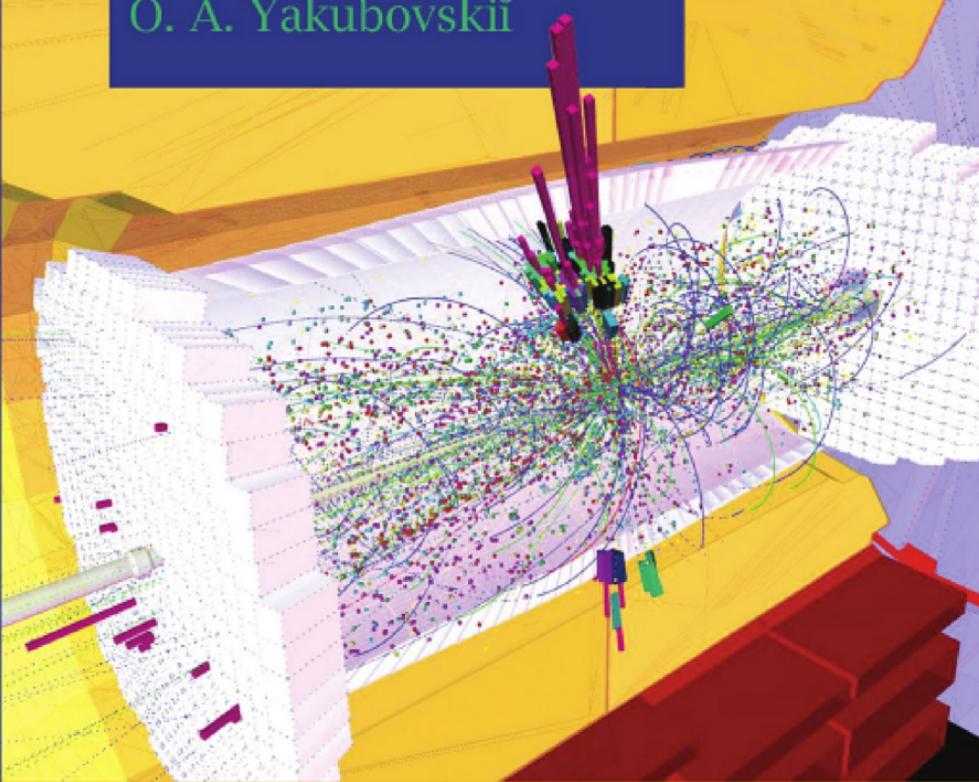


STUDENT MATHEMATICAL LIBRARY
Volume 47

Lectures on Quantum Mechanics for Mathematics Students

L. D. Faddeev
O. A. Yakubovskii



Lectures on Quantum Mechanics for Mathematics Students

STUDENT MATHEMATICAL LIBRARY
Volume 47

Lectures on Quantum Mechanics for Mathematics Students

**L. D. Faddeev
O. A. Yakubovskii**

**Translated by
Harold McFaden**



American Mathematical Society

Editorial Board

Gerald B. Folland Brad G Osgood (Chair)
Robin Forman Michael Starbird

The cover graphic was generated by Matt Strassler with help from Peter Skands. Processed through CMS by Albert De Roeck, Christophe Saout and Joanna Weng. Visualized by Ianna Osborne Copyright CERN

2000 *Mathematics Subject Classification.* Primary 81-01, 81Qxx.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-47

Library of Congress Cataloging-in-Publication Data

Faddeev, L D

[Lektsii po kvantovoi mekhanike dlia studentov-matematikov English]

Lectures on quantum mechanics for mathematical students / L D Faddeev,
O A Yakubovskii [English ed.]

p cm (Student mathematical library ; v 47)

ISBN 978-0-8218-4699-5 (alk paper)

1 Quantum theory I Yakubovskii, Oleg Aleksandrovich II Title

QC174.125 F3213 2009

530 12 -dc22

2008052385

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294, USA. Requests can also be made by e-mail to reprint-permission@ams.org

© 2009 by the American Mathematical Society All rights reserved

The American Mathematical Society retains all rights

except those granted to the United States Government

Printed in the United States of America

⊗ The paper used in this book is acid-free and falls within the guidelines

established to ensure permanence and durability

Visit the AMS home page at <http://www.ams.org/>

10 9 8 7 6 5 4 3 2 1 14 13 12 11 10 09

Contents

Preface	ix
Preface to the English Edition	xi
§1. The algebra of observables in classical mechanics	1
§2. States	6
§3. Liouville's theorem, and two pictures of motion in classical mechanics	13
§4. Physical bases of quantum mechanics	15
§5. A finite-dimensional model of quantum mechanics	27
§6. States in quantum mechanics	32
§7. Heisenberg uncertainty relations	36
§8. Physical meaning of the eigenvalues and eigenvectors of observables	39
§9. Two pictures of motion in quantum mechanics. The Schrödinger equation Stationary states	44
§10. Quantum mechanics of real systems. The Heisenberg commutation relations	49
§11. Coordinate and momentum representations	54
§12. “Eigenfunctions” of the operators Q and P	60
§13. The energy, the angular momentum, and other examples of observables	63

§14	The interconnection between quantum and classical mechanics. Passage to the limit from quantum mechanics to classical mechanics	69
§15.	One-dimensional problems of quantum mechanics. A free one-dimensional particle	77
§16	The harmonic oscillator	83
§17	The problem of the oscillator in the coordinate representation	87
§18.	Representation of the states of a one-dimensional particle in the sequence space l_2	90
§19.	Representation of the states for a one-dimensional particle in the space \mathcal{D} of entire analytic functions	94
§20.	The general case of one-dimensional motion	95
§21.	Three-dimensional problems in quantum mechanics. A three-dimensional free particle	103
§22	A three-dimensional particle in a potential field	104
§23	Angular momentum	106
§24.	The rotation group	108
§25.	Representations of the rotation group	111
§26.	Spherically symmetric operators	114
§27.	Representation of rotations by 2×2 unitary matrices	117
§28.	Representation of the rotation group on a space of entire analytic functions of two complex variables	120
§29.	Uniqueness of the representations D_j	123
§30.	Representations of the rotation group on the space $L^2(S^2)$. Spherical functions	127
§31	The radial Schrödinger equation	130
§32.	The hydrogen atom. The alkali metal atoms	136
§33.	Perturbation theory	147
§34.	The variational principle	154
§35.	Scattering theory. Physical formulation of the problem	157
§36	Scattering of a one-dimensional particle by a potential barrier	159

§37. Physical meaning of the solutions ψ_1 and ψ_2	164
§38 Scattering by a rectangular barrier	167
§39. Scattering by a potential center	169
§40. Motion of wave packets in a central force field	175
§41. The integral equation of scattering theory	181
§42. Derivation of a formula for the cross-section	183
§43. Abstract scattering theory	188
§44. Properties of commuting operators	197
§45. Representation of the state space with respect to a complete set of observables	201
§46. Spin	203
§47. Spin of a system of two electrons	208
§48. Systems of many particles. The identity principle	212
§49. Symmetry of the coordinate wave functions of a system of two electrons. The helium atom	215
§50. Multi-electron atoms. One-electron approximation	217
§51. The self-consistent field equations	223
§52. Mendeleev's periodic system of the elements	226
Appendix: Lagrangian Formulation of Classical Mechanics	231

Preface

This textbook is a detailed survey of a course of lectures given in the Mathematics-Mechanics Department of Leningrad University for mathematics students. The program of the course in quantum mechanics was developed by the first author, who taught the course from 1968 to 1973. Subsequently the course was taught by the second author. It has certainly changed somewhat over these years, but its goal remains the same: to give an exposition of quantum mechanics from a point of view closer to that of a mathematics student than is common in the physics literature. We take into account that the students do not study general physics. In a course intended for mathematicians, we have naturally aimed for a more rigorous presentation than usual of the mathematical questions in quantum mechanics, but not for full mathematical rigor, since a precise exposition of a number of questions would require a course of substantially greater scope.

In the literature available in Russian, there is only one book pursuing the same goal, and that is the American mathematician G. W. Mackey's book, *Mathematical Foundations of Quantum Mechanics*. The present lectures differ essentially from Mackey's book both in the method of presentation of the basics of quantum mechanics and in the selection of material. Moreover, these lectures assume somewhat less in the way of mathematical preparation of the students. Nevertheless, we have borrowed much both from Mackey's

book and from von Neumann's classical book, *Mathematical Foundations of Quantum Mechanics*.

The approach to the construction of quantum mechanics adopted in these lectures is based on the assertion that quantum and classical mechanics are different realizations of one and the same abstract mathematical structure. The features of this structure are explained in the first few sections, which are devoted to classical mechanics. These sections are an integral part of the course and should not be skipped over, all the more so because there is hardly any overlap of the material in them with the material in a course of theoretical mechanics. As a logical conclusion of our approach to the construction of quantum mechanics, we have a section devoted to the interconnection of quantum and classical mechanics and to the passage to the limit from quantum mechanics to classical mechanics.

In the selection of the material in the sections devoted to applications of quantum mechanics we have tried to single out questions connected with the formulation of interesting mathematical problems. Much attention here is given to problems connected with the theory of group representations and to mathematical questions in the theory of scattering. In other respects the selection of material corresponds to traditional textbooks on general questions in quantum mechanics, for example, the books of V. A. Fok or P. A. M. Dirac.

The authors are grateful to V. M. Babich, who read through the manuscript and made a number of valuable comments.

L. D. Faddeev and O. A. Yakubovskii

Preface to the English Edition

The history and the goals of this book are adequately described in the original Preface (to the Russian edition) and I shall not repeat it here. The idea to translate the book into English came from the numerous requests of my former students, who are now spread over the world. For a long time I kept postponing the translation because I hoped to be able to modify the book making it more informative. However, the recent book by Leon Takhtajan, *Quantum Mechanics for Mathematicians* (Graduate Studies in Mathematics, Volume 95, American Mathematical Society, 2008), which contains most of the material I was planning to add, made such modifications unnecessary and I decided that the English translation can now be published.

Just when the decision to translate the book was made, my coauthor Oleg Yakubovskii died. He had taught this course for more than 30 years and was quite devoted to it. He felt compelled to add some physical words to my more formal exposition. The Russian text, published in 1980, was prepared by him and can be viewed as a combination of my original notes for the course and his experience of teaching it. It is a great regret that he will not see the English translation.

Leon Takhtajan prepared a short appendix about the formalism of classical mechanics. It should play the role of introduction for students who did not take an appropriate course, which was obligatory at St. Petersburg University.

I want to add that the idea of introducing quantum mechanics as a deformation of classical mechanics has become quite fashionable nowadays. Of course, whereas the term “deformation” is not used explicitly in the book, the idea of deformation was a guiding principle in the original plan for the lectures.

I. D. Faddeev
St. Petersburg, November 2008

§ 1. The algebra of observables in classical mechanics

We consider the simplest problem in classical mechanics: the problem of the motion of a material point (a particle) with mass m in a force field $V(\mathbf{x})$, where $\mathbf{x}(x_1, x_2, x_3)$ is the radius vector of the particle. The force acting on the particle is

$$\mathbf{F} = -\operatorname{grad} V = -\frac{\partial V}{\partial \mathbf{x}}.$$

The basic physical characteristics of the particle are its coordinates x_1, x_2, x_3 and the projections of the velocity vector $\mathbf{v}(v_1, v_2, v_3)$. All the remaining characteristics are functions of \mathbf{x} and \mathbf{v} : for example, the momentum $\mathbf{p} = m\mathbf{v}$, the angular momentum $\mathbf{l} = \mathbf{x} \times \mathbf{p} = m\mathbf{x} \times \mathbf{v}$, and the energy $E = mv^2/2 + V(\mathbf{x})$.

The equations of motion of a material point in the Newton form are

$$(1) \quad m \frac{d\mathbf{v}}{dt} = -\frac{\partial V}{\partial \mathbf{x}}, \quad \frac{d\mathbf{x}}{dt} = \mathbf{v}.$$

It will be convenient below to use the momentum \mathbf{p} in place of the velocity \mathbf{v} as a basic variable. In the new variables the equations of motion are written as follows:

$$(2) \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial V}{\partial \mathbf{x}}, \quad \frac{d\mathbf{x}}{dt} = \frac{\mathbf{p}}{m}.$$

Noting that $\frac{\mathbf{p}}{m} = \frac{\partial H}{\partial \mathbf{p}}$ and $\frac{\partial V}{\partial \mathbf{x}} = \frac{\partial H}{\partial \mathbf{x}}$, where $H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x})$ is the Hamiltonian function for a particle in a potential field, we arrive at the equations in the Hamiltonian form

$$(3) \quad \frac{d\mathbf{x}}{dt} = \frac{\partial H}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{x}}.$$

It is known from a course in theoretical mechanics that a broad class of mechanical systems, and conservative systems in particular,

are described by the Hamiltonian equations

$$(4) \quad \dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad i = 1, 2, \dots, n$$

Here $H = H(q_1, \dots, q_n; p_1, \dots, p_n)$ is the Hamiltonian function, q_i and p_i are the generalized coordinates and momenta, and n is called the number of degrees of freedom of the system. We recall that for a conservative system, the Hamiltonian function H coincides with the expression for the total energy of the system in the variables q_i and p_i . We write the Hamiltonian function for a system of N material points interacting pairwise.

$$(5) \quad H = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i < j} V_{ij}(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^N V_i(\mathbf{x}_i).$$

Here the Cartesian coordinates of the particles are taken as the generalized coordinates q , the number of degrees of freedom of the system is $n = 3N$, and $V_{ij}(\mathbf{x}_i - \mathbf{x}_j)$ is the potential of the interaction of the i th and j th particles. The dependence of V_{ij} only on the difference $\mathbf{x}_i - \mathbf{x}_j$ is ensured by Newton's third law. (Indeed, the force acting on the i th particle due to the j th particle is $\mathbf{F}_{ij} = -\frac{\partial V_{ij}}{\partial \mathbf{x}_i} = \frac{\partial V_{ij}}{\partial \mathbf{x}_j} = -\mathbf{F}_{ji}$.) The potentials $V_i(\mathbf{x}_i)$ describe the interaction of the i th particle with the external field. The first term in (5) is the kinetic energy of the system of particles.

For any mechanical system all physical characteristics are functions of the generalized coordinates and momenta. We introduce the set \mathfrak{A} of real infinitely differentiable functions $f(q_1, \dots, q_n; p_1, \dots, p_n)$, which will be called observables.¹ The set \mathfrak{A} of observables is obviously a linear space and forms a real algebra with the usual addition and multiplication operations for functions. The real $2n$ -dimensional space with elements $(q_1, \dots, q_n; p_1, \dots, p_n)$ is called the phase space and is denoted by \mathcal{M} . Thus, the algebra of observables in classical mechanics is the algebra of real-valued smooth functions defined on the phase space \mathcal{M} .

We shall introduce in the algebra of observables one more operation, which is connected with the evolution of the mechanical system

¹We do not discuss the question of introducing a topology in the algebra of observables. Fortunately, most physical questions do not depend on this topology.

For simplicity the exposition to follow is conducted using the example of a system with one degree of freedom. The Hamiltonian equations in this case have the form

$$(6) \quad \dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}, \quad H = H(q, p).$$

The Cauchy problem for the system (6) and the initial conditions

$$(7) \quad q|_{t=0} = q_0, \quad p|_{t=0} = p_0$$

has a unique solution

$$(8) \quad q = q(q_0, p_0, t), \quad p = p(q_0, p_0, t).$$

For brevity of notation a point (q, p) in phase space will sometimes be denoted by μ , and the Hamiltonian equations will be written in the form

$$(9) \quad \dot{\mu} = v(\mu),$$

where $v(\mu)$ is the vector field of these equations, which assigns to each point μ of phase space the vector v with components $\frac{\partial H}{\partial p}$, $-\frac{\partial H}{\partial q}$.

The Hamiltonian equations generate a one-parameter commutative group of transformations

$$G_t: \mathcal{M} \rightarrow \mathcal{M}$$

of the phase space into itself,² where $G_t\mu$ is the solution of the Hamiltonian equations with the initial condition $G_t\mu|_{t=0} = \mu$. We have the equalities

$$(10) \quad G_{t+s} = G_t G_s = G_s G_t, \quad G_t^{-1} = G_{-t}.$$

In turn, the transformations G_t generate a family of transformations

$$U_t: \mathfrak{A} \rightarrow \mathfrak{A}$$

of the algebra of observables into itself, where

$$(11) \quad U_t f(\mu) = f_t(\mu) = f(G_t\mu)$$

In coordinates, the function $f_t(q, p)$ is defined as follows:

$$(12) \quad f_t(q_0, p_0) = f(q(q_0, p_0, t), p(q_0, p_0, t)).$$

²We assume that the Hamiltonian equations with initial conditions (7) have a unique solution on the whole real axis. It is easy to construct examples in which a global solution and, correspondingly, a group of transformations G_t do not exist. These cases are not interesting, and we do not consider them.

We find a differential equation that the function $f_t(q, p)$ satisfies. To this end, we differentiate the identity $f_{s+t}(\mu) = f_t(G_s \mu)$ with respect to the variable s and set $s = 0$:

$$\frac{\partial f_{s+t}(\mu)}{\partial s} \Big|_{s=0} = \frac{\partial f_t(\mu)}{\partial t},$$

$$\frac{\partial f_t(G_s \mu)}{\partial s} \Big|_{s=0} = \nabla f_t(\mu) \cdot v(\mu) = \frac{\partial f_t}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial f_t}{\partial p} \frac{\partial H}{\partial q}$$

Thus, the function $f_t(q, p)$ satisfies the differential equation

$$(13) \quad \frac{\partial f_t}{\partial t} = \frac{\partial H}{\partial p} \frac{\partial f_t}{\partial q} - \frac{\partial H}{\partial q} \frac{\partial f_t}{\partial p}$$

and the initial condition

$$(14) \quad f_t(q, p)|_{t=0} = f(q, p).$$

The equation (13) with the initial condition (14) has a unique solution, which can be obtained by the formula (12): that is, to construct the solutions of (13) it suffices to know the solutions of the Hamiltonian equations.

We can rewrite (13) in the form

$$(15) \quad \frac{df_t}{dt} = \{H, f_t\}.$$

where $\{H, f_t\}$ is the Poisson bracket of the functions H and f_t . For arbitrary observables f and g the Poisson bracket is defined by

$$\{f, g\} = \frac{\partial f}{\partial p} \frac{\partial g}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial g}{\partial p}.$$

and in the case of a system with n degrees of freedom

$$\{f, g\} = \sum_{i=1}^n \left(\frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} - \frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} \right).$$

We list the basic properties of Poisson brackets:

- 1) $\{f, g + \lambda h\} = \{f, g\} + \lambda \{f, h\}$ (linearity);
- 2) $\{f, g\} = -\{g, f\}$ (skew symmetry);
- 3) $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$ (Jacobi identity).
- 4) $\{f, gh\} = g\{f, h\} + \{f, g\}h$.

The properties 1), 2), and 4) follow directly from the definition of the Poisson brackets. The property 4) shows that the “Poisson bracket” operation is a derivation of the algebra of observables. Indeed, the Poisson bracket can be rewritten in the form

$$\{f, g\} = X_f g,$$

where $X_f = \frac{\partial f}{\partial p} \frac{\partial}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial}{\partial p}$ is a first-order linear differential operator, and the property 4) has the form

$$X_f gh = (X_f g)h + gX_f h.$$

The property 3) can be verified by differentiation, but it can be proved by the following argument. Each term of the double Poisson bracket contains as a factor the second derivative of one of the functions with respect to one of the variables; that is, the left-hand side of 3) is a linear homogeneous function of the second derivatives. On the other hand, the second derivatives of h can appear only in the sum $\{f, \{g, h\}\} + \{g, \{h, f\}\} = (X_f X_g - X_g X_f)h$, but a commutator of first-order linear differential operators is a first-order differential operator, and hence the second derivatives of h do not appear in the left-hand side of 3). By symmetry, the left-hand side of 3) does not contain second derivatives at all: that is, it is equal to zero.

The Poisson bracket $\{f, g\}$ provides the algebra of observables with the structure of a real Lie algebra.³ Thus, the set of observables has the following algebraic structure. The set \mathfrak{A} is:

- 1) a real linear space,
- 2) a commutative algebra with the operation fg ;
- 3) a Lie algebra with the operation $\{f, g\}$.

The last two operations are connected by the relation

$$\{f, gh\} = \{f, g\}h + g\{f, h\}.$$

The algebra \mathfrak{A} of observables contains a distinguished element, namely, the Hamiltonian function H , whose role is to describe the

³We recall that a linear space with a binary operation satisfying the conditions 1)-3) is called a Lie algebra

variation of observables with time:

$$\frac{df_t}{dt} = \{H, f_t\}.$$

We show that the mapping $U_t: \mathfrak{A} \rightarrow \mathfrak{A}$ preserves all the operations in \mathfrak{A} :

$$h = f + g \rightarrow h_t = f_t + g_t,$$

$$h = fg \rightarrow h_t = f_t g_t,$$

$$h = \{f, g\} \rightarrow h_t = \{f_t, g_t\};$$

that is, it is an automorphism of the algebra of observables. For example, we verify the last assertion. For this it suffices to see that the equation and the initial condition for h_t is a consequence of the equations and initial conditions for the functions f_t and g_t :

$$\begin{aligned}\frac{\partial h_t}{\partial t} &= \left\{ \frac{\partial f_t}{\partial t}, g_t \right\} + \left\{ f_t, \frac{\partial g_t}{\partial t} \right\} \\ &= \{\{H, f_t\}, g_t\} + \{f_t, \{H, g_t\}\} = \{H, \{f_t, g_t\}\} = \{H, h_t\}.\end{aligned}$$

Here we used the properties 2) and 4) of Poisson brackets. Furthermore,

$$h_t|_{t=0} = \{f_t, g_t\}|_{t=0} = \{f, g\}.$$

Our assertion now follows from the uniqueness of the solution of the equation (13) with the initial condition (14).

§ 2. States

The concept of a state of a system can be connected directly with the conditions of an experiment. Every physical experiment reduces to a measurement of the numerical value of an observable for the system under definite conditions that can be called the conditions of the experiment. It is assumed that these conditions can be reproduced multiple times, but we do not assume in advance that the measurement will give the same value of the observable when the experiment is repeated. There are two possible answers to the question of how to explain this uncertainty in the results of an experiment.

1) The number of conditions that are fixed in performing the experiments is insufficient to uniquely determine the results of the measurement of the observables. If the nonuniqueness arises only

for this reason, then at least in principle these conditions can be supplemented by new conditions, that is, one can pose the experiment more "cleanly", and then the results of all the measurements will be uniquely determined.

2) The properties of the system are such that in repeated experiments the observables can take different values independently of the number and choice of the conditions of the experiment.

Of course if 2) holds, then insufficiency of the conditions can only increase the nonuniqueness of the experimental results. We discuss 1) and 2) at length after we learn how to describe states in classical and quantum mechanics.

We shall consider that the conditions of the experiment determine the state of the system if conducting many repeated trials under these conditions leads to probability distributions for all the observables. In this case we speak of the measurement of an observable f for a system in the state ω . More precisely, a state ω on the algebra \mathfrak{A} of observables assigns to each observable f a probability distribution of its possible values, that is, a measure on the real line \mathbf{R} .

Let f be an observable and E a Borel set on the real line \mathbf{R} . Then the definition of a state ω can be written as

$$f, E \xrightarrow{\omega} \omega_f(E).$$

We recall the properties of a probability measure:

$$(1) \quad 0 \leq \omega_f(E) \leq 1, \quad \omega_f(\emptyset) = 0, \quad \omega_f(\mathbf{R}) = 1,$$

and if $E_1 \cap E_2 = \emptyset$, then $\omega_f(E_1 \cup E_2) = \omega_f(E_1) + \omega_f(E_2)$.

Among the observables there may be some that are functionally dependent, and hence it is necessary to impose a condition on the probability distributions of such observables. If an observable φ is a function of an observable f , $\varphi = \varphi(f)$, then this assertion means that a measurement of the numerical value of f yielding a value f_0 is at the same time a measurement of the observable φ and gives for it the numerical value $\varphi_0 = \varphi(f_0)$. Therefore, $\omega_f(E)$ and $\omega_{\varphi(f)}(E)$ are connected by the equality

$$(2) \quad \omega_{\varphi(f)}(E) = \omega_f(\varphi^{-1}(E)),$$

where $\varphi^{-1}(E)$ is the inverse image of E under the mapping φ .

A convex combination

$$(3) \quad \omega_f(E) = \alpha\omega_{1f}(E) + (1 - \alpha)\omega_{2f}(E), \quad 0 < \alpha < 1,$$

of probability measures has the properties (1) for any observable f and corresponds to a state which we denote by

$$(4) \quad \omega = \alpha\omega_1 + (1 - \alpha)\omega_2.$$

Thus, the states form a convex set. A convex combination (4) of states ω_1 and ω_2 will sometimes be called a mixture of these states. If for some state ω it follows from (4) that $\omega_1 = \omega_2 = \omega$, then we say that the state ω is indecomposable into a convex combination of different states. Such states are called *pure* states, and all other states are called *mixed* states.

It is convenient to take E to be an interval $(-\infty, \lambda]$ of the real axis. By definition, $\omega_f(\lambda) = \omega_f((-\infty, \lambda])$, and this is the distribution function of the observable f in the state ω . Numerically, $\omega_f(\lambda)$ is the probability of getting a value not exceeding λ when measuring f in the state ω . It follows from (1) that the distribution function $\omega_f(\lambda)$ is a nondecreasing function of λ with $\omega_f(-\infty) = 0$ and $\omega_f(+\infty) = 1$.

