this end, one may use the commonly known approximate values of the Planck constant $\hbar$ and the "conversion constant" $\hbar c$, namely

$$\hbar = 6.58 \times 10^{-22}\,\mathrm{MeV\,s}\,, \qquad \hbar c = 197\,\mathrm{MeV\,fm}\,,$$

where $1\,\mathrm{fm} = 10^{-13}\,\mathrm{cm}$ (fm stands for "fermi" or "femtometer"). Numerical values of observable quantities (such as decay rates or scattering cross sections) are then converted into ordinary units by setting

$$1\,\mathrm{MeV}^{-1} = 6.58 \times 10^{-22}\,\mathrm{s}\,,$$

or

$$1\,\mathrm{MeV}^{-1} = 197\,\mathrm{fm}\,.$$

While the natural system of units is universally accepted in the literature concerning quantum field theory and particle physics, there are three other conventions that may differ in various books, so one must emphasize what is our particular choice (to avoid any misunderstanding when comparing our results with other books or papers). First, the metric of the flat spacetime used throughout the present text is defined by

$$g_{\mu\nu} = g^{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)\,.$$

In other words, the metric we are employing here has the signature $(+\,-\,-\,-)$. Let us remark that such a choice seems to be prevalent in current literature; for instance, among the books that we cite in the list of relevant literature, only [13] and [18] use the metric with the inverse signature $(-\,+\,+\,+)$. Anyway, one should keep in mind that there is no question of which metric is "right" or "wrong"; its choice is just a matter of convention. Note also that readers specialized mostly in relativity and gravitation should not worry about our notation $g_{\mu\nu}$ for the metric tensor, which they got used to employ for the general case of curved Riemann space (and distinguish the case of the flat spacetime by using the symbol $\eta_{\mu\nu}$ or so). The notation used here is a common practice in the literature concerning relativistic quantum theory and particle physics, since in this area one is dealing just with flat spacetime (nevertheless, $\eta_{\mu\nu}$ is employed conventionally e.g. in the books [11, 13] or [29]).

Second, another important convention is that for the fifth Dirac gamma matrix γ_5. Here we use the definition

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3\,.$$

Again, this choice seems to be prevalent in the literature (note that within our list, the books [7] and [13] define γ_5 with opposite sign).

Finally, our convention for the fully antisymmetric Levi-Civita tensor is such that

$$\varepsilon_{0123} = +1\ .$$

In this case, one must admit that this is a minority choice, since the option prevalent in current literature is $\varepsilon^{0123} = +1$ (which corresponds to the sign change in contrast to our convention). So, the reader must be careful when comparing our formulae in Appendix C and elsewhere (see in particular (C.11)) with those presented in other textbooks. Note that the convention employed here agrees with the classic books by Bjorken and Drell [1, 2].

Chapter 1

Klein–Gordon and Dirac equations: brief history

The best known equation of quantum mechanics is undoubtedly the Schrödinger equation, which for a particle moving in an external field reads

$$i\hbar\frac{\partial\psi}{\partial t} = \left(-\frac{\hbar^2}{2m}\Delta + V(\vec{x})\right)\psi\,, \tag{1.1}$$

where Δ is the Laplace operator, $\Delta = \vec{\nabla}^2$, and $V(\vec{x})$ is the potential energy corresponding to an external force. The wave function $\psi = \psi(\vec{x},t)$ has the familiar interpretation: $|\psi(\vec{x},t)|^2$ represents the probability density for the particle localization at the point $\vec{x}$ and time t. Erwin Schrödinger published it in 1926 (and subsequently won the Nobel Prize in 1933). Let us consider first Eq. (1.1) for a free particle, i.e. for $V = 0$. There is a simple "correspondence principle" that may serve as a recipe for recovering the Schrödinger equation. Denoting the energy as E and momentum as $\vec{p}$, one may observe the correspondence

$$\begin{aligned} E &\longleftrightarrow i\hbar\frac{\partial}{\partial t}\,, \\ \vec{p} &\longleftrightarrow -i\hbar\vec{\nabla}\,, \end{aligned} \tag{1.2}$$

which leads from the usual non-relativistic relation between kinetic energy and momentum

$$E = \frac{\vec{p}^2}{2m} \tag{1.3}$$

to the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\vec{\nabla}^2\psi\,. \tag{1.4}$$

Let us stress as emphatically as possible that the correspondence (1.2) does not represent a derivation of the Schrödinger equation. This cannot be derived, it can only be postulated; this is what the founding fathers of quantum theory did. The meaning of the correspondence (1.2) is that it guarantees recovering the right relation between the energy and momentum (1.3) when the operators in (1.2) act on an appropriate wave function ψ, in particular the plane wave

$$\psi(\vec{x},t) \propto e^{-\frac{i}{\hbar}(Et-\vec{p}\cdot\vec{x})}\,. \tag{1.5}$$