The mathematical expectation (mean value) of an observable f in a state ω is defined by the formula⁴

$$\langle f | \omega \rangle = \int_{-\infty}^{\infty} \lambda d\omega_f(\lambda).$$

We remark that knowledge of the mathematical expectations for all the observables is equivalent to knowledge of the probability distributions. To see this, it suffices to consider the function $\theta(\lambda - f)$ of the observables, where $\theta(x)$ is the Heaviside function

$$\theta(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

It is not hard to see that

$$(5) \quad \omega_f(\lambda) = \langle \theta(\lambda - f) | \omega \rangle.$$

⁴The notation $\langle f | \omega \rangle$ for the mean value of an observable should not be confused with the Dirac notation often used in quantum mechanics for the scalar product $\langle \varphi, \psi \rangle$ of vectors

For the mean values of observables we require the following conditions, which are natural from a physical point of view:

- 1) $\langle C | \omega \rangle = C,$
- (6) 2) $\langle f + \lambda g | \omega \rangle = \langle f | \omega \rangle + \lambda \langle g | \omega \rangle,$
- 3) $\langle f^2 | \omega \rangle \geq 0.$

If these requirements are introduced, then the realization of the algebra of observables itself determines a way of describing the states. Indeed, the mean value is a positive linear functional on the algebra \mathfrak{A} of observables. The general form of such a functional is

$$(7) \quad \langle f | \omega \rangle = \int_{\mathcal{M}} f(p, q) d\mu_{\omega}(p, q),$$

where $d\mu_{\omega}(p, q)$ is the differential of the measure on the phase space, and the integral is over the whole of phase space. It follows from the condition 1) that

$$(8) \quad \int_{\mathcal{M}} d\mu_{\omega}(p, q) = \mu_{\omega}(\mathcal{M}) = 1.$$

We see that a state in classical mechanics is described by specifying a probability distribution on the phase space. The formula (7) can be rewritten in the form

$$(9) \quad \langle f | \omega \rangle = \int_{\mathcal{M}} f(p, q) \rho_{\omega}(p, q) dp dq;$$

that is, we arrive at the usual description in statistical physics of a state of a system with the help of the distribution function $\rho_{\omega}(p, q)$, which in the general case is a positive generalized function. The normalization condition of the distribution function has the form

$$(10) \quad \int_{\mathcal{M}} \rho(p, q) dp dq = 1.$$

In particular, it is easy to see that to a pure state there corresponds a distribution function

$$(11) \quad \rho(p, q) = \delta(q - p_0) \delta(p - p_0),$$

where $\delta(x)$ is the Dirac δ -function. The corresponding measure on the phase space is concentrated at the point (q_0, p_0) , and a pure state is defined by specifying this point of phase space. For this reason the

phase space \mathcal{M} is sometimes called the state space. The mean value of an observable f in the pure state ω is

$$(12) \quad \langle f | \omega \rangle = f(q_0, p_0).$$

This formula follows immediately from the definition of the δ -function:

$$(13) \quad f(q_0, p_0) = \int_{\mathcal{M}} f(q, p) \delta(q - q_0) \delta(p - p_0) dq dp.$$

In mechanics courses one usually studies only pure states, while in statistical physics one considers mixed states, with distribution function different from (11). But an introduction to the theory of mixed states from the very start is warranted by the following circumstances. The formulation of classical mechanics in the language of states and observables is nearest to the formulation of quantum mechanics and makes it possible to describe states in mechanics and statistical physics in a uniform way. Such a formulation will enable us to follow closely the passage to the limit from quantum mechanics to classical mechanics. We shall see that in quantum mechanics there are also pure and mixed states, and in the passage to the limit, a pure quantum state can be transformed into a mixed classical state, so that the passage to the limit is most simply described when pure and mixed states are treated in a uniform way.

We now explain the physical meaning of mixed and pure states in classical mechanics, and we find out why experimental results are not necessarily determined uniquely by the conditions of the experiment.

Let us consider a mixture

$$\omega = \alpha \omega_1 + (1 - \alpha) \omega_2, \quad 0 < \alpha < 1,$$

of the states ω_1 and ω_2 . The mean values obviously satisfy the formula

$$(14) \quad \langle f | \omega \rangle = \alpha \langle f | \omega_1 \rangle + (1 - \alpha) \langle f | \omega_2 \rangle.$$

The formulas (14) and (3) admit the following interpretation. The assertion that the system is in the state ω is equivalent to the assertion that the system is in the state ω_1 with probability α and in the state ω_2 with probability $(1 - \alpha)$. We remark that this interpretation is possible but not necessary.

The simplest mixed state is a convex combination of two pure states:

$$\rho(q, p) = \alpha\delta(q - q_1)\delta(p - p_1) + (1 - \alpha)\delta(q - q_2)\delta(p - p_2).$$

Mixtures of n pure states are also possible:

$$\rho(q, p) = \sum_{i=1}^n \alpha_i \delta(q - q_i) \delta(p - p_i), \quad \alpha_i > 0, \quad \sum_{i=1}^n \alpha_i = 1.$$

Here α_i can be interpreted as the probability of realization of the pure state given by the point q_i, p_i of phase space. Finally, in the general case the distribution function can be written as

$$\rho(q, p) = \int_{\mathcal{M}} \rho(q_0, p_0) \delta(q - q_0) \delta(p - p_0) dq_0 dp_0.$$

Such an expression leads to the usual interpretation of the distribution function in statistical physics: $\int_{\Omega} \rho(q, p) dq dp$ is the probability of observing the system in a pure state represented by a point in the domain Ω of phase space. We emphasize once more that this interpretation is not necessary, since pure and mixed states can be described in the framework of a unified formalism.

One of the most important characteristics of a probability distribution is the variance

$$(15) \quad \Delta_{\omega}^2 f = \langle (f - \langle f | \omega \rangle)^2 | \omega \rangle = \langle f^2 | \omega \rangle - \langle f | \omega \rangle^2.$$

We show that mixing of states leads to an increase in the variance. A more precise formulation of this assertion is contained in the inequalities

$$(16) \quad \Delta_{\omega}^2 f \geq \alpha \Delta_{\omega_1}^2 f + (1 - \alpha) \Delta_{\omega_2}^2 f,$$

$$(17) \quad \Delta_{\omega} f \Delta_{\omega} g \geq \alpha \Delta_{\omega_1} f \Delta_{\omega_1} g + (1 - \alpha) \Delta_{\omega_2} f \Delta_{\omega_2} g,$$

with equalities only if the mean values of the observables in the states ω_1 and ω_2 coincide:

$$\langle f | \omega_1 \rangle = \langle f | \omega_2 \rangle, \quad \langle g | \omega_1 \rangle = \langle g | \omega_2 \rangle.$$

In a weakened form, (16) and (17) can be written as

$$(18) \quad \Delta_{\omega}^2 f \geq \min(\Delta_{\omega_1}^2 f, \Delta_{\omega_2}^2 f),$$

$$(19) \quad \Delta_{\omega} f \Delta_{\omega} g \geq \min(\Delta_{\omega_1} f \Delta_{\omega_1} g, \Delta_{\omega_2} f \Delta_{\omega_2} g).$$

The proof uses the elementary inequality

$$(20) \quad \varphi(x) = \alpha + (1 - \alpha)x^2 - (\alpha + (1 - \alpha)x)^2 \geq 0. \quad -\infty < x < \infty.$$

with

$$(21) \quad \varphi(1) = 0, \quad \varphi(x) > 0 \quad \text{for } x \neq 1.$$

Using (14) and (20), we get that

$$\begin{aligned} \Delta_{\omega}^2 f &= \langle f^2 | \omega \rangle - \langle f | \omega \rangle^2 \\ &= \alpha \langle f^2 | \omega_1 \rangle + (1 - \alpha) \langle f^2 | \omega_2 \rangle - [\alpha \langle f | \omega_1 \rangle + (1 - \alpha) \langle f | \omega_2 \rangle]^2 \\ &\geq \alpha \langle f^2 | \omega_1 \rangle + (1 - \alpha) \langle f^2 | \omega_2 \rangle - \alpha \langle f | \omega_1 \rangle^2 - (1 - \alpha) \langle f | \omega_2 \rangle^2 \\ &= \alpha \Delta_{\omega_1}^2 f + (1 - \alpha) \Delta_{\omega_2}^2 f. \end{aligned}$$

The inequality (16) is proved. The inequality (17) is a consequence of (16). Indeed,

$$\begin{aligned} \Delta_{\omega}^2 f \Delta_{\omega}^2 g &\geq (\alpha \Delta_{\omega_1}^2 f + (1 - \alpha) \Delta_{\omega_2}^2 f)(\alpha \Delta_{\omega_1}^2 g + (1 - \alpha) \Delta_{\omega_2}^2 g) \\ &\geq [\alpha \Delta_{\omega_1} f \Delta_{\omega_1} g + (1 - \alpha) \Delta_{\omega_2} f \Delta_{\omega_2} g]^2. \end{aligned}$$

For pure states,

$$\rho(q, p) = \delta(q - q_0) \delta(p - p_0),$$

$$\langle f | \omega \rangle = f(q_0, p_0).$$

$$\Delta_{\omega}^2 f = f^2(q_0, p_0) - f^2(q_0, p_0) = 0;$$

that is, for pure states in classical mechanics the variance is zero. This means that for a system in a pure state, the result of a measurement of any observable is uniquely determined. A state of a classical system will be pure if by the time of the measurement the conditions of the experiment fix the values of all the generalized coordinates and momenta. It is clear that if a macroscopic body is regarded as a mechanical system of N molecules, where N usually has order 10^{23} , then no conditions in a real physical experiment can fix the values of q_0 and p_0 for all molecules, and the description of such a system with the help of pure states is useless. Therefore, one studies mixed states in statistical physics.

Let us summarize. In classical mechanics there is an infinite set of states of the system (pure states) in which all observables have completely determined values. In real experiments with systems of a

huge number of particles. mixed states arise. Of course, such states are possible also in experiments with simple mechanical systems. In this case the theory gives only probabilistic predictions.

§ 3. Liouville's theorem, and two pictures of motion in classical mechanics

We begin this section with a proof of an important theorem of Liouville. Let Ω be a domain in the phase space \mathcal{M} . Denote by $\Omega(t)$ the image of this domain under the action of a phase flow, that is, the set of points $G_t\mu$, $\mu \in \Omega$. Let $V(t)$ be the volume of $\Omega(t)$. Liouville's theorem asserts that

$$\frac{dV(t)}{dt} = 0.$$

Proof.

$$V(t) = \int_{\Omega(t)} d\mu = \int_{\Omega} \left| \frac{D(G_t\mu)}{D(\mu)} \right| d\mu, \quad d\mu = dq dp.$$

Here $D(G_t\mu)/D(\mu)$ denotes the Jacobi determinant of the transformation G_t . To prove the theorem, it suffices to show that

$$(1) \quad \frac{D(G_t\mu)}{D(\mu)} = 1$$

for all t . The equality (1) is obvious for $t = 0$. Let us now show that

$$(2) \quad \frac{d}{dt} \frac{D(G_t\mu)}{D(\mu)} = 0.$$

For $t = 0$ the formula (2) can be verified directly:

$$\begin{aligned} \frac{d}{dt} \frac{D(G_t\mu)}{D(\mu)} \Big|_{t=0} &= \left[\frac{D(\dot{q}, p)}{D(q, p)} + \frac{D(q, \dot{p})}{D(q, p)} \right] \Big|_{t=0} \\ &= \left(\frac{\partial \dot{q}}{\partial q} + \frac{\partial \dot{p}}{\partial p} \right) \Big|_{t=0} = \frac{\partial^2 H}{\partial q \partial p} - \frac{\partial^2 H}{\partial p \partial q} = 0. \end{aligned}$$

For $t \neq 0$ we differentiate the identity

$$\frac{D(G_{t+s}\mu)}{D(\mu)} = \frac{D(G_{t+s}\mu)}{D(G_t\mu)} \frac{D(G_t\mu)}{D(\mu)}$$

with respect to s and set $s = 0$, getting

$$\frac{d}{dt} \frac{D(G_t\mu)}{D(\mu)} = \left[\frac{d}{dt} \frac{D(G_s G_t \mu)}{D(G_t \mu)} \right] \Big|_{t=0} \frac{D(G_t \mu)}{D(\mu)} = 0.$$

Thus, (2) holds for all t . The theorem is proved. \square

We now consider the evolution of a mechanical system. We are interested in the time dependence of the mean values $\langle f | \omega \rangle$ of the observables. There are two possible ways of describing this dependence, that is, two pictures of the motion. We begin with the formulation of the so-called Hamiltonian picture. In this picture the time dependence of the observables is determined by the equation (1.15),⁵ and the states do not depend on time:

$$\frac{df_t}{dt} = \{H, f_t\}, \quad \frac{d\rho}{dt} = 0.$$

The mean value of an observable f in a state ω depends on the time according to the formula

$$(3) \quad \langle f_t | \omega \rangle = \int_{\mathcal{M}} f_t(\mu) \rho(\mu) d\mu = \int_{\mathcal{M}} f(G_t \mu) \rho(\mu) d\mu,$$

or, in more detail,

$$(4) \quad \langle f_t | \omega \rangle = \int_{\mathcal{M}} f(q(q_0, p_0, t), p(q_0, p_0, t)) \rho(q_0, p_0) dq_0 dp_0.$$

We recall that $q(q_0, p_0, t)$ and $p(q_0, p_0, t)$ are solutions of the Hamiltonian equations with initial conditions $q|_{t=0} = q_0$, $p|_{t=0} = p_0$.

For a pure state $\rho(q, p) = \delta(q - q_0)\delta(p - p_0)$, and it follows from (4) that

$$\langle f_t | \omega \rangle = f(q(q_0, p_0, t), p(q_0, p_0, t)).$$

This is the usual classical mechanics formula for the time dependence of an observable in a pure state.⁶ It is clear from the formula (4) that a state in the Hamiltonian picture determines the probability distribution of the initial values of q and p .

⁵In referring to a formula in previous sections the number of the corresponding section precedes the number of the formula

⁶In courses in mechanics it is usual to consider only pure states. Furthermore, no distinction is made between the dependence on time of an abstract observable in the Hamiltonian picture and the variation of its mean value

An alternative way of describing the motion is obtained if in (3) we make the change of variables $G_t\mu \rightarrow \mu$. Then

$$\begin{aligned}\int_{\mathcal{M}} f(G_t\mu) \rho(\mu) d\mu &= \int_{\mathcal{M}} f(\mu) \rho(G_{-t}\mu) \left| \frac{D(G_{-t}\mu)}{D(\mu)} \right| d\mu \\ &= \int_{\mathcal{M}} f(\mu) \rho_t(\mu) d\mu = \langle f | \omega_t \rangle.\end{aligned}$$

Here we have used the equality (1) and we have introduced the notation $\rho_t(\mu) = \rho(G_{-t}\mu)$. It is not hard to see that $\rho_t(\mu)$ satisfies the equation

$$(5) \quad \frac{d\rho_t}{dt} = -\{H, \rho_t\},$$

which differs from (1.15) by the sign in front of the Poisson bracket. The derivation of the equation (5) repeats that word-for-word for the equation (1.15), and the difference in sign arises because $G_{-t}\mu$ satisfies the Hamiltonian equations with reversed time. The picture of the motion in which the time dependence of the states is determined by (5), while the observables do not depend on time, is called the Liouville picture:

$$\frac{df}{dt} = 0, \quad \frac{d\rho_t}{dt} = -\{H, \rho_t\}.$$

The equation (5) is called Liouville's equation. It is obvious from the way the Liouville picture was introduced that

$$\langle f_t | \omega \rangle = \langle f | \omega_t \rangle.$$

This formula expresses the equivalence of the two pictures of motion. The mean values of the observables depend on time in the same way, and the difference between the pictures lies only in the way this dependence is described. We remark that it is common in statistical physics to use the Liouville picture.

§ 4. Physical bases of quantum mechanics

Quantum mechanics is the mechanics of the microworld. The phenomena it studies lie mainly beyond the limits of our perception, and therefore we should not be surprised by the seemingly paradoxical nature of the laws governing these phenomena.

It has not been possible to formulate the basic laws of quantum mechanics as a logical consequence of the results of some collection of fundamental physical experiments. In other words, there is so far no known formulation of quantum mechanics that is based on a system of axioms confirmed by experiment. Moreover, some of the basic statements of quantum mechanics are in principle not amenable to experimental verification. Our confidence in the validity of quantum mechanics is based on the fact that all the physical results of the theory agree with experiment. Thus, only consequences of the basic tenets of quantum mechanics can be verified by experiment, and not its basic laws. The main difficulties arising upon an initial study of quantum mechanics are apparently connected with these circumstances.

The creators of quantum mechanics were faced with difficulties of the same nature, though certainly much more formidable. Experiments most definitely pointed to the existence of peculiar quantum laws in the microworld, but gave no clue about the form of quantum theory. This can explain the truly dramatic history of the creation of quantum mechanics and, in particular, the fact that its original formulations bore a purely prescriptive character. They contained certain rules making it possible to compute experimentally measurable quantities, but a physical interpretation of the theory appeared only after a mathematical formalism of it had largely been created.

In this course we do not follow the historical path in the construction of quantum mechanics. We very briefly describe certain physical phenomena for which attempts to explain them on the basis of classical physics led to insurmountable difficulties. We then try to clarify what features of the scheme of classical mechanics described in the preceding sections should be preserved in the mechanics of the microworld and what can and must be rejected. We shall see that the rejection of only one assertion of classical mechanics, namely, the assertion that observables are functions on the phase space, makes it possible to construct a scheme of mechanics describing systems with behavior essentially different from the classical. Finally, in the following sections we shall see that the theory constructed is more general than classical mechanics, and contains the latter as a limiting case.

Historically, the first quantum hypothesis was proposed by Planck in 1900 in connection with the theory of equilibrium radiation. He succeeded in getting a formula in agreement with experiment for the spectral distribution of the energy of thermal radiation, conjecturing that electromagnetic radiation is emitted and absorbed in discrete portions, or quanta, whose energy is proportional to the frequency of the radiation:

$$(1) \quad E = h\omega,$$

where $\omega = 2\pi\nu$ with ν the frequency of oscillations in the light wave, and where $h = 1.05 \times 10^{-27}$ erg-sec is Planck's constant.⁷

Planck's hypothesis about light quanta led Einstein to give an extraordinarily simple explanation for the photoelectric effect (1905). The photoelectric phenomenon consists in the fact that electrons are emitted from the surface of a metal under the action of a beam of light. The basic problem in the theory of the photoelectric effect is to find the dependence of the energy of the emitted electrons on the characteristics of the light beam. Let V be the work required to remove an electron from the metal (the work function). Then the law of conservation of energy leads to the relation

$$h\omega = V + T,$$

where T is the kinetic energy of the emitted electron. We see that this energy is linearly dependent on the frequency and is independent of the intensity of the light beam. Moreover, for a frequency $\omega < V/h$ (the red limit of the photoelectric effect), the photoelectric phenomenon becomes impossible since $T \geq 0$. These conclusions, based on the hypothesis of light quanta, are in complete agreement with experiment. At the same time, according to the classical theory, the energy of the emitted electrons should depend on the intensity of the light waves, which contradicts the experimental results.

Einstein supplemented the idea of light quanta by introducing the momentum of a light quantum by the formula

$$(2) \quad \mathbf{p} = \hbar\mathbf{k}.$$

⁷In the older literature this formula is often written in the form $E = h\nu$, where the constant h in the latter formula obviously differs from the h in (1) by the factor 2π .

Here \mathbf{k} is the so-called wave vector, which has the direction of propagation of the light waves. The length k of this vector is connected with the wavelength λ , the frequency ω , and the velocity c of light by the relations

$$(3) \quad k = \frac{2\pi}{\lambda} = \frac{\omega}{c}.$$

For light quanta we have the formula

$$E = pc,$$

which is a special case of the relativity theory formula

$$E = \sqrt{p^2 c^2 + m^2 c^4}$$

for a particle with rest mass $m = 0$.

We remark that the historically first quantum hypotheses involved the laws of emission and absorption of light waves, that is, electrodynamics, and not mechanics. However, it soon became clear that discreteness of the values of a number of physical quantities was typical not only for electromagnetic radiation but also for atomic systems. The experiments of Franck and Hertz (1913) showed that when electrons collide with atoms, the energy of the electrons changes in discrete portions. The results of these experiments can be explained by the assumption that the energy of the atoms can have only definite discrete values. Later experiments of Stern and Gerlach in 1922 showed that the projection of the angular momentum of atomic systems on a certain direction has an analogous property. It is now well known that the discreteness of the values of a number of observables, though typical, is not a necessary feature of systems in the microworld. For example, the energy of an electron in a hydrogen atom has discrete values, but the energy of a freely moving electron can take arbitrary positive values. The mathematical apparatus of quantum mechanics had to be adapted to the description of observables taking both discrete and continuous values.

In 1911 Rutherford discovered the atomic nucleus and proposed a planetary model of the atom (his experiments on scattering of α particles on samples of various elements showed that an atom has a positively charged nucleus with charge Ze , where Z is the number of the element in the Mendeleev periodic table and e is the charge of an

electron, and that the size of the nucleus does not exceed 10^{-12} cm, while the atom itself has linear size of order 10^{-8} cm). The planetary model contradicts the basic tenets of classical electrodynamics. Indeed, when moving around the nucleus in classical orbits, the electrons, like all charged particles that are accelerating, must radiate electromagnetic waves. Thus, they must be losing their energy and must eventually fall into the nucleus. Therefore, such an atom cannot be stable, and this, of course, does not correspond to reality. One of the main problems of quantum mechanics is to account for the stability and to describe the structure of atoms and molecules as systems consisting of positively charged nuclei and electrons.

The phenomenon of diffraction of microparticles is completely surprising from the point of view of classical mechanics. This phenomenon was predicted in 1924 by de Broglie, who suggested that to a freely moving particle with momentum \mathbf{p} and energy E there corresponds (in some sense) a wave with wave vector \mathbf{k} and frequency ω , where

$$\mathbf{p} = \hbar \mathbf{k}, \quad E = \hbar \omega;$$

that is, the relations (1) and (2) are valid not only for light quanta but also for particles. A physical interpretation of de Broglie waves was later given by Born, but we shall not discuss it for the present. If to a moving particle there corresponds a wave, then, regardless of the precise meaning of these words, it is natural to expect that this implies the existence of diffraction phenomena for particles. Diffraction of electrons was first observed in experiments of Davisson and Germer in 1927. Diffraction phenomena were subsequently observed also for other particles.

We show that diffraction phenomena are incompatible with classical ideas about the motion of particles along trajectories. It is most convenient to argue using the example of a thought experiment concerning the diffraction of a beam of electrons directed at two slits,⁸ the scheme of which is pictured in Figure 1. Suppose that the electrons

⁸Such an experiment is a thought experiment, so the wavelength of the electrons at energies convenient for diffraction experiments does not exceed 10^{-7} cm, and the distance between the slits must be of the same order. In real experiments diffraction is observed on crystals, which are like natural diffraction lattices.

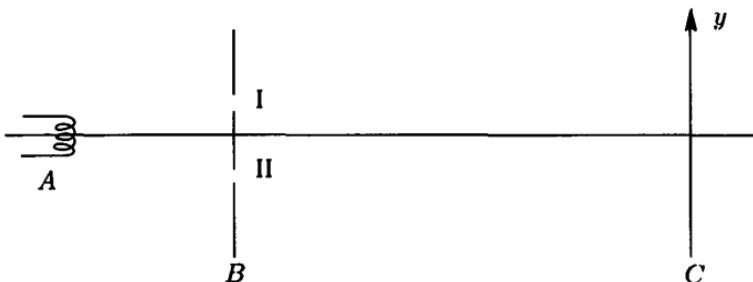


Figure 1

from the source A move toward the screen B and, passing through the slits I and II, strike the screen C.

We are interested in the distribution of electrons hitting the screen C with respect to the coordinate y . The diffraction phenomena on one and two slits have been thoroughly studied, and we can assert that the electron distribution $\rho(y)$ has the form *a* pictured in Figure 2 if only the first slit is open, the form *b* (Figure 2) if only the second slit is open, and the form *c* occurs when both slits are open. If we assume that each electron moves along a definite classical trajectory, then the electrons hitting the screen C can be split into two groups, depending on the slit through which they passed. For electrons in the first group it is completely irrelevant whether the second slit was open or not, and thus their distribution on the screen should be represented by the

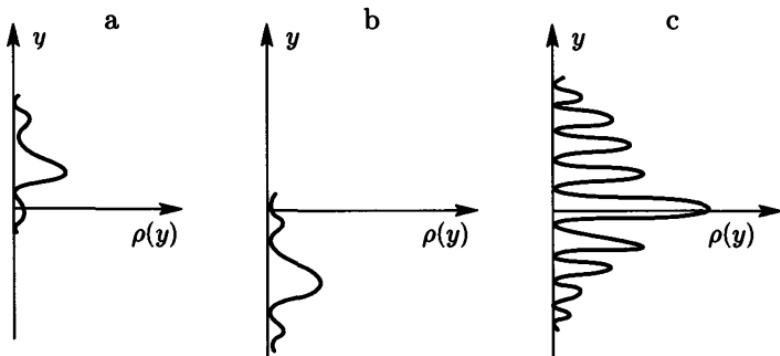


Figure 2

curve *a*; similarly, the electrons of the second group should have the distribution *b*. Therefore, in the case when both slits are open, the screen should show the distribution that is the sum of the distributions *a* and *b*. Such a sum of distributions does not have anything in common with the interference pattern in *c*. This contradiction indicates that under the conditions of the experiment described it is not possible to divide the electrons into groups according to the test of which slit they went through. Hence, we have to reject the concept of trajectory.

The question arises at once as to whether one can set up an experiment to determine the slit through which an electron has passed. Of course, such a formulation of the experiment is possible; for this it suffices to put a source of light between the screens B and C and observe the scattering of the light quanta by the electrons. In order to attain sufficient resolution, we have to use quanta with wavelength of order not exceeding the distance between the slits, that is, with sufficiently large energy and momentum. Observing the quanta scattered by the electrons, we are indeed able to determine the slit through which an electron passed. However, the interaction of the quanta with the electrons causes an uncontrollable change in their momenta, and consequently the distribution of the electrons hitting the screen must change. Thus, we arrive at the conclusion that we can answer the question as to which slit the electron passed through only at the cost of changing both the conditions and the final result of the experiment.

In this example we encounter the following general peculiarity in the behavior of quantum systems. The experimenter does not have the possibility of following the course of the experiment, since to do so would lead to a change in the final result. This peculiarity of quantum behavior is closely related to peculiarities of measurements in the microworld. Every measurement is possible only through an interaction of the system with the measuring device. This interaction leads to a perturbation of the motion of the system. In classical physics one always assumes that this perturbation can be made arbitrarily small, as is the case for the duration of the measurement. Therefore, the

simultaneous measurement of any number of observables is always possible.

A detailed analysis (which can be found in many quantum mechanics textbooks) of the process of measuring certain observables for microsystems shows that an increase in the precision of a measurement of observables leads to a greater effect on the system, and the measurement introduces uncontrollable changes in the numerical values of some of the other observables. This leads to the fact that a simultaneous precise measurement of certain observables becomes impossible in principle. For example, if one uses the scattering of light quanta to measure the coordinates of a particle, then the error of the measurement has the order of the wavelength of the light: $\Delta x \sim \lambda$. It is possible to increase the accuracy of the measurement by choosing quanta with a smaller wavelength, but then with a greater momentum $p = 2\pi\hbar/\lambda$. Here an uncontrollable change Δp of the order of the momentum of the quantum is introduced in the numerical values of the momentum of the particle. Therefore, the errors Δx and Δp in the measurements of the coordinate and the momentum are connected by the relation

$$\Delta x \Delta p \sim 2\pi\hbar.$$

A more precise argument shows that this relation connects only a coordinate and the momentum projection with the same index. The relations connecting the theoretically possible accuracy of a simultaneous measurement of two observables are called the Heisenberg uncertainty relations. They will be obtained in a precise formulation in the sections to follow. Observables on which the uncertainty relations do not impose any restrictions are simultaneously measurable. We shall see that the Cartesian coordinates of a particle are simultaneously measurable, as are the projections of its momentum, but this is not so for a coordinate and a momentum projection with the same index, nor for two Cartesian projections of the angular momentum. In the construction of quantum mechanics we must remember the possibility of the existence of quantities that are not simultaneously measurable.

After our little physical preamble we now try to answer the question posed above. what features of classical mechanics should be kept, and which ones is it natural to reject in the construction of a mechanics of the microworld? The main concepts of classical mechanics were the concepts of an observable and a state. The problem of a physical theory is to predict results of experiments, and an experiment is always a measurement of some characteristic of the system or an observable under definite conditions which determine the state of the system. Therefore, the concepts of an observable and a state must appear in any physical theory. From the point of view of the experimenter, to determine an observable means to specify a way of measuring it. We denote observables by the symbols a, b, c, \dots , and for the present we do not make any assumptions about their mathematical nature (we recall that in classical mechanics the observables are functions on the phase space). As before, we denote the set of observables by \mathfrak{A} .

It is reasonable to assume that the conditions of the experiment determine at least the probability distributions of the results of measurements of all the observables, and therefore it is reasonable to keep the definition in § 2 of a state. The states will be denoted by ω as before, the probability measure on the real axis corresponding to an observable a by $\omega_a(E)$, the distribution function of a in the state ω by $\omega_a(\lambda)$, and finally, the mean value of a in the state ω by $\langle \omega | a \rangle$.

The theory must contain a definition of a function of an observable. For the experimenter, the assertion that an observable b is a function of an observable a ($b = f(a)$) means that to measure b it suffices to measure a , and if the number a_0 is obtained as a result of measuring a , then the numerical value of the observable b is $b_0 = f(a_0)$. For the probability measures corresponding to a and $f(a)$, we have

$$(4) \quad \omega_{f(a)}(E) = \omega_a(f^{-1}(E))$$

for any states ω .

We note that all possible functions of a single observable a are simultaneously measurable, since to measure these observables it suffices to measure a . Below we shall see that in quantum mechanics this

example exhausts the cases of simultaneous measurement of observables; that is, if the observables b_1, b_2, \dots are simultaneously measurable, then there exist an observable a and functions f_1, f_2, \dots such that $b_1 = f_1(a), b_2 = f_2(a), \dots$

The set of functions $f(a)$ of an observable a obviously includes $f(a) = \lambda a$ and $f(a) = \text{const}$, where λ is a real number. The existence of the first of these functions shows that observables can be multiplied by real numbers. The assertion that an observable is a constant means that its numerical value in any state coincides with this constant.

We now try to make clear what meaning can be assigned to a sum $a + b$ and product ab of two observables. These operations would be defined if we had a definition of a function $f(a, b)$ of two variables. However, there arise fundamental difficulties here connected with the possibility of observables that are not simultaneously measurable. If a and b are simultaneously measurable, then the definition of $f(a, b)$ is completely analogous to the definition of $f(a)$. To measure the observable $f(a, b)$, it suffices to measure the observables a and b leading to the numerical value $f(a_0, b_0)$, where a_0 and b_0 are the numerical values of the observables a and b , respectively. For the case of observables a and b that are not simultaneously measurable, there is no reasonable definition of the function $f(a, b)$. This circumstance forces us to reject the assumption that observables are functions $f(q, p)$ on the phase space, since we have a physical basis for regarding q and p as not simultaneously measurable, and we shall have to look for observables among mathematical objects of a different nature.

We see that it is possible to define a sum $a + b$ and a product ab using the concept of a function of two observables only in the case when they are simultaneously measurable. However, another approach is possible for introducing a sum in the general case. We know that all information about states and observables is obtained by measurements; therefore, it is reasonable to assume that there are sufficiently many states to distinguish observables, and similarly, that there are sufficiently many observables to distinguish states.

More precisely, we assume that if

$$\langle a | \omega \rangle = \langle b | \omega \rangle$$

for all states ω , then the observables a and b coincide,⁹ and if

$$\langle a | \omega_1 \rangle = \langle a | \omega_2 \rangle$$

for all observables a , then the states ω_1 and ω_2 coincide.

The first of these assumptions makes it possible to define the sum $a + b$ as the observable such that

$$(5) \quad \langle a + b | \omega \rangle = \langle a | \omega \rangle + \langle b | \omega \rangle$$

for any state ω . We remark at once that this equality is an expression of a well-known theorem in probability theory about the mean value of a sum only in the case when the observables a and b have a common distribution function. Such a common distribution function can exist (and in quantum mechanics does exist) only for simultaneously measurable quantities. In this case the definition (5) of a sum coincides with the earlier definition. An analogous definition of a product is not possible, since the mean of a product is not equal to the product of the means, not even for simultaneously measurable observables.

The definition (5) of the sum does not contain any indication of a way to measure the observable $a + b$ involving known ways of measuring the observables a and b , and is in this sense implicit.

To give an idea of how much the concept of a sum of observables can differ from the usual concept of a sum of random variables, we present an example of an observable that will be studied in detail in what follows. Let

$$H = \frac{P^2}{2} + \frac{\omega^2 Q^2}{2}.$$

The observable H (the energy of a one-dimensional harmonic oscillator) is the sum of two observables proportional to the squares of the momentum and the coordinate. We shall see that these last observables can take any nonnegative numerical values, while the values of the observable H must coincide with the numbers $E_n = (n + 1/2)\omega$, where $n = 0, 1, 2, \dots$: that is, the observable H with discrete numerical values is a sum of observables with continuous values.

Thus, two operations are defined on the set \mathfrak{A} of observables: multiplication by real numbers and addition, and \mathfrak{A} thereby becomes

⁹This assumption allows us to regard an observable as specified if a real number is associated with each state

a linear space. Since real functions are defined on \mathfrak{A} , and in particular the square of an observable, there arises a natural definition of a product of observables:

$$(6) \quad a \circ b = \frac{(a+b)^2 - (a-b)^2}{4}.$$

We note that the product $a \circ b$ is commutative, but it is not associative in general. The introduction of the product $a \circ b$ turns the set \mathfrak{A} of observables into a real commutative algebra.

Recall that the algebra of observables in classical mechanics also contained a Lie operation—the Poisson bracket $\{f, g\}$. This operation appeared in connection with the dynamics of the system. With the introduction of such an operation each observable H generates a family of automorphisms of the algebra of observables:

$$U_t : \mathfrak{A} \rightarrow \mathfrak{A},$$

where $U_t f = f_t$: and f_t satisfies the equation

$$\frac{df_t}{dt} = \{H, f_t\}$$

and the initial condition

$$f_t|_{t=0} = f.$$

We recall that the mapping U_t is an automorphism in view of the fact that the Poisson bracket has the properties of a Lie operation. The fact that the observables in classical mechanics are functions on phase space does not play a role here. We assume that the algebra of observables in quantum mechanics also has a Lie operation; that is, associated with each pair of observables a, b is an observable $\{a, b\}$ with the properties

$$\begin{aligned} \{a, b\} &= -\{b, a\}, \\ \{\lambda a + b, c\} &= \lambda\{a, c\} + \{b, c\}, \\ \{a, b \circ c\} &= \{a, b\} \circ c + b \circ \{a, c\}, \\ \{a, \{b, c\}\} + \{b, \{c, a\}\} + \{c, \{a, b\}\} &= 0. \end{aligned}$$

Moreover, we assume that the connection of the Lie operation with the dynamics in quantum mechanics is the same as in classical mechanics. It is difficult to imagine a simpler and more beautiful way of describing the dynamics. Moreover, the same type of description of the dynamics

in classical and quantum mechanics allows us to hope that we can construct a theory that contains classical mechanics as a limiting case.

In fact, all our assumptions reduce to saying that in the construction of quantum mechanics it is reasonable to preserve the structure of the algebra of observables in classical mechanics, but to reject the realization of this algebra as functions on the phase space since we admit the existence of observables that are not simultaneously measurable.

Our immediate problem is to see that there is a realization of the algebra of observables that is different from the realization in classical mechanics. In the next section we present an example of such a realization, constructing a finite-dimensional model of quantum mechanics. In this model the algebra \mathfrak{A} of observables is the algebra of self-adjoint operators on the n -dimensional complex space \mathbf{C}^n . In studying this simplified model, we can follow the basic features of quantum theory. At the same time, after giving a physical interpretation of the model constructed, we shall see that it is too poor to correspond to reality. Therefore, the finite-dimensional model cannot be regarded as a definitive variant of quantum mechanics. However, it will seem very natural to improve this model by replacing \mathbf{C}^n by a complex Hilbert space.

§ 5. A finite-dimensional model of quantum mechanics

We show that the algebra \mathfrak{A} of observables can be realized as the algebra of self-adjoint operators on the finite-dimensional complex space \mathbf{C}^n .

Vectors in \mathbf{C}^n will be denoted by the Greek letters $\xi, \eta, \varphi, \psi, \dots$. Let us recall the basic properties of a scalar product:

- (1) $(\xi, \psi) = \overline{(\psi, \xi)},$
- (2) $(\xi + \lambda\eta, \psi) = (\xi, \psi) + \lambda(\eta, \psi),$
- (3) $(\xi, \xi) > 0 \quad \text{if } \xi \neq 0.$

Here λ is a complex number.

Vectors e_1, \dots, e_n form an orthonormal¹⁰ basis in \mathbf{C}^n if

$$(2) \quad (e_i, e_j) = \delta_{ij},$$

where δ_{ij} is the Kronecker delta symbol.

The expansion of an arbitrary vector ξ in the vectors of the basis e_1, \dots, e_n has the form

$$(3) \quad \xi = \sum_{i=1}^n \xi_i e_i, \quad \xi_i = (\xi, e_i).$$

The vector ξ is uniquely determined by the numbers ξ_1, \dots, ξ_n :

$$\xi \leftrightarrow (\xi_1, \dots, \xi_n).$$

If one chooses a basis, then one thereby chooses a concrete realization for vectors, or one specifies a representation. Let e_1, \dots, e_n be a basis. Then the vectors

$$(4) \quad e'_i = \sum_{k=1}^n U_{ik} e_k, \quad i = 1, 2, \dots, n,$$

also form a basis if the matrix $U = \{U_{ik}\}$ is invertible and

$$(5) \quad U_{ik}^{-1} = \bar{U}_{ki}.$$

A matrix whose elements satisfy the equality (5) is said to be unitary. The passage from one basis to another is a unitary transformation.

If a representation has been chosen, then the scalar product is given by the formula

$$(6) \quad (\xi, \eta) = \sum_{i=1}^n \xi_i \eta_i.$$

In the given basis, operators are represented by matrices:

$$(7) \quad A \leftrightarrow \{A_{ik}\}, \quad A_{ik} = (Ae_k, e_i).$$

Indeed, if $\eta = A\xi$, then

$$\eta_i = (A\xi, e_i) = \left(A \sum_{k=1}^n \xi_k e_k, e_i \right) = \sum_{k=1}^n \xi_k (Ae_k, e_i) = \sum_{k=1}^n A_{ik} \xi_k.$$

¹⁰In what follows, the word orthonormal will often be omitted, since we do not consider other bases

An operator A^* is called the adjoint of A if

$$(8) \quad (\Lambda\xi, \eta) = \xi(\Lambda^*, \eta)$$

for any pair of vectors ξ and η . Obviously, $\Lambda_{ik}^* = \overline{\Lambda}_{ki}$. An operator is said to be self-adjoint if $A^* = A$. For a self-adjoint operator, $\Lambda_{ik} = \overline{\Lambda}_{ki}$. It follows immediately from the definition of the adjoint of an operator that

$$(9) \quad (AB)^* = B^*A^*, \quad (\alpha A)^* = \alpha A^*,$$

where α is a complex number.

We now construct a realization of the algebra of observables. Let \mathfrak{A} be the set of self-adjoint operators on \mathbf{C}^n . In what follows, self-adjoint operators will often be called observables. The operations of addition and multiplication by a number are defined on the set of operators in the usual way. If $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, and $\lambda \in \mathbf{R}$, then $(A+B) \in \mathfrak{A}$ and $\lambda A \in \mathfrak{A}$, because $(A+B)^* = A+B$ and $(\lambda A)^* = \lambda A$. It is natural to regard these operations as operations of addition of observables and multiplication by a number.

Our next problem is to find out how to construct functions of observables. It can be assumed here that the usual definition of a function of an operator works. We can justify this assumption after we have learned how to construct probability distributions for observables in quantum mechanics. Then we shall be able to verify the formula $\omega_{f(A)}(E) = \omega_A(f^{-1}(E))$, which is equivalent to the general definition of a function of an observable as given in the preceding section.

We recall that there are several equivalent definitions of a function of an operator. If $f(x)$ has a power series expansion

$$(10) \quad f(x) = \sum_{n=0}^{\infty} c_n x^n$$

valid on the whole real axis, then $f(A)$ is defined by the formula

$$(11) \quad f(A) = \sum_{n=0}^{\infty} c_n A^n.$$

A second definition uses the existence for self-adjoint operators of an eigenvector basis.¹¹

$$A\varphi_i = a_i \varphi_i, \quad i = 1, \dots, n.$$

Here the φ_i are eigenvectors with $(\varphi_i, \varphi_j) = \delta_{ij}$, and the a_i are eigenvalues of the operator A . To define a linear operator $f(A)$ it suffices to define the result of the action of $f(A)$ on the vectors of a basis. By definition,

$$(12) \quad f(A) \varphi_i = f(a_i) \varphi_i.$$

In an eigenvector basis the matrix A is diagonal with the eigenvalues on the diagonal, that is, $A_{ij} = a_i \delta_{ij}$. In the same representation, $[f(A)]_{ij} = f(a_i) \delta_{ij}$. We remark that a self-adjoint operator corresponds to a real function, that is, $f(A) \in \mathfrak{A}$.

The operation $A \circ B$ is defined by the formula (4.6), which for self-adjoint operators has the form

$$(13) \quad A \circ B = \frac{(A + B)^2 - (A - B)^2}{4} = \frac{AB + BA}{2}.$$

The self-adjointness of the operator $A \circ B$ is obvious.

It remains for us to construct a Lie operation. To this end, we consider the commutator $[A, B] = AB - BA$ of operators A and B . The operation $[A, B]$ has the following properties:

- 1) $[A, B] = -[B, A]$,
- 2) $[A + \lambda B, C] = [A, C] + \lambda[B, C]$,
- (14) 3) $[A, B \circ C] = [A, B] \circ C + B \circ [A, C]$,
- 4) $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$.

All these properties can be verified directly. We remark that the property 3) is valid also for the nonsymmetrized product. Indeed,

$$[A, BC] = ABC - BCA + BAC - BAC = [A, B]C + B[A, C].$$

We see that the commutator has the properties of a Lie operation, but $[A, B]$ is not self-adjoint, that is, $[A, B] \notin \mathfrak{A}$. However, the expression

¹¹We recall that the eigenvalues of a self-adjoint operator are real and that eigenvectors corresponding to different eigenvalues are orthogonal. If an eigenvalue has multiplicity r , then there are r linearly independent eigenvectors corresponding to it, and they can always be chosen to be orthonormal.

$(i/h)[A, B]$, which differs from the commutator only by the imaginary factor i/h , does satisfy all the requirements. We remark that algebras \mathfrak{A} constructed with different constants h are not isomorphic. The choice of h can be made only after comparing the theory with experiment. Such a comparison shows that h coincides with Planck's constant. In what follows we use the notation

$$(15) \quad \{A, B\}_h = \frac{i}{h}[A, B]$$

and call $\{A, B\}_h$ the quantum Poisson bracket.

It is interesting to note that in classical mechanics we could have replaced the definition

$$\{f, g\} = \frac{\partial f}{\partial p} \frac{\partial g}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial g}{\partial p}$$

of the Poisson bracket by the definition

$$\{f, g\} = \alpha \left(\frac{\partial f}{\partial p} \frac{\partial g}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial g}{\partial p} \right),$$

where α is a real constant. It is not hard to see, however, that in classical mechanics this new definition of the Poisson bracket leads to an algebra that is isomorphic to the original algebra. Indeed, the change of variables $p = \sqrt{|\alpha|}p'$, $q = \operatorname{sgn} \alpha \sqrt{|\alpha|}q'$ returns us to the old definition.

An important role in quantum mechanics is played by the trace $\operatorname{Tr} A$ of an operator. By definition,

$$\operatorname{Tr} A = \sum_{i=1}^n A_{ii} = \sum_{i=1}^n (Ae_i, e_i).$$

We recall the basic properties of this operation. The trace does not depend on the choice of basis. In particular, if we take an eigenvector basis for an operator A , then

$$\operatorname{Tr} A = \sum_{i=1}^n a_i;$$

that is, the trace is the sum of the eigenvalues. If a_i is a multiple eigenvalue, then it appears as a term in the sum as many times as its multiplicity.

The trace of the product of two operators does not depend on the order of the factors:

$$\mathrm{Tr} AB = \mathrm{Tr} BA.$$

In the case of several factors, cyclic permutations can be taken under the sign of Tr .

$$\mathrm{Tr} ABC = \mathrm{Tr} BCA.$$

The trace of a self-adjoint matrix is a real number, since its eigenvalues are real.

§ 6. States in quantum mechanics

In this section we show how states are given in quantum mechanics. We recall that we kept the definition in § 2 of a state. There we also showed that specifying the probability distributions is equivalent to specifying the mean values for all observables. The arguments that led us to this result remain in force also in quantum mechanics since they did not use a concrete realization of the algebra of observables of classical mechanics. To specify a state means to specify on the algebra \mathfrak{A} of observables a functional $\langle \omega | A \rangle$ with the following properties:

- (1) 1) $\langle \lambda A + B | \omega \rangle = \lambda \langle A | \omega \rangle + \langle B | \omega \rangle,$
- 2) $\langle A^2 | \omega \rangle \geq 0,$
- 3) $\langle C | \omega \rangle = C,$
- 4) $\overline{\langle A | \omega \rangle} = \langle A | \omega \rangle.$

We have already discussed the property 1). The property 2) expresses the fact that for the observable A^2 , which has the sense of being non-negative, the mean value cannot be a negative number. The property 3) asserts that the mean value of the constant observable C in any state coincides with this constant. Finally, the property 4) says that the mean values are real. Thus, a state in quantum mechanics is a positive linear functional on the algebra \mathfrak{A} of self-adjoint operators. The general form of such a functional is

$$(2) \quad \langle A | \omega \rangle = \mathrm{Tr} MA,$$

where M is an operator on \mathbf{C}^n satisfying the conditions

- (3) 1) $M^* = M$,
- 2) $(M\xi, \xi) \geq 0$,
- 3) $\text{Tr } M = 1$.

The operator M is called the density matrix, and in quantum mechanics it plays the same role as the distribution function $\rho(q, p)$ does in classical mechanics.

We show that the properties of the density matrix follow from the properties formulated above of the averaging functional $\langle A | \omega \rangle$. Indeed, since the functional $\langle A | \omega \rangle$ is real,

$$\begin{aligned}\overline{\text{Tr } AM} &= \sum_{i,k=1}^n \overline{A_{ik} M_{ki}} = \sum_{i,k=1}^n A_{ki}^* M_{ik}^* \\ &= \sum_{i,k=1}^n A_{ki} M_{ik}^* = \text{Tr } AM^* = \text{Tr } AM.\end{aligned}$$

Setting $X = A_1 + iA_2$, where A_1 and A_2 are self-adjoint operators, we get from the last equality that

$$\text{Tr } XM = \text{Tr } XM^*,$$

where X is an arbitrary operator on \mathbf{C}^n . The property 1) follows at once from the arbitrariness of the operator X .

We now use the positivity of the functional $\langle A | \omega \rangle$:

$$\text{Tr } A^2 M \geq 0.$$

Let $A = P_\eta$, where P_η is the operator of projection on a normalized vector η ($\|\eta\| = 1$):

$$P_\eta \xi = (\xi, \eta) \eta.$$

To compute the trace it is convenient to use any basis in which the first basis vector is η ($e_1 = \eta, e_2, \dots, e_n$), so that

$$\text{Tr } P_\eta^2 M = \text{Tr } P_\eta M = (M\eta, \eta) \geq 0.$$

We see that the positivity of the density matrix is a necessary condition for the positivity of the functional $\langle A | \omega \rangle$. The sufficiency is

verified as follows.

$$\mathrm{Tr} A^2 M = \mathrm{Tr} A M A = \sum_{i=1}^n (A M A e_i, e_i) = \sum_{i=1}^n (M A e_i, A e_i) \geq 0,$$

since each term under the summation sign is nonnegative

Finally, the normalization condition $\mathrm{Tr} M = 1$ of the density matrix follows at once from the property 3) of the functional.

Thus, we have showed that states in quantum mechanics are described by positive self-adjoint operators with trace 1. (Recall that a state in classical mechanics is given by a nonnegative function $\rho(q, p)$ normalized by the condition $\int \rho(q, p) dq dp = 1$.)

Any operator M with the properties (3) describes some state of the system; that is, there is a one-to-one correspondence between the set of states and the set of density matrices M .

$$\omega \leftrightarrow M.$$

If M_1 and M_2 are operators with the properties (3), then it is obvious that a convex combination

$$M = \alpha M_1 + (1 - \alpha) M_2, \quad 0 < \alpha < 1,$$

of them also has these properties and hence corresponds to some state

$$\omega = \alpha \omega_1 + (1 - \alpha) \omega_2.$$

We see that the set of states in quantum mechanics is convex.

All the requirements of the density matrix are satisfied by a one-dimensional projection P_ψ ($\|\psi\| = 1$). Indeed,

- 1) $(P_\psi \xi, \eta) = (\xi, \psi)(\psi, \eta) = \overline{(\eta, \psi)}(\xi, \psi) = (\xi, P_\psi \eta),$
- 2) $(P_\psi \xi, \xi) = (\xi, \psi)(\psi, \xi) = |(\xi, \psi)|^2 \geq 0,$
- 3) $\mathrm{Tr} P_\psi = (\psi, \psi) = 1.$

At the end of the section we show that the state P_ψ cannot be decomposed into a convex combination of different states; that is, it is a pure state (Recall that in classical mechanics a pure state has a distribution function of the form $\rho(q, p) = \delta(q - q_0)\delta(p - p_0)$.) We note that a pure state is determined by the specification of a vector ψ , but there is not a one-to-one correspondence between pure states and vectors, because if ψ' differs from ψ by a numerical factor with

modulus 1, then $P_{\psi'} = P_\psi$. Thus, corresponding to a pure state is a class of vectors with norm 1 that differs from each other by a factor $e^{i\alpha}$ with $\alpha \in \mathbf{R}$.

The vector ψ is usually called a state vector, and the space in which the self-adjoint operators (observables) act is called the state space. For the present we take the space \mathbf{C}^n as the state space.

We show that any state in quantum mechanics is a convex combination of pure states. To this end, we observe that M , like any self-adjoint operator, has an eigenvector basis:

$$M\varphi_i = \mu_i \varphi_i.$$

Let us expand an arbitrary vector ξ in this basis and act on it by the operator M :

$$\begin{aligned}\xi &= \sum_{i=1}^n (\xi, \varphi_i) \varphi_i, \\ M\xi &= \sum_{i=1}^n \mu_i (\xi, \varphi_i) \varphi_i = \sum_{i=1}^n \mu_i P_{\varphi_i} \xi.\end{aligned}$$

Since ξ is arbitrary,

$$(4) \quad M = \sum_{i=1}^n \mu_i P_{\varphi_i}.$$

All the numbers μ_i are nonnegative since they are eigenvalues of a positive operator. Moreover, $\sum_{i=1}^n \mu_i = \text{Tr } M = 1$, and hence the state M is indeed a convex combination of the pure states P_{φ_i} . For the mean value of an observable in a pure state $\omega \leftrightarrow P_\psi$ the formula (2) takes the form

$$(5) \quad \langle A | \omega \rangle = (A\psi, \psi), \quad \|\psi\| = 1.$$

For the case of a mixed state $\omega \leftrightarrow M = \sum_{i=1}^n \mu_i P_{\varphi_i}$, we have

$$(6) \quad \langle A | \omega \rangle = \sum_{i=1}^n \mu_i (A\varphi_i, \varphi_i).$$

This formula shows that, as in classical mechanics, the assertion that a system is in a mixed state $M = \sum_{i=1}^n \mu_i P_{\varphi_i}$ is equivalent to the assertion that the system is in the pure state P_{φ_i} with probability μ_i for $i = 1, \dots, n$.

In concluding this section, we show that the state described by the density matrix $M = P_\psi$ is pure. We have to show that the equality

$$(7) \quad P_\psi = \alpha M_1 + (1 - \alpha) M_2, \quad 0 < \alpha < 1,$$

implies that $M_1 = M_2 = P_\psi$.

For the proof we use the fact that for a positive operator A and any vectors φ and ψ

$$(8) \quad |(A\varphi, \psi)|^2 \leq (A\varphi, \varphi)(A\psi, \psi).$$

It follows from (8) that $A\varphi = 0$ if $(A\varphi, \varphi) = 0$. (The inequality (8) is a condition for the positivity of a Hermitian form: $(A(a\varphi + b\psi), a\varphi + b\psi) = (A\varphi, \varphi)a\bar{a} + (A\varphi, \psi)a\bar{b} + (A\psi, \varphi)\bar{a}b + (A\psi, \psi)b\bar{b} \geq 0$, where a and b are complex numbers.)

Let \mathcal{H}_1 be the subspace orthogonal to the vector ψ . Then $P_\psi\varphi = 0$ if $\varphi \in \mathcal{H}_1$. Using (7) and the positivity of the operators M_1 and M_2 , we have

$$0 \leq \alpha(M_1\varphi, \varphi) \leq \alpha(M_1\varphi, \varphi) + (1 - \alpha)(M_2\varphi, \varphi) = (P_\psi\varphi, \varphi) = 0;$$

that is, $M_1\varphi = 0$. Since M_1 is self-adjoint, we get for any vector ξ that

$$(M_1\varphi, \xi) = (\varphi, M_1\xi) = 0, \quad \varphi \in \mathcal{H}_1,$$

hence, $M_1\xi = C_\xi\psi$, and in particular, $M_1\psi = C_\psi\psi$. Here C_ψ is a constant depending on the vector ξ . An arbitrary vector ξ can be represented in the form

$$\xi = (\xi, \psi)\psi + \varphi, \quad \varphi \in \mathcal{H}_1,$$

hence, $M_1\xi = C_\psi P_\psi\xi$, that is, $M_1 = C_\psi P_\psi$. Finally, we get from the condition $\text{Tr } M_1 = 1$ that $C_\psi = 1$, and thus $M_1 = P_\psi$ and $M_2 = P_\psi$.