In the same year when the non-relativistic equation (1.4) or (1.1) was postulated, a pertinent relativistic version was considered (preferably as a quantum mechanical equation for the electron). In that case, one has to use as a motivating hint the relation between the energy and momentum valid in special relativity, i.e.

$$E^2 = c^2\vec{p}^2 + m^2c^4 . \tag{1.6}$$

Then, using the correspondence (1.2), one gets immediately

$$-\hbar^2 \frac{\partial^2\psi}{\partial t^2} = \left(-\hbar^2 c^2 \vec{\nabla}^2 + m^2c^4\right)\psi , \tag{1.7}$$

and this can be recast in a more elegant form

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta + \frac{m^2c^2}{\hbar^2}\right)\psi = 0 . \tag{1.8}$$

Now, the differential operator in Eq. (1.8) is the familiar d'Alembert operator

$$\Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta , \tag{1.9}$$

and thus we end up with

$$\left(\Box + \frac{m^2c^2}{\hbar^2}\right)\psi(x) = 0 . \tag{1.10}$$

The equation (1.10) had been formulated in 1926 independently by several theorists: Erwin Schrödinger (who subsequently rejected it), Oskar Klein, Walter Gordon, and Vladimir Fock (or, better Fok: the name reads Фок in Russian). So, although it is apparently an equation with many parents, it is universally called the **Klein–Gordon equation**.

A remark is perhaps in order here. The constant appearing in Eq. (1.10) is the square of inverse of the Compton wavelength and one might wonder why it happens to be there, when (1.10) clearly has nothing to do with the famous Compton process (the photon scattering on a charged particle). The answer is guessed easily on dimensional grounds: the d'Alembert operator $\Box$ has, obviously, the dimension of inverse length squared, and any possible additive constant (with the same dimension) must be made of the fundamental constants of a relativistic quantum theory, i.e. c and $\hbar$ and, eventually, the relevant mass m. The combination $\hbar/mc$ is then the only possibility how to form a constant with the dimension of length (it is a refreshing simple exercise to show that such a combination of c, $\hbar$ and m is indeed unique).[2]

For convenience, let us now pass to the natural system of units with $\hbar = 1$, $c = 1$. Using the standard relativistic covariant notation, one then has

$$\left(\Box + m^2\right)\psi(x) = 0 , \tag{1.11}$$

with $\Box = \partial_\mu\partial^\mu$. The simplest solutions of Eq. (1.11) have the form of plane waves; for their description one may use two linearly independent exponentials

$$\begin{aligned} \psi_{(+)}(x) &= \text{const. } e^{-ip\cdot x} , \\ \psi_{(-)}(x) &= \text{const. } e^{ip\cdot x} , \end{aligned} \tag{1.12}$$

where $p \cdot x = p_0x_0 - \vec{p}\cdot\vec{x}$ (the logic of the chosen notation will become clear shortly). Inserting (1.12) into (1.11), one gets the condition

$$p^2 = m^2 , \tag{1.13}$$

[2] In fact, sticking to the traditional terminology, $\hbar/mc$ is the Compton wavelength divided by 2π.

i.e. $p_0^2 = \vec{p}^2 + m^2$. Without loss of generality, one may choose $p_0 > 0$,

$$p_0 = \sqrt{\vec{p}^2 + m^2}\,. \tag{1.14}$$

So, as expected, one recovers the correct relation between the energy and momentum of a particle with the mass m. Using the correspondence (1.2), one sees that the solution $\psi_{(+)}(x)$ describes a state with positive energy $E = p_0$ and momentum $\vec{p}$, while $\psi_{(-)}(x)$ carries negative energy $E = -p_0$ and momentum $-\vec{p}$. In any case, the four-component quantity p satisfying (1.13) is rightly called the four-momentum of the particle with the mass m. Thus we have encountered, for the first time, the problem of a wave function for the free particle with negative energy; we will see that this is a generic feature of the equations of relativistic quantum mechanics.