The converse can also be proved. If a state is pure, then there exists a vector $\psi \in \mathbf{C}^n$ with $\|\psi\| = 1$ such that $M = P_\psi$.

§ 7. Heisenberg uncertainty relations

In this section we show how the uncertainty relations mentioned in § 4 follow from the mathematical apparatus of quantum mechanics, and we give them a precise formulation.

Let us consider the variance of two observables A and B in a state ω . We recall that the variance of an observable is defined by

$$\Delta_\omega^2 A = \langle \omega | (A - A_{\text{avg}})^2 \rangle = \langle \omega | A^2 \rangle - \langle \omega | A \rangle^2,$$

where $A_{\text{avg}} = \langle \omega | A \rangle$. Let $\sqrt{\Delta_\omega^2 A}$ be denoted by $\Delta_\omega A$.

The uncertainty relations assert that for any state ω

$$(1) \quad \Delta_\omega A \Delta_\omega B \geq \frac{\hbar}{2} |\langle \{A, B\}_h | \omega \rangle|.$$

It suffices to prove the uncertainty relations for pure states. For mixed states they will then follow from the inequality (2.17), which holds also in quantum mechanics, since its derivation does not depend on the realization of the algebra of observables.

We begin with the obvious inequality

$$((A + i\alpha B)\psi, (A + i\alpha B)\psi) \geq 0, \quad \|\psi\| = 1, \quad \alpha \in \mathbf{R}.$$

Expanding the left-hand side, we get that

$$(A^2\psi, \psi) + \alpha^2(B^2\psi, \psi) + i\alpha((AB - BA)\psi, \psi) \geq 0.$$

Using the definition (5.15) of the quantum Poisson bracket and the formula (6.5) for the mean value of an observable in a pure state $\omega \leftrightarrow P_\psi$, we rewrite the last inequality in the form

$$\langle A^2 | \omega \rangle + \alpha^2 \langle B^2 | \omega \rangle - \alpha \hbar \langle \{A, B\}_h | \omega \rangle \geq 0$$

for any $\alpha \in \mathbf{R}$, and hence

$$\langle A^2 | \omega \rangle \langle B^2 | \omega \rangle \geq \frac{\hbar^2}{4} \langle \{A, B\}_h | \omega \rangle^2.$$

The uncertainty relations (1) follow immediately from this inequality if we make the substitutions $A \rightarrow A - A_{\text{avg}}$, $B \rightarrow B - B_{\text{avg}}$ and take into account that $\{A, B\}_h = \{(A - A_{\text{avg}}), (B - B_{\text{avg}})\}_h$.

The uncertainty relations show that the variances of two observables can be simultaneously zero only if the mean value of the Poisson bracket is zero. For commuting observables, $\{A, B\}_h = 0$, and the right-hand side is zero for all states. In this case the uncertainty relations do not impose any restrictions, and such observables are simultaneously measurable. Below we shall see that there really is a theoretically possible way of simultaneously measuring such observables. The strongest restrictions are imposed on the pairs of observables

such that $\langle \{A, B\}_h | \omega \rangle \neq 0$ for all ω , for example, if $\{A, B\}_h = \text{const}$. In this case there do not exist any states in which the variances of both observables are equal to zero. We shall see that for a coordinate and momentum projection with the same index,

$$\{P, Q\}_h = I,$$

where I is the identity operator. The uncertainty relations for these observables take the form

$$\Delta_\omega P \Delta_\omega Q \geq h/2;$$

that is, in nature there are no states in which a coordinate and momentum projection with the same index have simultaneously completely determined values. This assertion is valid for both mixed and pure states, and therefore the pure states in quantum mechanics, unlike those in classical mechanics, are not states with zero variance of all observables.

We can now go back to the question posed at the beginning of § 2: what can explain the nonuniqueness of the results of experiments? We have seen that in classical mechanics such nonuniqueness is connected solely with the conditions of the experiment. If in the conditions of the experiment we include sufficiently many preliminary measurements, then we can be certain that the system is in a pure state, and the results of any experiment are completely determined. In quantum mechanics the answer to this question turns out to be quite different. The uncertainty relations show that there is not even a theoretical possibility of setting up the experiment in such a way that the results of all measurements are uniquely determined by the conditions of the experiment. We are therefore forced to assume that the probabilistic nature of predictions in the microworld is connected with the physical properties of quantum systems.

These conclusions seem so unexpected that one might ask about the possibility of introducing so-called “hidden parameters” into the theory. One might suppose that the description of a state of the system with the help of the density matrix $M = P_\psi$ is not complete;

that is, in addition to P_ψ we should specify the values of some parameters x ("hidden parameters"),¹² and then the description becomes sufficient for the unique prediction of the results of any measurement. The probabilistic character of the predictions in a state $\omega \leftrightarrow P_\psi$ can then be explained by the fact that we do not know the values of the hidden parameters, and there is some probability distribution with respect to them. If the state determined by the pair (P_ψ, x) is denoted by ω_x , then our assumption reduces to the state ω being a convex combination of the states ω_x . We know that P_ψ does not decompose into a convex combination of operators with the properties of a density matrix; that is, there are no operators M corresponding to the states ω_x . The way of describing the states in quantum mechanics is determined by the choice of the algebra of observables. The assumption that there are states that do not correspond to any density matrices forces us to reject the assertion that the observables are self-adjoint operators, that is, to reject the basic assumption of quantum mechanics. Thus, we see that the introduction of hidden parameters into quantum mechanics is impossible without a radical restructuring of its foundations.

§ 8. Physical meaning of the eigenvalues and eigenvectors of observables

In this section we consider questions involving the physical interpretation of the theory. First of all we must learn how to construct distribution functions for observables in a given state. We know the general formula

$$(1) \quad \omega_A(\lambda) = \langle \theta(\lambda - A) | \omega \rangle,$$

where $\theta(x)$ is the Heaviside function. To construct a function of the observable $\theta(\lambda - A)$, we consider the equation

$$A\varphi_i = a_i\varphi_i.$$

¹²The "hidden parameters" x can be regarded as elements of some set X . We do not make any assumptions about the physical nature of these parameters, since it is irrelevant for the arguments to follow.

Just for simplicity of the arguments we assume for the present that the eigenvalues are distinct; that is, the spectrum of the operator A is simple, and we number the eigenvalues in increasing order: $a_1 < \dots < a_n$. Denote by P_{φ_i} the operators of projection on the eigenvectors. We introduce the operator

$$(2) \quad P_A(\lambda) = \sum_{a_i \leq \lambda} P_{\varphi_i}$$

(the subscript i under the summation sign takes the values such that $a_i \leq \lambda$). We show that

$$(3) \quad P_A(\lambda) = \theta(\lambda - A).$$

To prove this equality it suffices to show that the operators $P_A(\lambda)$ and $\theta(\lambda - A)$ act in the same way on the basis vectors φ_i . Using the definition of a function of an operator, we have

$$\theta(\lambda - A) \varphi_j = \theta(\lambda - a_j) \varphi_j = \begin{cases} \varphi_j, & \lambda \geq a_j, \\ 0, & \lambda < a_j. \end{cases}$$

On the other hand,

$$P_A(\lambda) \varphi_j = \sum_{a_i \leq \lambda} P_{\varphi_i} \varphi_j = \begin{cases} \varphi_j, & \lambda \geq a_j, \\ 0, & \lambda < a_j. \end{cases}$$

The last equality was written taking into account that $P_{\varphi_i} \varphi_j = \delta_{ij} \varphi_j$ and that the operator P_{φ_j} appears under the summation sign only if $\lambda \geq a_j$. It is easy to see that $P_A(\lambda)$ is a projection, that is, $P_A^*(\lambda) = P_A(\lambda)$ and $P_A^2(\lambda) = P_A(\lambda)$. Obviously, $P_A(\lambda) = 0$ for $\lambda < a_1$ and $P_A(\lambda) = 1$ for $\lambda \geq a_n$. The operator $P_A(\lambda)$ is called the spectral function of the operator A .

Now it is not hard to get an explicit form for the distribution function $\omega_A(\lambda)$, $\omega \leftrightarrow M$:

$$\omega_A(\lambda) = \langle P_A(\lambda) | \omega \rangle = \text{Tr } M P_A(\lambda) = \text{Tr } M \sum_{a_i \leq \lambda} P_{\varphi_i} = \sum_{a_i \leq \lambda} \text{Tr } M P_{\varphi_i},$$

and finally,

$$(4) \quad \omega_A(\lambda) = \sum_{a_i \leq \lambda} (M \varphi_i, \varphi_i).$$

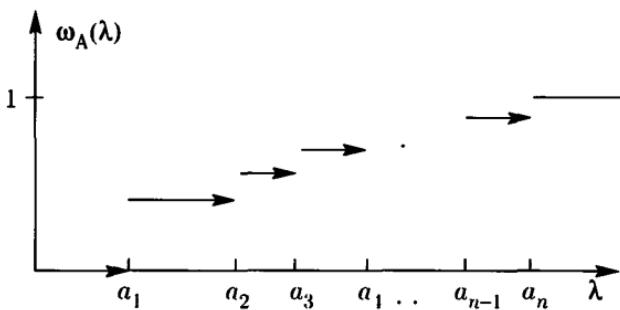


Figure 3

Recall that $(M\varphi_i, \varphi_i) \geq 0$. We see that $\omega_A(\lambda)$ is a piecewise constant function having jumps at the values of λ coinciding with eigenvalues, at which points it is right-continuous. Its graph is pictured in Figure 3

From the form of the distribution function it follows that the probability of obtaining a value of A that coincides only with one of the eigenvalues is nonzero. The probability W_i of getting the value a_i in a measurement is equal to the size of the jump of the function $\omega_A(\lambda)$ at this point:

$$(5) \quad W_i = (M\varphi_i, \varphi_i).$$

The sum of the W_i is 1, as should be expected, since

$$\sum_{i=1}^n W_i = \text{Tr } M = 1.$$

For a pure state $\omega \leftrightarrow P_\psi$, the formula (5) takes the form

$$(6) \quad W_i = |(\psi, \varphi_i)|^2.$$

Finally, if the system is in the state P_{φ_j} , that is, in the pure state determined by one of the eigenvectors of the observable A , then $W_i = \delta_{ij}$ in view of the formula (6). In such a pure state the number a_i is obtained with certainty in a measurement of A .

Thus far we have assumed that the operator A has a simple spectrum. Generalization of the results obtained to the case of a multiple spectrum when several eigenvectors $\varphi_i^{(1)}, \varphi_i^{(2)}, \dots, \varphi_i^{(r)}$ correspond to an eigenvalue a_i is not difficult. It suffices to replace the projections

P_{φ_i} by operators P_i projecting on the eigenspaces corresponding to the eigenvalues a_i :

$$P_i \psi = \sum_{k=1}^r (\psi, \varphi_i^{(k)}) \varphi_i^{(k)}.$$

Then the probability of getting the value a_i in a measurement of the observable A in the general case is given by the formula

$$(7) \quad W_i = \sum_{k=1}^r (M \varphi_i^{(k)}, \varphi_i^{(k)}),$$

and in the case of a pure state

$$(8) \quad W_i = \sum_{k=1}^r |(\psi, \varphi_i^{(k)})|^2.$$

We can now show that the usual definition of a function of an operator agrees with the concept in § 4 of a function of an observable. Indeed, the operators A and $f(A)$ have a common system of eigenvectors

$$\begin{aligned} A \varphi_i^{(k)} &= a_i \varphi_i^{(k)}, \\ f(A) \varphi_i^{(k)} &= f(a_i) \varphi_i^{(k)}, \end{aligned}$$

and the number $\sum_{k=1}^r (M \varphi_i^{(k)}, \varphi_i^{(k)})$ is the probability of getting the value a_i for the observable A in a measurement and at the same the probability of getting the value $f(a_i)$ for the observable $f(A)$. From this it follows at once that

$$\omega_{f(A)}(E) = \omega_A(f^{-1}(E)).$$

We remark also that in the states determined by the eigenvectors of A , the observables A and $f(A)$ simultaneously have completely determined values a_i and $f(a_i)$, respectively.

We can summarize the results connecting the mathematical apparatus of the theory with its physical interpretation as follows

- 1) An observable A in a state $\omega \leftrightarrow M$ has the mean value

$$\langle A | \omega \rangle = \text{Tr } MA$$

and the distribution function

$$\omega_A(\lambda) = \text{Tr } MP_A(\lambda).$$

For a pure state $M = P_\psi$, $\|\psi\| = 1$, we have

$$\begin{aligned}\langle A | \omega \rangle &= (A\psi, \psi), \\ \omega_A(\lambda) &= (P_A(\lambda)\psi, \psi).\end{aligned}$$

- 2) The set of eigenvalues of an observable A coincides with the set of possible results of a measurement of this observable.
- 3) In measuring an observable A , the probability W_i of obtaining a number coinciding with one of its eigenvalues is computed by the formula (7) in the general case and by the formula (8) for pure states.
- 4) The eigenvectors of an observable A determine the pure states, in which the observable with certainty takes a value equal to the corresponding eigenvalue.

We see that the model constructed satisfies many of the physical requirements of the mechanics of the microworld that were formulated in § 4. It admits the existence of observables that are not simultaneously measurable, and it explains the uncertainty relations. Observables having a discrete set of values can be described in a natural way in the framework of this model. On the other hand, we also see the limitations of such a model. Any observable cannot have more than n values, where n is the dimension of the state space \mathbf{C}^n , and hence this model does not let us describe observables with a continuous set of values. Therefore, it is hard to believe that such a model can describe real physical systems. However, these difficulties do not arise if we pass to $n = \infty$ and take the state space to be a complex Hilbert space and the observables to be self-adjoint operators acting in this space. We shall discuss the choice of a state space for concrete physical systems and the rules for constructing observables for them in § 11, but for the time being we continue to study our finite-dimensional model. A finite-dimensional model is convenient for us also in that we do not have to deal with the purely mathematical difficulties encountered in the spectral theory of unbounded self-adjoint operators acting in Hilbert space. We remark that the basic aspects of this theory were worked out by von Neumann in connection with the needs of quantum mechanics.

§ 9. Two pictures of motion in quantum mechanics. The Schrödinger equation. Stationary states

In this section we discuss questions of the dynamics of quantum systems. In essence, the basic assumption was made already in § 4 when we considered the necessity of introducing a Lie operation in the algebra of observables in quantum mechanics and the connection of this with the dynamics. Therefore, we begin directly by postulating the so-called Heisenberg picture, which is an analogue of the classical Hamiltonian picture. Just as in classical mechanics, the observables A depend on time while the states $\omega \leftrightarrow M$ do not:

$$(1) \quad \begin{aligned} \frac{dM}{dt} &= 0, \\ \frac{dA(t)}{dt} &= \{H, A(t)\}_h. \end{aligned}$$

Here H is the total energy operator of the system and is an analogue of the Hamiltonian function of classical mechanics. The operator H is sometimes called the Schrödinger operator. In form, the Heisenberg picture is exactly the same as the Hamiltonian picture, except that the right-hand side contains the quantum Poisson bracket instead of the classical one.

As in classical mechanics, the equation

$$(2) \quad \frac{dA(t)}{dt} = \{H, A(t)\}_h$$

together with the initial condition

$$(3) \quad A(t)|_{t=0} = A$$

gives a one-parameter family of automorphisms of the algebra \mathfrak{A} of observables.

The dependence of the mean value on the time is determined by the formula

$$\langle A(t) | \omega \rangle = \text{Tr } A(t) M.$$

The equation (2) with the initial condition (3) has a unique solution, which can be written in the form

$$(4) \quad A(t) = e^{\frac{i}{\hbar} H t} A e^{-\frac{i}{\hbar} H t}.$$

Indeed,

$$\frac{dA(t)}{dt} = \frac{i}{\hbar} [H, A(t)] = \{H, A(t)\}_h.$$

It is clear that the initial condition is satisfied.

The operator $U(t) = e^{-\frac{i}{\hbar} H t}$ appearing in the solution (4) is called the evolution operator. The evolution operator is a unitary operator in view of the self-adjointness of H :

$$U^*(t) = e^{\frac{i}{\hbar} H t} = U^{-1}(t).$$

The set of operators $U(t)$ forms a one-parameter group:

$$U(t_2) U(t_1) = U(t_1 + t_2), \quad U^{-1}(t) = U(-t).$$

If the observables H and A commute, then $\{H, A\}_h = 0$, and the mean value of A does not depend on the time. Such observables are called quantum integrals of motion.

In quantum mechanics there is a second equivalent picture of the motion which is an analogue of the classical picture of Liouville. To formulate this picture, we transform the formula for the mean value:

$$\begin{aligned} \langle A(t) | \omega \rangle &= \text{Tr } U^*(t) A U(t) M \\ &= \text{Tr } A U(t) M U^*(t) = \text{Tr } A M(t) = (A | \omega(t)), \end{aligned}$$

where

$$(5) \quad \omega(t) \leftrightarrow M(t) = U(t) M U^*(t)$$

and $M(t)$ is the unique solution of the equation

$$(6) \quad \frac{dM(t)}{dt} = -\{H, M(t)\}_h$$

with the initial condition

$$(7) \quad M(t)|_{t=0} = M.$$

We have arrived at the so-called Schrödinger picture. In this picture it turns out that the states depend on the time:

$$\begin{aligned} (8) \quad \frac{dM(t)}{dt} &= -\{H, M(t)\}_h, \\ \frac{dA}{dt} &= 0. \end{aligned}$$

From the very way it was introduced, the Schrödinger picture is equivalent to the Heisenberg picture, since the dependence of the mean values of the observables is the same in these pictures.

Let us now consider the dependence on time of the pure states in the Schrödinger picture. According to the general formula (5),

$$P_\psi(t) = U(t) P_\psi U^*(t).$$

We choose the dependence of the state vector $\psi(t)$ on time in such a way that

$$P_\psi(t) = P_{\psi(t)}.$$

We verify that this condition is satisfied by

$$(9) \quad \psi(t) = U(t) \psi.$$

For an arbitrary vector ξ ,

$$\begin{aligned} P_{\psi(t)} \xi &= (\xi, \psi(t)) \psi(t) = (\xi, U(t) \psi) U(t) \psi \\ &= U(t)(U^*(t)\xi, \psi) \psi = U(t)P_\psi U^*(t) \xi = P_\psi(t) \xi. \end{aligned}$$

Thus, in view of the formula (9), the dependence on time of the vector $\psi(t)$ guarantees the correct dependence on time of the pure state. We note that (9) does not and cannot necessarily follow from (5), since a state vector is determined by the state only up to a factor of modulus 1. In spite of this, it is always assumed in quantum mechanics that the dependence of state vectors on the time is determined by the formula (9). We remark that the evolution does not change the normalization of the state vector; that is,

$$\|\psi(t)\| = \|\psi\|,$$

since $U(t)$ is a unitary operator.

The vector $\psi(t)$ satisfies the equation

$$(10) \quad i\hbar \frac{d\psi(t)}{dt} = H\psi(t)$$

and the initial condition

$$\psi(t)|_{t=0} = \psi.$$

The equation (10) is called the Schrödinger equation and is the basic equation in quantum mechanics.

We now consider states that do not depend on the time in the Schrödinger picture. Such states are called stationary states. Obviously, the mean value of any observable and the probabilities of its definite values in stationary states do not depend on the time. The condition of stationarity follows at once from (6) and has the form

$$(11) \quad \{H, M\}_h = 0.$$

Let us consider pure stationary states. From (11) we have¹³

$$HP_{\psi(t)} = P_{\psi(t)}H.$$

We apply the left- and right-hand sides of this equality to the vector $\psi(t)$.

$$H\psi(t) = (H\psi(t), \psi(t)) \psi(t).$$

The number $E = (H\psi(t), \psi(t)) = \text{Tr } P_{\psi(t)}H$ does not depend on the time, and we see that for any t the vector $\psi(t)$ is an eigenvector of H with eigenvalue E . Therefore, the Schrödinger equation for the vector $\psi(t)$ takes the form

$$ih \frac{d\psi(t)}{dt} = E\psi(t).$$

The solution of this equation is

$$(12) \quad \psi(t) = \psi(0) e^{-\frac{i}{\hbar} Et}.$$

Thus, the pure stationary states are the states with definite energy, and a vector determining such a state depends on the time according to the formula (12).

The equation

$$H\varphi_i = E_i\varphi_i$$

is sometimes called the stationary (or time-independent) Schrödinger equation. The basic problems in quantum mechanics reduce to the solution of this equation. According to the general physical interpretation, the numbers E_i are the possible values of the energy (the energy levels) of the system. The state corresponding to the smallest value E_1 of the energy is called the ground state of the system, and the other states are called excited states. If the eigenvectors φ_i of

¹³We note that the independence of $P_{\psi(t)}$ from the time does not in any way imply that the vector $\psi(t)$, which satisfies the Schrödinger equation, does not depend on the time.

H are known, then it is easy to construct a solution of the Cauchy problem

$$ih \frac{d\psi(t)}{dt} = H\psi(t), \\ \psi(t)|_{t=0} = \psi,$$

for the Schrödinger equation. For this it suffices to expand the vector ψ in the eigenvector basis of the operator H ,

$$\psi = \sum_{i=1}^n c_i \varphi_i, \quad c_i = (\psi, \varphi_i),$$

and to use the formula (9). Then we get the solution of the Cauchy problem in the form

$$(13) \quad \psi(t) = \sum_{i=1}^n c_i \varphi_i e^{-\frac{i}{\hbar} E_i t}.$$

This formula is usually called the formula for the expansion of a solution of the Schrödinger equation in stationary states.

In the conclusion of this section we obtain the time-energy uncertainty relation. Setting $B = H$ in (7.1), we have

$$(14) \quad \Delta_\omega A \Delta_\omega H \geq \frac{\hbar}{2} |(\{A, H\}_h | \omega\rangle)|.$$

Recalling that $\frac{dA}{dt} = \{H, A\}_h$ in the Heisenberg picture and using the obvious equality $\frac{dA_{\text{avg}}}{dt} = (\frac{dA}{dt} | \omega\rangle)$, we rewrite the relation (14) as

$$\Delta_\omega A \Delta_\omega H \geq \frac{\hbar}{2} \left| \frac{dA_{\text{avg}}}{dt} \right|.$$

We introduce the time interval $\Delta_\omega t_A = \Delta_\omega A / |\frac{dA_{\text{avg}}}{dt}|$. In the time $\Delta_\omega t_A$ the mean value of the observable A_{avg} is displaced by the “width” $\Delta_\omega A$ of the distribution. Therefore, for the state ω and the observable A , $\Delta_\omega t_A$ is the characteristic time during which the distribution function $\omega_A(\lambda)$ can change noticeably. It follows from the inequality

$$\Delta_\omega t_A \Delta_\omega H \geq \hbar/2$$

that the set of values of $\Delta_\omega t_A$ for all possible observables A in the state ω is bounded from below.

Setting $\Delta t = \inf_A \Delta_\omega t_A$ and denoting $\Delta_\omega H$ by ΔE , we get that for any state ω .

$$(15) \quad \Delta E \Delta t \geq h/2.$$

This is the time-energy uncertainty relation. The physical meaning of this relation differs essentially from the meaning of the uncertainty relations (7.1). In (7.1), $\Delta_\omega A$ and $\Delta_\omega B$ are the uncertainties in the values of the observables A and B in the state ω at one and the same moment of time. In (15), ΔE is the uncertainty of the energy, and this does not depend on the time, and Δt characterizes the time over which the distribution of at least one of the observables can change noticeably. The smaller the time Δt , the more the state ω differs from a stationary state. For strongly nonstationary states, Δt is small and the uncertainty ΔE of the energy must be sufficiently large. On the other hand, if $\Delta E = 0$, then $\Delta t = \infty$. In this case the state is stationary.

§ 10. Quantum mechanics of real systems. The Heisenberg commutation relations

We have already mentioned that the finite-dimensional model constructed is too poor to correspond to reality and that it can be improved by taking the state space to be a complex Hilbert space \mathcal{H} and the observables to be self-adjoint operators acting in this space.

One can show that the basic formula for the mean value of an observable A in a state ω keeps its form

$$(1) \quad \langle A | \omega \rangle = \text{Tr } AM,$$

where M is a positive self-adjoint operator acting in \mathcal{H} with trace 1.

For any self-adjoint operator there is a spectral function $P_A(\lambda)$, that is, a family of projections with the following properties:

- 1) $P_A(\lambda) \leq P_A(\mu)$ for $\lambda < \mu$, that is, $P_A(\lambda)P_A(\mu) = P_A(\lambda)$;
- 2) $P_A(\lambda)$ is right-continuous, that is, $\lim_{\mu \rightarrow \lambda+0} P_A(\mu) = P_A(\lambda)$;
- 3) $P_A(-\infty) = \lim_{\lambda \rightarrow -\infty} P_A(\lambda) = 0$ and $P_A(\infty) = I$;
- 4) $P_A(\lambda)B = BP_A(\lambda)$ if B is any bounded operator commuting with A .

A vector φ belongs to the domain of the operator A ($\varphi \in D(A)$) if

$$\int_{-\infty}^{\infty} \lambda^2 d(P_A(\lambda) \varphi, \varphi) < \infty,$$

and then

$$(2) \quad A\varphi = \int_{-\infty}^{\infty} \lambda dP_A(\lambda) \varphi.$$

A function $f(A)$ of the operator A is defined by the formula

$$(3) \quad f(A) \varphi = \int_{-\infty}^{\infty} f(\lambda) dP_A(\lambda) \varphi.$$

The domain $D(f(A))$ of this operator is the set of elements φ such that

$$\int_{-\infty}^{\infty} |f(\lambda)|^2 d(P_A(\lambda) \varphi, \varphi) < \infty.$$

The spectrum of a self-adjoint operator is a closed subset of the real axis and consists of all the points of growth of the spectral function $P_A(\lambda)$. The jumps of this function correspond to the eigenvalues of the operator A , and $P_A(\lambda+0) - P_A(\lambda-0)$ is the projection on the eigenspace corresponding to the eigenvalue λ . The eigenvalues form the point spectrum. If the eigenvectors form a complete system, then the operator has a pure point spectrum. In the general case the space can be split into a direct sum of orthogonal and A -invariant subspaces \mathcal{H}_1 and \mathcal{H}_2 such that A acting in the first has a pure point spectrum and acting in the second does not have eigenvectors. The spectrum of the operator acting in the subspace \mathcal{H}_2 is said to be continuous.