In fact, there is another difficulty inherent in the Klein–Gordon equation. If one wants to implement the probabilistic interpretation of the wave function ψ, one should derive first a pertinent continuity equation connecting the probability density (for particle localization) and the density of probability current. Let us first remind the reader how one proceeds in the case of non-relativistic Schrödinger equation (1.4) (we are going to use the natural system of units, i.e. set $\hbar = 1$). We have the equations for ψ and ψ^*,

$$i\,\frac{\partial\psi}{\partial t} = -\frac{1}{2m}\vec{\nabla}^2\psi\,, \tag{1.15}$$

$$-i\,\frac{\partial\psi^*}{\partial t} = -\frac{1}{2m}\vec{\nabla}^2\psi^*\,. \tag{1.16}$$

Multiplying Eq. (1.15) by ψ^* and (1.16) by ψ, and subtracting the two equations, one gets immediately

$$i\,\frac{\partial}{\partial t}\,(\psi\psi^*) = -\frac{1}{2m}\left(\psi^*\,\vec{\nabla}^2\psi - \psi\,\vec{\nabla}^2\psi^*\right) = -\frac{1}{2m}\vec{\nabla}\left(\psi^*\,\vec{\nabla}\psi - \psi\,\vec{\nabla}\psi^*\right).$$

Thus one obtains the familiar result

$$\frac{\partial}{\partial t}\,\rho_{\text{Sch.}} + \vec{\nabla}\cdot\vec{j}_{\text{Sch.}} = 0\,, \tag{1.17}$$

with

$$\rho_{\text{Sch.}} = \psi\psi^* = |\psi|^2\,, \qquad \vec{j}_{\text{Sch.}} = \frac{1}{2mi}\left(\psi^*\,\vec{\nabla}\psi - \psi\,\vec{\nabla}\psi^*\right). \tag{1.18}$$

For the Klein–Gordon equation (1.11) one may try to proceed in a similar manner. To begin with, (1.11) is recast as

$$\frac{\partial^2\psi}{\partial t^2} = \vec{\nabla}^2\psi - m^2\psi\,, \tag{1.19}$$

and the same equation holds for ψ^*. Next, using the multiplication and subtraction trick as above, one gets first

$$\frac{\partial}{\partial t}\left(\psi^*\frac{\partial\psi}{\partial t} - \psi\frac{\partial\psi^*}{\partial t}\right) = \vec{\nabla}\left(\psi^*\,\vec{\nabla}\psi - \psi\,\vec{\nabla}\psi^*\right). \tag{1.20}$$

In order to make the left-hand side of Eq. (1.20) real, one has to include a factor of i; for getting quantities with the same dimension as in the case of the Schrödinger equation, one may write finally

$$\frac{\partial}{\partial t}\rho_{\text{KG}} + \vec{\nabla}\cdot\vec{j}_{\text{KG}} = 0\,,$$

where

$$\begin{aligned}\rho_{\mathrm{KG}} &= \frac{i}{2m}\left(\psi^*\frac{\partial\psi}{\partial t} - \psi\frac{\partial\psi^*}{\partial t}\right),\\ \vec{j}_{\mathrm{KG}} &= \frac{1}{2mi}\left(\psi^*\,\vec{\nabla}\psi - \psi\,\vec{\nabla}\psi^*\right).\end{aligned} \tag{1.21}$$

Obviously, in contrast to (1.18), the would-be probability density ρ_{KG} in (1.21) is not *a priori* positive. In particular, it is easy to see that for $\psi_{(+)}$ shown in (1.12) the expression for ρ_{KG} is positive, while for $\psi_{(-)}$ one gets a negative value of ρ_{KG}. This, of course, is a serious flaw. On the top of that, it has soon become clear that the Klein–Gordon equation is not viable as an equation for the electron, because it cannot incorporate a description of intrinsic angular momentum, the spin (note that the concept of electron spin appeared on the physical stage in 1925, when it was introduced by George Uhlenbeck and Samuel Goudsmit — surprisingly, they have never received the Nobel Prize for it).

Despite the above-mentioned difficulty with the interpretation of the probability density, Klein–Gordon equation, as an equation of relativistic quantum mechanics, does have some limited applicability for the description of spinless particles (for more details, see e.g. the book [1]). However, we will exploit this equation fully later on, within the framework of field theory.

Anyway, it is clear that a topical question that certainly resonated in minds of quantum theorists in the second half of 1920s was: So, what is the right relativistic quantum equation for electron? The problem was resolved in 1928 by Paul Dirac. His solution was, at that time, quite astonishing and this historical breakthrough is thus worth recapitulating here (for the original paper, see [32]).

As we have already stressed, a major flaw of the Klein–Gordon equation is the non-positivity of the would-be probability density in (1.21). It is clear what is the source of this inherent feature of the ρ_{KG}: the equation (1.19) is of the second order in time and thus a time derivative emerges necessarily in (1.21). Thus, it is desirable to have an equation that would be of the first order with respect to time. To ensure the relativistic covariance, it should also be of the first order in space variables (time and space coordinates are treated on an equal footing in Lorentz transformations). In any case, one has to maintain the relativistic relation between energy and momentum (1.6) (for a moment, we come back to ordinary units). For his purpose, Dirac took the square root of (1.6) by linearizing it as follows,

$$E = c\alpha^j p^j + \beta mc^2\,, \tag{1.22}$$

where α^j, $j = 1, 2, 3$, and β are some constant coefficients (summation over the index j is understood here, so $\alpha^j p^j$ can also be written as $\vec{\alpha}\cdot\vec{p}$). Now, employing the correspondence (1.2) one arrives at the equation

$$i\hbar\frac{\partial\psi}{\partial t} = \left(-i\hbar c\alpha^j\nabla^j + \beta mc^2\right)\psi\,. \tag{1.23}$$

The consistency condition for such an equation is that upon squaring it, one should recover the Klein–Gordon equation (which corresponds trivially to the energy–momentum relation (1.6)). Before squaring Eq. (1.23) one must clarify a simple point: If one has an equation $A\psi = B\psi$ with A, B being some operators, it does not imply automatically that $A^2\psi = B^2\psi$. Indeed, if A and B do not commute, the latter identity is not guaranteed. However, if $AB = BA$, then obviously $A\psi = B\psi \Rightarrow A^2\psi = AB\psi = BA\psi = B^2\psi$. Eq. (1.23) clearly corresponds to the case $[A, B] = 0$, since the time derivative commutes with ∇^j on the right-hand side. One may now square Eq. (1.23) with confidence; the only caveat is that one must not assume *a priori* that the coefficients α^j, β commute (they cannot be ordinary numbers). Thus, one gets

$$-\hbar^2\frac{\partial^2\psi}{\partial t^2} = \Big[-\hbar^2c^2\big(\vec{\alpha}\cdot\vec{\nabla}\big)\big(\vec{\alpha}\cdot\vec{\nabla}\big) - i\hbar c\cdot mc^2\left(\vec{\alpha}\beta+\beta\vec{\alpha}\right)\cdot\vec{\nabla} + \beta^2m^2c^4\Big]\psi\,. \tag{1.24}$$

In Eq. (1.24) one has

$$\big(\vec{\alpha}\cdot\vec{\nabla}\big)\big(\vec{\alpha}\cdot\vec{\nabla}\big) = \alpha^j\alpha^k\nabla^j\nabla^k = \frac{1}{2}\big\{\alpha^j,\alpha^k\big\}\nabla^j\nabla^k + \frac{1}{2}\big[\alpha^j,\alpha^k\big]\nabla^j\nabla^k\,, \tag{1.25}$$

but the last term in (1.25) vanishes, since $\nabla^j\nabla^k = \nabla^k\nabla^j$. In order to turn Eq. (1.24) into the form of the Klein–Gordon equation, the coefficients α^j must obviously satisfy the identities

$$\begin{aligned}\big\{\alpha^j,\alpha^k\big\} &= 2\delta^{jk}\,,\\ \big\{\beta,\alpha^j\big\} &= 0\,,\\ \beta^2 &= 1\,.\end{aligned} \tag{1.26}$$

So, it is clear that α^j and β must be matrices rather than ordinary numbers, as we have rightly anticipated before. Moreover, the equation (1.23) has a "Schrödinger-like" form; the operator on its right-hand side could be interpreted as a Hamiltonian that should be Hermitian (self-adjoint). It means that one should impose an additional constraint on α^j and β, namely

$$\big(\alpha^j\big)^\dagger = \alpha^j\,, \qquad \beta^\dagger = \beta\,. \tag{1.27}$$

Now the question is, what can be matrices satisfying (1.26) and (1.27). First of all, it is not difficult to show that the dimension of such matrices must be even. Indeed, (1.26) means, in particular, that

$$\alpha^j\alpha^k = -\alpha^k\alpha^j \qquad \text{for} \quad j\neq k\,, \tag{1.28}$$

and

$$\big(\alpha^j\big)^2 = 1 \qquad \text{for} \quad j=1,2,3\,. \tag{1.29}$$

Let us now consider the determinants of the matrix products in (1.28). One has

$$\det\alpha^j\,\det\alpha^k = \det(-\mathbb{1})\,\det\alpha^k\,\det\alpha^j = (-1)^d\,\det\alpha^j\,\det\alpha^k\,, \tag{1.30}$$

where d is the dimension of matrices in question. Obviously, $\det\alpha^j \neq 0$ because of (1.29), so (1.30) implies $(-1)^d = 1$, i.e. d is even.