It follows from the basic formula (1) that in the general case the distribution function of A in the state $\omega \leftrightarrow M$ has the form

$$(4) \quad \omega_A(\lambda) = \text{Tr } P_A(\lambda) M,$$

while for pure states $M = P_\psi$, $\|\psi\| = 1$, we have

$$(5) \quad \omega_A(\lambda) = (P_A(\lambda) \psi, \psi).$$

Unlike in the finite-dimensional model, the function $\omega_A(\lambda)$ is not necessarily a step function. The set of admissible values of an observable A coincides with the set of points of growth of the function $\omega_A(\lambda)$ for all possible states ω . Therefore, we can assert that the set of possible results of measurements of A coincides with its spectrum. We see

that the theory makes it possible to describe observables with both a discrete and a continuous set of values.

Our problem now is to give rules for choosing the state spaces and to learn how to construct the basic observables for real physical systems. Here we shall describe quantum systems having a classical analogue. The problem is posed as follows. Suppose that we have a classical system, that is, we are given its phase space and Hamiltonian function. We must find a quantum system, that is, construct a state space and a Schrödinger operator, in such a way that a one-to-one relation $f \leftrightarrow A_f$ is established between the classical observables (functions on the phase space) and quantum observables (operators acting in the state space). Furthermore, the Hamiltonian function must correspond to the Schrödinger operator. This one-to-one relation certainly cannot be an isomorphism $fg \leftarrow / \rightarrow A_f \circ A_g$, $\{f, g\} \leftarrow / \rightarrow \{A_f, A_g\}_h$ (it is therefore that quantum mechanics differs from classical mechanics), but it must become an isomorphism as $\hbar \rightarrow 0$ (this ensures that quantum mechanics approaches classical mechanics in the limit). The quantum observables A_f usually have the same names as the classical observables f . We remark that we must not exclude the possibility of the existence of quantum systems that do not have simple classical analogues. For such systems there can be observables that do not correspond to any function of the generalized coordinates and momenta.

The correspondence rules and the approach to the classical mechanics limit will be described at length in § 14. For the present we establish a correspondence only between the most important observables, and we show how to construct the state space for the simplest systems.

Let us first consider a material point. Its phase space is six-dimensional, and a point in it is determined by specifying three Cartesian coordinates q_1, q_2, q_3 and three momentum projections p_1, p_2, p_3 . It is not hard to compute the classical Poisson brackets for any pairs of these observables:

$$\{q_i, q_j\} = 0, \quad \{p_i, p_j\} = 0, \quad \{p_i, q_j\} = \delta_{ij}, \quad i, j = 1, 2, 3.$$

For a particle in quantum mechanics we introduce six observables $Q_1, Q_2, Q_3, P_1, P_2, P_3$ for which the quantum Poisson brackets have the same values:

$$(6) \quad \{Q_i, Q_j\}_h = 0, \quad \{P_i, P_j\}_h = 0, \quad \{P_i, Q_j\}_h = I\delta_{ij}.$$

where I is the identity operator. These observables will be called the coordinates and momentum projections.

Below we shall justify this association of operators with coordinates and momenta by the following circumstances. After studying the observables determined by the relations (6), we shall see that they have many of the properties of classical coordinates and momenta. For example, for free particles the projections of the momentum are integrals of motion, and the mean values of the coordinates depend linearly on the time (uniform rectilinear motion).

Further, the correspondence $q_i \leftrightarrow Q_i, p_i \leftrightarrow P_i, i = 1, 2, 3$, is fundamental for us for the construction of a general correspondence $f(q, p) \leftrightarrow A_f$. The quantum observables defined in this way, including the coordinates and momenta, turn into the corresponding classical observables in the limit.¹⁴ Finally, the correspondence rules $f \leftrightarrow A_f$ make the theory completely concrete, and its results can be verified in experiments. It is the agreement of theory with experiment that should be regarded as the definitive justification of the assumption (6) and of all of quantum mechanics.

The conditions (6) are called the Heisenberg quantization conditions. Together with (7.1) they at once imply the Heisenberg uncertainty relations for the coordinates and momenta. We have already discussed these relations.

The formulas (6) are often written for the commutators:

$$(7) \quad [Q_j, Q_k] = 0, \quad [P_j, P_k] = 0, \quad [Q_j, P_k] = i\hbar\delta_{jk}.$$

These relations are called the Heisenberg commutation relations.

¹⁴The exact meaning of this assertion is discussed in § 14

We now state without proof the remarkable von Neumann–Stone theorem.¹⁵

For the system of relations (7) there exists a unique irreducible representation by operators acting in a Hilbert space (unique up to a unitary transformation).

We recall what an irreducible representation is. One usually uses two equivalent definitions:

1) A representation of the relations (7) is said to be irreducible if there does not exist an operator other than a multiple CI of the identity that commutes with all the operators Q_i and P_i .

2) A representation is said to be irreducible if there does not exist a nontrivial subspace \mathcal{H}_0 of \mathcal{H} that is invariant under all the operators Q_i and P_i .

We verify the equivalence of these definitions. If there is a nontrivial subspace \mathcal{H}_0 invariant under Q_i and P_i , then the projection $P_{\mathcal{H}_0}$ onto this subspace commutes with Q_i and P_i , and obviously $P_{\mathcal{H}_0} \neq CI$.

If there is an operator $A \neq CI$ commuting with all the Q_i and P_i , then A^* and hence also $A + A^*$ commute with Q_i and P_i . Therefore, we can assume from the very beginning that A is self-adjoint. The self-adjoint operator A has a spectral function $P_A(\lambda)$ which for some λ is different from zero and CI . The operator $P_A(\lambda)$ commutes with Q_i and P_i , and thus the subspace \mathcal{H}_0 onto which it projects is invariant with respect to Q_i and P_i .

It follows from the von Neumann–Stone theorem that if we find some representation of the Heisenberg commutation relations and prove that it is irreducible, then all other irreducible representations will differ from it by a unitary transformation.

We know that the physical interpretation of the theory is based on the formula $\langle A | \omega \rangle = \text{Tr } AM$ for the mean value. The right-hand side of this formula does not change under a unitary transformation

¹⁵The formulation given is not precise, since there are subtleties connected with the unboundedness of the operators P and Q that we are not going to discuss. Similar difficulties do not arise if instead of the operators P and Q themselves we consider bounded functions of these operators. We present a precise formulation of the von Neumann–Stone theorem in § 13.

of all the operators. Therefore, the physical results of the theory do not depend on which of the unitarily equivalent representations we choose.

§ 11. Coordinate and momentum representations

We describe the so-called coordinate representation most often used in quantum mechanics. It is sometimes called the Schrödinger representation. As the state space for a material point in this representation, we choose the space $L^2(\mathbf{R}^3)$ of square-integrable complex-valued functions $\varphi(\mathbf{x}) = \varphi(x_1, x_2, x_3)$. The scalar product is defined by

$$(1) \quad (\varphi_1, \varphi_2) = \int_{\mathbf{R}^3} \varphi_1(\mathbf{x}) \overline{\varphi_2(\mathbf{x})} d\mathbf{x}.$$

The operators Q_j and P_j , $j = 1, 2, 3$, are introduced by the formulas

$$(2) \quad \begin{aligned} Q_j \varphi(\mathbf{x}) &= x_j \varphi(\mathbf{x}), \\ P_j \varphi(\mathbf{x}) &= \frac{\hbar}{i} \frac{\partial}{\partial x_j} \varphi(\mathbf{x}). \end{aligned}$$

These are unbounded operators, and this corresponds to the physical meaning of the observables, since the numerical values of the coordinates and momenta are unbounded. The operators Q_j and P_j themselves and all their positive integer powers are defined on the domain D formed by the infinitely differentiable functions that decrease faster than any power. Obviously, D is dense in \mathcal{H} . It can be shown that the operators Q_j and P_j have unique self-adjoint extensions upon closure. It is easy to verify that these operators are symmetric. For example, for P_1 and any functions $\psi, \varphi \in D$ we get by integrating by parts that

$$\begin{aligned} (P_1 \varphi, \psi) &= \int_{\mathbf{R}^3} d\mathbf{x} \frac{\hbar}{i} \frac{\partial \varphi(\mathbf{x})}{\partial x_1} \overline{\psi(\mathbf{x})} = \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_3 \frac{\hbar}{i} \varphi(\mathbf{x}) \overline{\psi(\mathbf{x})} \Big|_{x_1=-\infty}^{x_1=\infty} \\ &\quad - \int_{\mathbf{R}^3} d\mathbf{x} \varphi(\mathbf{x}) \frac{\hbar}{i} \overline{\frac{\partial \psi(\mathbf{x})}{\partial x_1}} = \int_{\mathbf{R}^3} d\mathbf{x} \varphi(\mathbf{x}) \frac{\hbar}{i} \overline{\frac{\partial \psi(\mathbf{x})}{\partial x_1}} = (\varphi, P_1 \psi) \end{aligned}$$

Let us verify the commutation relations (10.7). The first two of them are obvious, and the third follows from the equalities

$$\begin{aligned} P_k Q_j \varphi(\mathbf{x}) &= \frac{\hbar}{i} \frac{\partial}{\partial x_k} (x_j \varphi(\mathbf{x})) = \frac{\hbar}{i} \delta_{jk} \varphi(\mathbf{x}) + x_j \frac{\hbar}{i} \frac{\partial}{\partial x_k} \varphi(\mathbf{x}), \\ Q_j P_k \varphi(\mathbf{x}) &= x_j \frac{\hbar}{i} \frac{\partial}{\partial x_k} \varphi(\mathbf{x}). \end{aligned}$$

Thus, we have indeed constructed a representation of the relations (10.7) by linear operators. The irreducibility of this representation will be proved later.

We present one more example of a representation (the momentum representation). In this representation the state space \mathcal{H} is also the complex space $L^2(\mathbf{R}^3)$ with elements $\varphi(\mathbf{p})$, and the operators Q_j and P_j , $j = 1, 2, 3$, are defined as follows:

$$(3) \quad \begin{aligned} Q_j \varphi(\mathbf{p}) &= i\hbar \frac{\partial}{\partial p_j} \varphi(\mathbf{p}), \\ P_j \varphi(\mathbf{p}) &= p_j \varphi(\mathbf{p}). \end{aligned}$$

The commutation relations (10.7) are verified in the same way as for the coordinate representation, and the proof of irreducibility is the same. By the von Neumann Stone theorem, these representations are unitarily equivalent; that is, there is a unitary transformation¹⁶

$$\varphi(\mathbf{x}) \xrightarrow{W} \varphi(\mathbf{p})$$

such that the operators defined by (3) are transformed into the operators (2). It is not hard to see that the Fourier transformation

$$(4) \quad \varphi(\mathbf{x}) = \left(\frac{1}{2\pi\hbar} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{\frac{i}{\hbar} \mathbf{px}} \varphi(\mathbf{p}) d\mathbf{p},$$

$$(5) \quad \varphi(\mathbf{p}) = \left(\frac{1}{2\pi\hbar} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{-\frac{i}{\hbar} \mathbf{px}} \varphi(\mathbf{x}) d\mathbf{x}$$

is such a transformation.

The unitarity of the Fourier transformation follows from the Parseval equality

$$\int_{\mathbf{R}^3} |\varphi(\mathbf{x})|^2 d\mathbf{x} = \int_{\mathbf{R}^3} |\varphi(\mathbf{p})|^2 d\mathbf{p}.$$

¹⁶We intentionally use one and the same symbol to denote the different functions $\varphi(\mathbf{x})$ and $\varphi(\mathbf{p})$, since both these functions represent one and the same state φ .

Applying the operator $\frac{\hbar}{i} \frac{\partial}{\partial x_j}$ to the expression (4), we get that

$$\frac{\hbar}{i} \frac{\partial}{\partial x_j} \left(\frac{1}{2\pi\hbar} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{\frac{i}{\hbar} \mathbf{px}} \varphi(\mathbf{p}) d\mathbf{p} = \left(\frac{1}{2\pi\hbar} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{\frac{i}{\hbar} \mathbf{px}} p_j \varphi(\mathbf{p}) d\mathbf{p}.$$

We see that under the unitary transformation W , the operator $\frac{\hbar}{i} \frac{\partial}{\partial x_j}$ is transformed into the operator of multiplication by the variable p_j . It can be shown similarly that

$$x_j \varphi(\mathbf{x}) \xrightarrow{W} i\hbar \frac{\partial}{\partial p_j} \varphi(\mathbf{p}).$$

Thus, we have proved the unitary equivalence of the coordinate and the momentum representations.

We write the formulas for the transformation of integral operators in passing from the coordinate to the momentum representation:

$$A\varphi(\mathbf{x}) = \int_{\mathbf{R}^3} A(\mathbf{x}, \mathbf{y}) \varphi(\mathbf{y}) d\mathbf{y},$$

$$\begin{aligned} A\varphi(\mathbf{p}) &= \left(\frac{1}{2\pi\hbar} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{-\frac{i}{\hbar} \mathbf{px}} A\varphi(\mathbf{x}) d\mathbf{x} \\ &= \left(\frac{1}{2\pi\hbar} \right)^3 \int_{\mathbf{R}^9} e^{-\frac{i}{\hbar} \mathbf{px}} A(\mathbf{x}, \mathbf{y}) e^{\frac{i}{\hbar} \mathbf{qy}} \varphi(\mathbf{q}) d\mathbf{x} d\mathbf{y} d\mathbf{q} \\ &= \int_{\mathbf{R}^3} A(\mathbf{p}, \mathbf{q}) \varphi(\mathbf{q}) d\mathbf{q}, \end{aligned}$$

where

$$(6) \quad A(\mathbf{p}, \mathbf{q}) = \left(\frac{1}{2\pi\hbar} \right)^3 \int_{\mathbf{R}^6} e^{-\frac{i}{\hbar} \mathbf{px}} A(\mathbf{x}, \mathbf{y}) e^{\frac{i}{\hbar} \mathbf{qy}} d\mathbf{x} d\mathbf{y}.$$

If an operator in the coordinate representation is the operator of multiplication by the function $f(\mathbf{x})$, then it can be regarded as the integral operator with kernel

$$A(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) \delta(\mathbf{x} - \mathbf{y}).$$

In the momentum representation the kernel of such an operator depends only on the difference of the variables and has the form

$$(7) \quad A(\mathbf{p}, \mathbf{q}) = \left(\frac{1}{2\pi\hbar} \right)^3 \int_{\mathbf{R}^3} f(\mathbf{x}) e^{\frac{i}{\hbar} \mathbf{x}(\mathbf{q} - \mathbf{p})} d\mathbf{x}.$$

The transformation formulas from the momentum representation to the coordinate representation differ by the sign in the exponent:

$$(8) \quad A(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{2\pi\hbar} \right)^3 \int_{\mathbf{R}^6} e^{\frac{i}{\hbar} \mathbf{p}\cdot\mathbf{x}} A(\mathbf{p}, \mathbf{q}) e^{-\frac{i}{\hbar} \mathbf{q}\cdot\mathbf{y}} d\mathbf{p} d\mathbf{q},$$

and if $A(\mathbf{p}, \mathbf{q}) = F(\mathbf{p})\delta(\mathbf{p} - \mathbf{q})$, then

$$(9) \quad A(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{2\pi\hbar} \right)^3 \int_{\mathbf{R}^3} F(\mathbf{p}) e^{\frac{i}{\hbar} \mathbf{p}(\mathbf{x}-\mathbf{y})} d\mathbf{p}.$$

We now explain the physical meaning of the functions $\varphi(\mathbf{x})$ and $\varphi(\mathbf{p})$, which are usually called wave functions. We begin with the wave function in the coordinate representation $\varphi(\mathbf{x})$. In this representation the operators Q_i ($i = 1, 2, 3$) are the operators of multiplication by the variables x_i ; that is, the coordinate representation is an eigenrepresentation for the operators Q_i . To see this we make a little mathematical digression.

To give a function representation of the Hilbert space \mathcal{H} means to specify a one-to-one correspondence between vectors in this space and functions of a real variable, $\varphi \leftrightarrow \varphi(x)$, and to specify a measure $dm(x)$ on the real axis such that

$$(\varphi, \psi) = \int_{\mathbf{R}} \varphi(x) \overline{\psi(x)} dm(x).$$

The values of $\varphi(x)$ can be complex numbers or functions of other variables, that is, they are elements of some additional space, and $\varphi(x)\overline{\psi(x)}$ then denotes the scalar product in this additional space. In the first case the representation is said to be simple, and in the second to be multiple. Two functions $\varphi(x)$ are regarded as the same if they differ only on a set of m -measure zero.

A representation is called a spectral representation for an operator A if the action of this operator reduces to multiplication by some function $f(x)$.

$$A\varphi \leftrightarrow f(x)\varphi(x)$$

A representation is called a direct representation or eigenrepresentation for an operator A if $f(x) = x$:

$$A\varphi \leftrightarrow x\varphi(x)$$

The spectrum of the operator A can be described as the support of the distribution function of the measure. The jumps of the measure correspond to the eigenvalues. If the measure is absolutely continuous, then the spectrum is continuous.

We recall that in the finite-dimensional space \mathbf{C}^n we defined an eigenrepresentation for an operator A to be a representation in which the vectors φ were given by the coefficients φ_i of the expansion in an eigenvector basis of A . Suppose that the spectrum of A is simple:

$$A\psi_i = a_i \psi_i,$$

$$\varphi = \sum_{i=1}^n \varphi_i \psi_i, \quad \varphi_i = (\varphi, \psi_i).$$

With a vector φ one can associate a function $\varphi(x)$ such that $\varphi(a_i) = \varphi_i$ (the values of $\varphi(x)$ for $x \neq a_i$ can be arbitrary). The distribution function of the measure is chosen to be piecewise constant with unit jumps at $x = a_i$. The operator A can then be given by the formula

$$A\varphi \leftrightarrow x\varphi(x),$$

since this expression is equivalent to the usual one:

$$A \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix} = \begin{pmatrix} a_1 \varphi_1 \\ \vdots \\ a_n \varphi_n \end{pmatrix}.$$

Corresponding to a multiple eigenvalue a_i are several eigenvectors. We can introduce the function $\varphi(x)$ whose value at the point a_i is the vector with components φ_{ik} , $k = 1, 2, \dots, r_i$. In this case the representation is multiple, and its multiplicity is equal to the multiplicity r_i of the eigenvalue a_i .

We see that in an eigenrepresentation an operator is an analogue of a diagonal matrix on \mathbf{C}^n .

In an eigenrepresentation for an operator A it is easy to describe a function $f(A)$:

$$(10) \quad f(A)\varphi \leftrightarrow f(x)\varphi(x).$$

In particular, for the spectral function $P_A(\lambda) = \theta(\lambda - A)$ the formula (10) gives

$$P_A(\lambda) \varphi \leftrightarrow \theta(\lambda - x) \varphi(x) = \begin{cases} \varphi(x), & x \leq \lambda, \\ 0, & x > \lambda. \end{cases}$$

The spectrum of A will be completely described if for A we can construct a spectral representation of it. A transformation carrying a given representation into a spectral representation is called a spectral transformation.

Let us return to the coordinate representation. We see that it really is a multiple eigenrepresentation for the three operators Q_1, Q_2, Q_3 :

$$Q_1 \varphi \leftrightarrow x_1 \varphi(x_1).$$

The value of $\varphi(x_1)$ is a vector in an additional space, in this case a function of the variables x_2 and x_3 . The scalar product is defined by the formula

$$(\varphi, \psi) = \int_{\mathbf{R}} \varphi(x_1) \overline{\psi(x_1)} dx_1,$$

that is, in the coordinate representation the measure $m(x)$ is Lebesgue measure, and the support of its distribution function is the whole real axis, and hence the spectrum of the coordinate is continuous and fills the whole axis.

The distribution function of the coordinate Q_1 in a pure state $\omega \leftrightarrow P_\varphi$ has the form

$$\omega_{Q_1}(\lambda) = (P_{Q_1}(\lambda) \varphi, \varphi) = \int_{-\infty}^{\lambda} \varphi(x_1) \overline{\varphi(x_1)} dx_1,$$

from which it follows that

$$\varphi(x_1) \overline{\varphi(x_1)} = \int_{\mathbf{R}^2} \varphi(x_1, x_2, x_3) \overline{\varphi(x_1, x_2, x_3)} dx_2 dx_3$$

is the density of the distribution function of the coordinate Q_1 . The expressions for the densities of the distribution functions of Q_2 and Q_3 are written similarly. It is natural to expect that $|\varphi(x)|^2$ is the density of the distribution function common for all the coordinates; that is, the probability of finding a particle in a region Ω of three-dimensional space is determined by the expression $\int_{\Omega} |\varphi(x)|^2 dx$. We

prove this assertion in the section devoted to systems of commuting observables.

The momentum representation is an eigenrepresentation for the three operators P_1, P_2, P_3 , and $|\varphi(\mathbf{p})|^2$ is the density of the distribution function common for the three projections of the momentum.

We can now look at the uncertainty relations for the coordinates and momenta from a new point of view. We see that these relations are explained by a well-known property of the Fourier transformation. The more strongly the function $\varphi(\mathbf{x})$ is concentrated, and thus the smaller the uncertainties $\Delta_\omega Q_i$ of the coordinates are, the more the Fourier transform $\varphi(\mathbf{p})$ is spread out, and thus the greater the uncertainties $\Delta_\omega P_i$ of the momenta.

§ 12. “Eigenfunctions” of the operators Q and P

We now consider the equations for the eigenvectors of the operators Q and P . For simplicity we consider a particle with one degree of freedom. In the coordinate representation these equations have the form

$$(1) \quad x\varphi_{x_0}(x) = x_0\varphi_{x_0}(x).$$

$$(2) \quad \frac{\hbar}{i} \frac{d}{dx} \varphi_p(x) = p\varphi_p(x).$$

Solving them, we get that

$$(3) \quad \varphi_{x_0}(x) = \delta(x - x_0).$$

$$(4) \quad \varphi_p(x) = \left(\frac{1}{2\pi\hbar} \right)^{\frac{1}{2}} e^{\frac{i}{\hbar} px}.$$

(The first formula follows at once from the property $x\delta(x) = 0$ of the δ -function, and the second is obvious. The choice of the normalization constant will be clear from what follows.)

Although an “eigenfunction” $\varphi_{x_0}(x)$ of the coordinate operator is a generalized function while $\varphi_p(x)$ is an ordinary function, they have in common that they both fail to be square integrable; that is, they do not belong to L^2 . The operators Q and P do not have eigenfunctions in the usual sense of the word.

To understand how the functions $\varphi_x(x)$ and $\varphi_p(x)$ are connected with the spectrum of the operators Q and P , we recall the meaning of eigenvectors in \mathbf{C}^n . The problem of finding the eigenvalues and eigenvectors of a matrix A is connected with the problem of diagonalizing this matrix by a similarity transformation, or in other words, with the transformation of an operator from some initial representation to an eigenrepresentation. A self-adjoint operator A on \mathbf{C}^n always has a basis consisting of eigenvectors:

$$A\varphi_i = a_i \varphi_i, \quad (\varphi_i, \varphi_k) = \delta_{ik}.$$

We look for the eigenvectors in the initial representation, in which

$$\begin{aligned}\xi &\leftrightarrow (\xi_1, \dots, \xi_n), \\ \varphi_i &\leftrightarrow (\varphi_1^{(i)}, \dots, \varphi_n^{(i)}).\end{aligned}$$

The components of the vector ξ in an eigenrepresentation of A are computed from the formula

$$(5) \quad \xi'_i = (\xi, \varphi_i) = \sum_{k=1}^n \xi_k \overline{\varphi_k^{(i)}}.$$

We see that the matrix consisting of the numbers $U_{ik} = \overline{\varphi_k^{(i)}}$ conjugate to the components of the eigenvectors implements a spectral transformation:

$$\xi'_i = \sum_{k=1}^n U_{ik} \xi_k.$$

The matrix U is unitary, since

$$\sum_{k=1}^n U_{ik} U_{kj}^* = \sum_{k=1}^n U_{ik} \overline{U_{jk}} = \sum_{k=1}^n \overline{\varphi_k^{(i)}} \varphi_k^{(j)} = (\varphi_j, \varphi_i) = \delta_{ij}.$$

If in (5) we formally replace φ_i by φ_p and ξ'_i by $\xi(p)$, then we get that

$$\xi(p) = (\xi, \varphi_p).$$

or

$$(6) \quad \xi(p) = \left(\frac{1}{2\pi h} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{ipx}{h}} \xi(x) dx.$$

This formula is already familiar to us. We see that the function $U(p, x) = \overline{\varphi_p(x)}$ is the kernel of a unitary operator carrying the coordinate representation into the momentum representation (the eigenrepresentation for the operator P). An “eigenfunction” of the coordinate operator Q can be interpreted in the same way. The coordinate representation is an eigenrepresentation for Q ; therefore, the operator implementing the spectral transformation must be the identity, and $\overline{\varphi_{x_0}(x)} = \delta(x - x_0)$ is the kernel of the identity operator.

We see from these examples that even though the “eigenfunctions” of the continuous spectrum are not eigenelements in the usual sense of the word, their connection with the spectral transformation remains the same as for eigenvectors in a finite-dimensional space.

We remark that there is a construction enabling one to give solutions of the equation

$$A\varphi_\lambda = \lambda\varphi_\lambda$$

the precise meaning of eigenvectors even for the case when φ_λ does not belong to \mathcal{H} . To this end we introduce a broader space $\Phi^* \supset \mathcal{H}$ whose elements are linear functionals defined on some space $\Phi \subset \mathcal{H}$. The pair of spaces Φ, Φ^* can be constructed in such a way that each self-adjoint operator acting in \mathcal{H} has in Φ^* a complete system of eigenvectors. The eigenelements of A in Φ^* and not in \mathcal{H} are called generalized eigenelements. If A has a simple spectrum, then any $\psi \in \Phi$ has an eigenelement decomposition

$$(7) \quad \psi = \int_{\mathbf{R}} c(\lambda) \varphi_\lambda d\lambda.$$

The Fourier transformation can be interpreted as an eigenfunction decomposition of the momentum operator $\varphi_p(x)$:

$$(8) \quad \varphi(x) = \left(\frac{1}{2\pi\hbar} \right)^{\frac{1}{2}} \int_{\mathbf{R}} \varphi(p) e^{ipx/\hbar} dp$$

The normalization constant $(1/2\pi\hbar)^{1/2}$ in the expression for an eigenfunction of the momentum corresponds to the normalization

$$(9) \quad \int_{\mathbf{R}} \varphi_p(x) \overline{\varphi_{p'}(x)} dx = \delta(p - p').$$

This formula is a consequence of the well-known integral representation for the δ -function:

$$(10) \quad \int_{\mathbf{R}} e^{ikx} dx = 2\pi\delta(k).$$

In the last two formulas the integrals are understood in the principal value sense. The formula (9) is an analogue of the equality

$$(\varphi_i, \varphi_k) = \delta_{ik}$$

for eigenvectors of the discrete spectrum.

We now construct the spectral function of the momentum operator. Only here we use the letter Π for the spectral function, so there will be no confusion with the momentum operator.

In the momentum representation the operator $\Pi_p(\lambda)$ is the operator of multiplication by a function:

$$\Pi_p(\lambda) \varphi(p) = \theta(\lambda - p) \varphi(p).$$

Let us pass to the coordinate representation. Using the formula (11.9) for the one-dimensional case, we get that

$$\begin{aligned} \Pi_p(\lambda)(x, y) &= \frac{1}{2\pi\hbar} \int_{\mathbf{R}} \theta(\lambda - p) e^{\frac{i}{\hbar} p(x-y)} dp \\ &= \frac{1}{2\pi\hbar} \int_{-\infty}^{\lambda} e^{\frac{i}{\hbar} p(x-y)} dp = \int_{-\infty}^{\lambda} \varphi_p(x) \overline{\varphi_p(y)} dp. \end{aligned}$$

The derivative of the spectral function with respect to the parameter λ is called the spectral density. The kernel of this operator has the form

$$\frac{d}{d\lambda} \Pi_p(\lambda)(x, y) = \varphi_\lambda(x) \overline{\varphi_\lambda(y)} = \frac{1}{2\pi\hbar} e^{\frac{i}{\hbar} \lambda(x-y)}.$$

We see that this is an analogue of the projection on a one-dimensional eigenspace.

§ 13. The energy, the angular momentum, and other examples of observables

We now proceed to a study of more complicated observables. Our problem is to associate with an arbitrary classical observable $f(p, q)$ a quantum analogue A_f of it. We would like to set $A_f = f(P, Q)$,

but there is no general definition of a function on noncommuting operators. For example, it is no longer clear which of the operators Q^2P , QPQ , or PQ^2 one should associate with the classical observable q^2p . However, for the most important observables such difficulties do not arise in general, since these observables are sums of functions of the mutually commuting components Q_1, Q_2, Q_3 and P_1, P_2, P_3 .

We present some examples.

The kinetic energy of a particle in classical mechanics is

$$T = \frac{p_1^2 + p_2^2 + p_3^2}{2m}.$$

The corresponding kinetic energy operator has the form

$$T = \frac{P_1^2 + P_2^2 + P_3^2}{2m},$$

and in the coordinate representation

$$(1) \quad T\varphi(\mathbf{x}) = -\frac{\hbar^2}{2m}\Delta\varphi(\mathbf{x}),$$

where $\Delta = \frac{\delta^2}{\delta x_1^2} + \frac{\delta^2}{\delta x_2^2} + \frac{\delta^2}{\delta x_3^2}$ is the Laplace operator. In the momentum representation the operator T is the operator of multiplication by the corresponding function:

$$T\varphi(\mathbf{p}) = \frac{p_1^2 + p_2^2 + p_3^2}{2m}\varphi(\mathbf{p}).$$

In exactly the same way it is easy to introduce the potential energy operator $V(Q_1, Q_2, Q_3)$. In the coordinate representation, V is the operator of multiplication by the function

$$(2) \quad V\varphi(\mathbf{x}) = V(\mathbf{x})\varphi(\mathbf{x}),$$

and in the momentum representation it is the integral operator with kernel

$$V(\mathbf{p}, \mathbf{q}) = \left(\frac{1}{2\pi\hbar}\right)^3 \int_{\mathbf{R}^3} V(\mathbf{x}) e^{\frac{i}{\hbar}\mathbf{x}(\mathbf{q}-\mathbf{p})} d\mathbf{x}.$$

The total energy operator (the Schrödinger operator) is defined as

$$II = T + V.$$

We write the Schrödinger equation in detail

$$(3) \quad i\hbar \frac{d\psi(t)}{dt} = II\psi(t).$$

In the coordinate representation the Schrödinger equation for a particle is the partial differential equation

$$(4) \quad i\hbar \frac{\partial \psi(\mathbf{x}, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi(\mathbf{x}, t) + V(\mathbf{x}) \psi(\mathbf{x}, t),$$

and in the momentum representation the integro-differential equation

$$ih \frac{\partial \psi(\mathbf{p}, t)}{\partial t} = \frac{\mathbf{p}^2}{2m} \psi(\mathbf{p}, t) + \int_{\mathbf{R}^3} V(\mathbf{p}, \mathbf{q}) \psi(\mathbf{q}, t) d\mathbf{q}.$$

Along with (4) we consider the equation for the complex conjugate wave function:

$$(5) \quad -ih \frac{\partial \overline{\psi(\mathbf{x}, t)}}{\partial t} = -\frac{\hbar^2}{2m} \Delta \overline{\psi(\mathbf{x}, t)} + V(\mathbf{x}) \overline{\psi(\mathbf{x}, t)}.$$

Multiplying the equation (4) by $\overline{\psi}$ and the equation (5) by ψ and subtracting one from the other, we get that

$$\begin{aligned} ih \left(\bar{\psi} \frac{\partial \psi}{\partial t} + \psi \frac{\partial \bar{\psi}}{\partial t} \right) &= -\frac{\hbar^2}{2m} (\bar{\psi} \Delta \psi - \psi \Delta \bar{\psi}) \\ &= -\frac{\hbar^2}{2m} \operatorname{div}(\psi \operatorname{grad} \bar{\psi} - \bar{\psi} \operatorname{grad} \psi), \end{aligned}$$

or

$$(6) \quad \frac{\partial |\psi|^2}{\partial t} + \operatorname{div} \mathbf{j} = 0,$$

where

$$\mathbf{j} = \frac{i\hbar}{2m} (\psi \operatorname{grad} \bar{\psi} - \bar{\psi} \operatorname{grad} \psi).$$

The equation (6) is the equation of continuity and expresses the law of conservation of probability. The vector \mathbf{j} is called the density vector of the probability flow. It is clear from (6) that \mathbf{j} has the following meaning: $\int_S j_n dS$ is the probability that a particle crosses the surface S per unit of time.

The angular velocity plays an important role in both classical and quantum mechanics. In classical mechanics this observable is a vector whose projections in a Cartesian coordinate system have the form

$$l_1 = q_2 p_3 - q_3 p_2,$$

$$l_2 = q_3 p_1 - q_1 p_3,$$

$$l_3 = q_1 p_2 - q_2 p_1.$$

In quantum mechanics the operators

$$(7) \quad L_k = Q_l P_m - Q_m P_l$$

are the projections of the angular momentum. Here k, l, m are cyclic permutations of the numbers 1, 2, 3. The right-hand side of (7) contains only products of coordinates and momentum projections with different indices, that is, products of commuting operators. It is interesting to note that the operators L_k have the same form in the momentum and coordinate representations:

$$(8) \quad \begin{aligned} L_k \varphi(\mathbf{x}) &= \frac{\hbar}{i} \left(x_l \frac{\partial}{\partial x_m} - x_m \frac{\partial}{\partial x_l} \right) \varphi(\mathbf{x}), \\ L_k \varphi(\mathbf{p}) &= \frac{\hbar}{i} \left(p_l \frac{\partial}{\partial p_m} - p_m \frac{\partial}{\partial p_l} \right) \varphi(\mathbf{p}). \end{aligned}$$

The properties of the operators L_k will be studied at length below.

Quantum mechanics can also describe systems more complicated than just a material point, of course. For example, for a system of N material points the state space in the coordinate representation is the space¹⁷ $L^2(\mathbf{R}^{3N})$ of functions $\varphi(\mathbf{x}_1, \dots, \mathbf{x}_N)$ of N vector variables. The Schrödinger operator for such a system (the analogue of the Hamiltonian functions (1.5)) has the form

$$(9) \quad H = - \sum_{i=1}^N \frac{\hbar^2}{2m_i} \Delta_i + \sum_{i < j}^N V_{ij}(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^N V_i(\mathbf{x}).$$

The physical meaning of the terms here is the same as in classical mechanics.

We consider two more operators for a material point which turn out to be useful for discussing the interconnection of quantum and classical mechanics:

$$(10) \quad U(\mathbf{u}) = e^{-i(u_1 P_1 + u_2 P_2 + u_3 P_3)},$$

$$(11) \quad V(\mathbf{v}) = e^{-i(v_1 Q_1 + v_2 Q_2 + v_3 Q_3)},$$

¹⁷Below we shall see that for identical particles, the state space \mathcal{H} coincides with the subspace $L_S^2 \subset L^2$ of functions with a definite symmetry

where $\mathbf{u}(u_1, u_2, u_3)$ and $\mathbf{v}(v_1, v_2, v_3)$ are real parameters. The operator $V(\mathbf{v})$ in the coordinate representation is the operator of multiplication by a function:

$$V(\mathbf{v}) \varphi(\mathbf{x}) = e^{i\mathbf{vx}} \varphi(\mathbf{x}).$$

We now explain the meaning of the operator $U(\mathbf{u})$. For simplicity we consider the one-dimensional case

$$U(u) = e^{-iuP}.$$

Let $\varphi(u, x) = e^{-iuP} \varphi(x)$. Differentiating $\varphi(u, x)$ with respect to the parameter u , we have

$$\frac{\partial \varphi(u, x)}{\partial u} = -iP \varphi(u, x),$$

or

$$\frac{\partial \varphi(u, x)}{\partial u} = -h \frac{\partial \varphi(u, x)}{\partial x}.$$

To find the function $\varphi(u, x)$, we must solve this equation with the initial condition

$$\varphi(u, x)|_{u=0} = \varphi(x).$$

The unique solution obviously has the form

$$\varphi(u, x) = \varphi(x - hu).$$

We see that in the coordinate representation the operator $U(\mathbf{u})$ is the operator translating the argument of the function $\varphi(x)$ by the quantity $-hu$. In the three-dimensional case

$$(12) \quad U(\mathbf{u}) \varphi(\mathbf{x}) = \varphi(\mathbf{x} - \mathbf{uh}).$$

The operators $U(\mathbf{u})$ and $U(\mathbf{v})$ are unitary, because the operators P_1, P_2, P_3 and Q_1, Q_2, Q_3 are self-adjoint. Let us find the commutation relations for the operators $U(\mathbf{u})$ and $U(\mathbf{v})$. In the coordinate representation

$$V(\mathbf{v}) U(\mathbf{u}) \varphi(\mathbf{x}) = e^{-i\mathbf{vx}} \varphi(\mathbf{x} - \mathbf{uh}),$$

$$U(\mathbf{u}) V(\mathbf{v}) \varphi(\mathbf{x}) = e^{-i\mathbf{v}(\mathbf{x}-\mathbf{uh})} \varphi(\mathbf{x} - \mathbf{uh}).$$

From these equalities it follows at once that

$$(13) \quad U(\mathbf{u}) V(\mathbf{v}) = V(\mathbf{v}) U(\mathbf{u}) e^{i\mathbf{vu}h}.$$

Of course, the relations (13) do not depend on the representation. We note further the formulas

$$\begin{aligned} U(\mathbf{u}_1)U(\mathbf{u}_2) &= U(\mathbf{u}_1 + \mathbf{u}_2), \\ V(\mathbf{v}_1)V(\mathbf{v}_2) &= V(\mathbf{v}_1 + \mathbf{v}_2); \end{aligned}$$

that is, the sets of operators $U(\mathbf{u})$ and $V(\mathbf{v})$ form groups: denote these groups by U and V

We can now give a precise formulation of von Neumann's theorem. For simplicity we confine ourselves to a system with a single degree of freedom

Theorem. *Let U and V be one-parameter groups of unitary operators $U(u)$ and $V(v)$ acting in a Hilbert space \mathcal{H} and satisfying the condition*

$$(14) \quad U(u)V(v) = V(v)U(u)e^{ivuh}$$

Then \mathcal{H} can be represented as a direct sum

$$(15) \quad \mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \dots,$$

where each \mathcal{H}_l is carried into itself by all the operators $U(u)$ and $V(v)$, and each \mathcal{H}_l can be mapped unitarily onto $L^2(\mathbf{R})$ in such a way that the operators $V(v)$ pass into the operators $\psi(x) \rightarrow e^{-ivx}\psi(x)$ and the operators $U(u)$ pass into the operators $\psi(x) \rightarrow \psi(x - uh)$.

One can say that a representation of the relations (14) by unitary operators acts in the space \mathcal{H} . If this representation is irreducible, then the sum (15) contains only one term.

In the conclusion of this section we prove the irreducibility of the coordinate representation for P and Q .

Let K be an operator commuting with Q and P :

$$[K, Q] = 0, \quad [K, P] = 0.$$

It follows from the second equality that

$$KU(u) = U(u)K.$$

Applying the operators on both sides of the equality to an arbitrary function $\varphi(x)$, we get that

$$\int_{\mathbf{R}} K(x, y) \varphi(y - uh) dy = \int_{\mathbf{R}} K(x - uh, y) \varphi(y) dy.$$

The substitution $y - uh \rightarrow y$ on the left-hand side lets us rewrite this equality as

$$\int_{\mathbf{R}} K(x, y + uh) \varphi(y) dy = \int_{\mathbf{R}} K(x - uh, y) \varphi(y) dy.$$

Since φ is arbitrary,

$$K(x, y + uh) = K(x - uh, y).$$

from which it is obvious that the kernel $K(x, y)$ depends only on the difference $x - y$, that is,

$$K(x, y) = k(x - y).$$

We now use the fact that K commutes with the coordinate operator:

$$\int_{-\infty}^{\infty} xk(x - y) \varphi(y) dy = \int_{-\infty}^{\infty} k(x - y) y \varphi(y) dy,$$

from which it follows that

$$(x - y) k(x - y) = 0$$

The solution of this equation has the form

$$k(x - y) = c\delta(x - y),$$

and the function on the right-hand side is the kernel of the operator CI .

Accordingly, we have shown that any operator commuting with Q and P is a multiple of the identity.

§ 14. The interconnection between quantum and classical mechanics. Passage to the limit from quantum mechanics to classical mechanics

We know that the behavior of a macroscopic system is described well by classical mechanics, so in the passage to a macro-object, quantum mechanics must reduce to the same results. A criterion for the "quantumness" of the behavior of a system can be the relative magnitude of Planck's constant $h = 1.05 \cdot 10^{-27}$ erg-sec and the system's characteristic numerical values for observables having the dimension of action. If Planck's constant is negligibly small in comparison with the values

of such observables, then the system should have a classical behavior. The passage from quantum mechanics to classical mechanics can be described formally as the passage to the limit as $\hbar \rightarrow 0$.¹⁸ Of course, under such a passage to the limit, the methods for describing the mechanical systems remain different (operators acting in a Hilbert space cannot be transformed into functions on phase space), but the physical results of quantum mechanics as $\hbar \rightarrow 0$ must coincide with the classical results.

We again consider a system with one degree of freedom. Our presentation will be conducted according to the following scheme.

We first use a certain rule to associate self-adjoint operators A_f with real functions $f(p, q)$ on the phase space ($f \rightarrow A_f$). Then we find an inversion formula enabling us to recover the function $f(p, q)$ from the operator A_f ($A_f \rightarrow f$). Thereby, we shall have established a one-to-one relation between real functions on the phase space and self-adjoint operators acting in \mathcal{H} , ($f \leftrightarrow A_f$). The correct formula turns out to be

$$(1) \quad \text{Tr } A_f = \int_{\mathcal{M}} f(p, q) \frac{dp dq}{2\pi\hbar}.$$

Finally, we determine what functions on the phase space correspond to the product $A_f \circ A_g$ and to the quantum Poisson bracket $\{A_f, A_g\}$. We shall see that these functions do not coincide with the product fg and to the classical Poisson bracket $\{f, g\}$, but do tend to them in the limit as $\hbar \rightarrow 0$. Thus, we shall see that the algebra of observables in quantum mechanics is not isomorphic to the algebra of observables in classical mechanics, but the one-to-one correspondence $f \leftrightarrow A_f$ does become an isomorphism as $\hbar \rightarrow 0$.

Suppose that a quantum system with Schrödinger operator H is in a state M , and let A_f be some observable for this system. The one-to-one correspondence described enables us to associate with the operators H , M , and A_f , a Hamiltonian function $H(p, q)$, a function $\rho_1(p, q)$, and an observable $f(p, q)$. Let $\rho(p, q) = \rho_1(p, q)/2\pi\hbar$. It

¹⁸Here there is an analogy with the connection between relativistic and classical mechanics. Relativistic effects can be ignored if the velocities characteristic for the system are much less than the velocity c of light. The passage from relativistic mechanics to classical mechanics can be regarded formally as the limit as $c \rightarrow \infty$.

follows from (1) that

$$(2) \quad \text{Tr } M = \int_{\mathcal{M}} \rho(p, q) dp dq = 1,$$

$$(3) \quad \text{Tr } MA_f \xrightarrow[h \rightarrow 0]{} \int_{\mathcal{M}} f(p, q) \rho(p, q) dp dq.$$

The formula (2) shows that the function $\rho(p, q)$ has the correct normalization, and the formula (3) asserts that in the limit as $h \rightarrow 0$ the mean value of an observable in quantum mechanics coincides with the mean value of the corresponding classical observable.¹⁹ Further, suppose that the evolution of the quantum system is described by the Heisenberg picture and the correspondence $A_f(t) \leftrightarrow f(t)$ has been established for an arbitrary moment of time t . We show that as $h \rightarrow 0$ the classical observable $f(t)$ depends correctly on the time. The operator $A_f(t)$ satisfies the equation

$$(4) \quad \frac{dA_f(t)}{dt} = \{H, A_f(t)\}_h.$$

It follows from the linearity of the correspondence $f \leftrightarrow A_f$ that $\frac{df}{dt} \leftrightarrow \frac{dA_f}{dt}$, and moreover, $\{H, A_f\}_h \leftrightarrow \{H, f\}$ as $h \rightarrow 0$, so that as $h \rightarrow 0$ the classical equation

$$\frac{df}{dt} = \{H, f\}$$

is a consequence of the quantum equation (4).

We proceed to the details. Let us consider a function $f(p, q)$ and denote by $\hat{f}(u, v)$ its Fourier transform:²⁰

$$(5) \quad f(q, p) = \frac{1}{2\pi} \int_{\mathbf{R}^2} \hat{f}(u, v) e^{-iqv} e^{-ipu} du dv,$$

$$(6) \quad \hat{f}(u, v) = \frac{1}{2\pi} \int_{\mathbf{R}^2} f(p, q) e^{iqv} e^{ipu} dq dp.$$

¹⁹We underscore once more that the left-hand side of (3) does not coincide with the right-hand side when $h \neq 0$, since a function different from $f(p, q)\rho_1(p, q)$ corresponds to the product MA_f . Moreover, we note that for $h \neq 0$ the function $\rho(p, q)$ can fail to be positive, that is, it does not correspond to a classical state, but in the limit as $h \rightarrow 0$ it follows from (3) that $\rho(p, q)$ satisfies all the requirements of a classical distribution function

²⁰We omit the stipulations which should be made in order for all the subsequent transformations to be completely rigorous

The Fourier transform of a real function $f(p, q)$ has the property that

$$(7) \quad \overline{\hat{f}(u, v)} = \hat{f}(-u, -v).$$

To construct the operator A_f corresponding to the function $f(p, q)$, we would like to replace the variables q and p in (5) by the operators Q and P . However, it is not immediately clear in what order we should write the noncommuting factors $V(v)$ and $U(u)$, which satisfy the commutation relations

$$(8) \quad U(u) V(v) = V(v) U(u) e^{ihuv}.$$

A natural recipe (but by no means the only one) ensuring self-adjointness was proposed by Weyl and has the form

$$(9) \quad A_f = \frac{1}{2\pi} \int_{\mathbf{R}^2} \hat{f}(u, v) V(v) U(u) e^{\frac{ihuv}{2}} du dv.$$

The appearance of the extra factor $e^{ihuv/2}$ is connected with the non-commutativity of $V(v)$ and $U(u)$ and ensures that the operator A_f is self-adjoint. Indeed, using (7) and (8), we have²¹

$$\begin{aligned} A_f^* &= \frac{1}{2\pi} \int \overline{\hat{f}(u, v)} U^*(u) V^*(v) e^{-\frac{ihuv}{2}} du dv \\ &= \frac{1}{2\pi} \int \hat{f}(-u, -v) U(-u) V(-v) e^{-\frac{ihuv}{2}} du dv \\ &= \frac{1}{2\pi} \int \hat{f}(u, v) U(u) V(v) e^{-\frac{ihuv}{2}} du dv \\ &= \frac{1}{2\pi} \int \hat{f}(u, v) V(v) U(u) e^{\frac{ihuv}{2}} du dv = A_f. \end{aligned}$$

At the end of this section we verify that the operators $f(Q)$ and $f(P)$ correspond to the functions $f(q)$ and $f(p)$ by virtue of the formula (9), in complete agreement with the assumptions made earlier.

We now find an inversion formula. In computing the trace of an operator K we shall use the formula

$$(10) \quad \text{Tr } K = \int K(x, x) dx.$$

²¹Below we omit the symbol \mathbf{R}^n in the expression for an integral over the n -dimensional space

where $K(x, y)$ is the kernel of K . We find the kernel of the operator $V(v)U(u)$. It follows from the formula

$$V(v)U(u)\varphi(x) = e^{-ivx}\varphi(x - uh)$$

that the kernel of the integral operator $V(v)U(u)$ is the function

$$V(v)U(u)(x, x') = e^{-ivx}\delta(x - uh - x').$$

By (10),

$$(11) \quad \text{Tr } V(v)U(u) = \int_{-\infty}^{\infty} e^{-ivx}\delta(-uh)dx = \frac{2\pi}{h}\delta(v)\delta(u).$$

We verify that the inversion formula is

$$(12) \quad \hat{f}(u, v) = h \text{Tr } A_f V(-v) U(-u) e^{\frac{iuhu}{2}}.$$

To this end we compute the right-hand side of the equality, using (8) and (11):

$$\begin{aligned} & \text{Tr } A_f V(-v) U(-u) e^{\frac{iuhu}{2}} \\ &= \frac{1}{2\pi} \text{Tr} \int \hat{f}(u', v') V(v') U(u') e^{\frac{iuh'v'}{2}} V(-v) U(-u) e^{\frac{iuhu}{2}} du' dv' \\ &= \frac{1}{2\pi} \text{Tr} \int \hat{f}(u', v') V(v') V(-v) U(u') U(-u) e^{\frac{iuh}{2}(u'v' + uv - 2u'v)} du' dv' \\ &= \frac{1}{2\pi} \text{Tr} \int \hat{f}(u', v') V(v' - v) U(u' - u) e^{\frac{iuh}{2}(u'v' + uv - 2u'v)} du' dv' \\ &= \frac{1}{h} \int \hat{f}(u', v') \delta(v' - v) \delta(u' - u) e^{\frac{iuh}{2}(u'v' + uv - 2u'v)} du' dv' \\ &= \frac{1}{h} \hat{f}(u, v) \end{aligned}$$

We set $u = v = 0$ in (12):

$$\hat{f}(0, 0) = h \text{Tr } A_f.$$

On the other hand,

$$\hat{f}(0, 0) = \frac{1}{2\pi} \int f(p, q) dp dq,$$

and we obtain the formula (1).

It now remains for us to find the functions on phase space corresponding to $A_f \circ A_g$ and $\{A_f, A_g\}_h$ using (12), and to see that these

functions tend to fg and $\{f, g\}$ as $h \rightarrow 0$. We first find the function $F(p, q)$ corresponding to the nonsymmetric product $A_f A_g$. Its Fourier transform is

$$\begin{aligned} \widehat{F}(u, v) &= h \operatorname{Tr} A_f A_g V(-v) U(-u) e^{\frac{i h u v}{2}} \\ &= h \operatorname{Tr} \left(\frac{1}{2\pi} \right)^2 \int du_1 dv_1 du_2 dv_2 \widehat{f}(u_1, v_1) \widehat{g}(u_2, v_2) V(v_1) U(u_1) e^{\frac{i h u_1 v_1}{2}} \\ &\quad \times V(v_2) U(u_2) e^{\frac{i h u_2 v_2}{2}} V(-v) U(-u) e^{\frac{i h u v}{2}} \\ &= h \operatorname{Tr} \left(\frac{1}{2\pi} \right)^2 \int du_1 dv_1 du_2 dv_2 \widehat{f}(u_1, v_1) \widehat{g}(u_2, v_2) \\ &\quad \times V(v_1 + v_2 - v) U(u_1 + u_2 - u) \\ &\quad \times \exp \left[\frac{i h}{2} (u_1 v_1 + u_2 v_2 + uv + 2u_1 v_2 - 2u_1 v - 2u_2 v) \right]. \end{aligned}$$

Finally,

(13)

$$\begin{aligned} A_f A_g \leftrightarrow \frac{1}{2\pi} \int du_1 dv_1 du_2 dv_2 \widehat{f}(u_1, v_1) \widehat{g}(u_2, v_2) \delta(v_1 + v_2 - v) \\ \times \delta(u_1 + u_2 - u) e^{\frac{i h}{2} (u_1 v_2 - u_2 v_1)}. \end{aligned}$$

The exponent in the exponential was transformed with the δ -functions under the integral sign taken into account. We recall that the Fourier transform $\widehat{\Phi}(u, v)$ of the product $\Phi(p, q) = f(p, q)g(p, q)$ of the two functions is the convolution of the Fourier transforms of the factors:

$$\begin{aligned} \widehat{\Phi}(u, v) &= \frac{1}{2\pi} \int du_1 dv_1 du_2 dv_2 \\ &\quad \times \widehat{f}(u_1, v_1) \widehat{g}(u_2, v_2) \delta(v_1 + v_2 - v) \delta(u_1 + u_2 - u). \end{aligned}$$

The function $\widehat{F}(u, v)$ differs from the function $\widehat{\Phi}(u, v)$ by the factor $e^{\frac{i h}{2} (u_1 v_2 - u_2 v_1)}$ under the integral sign. This factor depends on the order of the operators A_f and A_g , and therefore the operators $A_f A_g$ and $A_g A_f$ correspond to different functions on the phase space. In the limit as $h \rightarrow 0$ we have $e^{\frac{i h}{2} (u_1 v_2 - u_2 v_1)} \rightarrow 1$ and $F(p, q) \rightarrow \Phi(p, q)$. Of course, this assertion is valid also for the function $F_s(p, q)$ corresponding to the symmetrized product $A_f \circ A_g$.