The simplest choice would be $d = 2$, but it does not work; the point is that there are not four mutually anticommuting 2×2 matrices. Indeed, for α^j, $j = 1,2,3$, one could take the Pauli matrices σ^j, but then there is no non-trivial β that would anticommute with them. Proving this statement independently is left to the reader as an instructive algebraic exercise.

The next try is $d = 4$ and we will see shortly that it does work. Needless to say, it then also means that the wave function ψ in Eq. (1.23) has four components. Before showing an explicit example of 4×4 matrices satisfying (1.26) and (1.27), let us mention another general property of the matrices in question. It is easy to show that matrices α^j and β are traceless,

$$\begin{aligned}\operatorname{Tr}\alpha^j &= 0\,, \qquad j=1,2,3\,,\\ \operatorname{Tr}\beta &= 0\,.\end{aligned} \tag{1.31}$$

Let us prove e.g. the first identity (1.31). Since $\beta^2 = 1$, one may write

$$\operatorname{Tr}\alpha^j = \operatorname{Tr}\big(\beta^2\alpha^j\big) = \operatorname{Tr}\big(\beta\alpha^j\beta\big) = -\operatorname{Tr}\big(\alpha^j\beta^2\big) = -\operatorname{Tr}\alpha^j\,,$$

so that indeed $\mathrm{Tr}\,\alpha^j = 0$. Note that we have utilized just the trace cyclicity and the anticommutation property $\beta\alpha^j = -\alpha^j\beta$. The second identity (1.31) can be proved in the similar way, employing the same trick with e.g. $\left(\alpha^1\right)^2 = 1$.

Finally, let us display an explicit example of the 4×4 matrices satisfying (1.26) and (1.27). They are

$$\alpha^j = \begin{pmatrix} 0 & \sigma^j \\ \sigma^j & 0 \end{pmatrix}, \qquad j = 1, 2, 3, \qquad \beta = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}, \tag{1.32}$$

where σ^j are the familiar Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.33}$$

and $\mathbb{1}$ stands for the 2×2 unit matrix. It is straightforward to verify that the matrices (1.32) have indeed the required properties. The representation (1.32) is used frequently in practical calculations and is called the **standard representation**.

In the next chapter we will see that the magic Dirac's trick of taking the square root of the energy–momentum relation (1.6) in terms of 4×4 matrix coefficients leads indeed to a successful description of the electron. The great leap from the simple kinematical relation (1.6) to the deep quantum equation with rich physical contents makes the Dirac equation one of the most remarkable achievements of the 20th century physics. Note that Dirac received the Nobel Prize in 1933 together with E. Schrödinger. Many historical details concerning the Dirac's discovery can be found in the book [9].

Chapter 2

Physical contents of Dirac equation: preliminary discussion

As we have noted in the preceding chapter, the prime motivation for finding an alternative to the Klein–Gordon equation was the requirement that the probability defined in terms of a quantum mechanical wave function should be positive. So, let us now examine this problem for the Dirac equation; for convenience, we return to the natural units. Eq. (1.23) then reads

$$i\,\frac{\partial\psi}{\partial t} = -i\vec{\alpha}\cdot\vec{\nabla}\psi + \beta m\psi \tag{2.1}$$

(we will use the standard representation (1.32) in what follows). Let us recall that ψ is a four-component wave function that is conventionally written as a column

$$\psi(x) = \begin{pmatrix}\psi_1(x)\\ \psi_2(x)\\ \psi_3(x)\\ \psi_4(x)\end{pmatrix}. \tag{2.2}$$

Upon Hermitian conjugation of Eq. (2.1) one has

$$-i\,\frac{\partial\psi^\dagger}{\partial t} = i\,\vec{\nabla}\psi^\dagger\vec{\alpha} + m\psi^\dagger\beta\,, \tag{2.3}$$

where $\psi^\dagger = (\psi_1^*, \psi_2^*, \psi_3^*, \psi_4^*)$, and we have utilized the hermiticity property (1.27) of $\vec{\alpha}$ and β. Multiplying Eq. (2.1) by $\psi^\dagger$ from the left and (2.3) by ψ from the right, and taking then the difference of the two equations, one gets immediately

$$\frac{\partial}{\partial t}(\psi^\dagger\psi) + \vec{\nabla}(\psi^\dagger\vec{\alpha}\psi) = 0\,, \tag{2.4}$$

which is the anticipated continuity equation. Thus we may identify the probability density and the probability current as

$$\rho_{\text{Dirac}} = \psi^\dagger\psi\,, \qquad \vec{j}_{\text{Dirac}} = \psi^\dagger\vec{\alpha}\psi\,. \tag{2.5}$$