We denote by $G(p, q)$ the function corresponding to the quantum Poisson bracket

$$\{A_f, A_g\}_h = \frac{i}{h} (A_f A_g - A_g A_f).$$

From (13) we get that

$$\begin{aligned}\widehat{G}(u, v) &= \frac{i}{2\pi h} \int du_1 dv_1 du_2 dv_2 \hat{f}(u_1, v_1) \hat{g}(u_2, v_2) \delta(v_1 + v_2 - v) \\ &\quad \times \delta(u_1 + u_2 - u) \left[e^{\frac{ih}{2}(u_1v_2 - u_2v_1)} - e^{-\frac{ih}{2}(u_2v_1 - u_1v_2)} \right] \\ &= \frac{1}{\pi h} \int du_1 dv_1 du_2 dv_2 \hat{f}(u_1, v_1) \hat{g}(u_2, v_2) \delta(v_1 + v_2 - v) \\ &\quad \times \delta(u_1 + u_2 - u) \sin \frac{h}{2}(u_2v_1 - u_1v_2),\end{aligned}$$

and as $h \rightarrow 0$

$$\begin{aligned}\widehat{G}(u, v) &\rightarrow \frac{1}{\pi} \int du_1 dv_1 du_2 dv_2 (u_2v_1 - u_1v_2) \hat{f}(u_1, v_1) \hat{g}(u_2, v_2) \\ &\quad \times \delta(v_1 + v_2 - v) \delta(u_1 + u_2 - u).\end{aligned}$$

The integral on the right-hand side is the Fourier transform of the classical Poisson bracket

$$\{f, g\} = \frac{\partial f}{\partial p} \frac{\partial g}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial g}{\partial p},$$

since the functions $-iv\hat{f}$ and $-iu\hat{f}$ are the Fourier transforms of the respective derivatives $\frac{\partial f}{\partial q}$ and $\frac{\partial f}{\partial p}$.

Thus, we have verified all the assertions made at the beginning of the section.

In conclusion we present some examples of computations using the formulas in this section.

Let us find a formula for the kernel of the operator A_f in the coordinate representation. Using the formula

$$V(v) U(u)(x, x') = e^{ivx} \delta(x - uh - x'),$$

we get that

$$(14) \quad A_f(x, x') = \frac{1}{2\pi} \int \hat{f}(u, v) e^{-iu x} \delta(x - uh - x') e^{\frac{iuhv}{2}} du dv,$$

$$A_f(x, x') = \frac{1}{2\pi h} \int \hat{f}\left(\frac{x-x'}{h}, v\right) e^{-\frac{v}{2}(x+x')} dv.$$

We show that $f(q) \leftrightarrow f(Q)$. For such a function on the phase space,

$$\hat{f}(u, v) = \frac{1}{2\pi} \int f(q) e^{ivq} e^{iup} dq dp = \tilde{f}(v) \delta(u).$$

Here $\tilde{f}(v)$ denotes the Fourier transform of the function $f(q)$ of a single variable. Further, by (14),

$$A_f(x, x') = \frac{1}{2\pi h} \int \delta\left(\frac{x-x'}{h}\right) \tilde{f}(v) e^{-\frac{v}{2}(x+x')} dv$$

$$= f\left(\frac{x+x'}{2}\right) \delta(x - x') = f(x) \delta(x - x'),$$

and the operator with this kernel is the operator of multiplication by the function $f(x)$. In exactly the same way it is easy to verify that $f(p) \leftrightarrow f(P)$ in the momentum representation.

We shall get an explicit formula for finding the classical state corresponding to the limit of a pure quantum state as $h \rightarrow 0$.

Using the inversion formula, we find $\rho_1(q, p)$ corresponding to the operator P_ψ , and we construct $\rho(q, p) = \rho_1(q, p)/(2\pi h)$. The vector ψ is assumed to be given in the coordinate representation. From the formula

$$V(-v) U(-u) P_\psi \xi(x) = e^{ivx} \int \xi(x') \overline{\psi(x')} dx' \psi(x + uh).$$

it follows that

$$V(-v) U(-u) P_\psi(x, x') = e^{ivx} \psi(x + uh) \overline{\psi(x')}$$

and

$$\hat{\rho}(u, v) = \frac{1}{2\pi h} h \operatorname{Tr} V(-v) U(-u) P_\psi e^{\frac{iuhv}{2}}$$

$$= \frac{1}{2\pi} \int e^{ivx} \psi(x + uh) \overline{\psi(x)} e^{\frac{iuhv}{2}} dx$$

If we introduce the function²²

$$(15) \quad F(x, u) = \frac{1}{2\pi} \lim_{h \rightarrow 0} \psi(x + uh) \overline{\psi(x)},$$

then

$$\hat{\rho}(u, v) = \int e^{ivx} F(x, u) dx$$

is the Fourier transform of the classical distribution function corresponding to the limit of the state P_ψ as $h \rightarrow 0$. The distribution function $\rho(q, p)$ itself is given by

$$(16) \quad \rho(p, q) = \int e^{-ipu} F(q, u) du.$$

Suppose that $\psi(x)$ is a continuous function and does not depend on h as a parameter. Then

$$F(x, u) = \frac{1}{2\pi} |\psi(x)|^2$$

and

$$\rho(p, q) = |\psi(q)|^2 \delta(p).$$

The state of a particle at rest with density $|\psi(q)|^2$ of the coordinate distribution function corresponds to such a quantum state in the limit as $h \rightarrow 0$.

Let $\psi(x) = \varphi(x)e^{\frac{ip_0x}{\hbar}}$, where $\varphi(x)$ is independent of h and is continuous. In this case in the limit as $h \rightarrow 0$, we arrive at the classical state with distribution function

$$\rho(p, q) = |\varphi(q)|^2 \delta(p - p_0).$$

We see from these examples that in the limit as $h \rightarrow 0$ a mixed classical state can correspond to a pure state in quantum mechanics.

§ 15. One-dimensional problems of quantum mechanics. A free one-dimensional particle

In §§ 15–20 we consider one-dimensional problems of quantum mechanics. The Hamiltonian function for a particle with one degree of

²²This limit is not always trivial, since in physically interesting cases $\psi(x)$ usually depends on h as a parameter (see the example below)

freedom in a potential field has the form

$$H(p, q) = \frac{p^2}{2m} + V(q).$$

Corresponding to this Hamiltonian function is the Schrödinger operator

$$H = \frac{P^2}{2m} + V(Q).$$

For a free particle $V = 0$. We begin with a study of this simplest problem in quantum mechanics. Let us find the spectrum of the Schrödinger operator

$$(1) \quad H = \frac{P^2}{2m}.$$

The equation for the eigenvectors has the form

$$(2) \quad \frac{P^2}{2m} \psi = E\psi,$$

or, in the coordinate representation,

$$(3) \quad -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = E\psi.$$

It is convenient to use a system of units in which $\hbar = 1$ and $m = 1/2$. Then for $E > 0$ the equation (2) takes the form

$$(4) \quad \psi'' + k^2\psi = 0, \quad k^2 = E, \quad k > 0.$$

The last equation has two linearly independent solutions

$$(5) \quad \psi_{\pm k}(x) = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} e^{\pm ikx}.$$

We see that to each value $E > 0$ there correspond two eigenfunctions of the operator H . For $E < 0$ the equation (3) does not have solutions bounded on the whole real axis. Thus, the spectrum of the Schrödinger operator (1) is continuous and positive and has multiplicity two.

The functions (5) are at the same time the eigenfunctions of the momentum operator $P = -i\frac{d}{dx}$ that correspond to the eigenvalues $\pm k$. We note that the functions (5) do not belong to $L^2(\mathbf{R})$ and,

therefore, are not eigenfunctions in the ordinary sense. For subsequent formulas it is convenient to assume that $-\infty < k < \infty$ and $k = \pm\sqrt{E}$. Then both the solutions (5) have the form

$$(6) \quad \psi_k(x) = \left(\frac{1}{2\pi} \right)^{\frac{1}{2}} e^{ikx}.$$

The normalizing factor in (6) was chosen so that

$$(7) \quad \int_{\mathbf{R}} \psi_k(x) \overline{\psi_{k'}(x)} dx = \delta(k - k').$$

The solutions of (2) in the momentum representation are obtained just as simply. For $m = 1$, (2) has the form

$$(8) \quad p^2 \psi = k^2 \psi,$$

and its solutions are the functions

$$(9) \quad \psi_k(p) = \delta(p - k), \quad k = \pm\sqrt{E}.$$

The normalization of the functions (9) is chosen the same as for the functions (6):

$$\int_{\mathbf{R}} \psi_k(p) \overline{\psi_{k'}(p)} dp = \int_{\mathbf{R}} \delta(p - k) \delta(p - k') dp = \delta(k - k').$$

To explain the physical meaning of the eigenfunctions ψ_k , we construct a solution of the nonstationary Schrödinger equation for a free particle

$$i \frac{d\psi(t)}{dt} = H\psi(t)$$

under the initial conditions

$$\psi(0) = \varphi, \quad \|\varphi\| = 1.$$

This problem is simplest to solve in the momentum representation

$$i \frac{\partial \psi(p, t)}{\partial t} = p^2 \psi(p, t),$$

$$\psi(p, 0) = \varphi(p), \quad \int_{\mathbf{R}} |\varphi(p)|^2 dp = 1.$$

Obviously,

$$\psi(p, t) = \varphi(p) e^{-ip^2 t}.$$

In the coordinate representation the same state is described by the function

(10)

$$\psi(x, t) = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \int_{\mathbf{R}} \varphi(p) e^{i(px - p^2 t)} dp = \int_{\mathbf{R}} \varphi(k) \psi_k(x) e^{-ik^2 t} dk.$$

It is easy to verify that the normalization of the function $\psi(x, t)$ does not depend on the time:

$$\int_{\mathbf{R}} |\psi(x, t)|^2 dx = \int_{\mathbf{R}} |\varphi(p)|^2 dp = 1.$$

The formula (10) can be regarded also as a decomposition of the solution $\psi(x, t)$ of the Schrödinger equation with respect to stationary states, that is, as a generalization of the formula (9.13) to the case of a continuous spectrum. The role of the coefficients c_n is played here by the function $\varphi(k)$.

The state described by the functions $\psi(x, t)$ or $\psi(p, t)$ has the simplest physical meaning in the case when $\varphi(p)$ is nonzero in some small neighborhood of the point p_0 . States of this type are generally called wave packets. The graph of the function $|\varphi(p)|^2$ is pictured in Figure 4.

We recall that $|\psi(p, t)|^2 = |\varphi(p)|^2$ is the density of the momentum distribution function. The fact that this density does not depend on the time is a consequence of the law of conservation of momentum for a free particle. If the distribution $|\varphi(p)|^2$ is concentrated in a small neighborhood of the point p_0 , then the state $\psi(t)$ can be interpreted as a state with almost precisely given momentum.

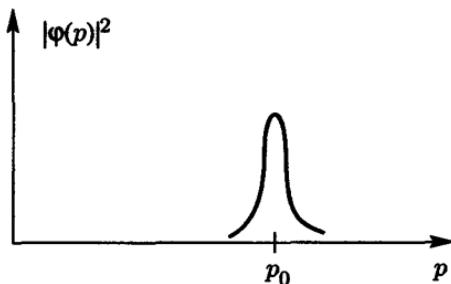


Figure 4

The function $|\psi(x, t)|^2$ is the density of the coordinate distribution function. Let us trace its evolution in time. First of all it follows from the Riemann-Lebesgue theorem²³ that $\psi(x, t) \rightarrow 0$ as $|t| \rightarrow \infty$ for smooth $\varphi(p)$, and therefore the integral $\int_{\Omega} |\psi(x, t)|^2 dx$ over any finite region Ω of the real axis tends to zero. This means that as $|t| \rightarrow \infty$, the particle leaves any finite region; that is, the motion is infinite.

In order to follow more closely the motion of the particle for large $|t|$, we use the stationary phase method. This method can be used to compute integrals of the form

$$(11) \quad I(N) = \int_a^b F(x) e^{iNf(x)} dx$$

asymptotically as $N \rightarrow \infty$. For the case of smooth functions $f(x)$ and $F(x)$ and when there is a single point \tilde{x} in (a, b) , where $f(x)$ is stationary ($f'(\tilde{x}) = 0$, $f''(\tilde{x}) \neq 0$), we have the formula

$$(12) \quad I(N) = \left(\frac{2\pi}{N|f''(\tilde{x})|} \right)^{\frac{1}{2}} F(\tilde{x}) \exp \left(iNf(\tilde{x}) + \frac{i\pi}{4} \operatorname{sgn} f''(\tilde{x}) \right) + O\left(\frac{1}{N}\right).$$

We rewrite the expression for $\psi(x, t)$ in the form

$$\psi(x, t) = \left(\frac{1}{2\pi} \right)^{\frac{1}{2}} \int_{\mathbb{R}} \varphi(p) e^{i|t|\left[\mp p^2 + \frac{px}{|t|}\right]} dp.$$

The upper sign in the exponential corresponds to $t > 0$ and the lower to $t < 0$. We find the asymptotics of $\psi(x, t)$ as $t \rightarrow \pm\infty$ from the formula (12). Then the stationarity is found from the condition

$$f'(\tilde{p}) = \mp 2\tilde{p} + \frac{x}{|t|} = 0,$$

from which $\tilde{p} = \frac{x}{2t}$. Further, $f''(\tilde{p}) = \mp 2$, and as $t \rightarrow \pm\infty$

$$(13) \quad \psi(x, t) = \varphi\left(\frac{x}{2t}\right) \left(\frac{1}{2|t|}\right)^{\frac{1}{2}} e^{ix(x,t)} + O\left(\frac{1}{|t|}\right),$$

where χ is a real function whose form is not important for what follows.

²³This theorem asserts that $\int_{-\infty}^{\infty} f(x) e^{iNx} dx \rightarrow 0$ as $N \rightarrow \infty$ if the function $f(x)$ is piecewise continuous and absolutely integrable on the whole axis $-\infty < x < \infty$.

By assumption, $\varphi(p)$ is nonzero only in a small neighborhood of the point p_0 , therefore, $|\psi(x, t)|^2 = |\varphi(\frac{x}{2t})|^2 \frac{1}{2|t|} + O(\frac{1}{|t|^{3/2}})$ is nonzero only near the point $x = 2p_0t$. From this relation it is clear that the region with a nonzero probability of finding a particle moves along the x -axis with constant velocity $v = 2p_0$; that is, the classical connection $p = mv$ between the momentum and the velocity remains valid (recall that we set $m = 1/2$). Further, from the asymptotic expression for $|\psi(x, t)|^2$ it is clear that the square root $\Delta_\omega x$ of the variance of the coordinate for large values of $|t|$ satisfies the relation

$$\Delta_\omega x \cong 2\Delta_\omega p|t|$$

or

$$\Delta_\omega x \cong \Delta_\omega v|t|.$$

This means that the region in which there is a large probability of finding the particle will spread out with velocity $\Delta_\omega v$ while moving along the x -axis.

Even without computations it is not hard to see that the behavior of a classical free particle in a state with nonzero variance of coordinates and momenta will be exactly the same. Thus, quantum mechanics leads to almost the same results for a free particle as does classical mechanics. The only difference is that, according to the uncertainty relations, there are no states with zero variance of coordinates and momenta in quantum mechanics.

We mention some more formal properties of a solution of the Schrödinger equation for a free particle. It follows from (13) that

$$\psi(x, t) = O\left(\frac{1}{|t|^{1/2}}\right);$$

that is, for any x

$$|\psi(x, t)| < \frac{C}{|t|^{1/2}},$$

where C is a constant. It is thus natural to expect that the vector $\psi(t)$ tends weakly to zero as $|t| \rightarrow \infty$. It is simplest to prove this assertion in the momentum representation. For an arbitrary vector $f \in L^2(\mathbf{R})$,

$$(\psi(t), f) = \int_{\mathbf{R}} \varphi(p) e^{-ip^2 t} \overline{f(p)} dp.$$

In view of the Riemann–Lebesgue theorem, the last integral tends to zero as $|t| \rightarrow \infty$.

In the coordinate representation the weak convergence to zero of $\psi(t)$ has a very simple meaning. The constant vector f is given by a function which is markedly different from zero only in some finite region Ω , and the region in which $\psi(x, t)$ is nonzero goes to infinity while spreading out. Therefore, $(\psi(t), f) \rightarrow 0$ as $|t| \rightarrow \infty$.

We verify that the asymptotic behavior of the expression (13) for the function $\psi(x, t)$ has the correct normalization. As $|t| \rightarrow \infty$, we have

$$\int_{\mathbf{R}} |\psi(x, t)|^2 dx \cong \int_{\mathbf{R}} \frac{1}{2|t|} \left| \varphi\left(\frac{x}{2t}\right) \right|^2 dx = \int_{\mathbf{R}} |\varphi(k)|^2 dk = 1.$$

§ 16. The harmonic oscillator

In classical mechanics a harmonic oscillator is defined to be a system with the Hamiltonian function

$$H(p, q) = \frac{p^2}{2m} + \frac{m\omega^2}{2}q^2.$$

The parameter $\omega > 0$ has the sense of the frequency of oscillations.

The Schrödinger operator of the corresponding quantum mechanical system is

$$(1) \quad H = \frac{P^2}{2} + \frac{\omega^2 Q^2}{2}.$$

We use a system of units in which $m = 1$ and $\hbar = 1$. Our problem is to find the eigenvectors and eigenvalues of H . We solve this problem by using only the Heisenberg uncertainty relations for the operators P and Q and not by passing to a concrete representation. To this end we introduce the operators

$$(2) \quad \begin{aligned} a &= \frac{1}{\sqrt{2\omega}}(\omega Q + iP), \\ a^* &= \frac{1}{\sqrt{2\omega}}(\omega Q - iP). \end{aligned}$$

Using the Heisenberg relation

$$(3) \quad [Q, P] = i,$$

we get that

$$\omega aa^* = \frac{1}{2}(\omega^2 Q^2 + P^2) + \frac{i\omega}{2}(-QP + PQ) = H + \frac{\omega}{2},$$

$$\omega a^*a = \frac{1}{2}(\omega^2 Q^2 + P^2) + \frac{i\omega}{2}(QP - PQ) = H - \frac{\omega}{2},$$

or

$$(4) \quad H = \omega aa^* - \frac{\omega}{2},$$

$$(5) \quad H = \omega a^*a + \frac{\omega}{2}.$$

From (4) and (5) we find at once the commutation relation for the operators a and a^* :

$$(6) \quad [a, a^*] = 1.$$

From this it is easy to verify by induction that

$$(7) \quad [a, (a^*)^n] = n(a^*)^{n-1}.$$

Finally, we need the commutation relations of the operators a and a^* with H . To compute the commutator $[H, a]$, it suffices to multiply the formula (4) by a from the right, multiply (5) by a from the left, and find the difference between the expressions obtained:

$$(8) \quad [H, a] = -\omega a.$$

Similarly,

$$(9) \quad [H, a^*] = \omega a^*.$$

Let us now pass to a study of the spectrum of the operator H . Suppose that there is at least one eigenvalue E of H . Let ψ_E denote a corresponding eigenvector. We show that the eigenvalues E are bounded from below. By assumption, the vector ψ_E satisfies the equation $H\psi_E = E\psi_E$, which by (5) can be rewritten in the form

$$\left(\omega a^*a + \frac{\omega}{2} \right) \psi_E = E\psi_E.$$

Multiplying this equality from the left by ψ_E , we get that

$$\omega \|a\psi_E\|^2 + \frac{\omega}{2} \|\psi_E\|^2 = E\|\psi_E\|^2,$$

which implies at once that $E \geq \omega/2$, with equality possible only if $a\psi_E = 0$. From the expression (5) for H it is clear that if some

vector ψ satisfies the condition $a\psi = 0$, then it is an eigenvector of H corresponding to the eigenvalue $\omega/2$.

We show how to construct new eigenvectors from an arbitrary eigenvector ψ_E . Let us compute the expression $Ha\psi_E$ using (8):

$$Ha\psi_E = aH\psi_E - \omega a\psi_E = (E - \omega)a\psi_E.$$

It is clear from the last relation that either $a\psi_E$ is an eigenvector corresponding to the eigenvalue $E - \omega$, or $a\psi_E = 0$. If $a\psi_E \neq 0$, then either $a^2\psi_E$ is an eigenvector with eigenvalue $E - 2\omega$, or $a^2\psi_E = 0$. Thus, from an arbitrary eigenvector ψ_E we can construct a sequence of eigenvectors $\psi_E, a\psi_E, \dots, a^N\psi_E$ corresponding to the eigenvalues $E, E - \omega, \dots, E - N\omega$. However, this sequence cannot be infinite, because the eigenvalues of H are bounded from below by the number $\omega/2$. Therefore, there exists an $N \geq 0$ such that $a^N\psi_E \neq 0$ and $a^{N+1}\psi_E = 0$. Let the vector $a^N\psi_E$ be denoted by ψ_0 . This vector satisfies the equations

$$(10) \quad a\psi_0 = 0, \quad H\psi_0 = \frac{\omega}{2}\psi_0.$$

We see that the assumption that there is at least one eigenvector ψ_E is equivalent to the assumption that there is a vector satisfying (10). The vector ψ_0 describes the ground state of the oscillator, that is, the state with the smallest energy $\omega/2$.

Let us examine how the operator a^* acts on eigenvectors of H . Using (9), we get that

$$(11) \quad Ha^*\psi_E = a^*H\psi_E + \omega a^*\psi_E = (E + \omega)a^*\psi_E.$$

We point out the fact that $a^*\psi_E$ cannot be the zero vector. Indeed, it is obvious from the expression (4) for H that a vector satisfying the equation $a^*\psi = 0$ is an eigenvector of H with eigenvalue $-\omega/2$, which is impossible, since $E \geq \omega/2$. Therefore, it follows from (15.11) that $a^*\psi_E$ is an eigenvector of H with eigenvalue $(E + \omega)$. Similarly, $(a^*)^2\psi_E$ is an eigenvector with eigenvalue $(E + 2\omega)$. Beginning such a construction from the vector ψ_0 , we get an infinite sequence of eigenvectors $\psi_0, a^*\psi_0, \dots, (a^*)^n\psi_0, \dots$ corresponding to the eigenvalues $\omega/2, \omega/2 + \omega, \dots, (n+1/2)\omega, \dots$. Let the vector ψ_0 be normalized:

$\|\psi_0\| = 1$. We compute the norm of the vector $(a^*)^n \psi_0$:

$$\begin{aligned}\|(a^*)^n \psi_0\|^2 &= ((a^*)^n \psi_0, (a^*)^n \psi_0) \\ &= (\psi_0, a^{n-1} (a^*)^n a \psi_0) + n(\psi_0, a^{n-1} (a^*)^{n-1} \psi_0) \\ &= n \|(a^*)^{n-1} \psi_0\|^2 = \cdots = n! \|\psi_0\|^2 = n!.\end{aligned}$$

Here we used (7) and (10).

Thus, the sequence of normalized eigenvectors of the operator H can be given by the formula

$$(12) \quad \psi_n = \frac{1}{\sqrt{n!}} (a^*)^n \psi_0, \quad n = 0, 1, 2, \dots.$$

The orthogonality of the vectors ψ_n corresponding to different eigenvalues follows from general considerations, but it can also be verified directly:

$$(\psi_k, \psi_n) = \frac{1}{\sqrt{n! k!}} (\psi_0, a^k (a^*)^n \psi_0) = 0 \quad \text{for } k \neq n.$$

The last equality is obtained from (7) and (10).

We now discuss the question of uniqueness of the vector ψ_0 . Let \mathcal{H} be the Hilbert space spanned by the orthonormal system of vectors ψ_n . The elements $\varphi \in \mathcal{H}$ have the form

$$(13) \quad \varphi = \sum_n C_n \frac{(a^*)^n}{\sqrt{n!}} \psi_0, \quad \sum_n |C_n|^2 < \infty.$$

In this space we get a realization of the Heisenberg commutation relations if, in correspondence with (2), we let

$$(14) \quad Q = \frac{a + a^*}{\sqrt{2\omega}}, \quad P = \frac{\sqrt{\omega}(a - a^*)}{i\sqrt{2}}.$$

The relation (3) is then a consequence of the formula (6). The space \mathcal{H} is invariant under the action of the operators P and Q (more precisely, under the action of bounded functions $f(P)$ and $f(Q)$, for example, $U(u)$ and $V(v)$), and does not contain subspaces with the same property. Therefore, the representation of Q and P acting in \mathcal{H} by (14) is irreducible.

If in some representation there exist two vectors ψ_0 and ψ'_0 satisfying the equation (10), then together with \mathcal{H} we can construct a space \mathcal{H}' similarly from the vector ψ'_0 . This space will be invariant

with respect to P and Q ; that is, a representation in which there is more than one solution of (10) will be reducible.

§ 17. The problem of the oscillator in the coordinate representation

In the last section we found the spectrum of the Schrödinger operator for the oscillator in a purely algebraic way, under the single assumption that the operator H has at least one eigenvector. This is equivalent to the existence of a solution of the equation $a\psi_0 = 0$. We show that in the coordinate representation there does indeed exist a unique solution of this equation. Let us write the operators a and a^* in the coordinate representation:

$$(1) \quad a = \frac{1}{\sqrt{2\omega}} \left(\omega x + \frac{d}{dx} \right),$$

$$(2) \quad a^* = \frac{1}{\sqrt{2\omega}} \left(\omega x - \frac{d}{dx} \right).$$

The equation $a\psi_0 = 0$ takes the form

$$\omega x\psi_0(x) + \psi'_0(x) = 0.$$

Separating the variables, we get that

$$\frac{d\psi_0}{\psi_0} = -\omega x dx$$

and

$$\psi_0(x) = Ce^{-\frac{\omega x^2}{2}}.$$

The constant C is found from the normalization condition

$$\int_{-\infty}^{\infty} |\psi_0(x)|^2 dx = |C|^2 \int_{-\infty}^{\infty} e^{-\omega x^2} dx = |C|^2 \left(\frac{\pi}{\omega}\right)^{\frac{1}{2}} = 1.$$

This condition is satisfied by $C = (\omega/\pi)^{1/4}$, and the normalized eigenfunction for the ground state has the form

$$\psi_0(x) = \left(\frac{\omega}{\pi}\right)^{\frac{1}{4}} e^{-\frac{\omega x^2}{2}}.$$

The eigenfunctions $\psi_n(x)$ for the excited states are found by the formula (16.12) using (2):

$$\psi_n(x) = \left(\frac{\omega}{\pi}\right)^{\frac{1}{4}} \frac{(2\omega)^{-\frac{n}{2}}}{\sqrt{n!}} \left(\omega x - \frac{d}{dx}\right)^n e^{-\frac{\omega x^2}{2}}.$$

Obviously, these functions have the form

$$\psi_n(x) = P_n(x) e^{-\frac{\omega x^2}{2}},$$

where $P_n(x)$ is an n th-degree polynomial. It can be shown that $P_n(x) = H_n(\sqrt{\omega}x)$, where $H_n(\xi)$ is the n th Chebyshev–Hermite polynomial. It is known that the system of functions $H_n(\xi)e^{-\frac{\xi^2}{2}}$ is complete in $L^2(\mathbf{R})$. This assertion follows in general from the irreducibility of the coordinate representation and results in the preceding section.