The positivity of the ρ_{Dirac} is obvious, since

$$\psi^\dagger\psi = |\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + |\psi_4|^2\,. \tag{2.6}$$

This is an expected result, due to the fact that the Dirac equation (2.1) is, in a sense, "square root of Klein–Gordon equation"; more precisely, it is an evolution equation of the 1st order in time, having the form

$$i\,\frac{\partial\psi}{\partial t} = H\,\psi\,, \tag{2.7}$$

where H is the Dirac Hamiltonian

$$H = -i\vec{\alpha} \cdot \vec{\nabla} + \beta m \, . \tag{2.8}$$

Thus, the time evolution is generated by an operator of energy, as it should be, in accordance with the general principles of quantum theory.

A next issue is the angular momentum. Let us start with orbital angular momentum, defined in the standard way as $\vec{L} = \vec{x} \times \vec{p}$, where $\vec{p}$ is the (linear) momentum $\vec{p} = -i\vec{\nabla}$. As we know, $\vec{L}$ commutes with the non-relativistic Hamiltonian in the Schrödinger equation (1.4). For the Dirac Hamiltonian (2.8) one gets, employing the canonical commutation relation $[x^j, p^k] = i\delta^{jk}$,

$$[H, \vec{L}] = -i(\vec{\alpha} \times \vec{p}) \, . \tag{2.9}$$

Let us remark that the vector product in (2.9) is defined formally as usual, i.e.

$$(\vec{\alpha} \times \vec{p})^j = \varepsilon^{jkl} \alpha^k p^l \, .$$

So, apparently, there is something missing, since any decent angular momentum should be an integral of motion for the free particle, i.e. the corresponding operator should commute with the Hamiltonian. In other words, the fact that $[H, \vec{L}] \neq 0$ is a hint that we are on the right track towards the electron spin. A good candidate for such an additional ingredient of the full angular momentum is guessed quite easily. Let us consider the 4×4 matrices

$$\vec{S} = \frac{1}{2}\vec{\Sigma} \, , \qquad \vec{\Sigma} = \begin{pmatrix} \vec{\sigma} & 0 \\ 0 & \vec{\sigma} \end{pmatrix} , \tag{2.10}$$

and recall that the Pauli matrices have the commutation relations

$$[\sigma_j, \sigma_k] = 2i\varepsilon_{jkl}\sigma_l \, . \tag{2.11}$$

This means that the matrices $\vec{S}$ defined by (2.10) satisfy

$$[S_j, S_k] = i\varepsilon_{jkl}S_l \, , \tag{2.12}$$

which, of course, is a set of commutation relations for components of an angular momentum. Needless to say, the matrices $\vec{S}$ possess eigenvalues $\pm 1/2$ (because $(\sigma_j)^2 = 1$ for $j = 1, 2, 3$). Now we may evaluate the commutator $[H, \vec{S}]$. Clearly, $\vec{S}$ commutes with the diagonal matrix β (see (1.32)). Concerning the commutator involving $\vec{\alpha}$, one gets first

$$[\alpha^j, \Sigma^k] = \begin{pmatrix} 0 & 2i\varepsilon^{jkl}\sigma^l \\ 2i\varepsilon^{jkl}\sigma^l & 0 \end{pmatrix} ,$$

so that

$$[H, \Sigma^k] = 2i(\vec{\alpha} \times \vec{p})^k \, . \tag{2.13}$$

Summarizing the results of our simple algebraic exercise, we have

$$\begin{aligned} [H, \vec{L}] &= -i(\vec{\alpha} \times \vec{p}) \, , \\ [H, \vec{S}] &= i(\vec{\alpha} \times \vec{p}) \, , \end{aligned} \tag{2.14}$$

and thus

$$[H, \vec{J}] = 0 \, , \tag{2.15}$$

with

$$\vec{J} = \vec{L} + \vec{S} \,. \tag{2.16}$$

Thus, in such a straightforward manner we have recovered the electron spin as a part of the conserved total angular momentum (2.16).

Let us now recall the problem of negative energy solutions of the Klein–Gordon equation, mentioned in the preceding chapter (cf. (1.12)). One may wonder whether the Dirac equation suffers an analogous difficulty. For clarifying this point, we are going to consider the solution of Eq. (2.1) in the form of a plane wave involving the usual factor $\exp\left[-i(Et - \vec{p}\cdot\vec{x})\right]$. To make our discussion as simple as possible, we will restrict ourselves to the case of a particle at rest, i.e. set $\vec{p} = 0$. Eq. (2.1) is then reduced to

$$i\,\frac{\partial\psi}{\partial t} = \beta m \psi \,. \tag{2.17}$$

Taking into account the block diagonal structure of the matrix β ((1.32), it is useful to split the ψ as

$$\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}, \tag{2.18}$$

where φ and χ are two-component column vectors. Eq. (2.17) is then recast as

$$i\,\frac{\partial\varphi}{\partial t} = m\varphi \,, \tag{2.19}$$

$$i\,\frac{\partial\chi}{\partial t} = -m\chi \,. \tag{2.20}$$

Thus, two linearly independent solutions of Eq. (2.19) may be written e.g. as

$$\varphi_{(1)} = e^{-imt}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \varphi_{(2)} = e^{-imt}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.21}$$

and similarly for (2.20),

$$\chi_{(1)} = e^{imt}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \chi_{(2)} = e^{imt}\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.22}$$

In this way, we obtain a set of four independent solutions of Eq. (2.1)

$$\psi_{(1)} = e^{-imt}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \psi_{(2)} = e^{-imt}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \psi_{(3)} = e^{imt}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \psi_{(4)} = e^{imt}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.23}$$

Obviously, $\psi_{(1)}$ and $\psi_{(2)}$ correspond to the positive rest energy $E = m$, while $\psi_{(3)}$ and $\psi_{(4)}$ carry negative energy $E = -m$ (they are also characterized by the two possible spin projections to the third axis, up and down $(\pm 1/2)$). It is interesting to notice that in the considered case, the existence of the negative energy solutions is a consequence of the specific structure of the matrix β. If β were 4×4 unit matrix, we would have only a solution with positive energy. But, alas, β can never be the unit matrix because of the required anticommutation relations (1.26). As we have already noted in the preceding chapter, the appearance of negative energy solutions is a generic feature of the equations of relativistic quantum mechanics. We will discuss the plane-wave solutions of Dirac equation in detail later on.

The last topic that we are going to discuss here is a derivation of the spin magnetic moment of the electron. Soon after the birth of relativistic quantum mechanics this was indeed one of the most remarkable achievements of the Dirac theory, so it certainly deserves a detailed exposition.

To this end, one has to start with the Dirac equation for the electron in an external electromagnetic field. Using the scalar potential ϕ and vector potential $\vec{A}$, one may write the relevant equation as

$$i\,\frac{\partial\psi}{\partial t} = \left[\vec{\alpha}\cdot(-i\vec{\nabla} - e\vec{A}) + e\phi + \beta m\right]\psi\,. \tag{2.24}$$

Note that the form (2.24) represents the so-called **minimal electromagnetic interaction** and satisfies certainly the requirement of gauge invariance (invariance under gauge transformations of the potentials ϕ and $\vec{A}$). In fact, it is not the most general choice, but coincides with the recipe to be employed later on, in quantum electrodynamics. More comments on a possible extension of the gauge invariant electromagnetic interaction within the framework of Dirac equation are deferred to the Chapter 13.

Our ultimate goal is to get the non-relativistic two-component **Pauli equation**, from which one can extract easily the value of the magnetic moment in question. For this purpose, we will separate upper and lower components of the wave function ψ as

$$\psi = \begin{pmatrix}\widetilde{\varphi}\\ \widetilde{\chi}\end{pmatrix}. \tag{2.25}$$

Then, denoting

$$-i\vec{\nabla} - e\vec{A} = \vec{\pi}\,, \tag{2.26}$$

Eq. (2.24) is recast as a pair of coupled two-component equations

$$\begin{aligned} i\,\frac{\partial\widetilde{\varphi}}{\partial t} &= (\vec{\sigma}\cdot\vec{\pi})\widetilde{\chi} + (e\phi + m)\widetilde{\varphi}\,,\\ i\,\frac{\partial\widetilde{\chi}}{\partial t} &= (\vec{\sigma}\cdot\vec{\pi})\widetilde{\varphi} + (e\phi - m)\widetilde{\chi}\,. \end{aligned} \tag{2.27}$$

Throughout our calculation we will have in mind a situation close to the non-relativistic limit; thus, it is convenient to factorize in the wave function a part corresponding to the rest energy. (cf. (2.21)), i.e. introduce the Ansatz

$$\begin{pmatrix}\widetilde{\varphi}\\ \widetilde{\chi}\end{pmatrix} = e^{-imt}\begin{pmatrix}\varphi\\ \chi\end{pmatrix}. \tag{2.28}$$

Inserting (2.28) into Eq. (2.27) one gets, after a simple manipulation,

$$i\,\frac{\partial\varphi}{\partial t} = (\vec{\sigma}\cdot\vec{\pi})\chi + e\phi\varphi\,, \tag{2.29a}$$

$$i\,\frac{\partial\chi}{\partial t} = (\vec{\sigma}\cdot\vec{\pi})\varphi + e\phi\chi - 2m\chi\,. \tag{2.29b}$$

We consider weak fields, in particular $e\phi \ll m$, as well as a small kinetic energy; the latter assumption may be expressed, technically, as

$$\frac{\partial\chi}{\partial t} \ll m\chi\,.$$

Thus, in Eq. (2.29b) we will neglect $\partial\chi/\partial t$ and $e\phi\chi$ in comparison with $2m\chi$. Consequently, the function χ can be approximately written as

$$\chi \doteq \frac{1}{2m}(\vec{\sigma}\cdot\vec{\pi})\varphi\,. \tag{2.30}$$

Using the last expression in Eq. (2.29a), we have

$$i\,\frac{\partial\varphi}{\partial t}=\frac{1}{2m}(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})\varphi+e\phi\varphi\,. \tag{2.31}$$

To work out the right-hand side of Eq. (2.31), one may utilize the familiar identity for Pauli matrices

$$\sigma_j\sigma_k=\delta_{jk}\cdot\mathbb{1}+i\varepsilon_{jkl}\sigma_l\,. \tag{2.32}$$

From (2.32) one then gets

$$(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})=\vec{\pi}^2+i\vec{\sigma}\cdot(\vec{\pi}\times\vec{\pi})\,. \tag{2.33}$$

One must treat the vector product carefully, since $\vec{\pi}$ is a differential operator. So, one has to evaluate it by letting it act on an arbitrary test function f; one obtains, after some manipulations,

$$(\vec{\pi}\times\vec{\pi})^j f=ie(\vec{\nabla}\times\vec{A})^j f\,,$$

so that

$$\vec{\pi}\times\vec{\pi}=ie(\vec{\nabla}\times\vec{A})=ie\vec{B}\,, \tag{2.34}$$

where $\vec{B}$ is the magnetic field (the reader is encouraged to reproduce independently the result (2.34)). In total, we thus have

$$(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})=(-i\vec{\nabla}-e\vec{A})^2-e\vec{\sigma}\cdot\vec{B}\,.$$

The two-component equation (2.31) thus becomes

$$i\,\frac{\partial\varphi}{\partial t}=\left[\frac{1}{2m}(\vec{p}-e\vec{A})^2+e\phi-\frac{e}{2m}\vec{\sigma}\cdot\vec{B}\right]\varphi\,, \tag{2.35}$$

and this is the anticipated Pauli equation. Obviously, the last term in the square brackets represents an interaction of magnetic moment with magnetic field $\vec{B}$. Since the Pauli matrices have eigenvalues ± 1, one may conclude that the value of the magnetic moment in question is $e/(2m)$ (i.e. one **Bohr magneton**). Note that Wolfgang Pauli formulated Eq. (2.35) in 1927 as a phenomenological description of the electron moving in an external field; he then used the empirically known value of the spin magnetic moment. The derivation described above is actually a **prediction** of the relevant value, made on the basis of a more fundamental equation (though restricted to the minimal electromagnetic interaction). This is why the result (2.35) obtained as a non-relativistic approximation of Dirac equation is extolled as a true achievement.

One more remark is in order here. Magnetic moment of a particle is usually characterized also by its **gyromagnetic ratio**, which is the ratio of the magnetic moment to the angular momentum. It is a well-known fact that for the orbital motion, the gyromagnetic ratio is equal to $e/(2m)$ (this holds both in classical and in quantum theory). For the spin magnetic moment we obviously have the gyromagnetic ratio e/m, since the magnitude of the spin projection is $1/2$. Thus, the spin magnetic moment of electron does not obey the "normal" rule and differs from it by a dimensionless factor called simply ***g*-factor**, here equal to 2. The g-factor has become a usual way of description of intrinsic magnetic moment of subatomic particles.

The above-described elegant derivation of the electron spin magnetic moment, in particular the natural explanation of the "anomalous" value $g=2$ for the g-factor was certainly a great success of the Dirac theory in 1928. In fact, even more remarkable was the continuation of this success story some 20 years later. It turned out that quantum electrodynamics (QED) leads to a tiny correction to the Dirac's prediction. The correction is of relative order of one per-mille; it was found experimentally in 1947 and subsequently calculated theoretically by Julian Schwinger, one of the founding fathers of modern QED. This achievement corroborated strongly the QED as the relevant model of quantum field theory capable to describe the most subtle electromagnetic phenomena. We will discuss this topic in detail in Chapter 50.