The function $|\psi_n(x)|^2$ is the density of the coordinate distribution function in the n th state of the oscillator. It is interesting to compare this distribution with the corresponding classical distribution. The solution of the classical problem of an oscillator has the form

$$(3) \quad x(t) = A \sin(\omega t + \varphi),$$

where A is the amplitude of oscillation and φ is the initial phase. The energy of the oscillations can be computed from the formula

$$E = \frac{\dot{x}^2}{2} + \frac{\omega^2 x^2}{2}$$

and is equal to $\omega^2 A^2 / 2$. The corresponding density of the coordinate distribution function has the form

$$(4) \quad F(x) = \delta(x - x(t)).$$

It is clear that a stationary state of a quantum oscillator can under no conditions pass into a classical pure state given by the formula (3). It is natural to assume that the limit of a quantum state will be a classical mixed state that is a convex combination of solutions (3) with random phases φ . For such a state the density of the coordinate

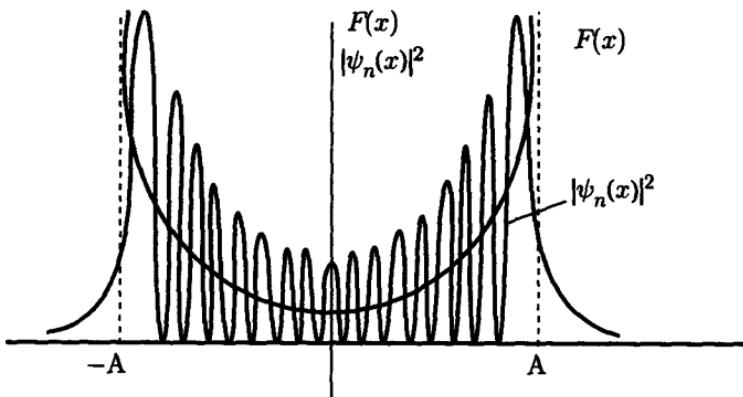


Figure 5

distribution function is obtained by averaging (4):

$$\begin{aligned}
 F(x) &= \frac{1}{2\pi} \int_0^{2\pi} \delta(A \sin(\omega t + \varphi) - x) d\varphi \\
 &= \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \delta(A \sin \varphi - x) d\varphi \\
 &= \frac{1}{\pi} \int_{-A}^A \frac{\delta(x - y) dy}{\sqrt{A^2 - y^2}} \\
 &= \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{A^2 - x^2}} & x \in (-A, A), \\ 0, & x \notin (-A, A), \end{cases}
 \end{aligned}$$

or, briefly,

$$F(x) = \frac{\theta(A^2 - x^2)}{\pi \sqrt{A^2 - x^2}}.$$

The graphs of the functions $F(x)$ and $|\psi_n(x)|^2$ are shown in Figure 5 for sufficiently large n under the condition that²⁴ $E_n = \hbar(n + 1/2)\omega = \omega^2 A^2 / 2$. As $n \rightarrow \infty$, the quantum distribution will tend to the classical distribution. The condition $n \rightarrow \infty$ can be satisfied for a given energy E if $\hbar \rightarrow 0$.

We point out an important difference between the functions $F(x)$ and $|\psi_n(x)|^2$: $F(x) = 0$ for $|x| > A$; that is, the classical particle is

²⁴Here for convenience we write out Planck's constant \hbar explicitly, not assuming that $\hbar = 1$

always between $-A$ and A (in the classically allowed region), while the function $|\psi_n(x)|^2$ does not vanish for $|x| > A$. This means that the quantum particle can be observed with finite probability in the classically forbidden region. For an arbitrary potential $V(x)$ this region is determined by the condition that the total energy E is $< V(x)$, which corresponds to negative values of the classical kinetic energy.

§ 18. Representation of the states of a one-dimensional particle in the sequence space l_2

According to the results in § 16, any vector $\varphi \in \mathcal{H}$ can be expanded in a series of eigenvectors of the energy operator of the harmonic oscillator:

$$\varphi = \sum_n C_n \frac{(a^*)^n}{\sqrt{n!}} \psi_0, \quad \sum_n |C_n|^2 < \infty.$$

Each vector $\varphi \in \mathcal{H}$ is uniquely determined by the coefficient sequence $\{C_n\}$, that is,

$$\varphi \leftrightarrow \{C_n\}.$$

If $\psi \leftrightarrow \{B_n\}$, then

$$(\psi, \varphi) = \sum_n B_n \bar{C}_n$$

in view of the orthonormality of the system of vectors

$$\psi_n = \frac{(a^*)^n}{\sqrt{n!}} \psi_0.$$

Let us consider how the operators act in such a representation. For this it suffices to construct the operators a and a^* . Suppose that $\varphi \leftrightarrow \{C_n\}$ and $a\varphi \leftrightarrow \{C'_n\}$. We must express C'_n in terms of C_n .

This can be done as follows:

$$\begin{aligned}
 a\varphi &= a \sum_n C_n \frac{(a^*)^n}{\sqrt{n!}} \psi_0 \\
 &= \sum_n C_n \frac{n(a^*)^{n-1}}{\sqrt{n!}} \psi_0 \\
 (1) \quad &= \sum_n \sqrt{n} C_n \frac{(a^*)^{n-1}}{\sqrt{(n-1)!}} \psi_0 \\
 &= \sum_n \sqrt{n+1} C_{n+1} \frac{(a^*)^n}{\sqrt{n!}} \psi_0.
 \end{aligned}$$

The calculation used the relations $[a, (a^*)^n] = n(a^*)^{n-1}$ and $a\psi_0 = 0$, and the last equality is obtained by the change $n \rightarrow n + 1$ of the summation index. It is clear from (1) that

$$(2) \quad C'_n = \sqrt{n+1} C_n.$$

Similarly,

$$\begin{aligned}
 a^*\varphi &= a^* \sum_n C_n \frac{(a^*)^n}{\sqrt{n!}} \psi_0 \\
 &= \sum_n C_n \frac{(a^*)^{n+1}}{\sqrt{n!}} \psi_0 \\
 &= \sum_n \sqrt{n+1} C_n \frac{(a^*)^{n+1}}{\sqrt{(n+1)!}} \psi_0 \\
 &= \sum_n \sqrt{n} C_{n-1} \frac{(a^*)^n}{\sqrt{n!}} \psi_0.
 \end{aligned}$$

Therefore, if $a^*\varphi \leftrightarrow \{C''_n\}$, then

$$(3) \quad C''_n = \sqrt{n} C_{n-1}.$$

The formulas (2) and (3) become especially intuitive in matrix notation. Let us write the sequence $\{C_n\}$ as a column:

$$\varphi \leftrightarrow \begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \end{pmatrix}.$$

Then the operators a and a^* can be written as the matrices

(4)

$$a \leftrightarrow \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad a^* \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & \dots \\ \sqrt{1} & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

We verify that these expressions are equivalent to the relations (2) and (3). Indeed,

$$a\varphi \leftrightarrow \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \sqrt{1}C_1 \\ \sqrt{2}C_2 \\ \sqrt{3}C_3 \\ \vdots \end{pmatrix},$$

which coincides with (2); the connection between the formulas (3) and (4) can be verified similarly. The commutation relation $[a, a^*] = 1$ can be verified immediately for a and a^* in the representation (4). The eigenvectors of the operator H in this representation have the form

$$\psi_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \quad \psi_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \quad \psi_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \end{pmatrix}, \dots$$

The vector ψ_0 obviously satisfies the equation $a\psi_0 = 0$.

The operators a^* and a are often called the operators of creation and annihilation of excitation, due to the fact that a^* carries a state with energy E to a state with greater energy $E + \omega$, while a carries a state with energy E to a state with energy $E - \omega$ (the ground state ψ_0 is annihilated by a). Sometimes the so-called excitation number operator $N = a^*a$ is introduced. In our representation it has the form

$$N \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 2 & 0 & \dots \\ 0 & 0 & 0 & 3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The eigenvalues of this operator coincide with the ordinal number of the excited state ψ_n .

Finally, we write the operators H , P , and Q in this representation:

$$(5) \quad H = \omega a^* a + \frac{\omega}{2} \leftrightarrow \begin{pmatrix} \frac{\omega}{2} & 0 & 0 & \dots \\ 0 & \frac{3}{2}\omega & 0 & \dots \\ 0 & 0 & \frac{5}{2}\omega & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix},$$

$$(6) \quad P = \frac{\sqrt{\omega}(a - a^*)}{i\sqrt{2}} \leftrightarrow \frac{\sqrt{\omega}}{i\sqrt{2}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ -\sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & -\sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & -\sqrt{3} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

$$(7) \quad Q = \frac{a + a^*}{\sqrt{2\omega}} \leftrightarrow \frac{1}{\sqrt{2\omega}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

It is clear from these formulas that H is represented by a diagonal matrix; that is, this representation is an eigenrepresentation for the Schrödinger operator of the harmonic oscillator. Further, it is at once obvious that P and Q are self-adjoint operators, and it can easily be verified that they satisfy the commutation relation $[Q, P] = i$.

In connection with the representation of states in the space l_2 , we would like to say something about the original Heisenberg matrix formulation of quantum mechanics. Finding the admissible values of the energy of a system was a basic problem in the initial stage of the development of quantum mechanics. For a system with one degree of freedom, Heisenberg proposed the following recipe. He considered a classical system with Hamiltonian function $H(q, p) = p^2/2m + V(q)$. Self-adjoint matrices Q and P satisfying the relation $[Q, P] = i$ were constructed (such matrices are not uniquely determined). The matrix $H = P^2/2m + V(Q)$ was constructed next. The last step consisted in diagonalization of this matrix, and the eigenvalues of H were identified with the admissible values of the energy.

The formulas (6) and (7) give an example of matrices P and Q satisfying the Heisenberg commutation relations. These matrices

were chosen so that the matrix of the operator H for the oscillator is at once diagonal. However, for an arbitrary H it is not possible to avoid this diagonalization step.

We see that the Heisenberg formulation essentially coincides with the contemporary formulation of quantum mechanics if \mathcal{D} is taken as the state space.

§ 19. Representation of the states for a one-dimensional particle in the space \mathcal{D} of entire analytic functions

We consider the set of functions of a complex variable of the form

$$(1) \quad f(z) = \sum_{n=0}^{\infty} C_n \frac{z^n}{\sqrt{n!}}, \quad \sum_n |C_n|^2 < \infty.$$

This set \mathcal{D} of functions becomes a Hilbert space if the scalar product is defined by

$$(2) \quad (f_1, f_2) = \frac{1}{\pi} \int f_1(z) \overline{f_2(z)} e^{-|z|^2} d\mu(z).$$

The integral is taken over the complex plane, and $d\mu(z) = dx dy$.

Let us verify that the functions $z^n/\sqrt{n!}$ form an orthonormal basis in \mathcal{D} . For this we compute the integral

$$I_{nm} = \int z^n \bar{z}^m e^{-|z|^2} d\mu(z) = \int_0^\infty \rho d\rho \int_0^{2\pi} d\varphi \rho^{n+m} e^{i\varphi(n-m)} e^{-\rho^2}.$$

For $n \neq m$ we have $I_{mn} = 0$ due to the integration with respect to φ . For $n = m$,

$$I_{nn} = 2\pi \int_0^\infty \rho^{2n+1} e^{-\rho^2} d\rho = \pi \int_0^\infty t^n e^{-t} dt = \pi n!.$$

An arbitrary state $\varphi \in \mathcal{H}$ with $\varphi = \sum_n C_n \frac{(a^*)^n}{\sqrt{n!}} \psi_0$ can be represented by the function $f(z) = \sum_n C_n \frac{z^n}{\sqrt{n!}}$: $\varphi \leftrightarrow f(z)$. The eigenvectors ψ_n of the oscillator are represented by the basis functions $z^n/\sqrt{n!}$.

We consider how the operators a and a^* act in this representation. Using the computations which led us to the formulas (18.2) and (18.3),

we can write the vectors $a\varphi$ and $a^*\varphi$ in the form

$$a\varphi = \sum_n C_n \frac{n(a^*)^{n-1}}{\sqrt{n!}} \psi_0, \quad a^*\varphi = \sum_n C_n \frac{(a^*)^{n+1}}{\sqrt{n!}} \psi_0.$$

These vectors are represented by the functions

$$\begin{aligned} a\varphi &\leftrightarrow \sum_n C_n \frac{nz^{n-1}}{\sqrt{n!}} = \frac{d}{dz} f(z), \\ a^*\varphi &\leftrightarrow \sum_n C_n \frac{z^{n+1}}{\sqrt{n!}} = zf(z); \end{aligned}$$

that is, for a and a^* we have obtained the representation

$$a \leftrightarrow \frac{d}{dz}, \quad a^* \leftrightarrow z.$$

Here are the corresponding formulas for the operators Q , P , H :

$$\begin{aligned} Q &\leftrightarrow \frac{1}{\sqrt{2\omega}} \left(\frac{d}{dz} + z \right), \\ P &\leftrightarrow \frac{\sqrt{\omega}}{i\sqrt{2}} \left(\frac{d}{dz} - z \right), \\ H &\leftrightarrow \omega \left(z \frac{d}{dz} + \frac{1}{2} \right). \end{aligned}$$

All the basic relations ($[a, a^*] = 1$, $[Q, P] = i$, $H\psi_n = \omega(n+1/2)\psi_n$) can easily be verified in this representation. The representation constructed can turn out to be convenient if the observables under study are polynomials in Q and P .

§ 20. The general case of one-dimensional motion

In the preceding sections we considered two one-dimensional problems of quantum mechanics: the problem of a free particle and the problem of a harmonic oscillator. The free particle gives us an example of a system with a continuous spectrum for the Schrödinger operator, and the harmonic oscillator gives an example of a system with a pure point spectrum. In most real physical problems the spectrum is

more complicated. Let us consider the problem of the spectrum of a Schrödinger operator

$$H = -\frac{d^2}{dx^2} + V(x)$$

with very general assumptions about the potential.

The forces acting on the particle usually are appreciably different from zero in some finite region on the x -axis and tend to zero as $|x| \rightarrow \infty$; therefore, it is most common to encounter potentials $V(x)$ that tend to constant values as $x \rightarrow \pm\infty$. For simplicity we confine ourselves to the case when the potential is exactly equal to constants for $x < -a$ and $x > a$. By the arbitrariness in the definition of a potential, we can always take one of these constants equal to zero.

We consider Schrödinger equations

$$(1) \quad \psi'' + E\psi = V(x)\psi$$

under the condition that $V(x)$ is a continuous function on the real axis with $V(x) = 0$ for $x < -a$ and $V(x) = V_0$ for $x > a$. For definiteness we assume that $V_0 > 0$. The graph of the potential is pictured in Figure 6.

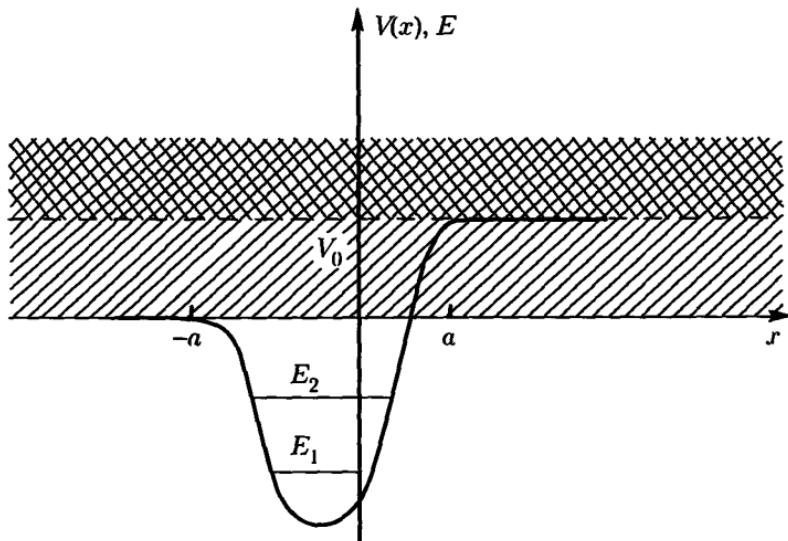


Figure 6

For $x < -a$ and $x > a$ the Schrödinger equation (1) simplifies:

$$(2) \quad \psi'' + E\psi = 0, \quad x < -a,$$

$$(3) \quad \psi'' + (E - V_0)\psi = 0, \quad x > a.$$

For any values of E there are two linearly independent solutions of (1), which we denote by ψ_1 and ψ_2 . The general solution of this equation is

$$(4) \quad \psi = C_1\psi_1 + C_2\psi_2.$$

In our study of the spectrum of the operator H we are interested in either square-integrable solutions of (1) that are eigenfunctions of H or solutions that are bounded on the whole real axis. With the help of the latter we can describe the continuous spectrum of H .

We now consider three cases.

1) $E < 0$.

We rewrite the equations (2) and (3) in the form

$$\psi'' - \kappa^2\psi = 0, \quad x < -a, \quad \kappa^2 = -E > 0, \quad \kappa > 0;$$

$$\psi'' - \kappa_1^2\psi = 0, \quad x > a. \quad \kappa_1^2 = -(E - V_0) > 0, \quad \kappa_1 > 0.$$

The functions $e^{\pm\kappa x}$ for $x < -a$ and $e^{\pm\kappa_1 x}$ for $x > a$ are linearly independent solutions of these equations. Therefore, an arbitrary solution (4) of the equation (1) in the region $x < -a$ has the form $C'_1 e^{-\kappa x} + C''_2 e^{\kappa x}$, and in the region $x > a$ the form $C''_1 e^{-\kappa_1 x} + C''_2 e^{\kappa_1 x}$. Here C'_1 , C''_2 , C''_1 , and C''_2 are constants that depend linearly on C_1 and C_2 in the formula (4). The solution ψ will be square integrable under the conditions $C'_1 = 0$ and $C''_2 = 0$. Just one of these conditions is enough for ψ to be determined up to a numerical factor. From the condition $C'_1 = 0$, one can find the ratio of the coefficients C_1 and C_2 , which will depend on the parameter E :

$$\frac{C_1}{C_2} = F_1(E).$$

Similarly, from the condition $C''_2 = 0$, we get that

$$\frac{C_1}{C_2} = F_2(E).$$

A square-integrable solution ψ will exist only for values of E such that

$$F_1(E) = F_2(E).$$

The roots of this algebraic equation, if they exist, are the eigenvalues of the operator H . From the foregoing it is natural to expect the presence of a simple point spectrum for $E < 0$.

2) $0 < E < V_0$.

We write the equations (2) and (3) in the form

$$\begin{aligned} \psi'' + k^2\psi = 0, \quad x < -a, \quad k^2 = E > 0, \quad k > 0; \\ \psi'' - \kappa_1^2\psi = 0, \quad x > a, \quad \kappa_1^2 = -(E - V_0) > 0, \quad \kappa_1 > 0. \end{aligned}$$

The functions $e^{\pm ikx}$ for $x < -a$ and $e^{\pm \kappa_1 x}$ for $x > a$ are linearly independent solutions. It is immediately obvious that there are no square-integrable solutions, but a bounded solution can be constructed if we choose C_1/C_2 so that ψ has the form $C_1''e^{-\kappa_1 x}$ for $x > a$. Therefore, in the interval $0 < E < V_0$ the spectrum is simple and continuous.

3) $E > V_0$.

In this case both equations (2) and (3) have oscillatory solutions ($e^{\pm ikx}$ for $x < -a$ and $e^{\pm ik_1 x}$ for $x > a$, $k_1^2 = E - V_0$), hence any solution of (1) is bounded, and there are no square-integrable solutions. For $E > V_0$ the spectrum of H is continuous and has multiplicity two.

In Figure 6 the eigenvalues of H are represented by horizontal lines, the ordinary shading shows the region of the simple continuous spectrum, and the double shading shows the region of the spectrum of multiplicity two.

We discuss the physical meaning of the solutions of (1). The square-integrable solutions describe the stationary states with energy equal to an eigenvalue. These functions decay exponentially as $|x| \rightarrow \infty$, and thus the probability of observing the particle outside some finite region is close to zero. It is clear that such states correspond to a finite motion of the particle. The eigenfunctions of the continuous spectrum do not have an immediate physical meaning, since they do not belong to the state space. However, with their help we can construct states of wave packet type which we considered for a free particle. These states can be interpreted as states with

almost given energy. A study of the evolution of such states shows that they describe a particle that goes to infinity (infinite motion) as $|t| \rightarrow \infty$. We shall return to this question when we study the theory of scattering.

In classical mechanics, as in quantum mechanics, the motion is finite for $E < 0$ and infinite for $E > 0$. For $0 < E < V_0$ the particle can go to infinity in one direction, and for $E > V_0$ in two directions. We direct attention to the fact that the multiplicity of the continuous spectrum coincides with the number of directions in which the particle can go to infinity.

In the example of a particle in a one-dimensional potential well we consider the question of the classical limit of quantum stationary states. For the computation of the limit (14.15) it is convenient to use the asymptotic form of the solution of the Schrödinger equation as $\hbar \rightarrow 0$. Methods for constructing asymptotic solutions of the Schrödinger equation as $\hbar \rightarrow 0$ are called quasi-classical methods. We use one such method: the Wentzel Kramers–Brillouin (WKB) method.

We write the Schrödinger equation in the form

$$(5) \quad \psi'' + \frac{E - V(x)}{\hbar^2} \psi = 0.$$

Here, as earlier, we use a system of units in which $m = 1/2$. Setting $\psi(x) = \exp\left[\frac{i}{\hbar} \int_{x_0}^x g(x) dx\right]$, we get an equation for $g(x)$:

$$(6) \quad i\hbar g' - g^2 + E - V = 0.$$

The solution of this equation will be sought in the form of a power series in \hbar/i :

$$(7) \quad g(x) = \sum_{k=0}^{\infty} \left(\frac{\hbar}{i}\right)^k g_k(x).$$

Substituting (7) in (5), we have

$$-\sum_{n=1}^{\infty} \left(\frac{\hbar}{i}\right)^n g'_{n-1} - \sum_{n=1}^{\infty} \sum_{k=0}^n \left(\frac{\hbar}{i}\right)^n g_k g_{n-k} - g_0^2 + E - V = 0.$$

Equating the coefficients of $(h/i)^n$, we get a system of recursive equations for the functions $g_k(x)$:

$$(8) \quad g_0^2 = E - V.$$

$$(9) \quad g'_{n-1} = - \sum_{k=0}^n g_k g_{n-k}.$$

From (8) we find $g_0(x)$:

$$g_0(x) = \pm \sqrt{E - V(x)} = \pm p(x).$$

Here $p(x)$ with $E \geq V(x)$ is the classical expression for the absolute value of the momentum of a particle with energy E in the field $V(x)$. For $E < V(x)$ the function $p(x)$ becomes purely imaginary. Setting $n = 1$, we get from (9) that

$$g'_0 = -2g_0 g_1, \quad g_1 = -\frac{1}{2} \frac{g'_0}{g_0} = -\frac{1}{2} \frac{d}{dx} \log |p(x)|.$$

Confining ourselves to these terms of the expansion (7), we get the asymptotic form as $h \rightarrow 0$ for the two linearly independent solutions of the Schrödinger equation:

$$(10) \quad \psi_{1,2} = \frac{1}{\sqrt{|p(x)|}} \exp \left[\pm \frac{i}{h} \int_{x_0}^x p(x) dx \right] + O(h)$$

with $p(x) \neq 0$. The functions (10) are sometimes called WKB-solutions of the Schrödinger equation.

The precise theory of the WKB method is fairly complicated. It is known that in the general case the series (10.7) diverges and is an asymptotic series. A finite number of terms of this series enables us to construct a good approximation for the function ψ if Planck's constant h can be regarded as sufficiently small under the conditions of the specific problem.

In what follows we assume that $V(x) = 0$ for $|x| \geq a$ and that $E < 0$. We suppose that for $\min_x V(x) < E < 0$ there are two points x_1 and x_2 ($-a \leq x_1 < x_2 \leq a$) satisfying the condition $E - V(x) = 0$. These are so-called turning points, at which according to classical mechanics the particle reverses direction. It is not hard to see that in the classically forbidden region ($x < x_1$ or $x > x_2$) one of the WKB-solutions increases exponentially while the other decays upon

going farther from the turning point into the depth of the forbidden region. For $|x| > a$, the WKB-solutions coincide with the exact solutions and have the form $e^{\pm \kappa x}$, where $E = -\kappa^2$. We recall that the eigenfunctions of the discrete spectrum of H decrease exponentially as $x \rightarrow \pm\infty$. As $\hbar \rightarrow 0$, an eigenfunction must coincide in the allowed region with some linear combination $C_1\psi_1 + C_2\psi_2$ of the WKB-solutions. The construction of such a linear combination is a fairly complicated problem since the WKB-solutions become meaningless at the turning points. It can be shown that the conditions for decrease of the function $\psi(x)$ hold as $x \rightarrow -\infty$ if

$$(11) \quad \psi(x) = \frac{C}{\sqrt{p(x)}} \sin \left(\frac{1}{\hbar} \int_{x_1}^x p(x) dx + \frac{\pi}{4} \right) + O(\hbar).$$

Similarly, it follows from the conditions of decrease as $x \rightarrow +\infty$ that

$$(12) \quad \psi(x) = \frac{C'}{\sqrt{p(x)}} \sin \left(\frac{1}{\hbar} \int_x^{x_2} p(x) dx + \frac{\pi}{4} \right) + O(\hbar).$$

These two expressions for $\psi(x)$ coincide if

$$(13) \quad \int_{x_1}^{x_2} p(x) dx = \pi \hbar \left(n + \frac{1}{2} \right), \quad n = 0, 1, 2, \dots$$

This condition determines the energy eigenvalues in the quasi-classical approximation and corresponds to the Bohr-Sommerfeld quantization rule in the old quantum theory.

We proceed to the computation of the classical limit of a quantum state. The limit as $\hbar \rightarrow 0$ can be found under various conditions. For example, we can consider a state corresponding to the energy eigenvalue E_n for a fixed value of n in the condition (13). It is easy to see that then $E_n \rightarrow V_0 = \min_x V(x)$ as $\hbar \rightarrow 0$, and the state of a particle at rest on the bottom of the potential well is obtained in the limit. Let us analyze a more interesting case: $\hbar \rightarrow 0$, $n \rightarrow \infty$, and the energy E remains constant. (We remark that in the given case the integral on the left-hand side of (13) does not depend on \hbar , and $\hbar \rightarrow 0$, running through some sequence of values.) Substituting the asymptotic expression (11) for $\psi(x)$ in the formula (14.15), we get

that

$$\begin{aligned}
 F(x, u) &= \frac{1}{2\pi} \lim_{h \rightarrow 0} \psi(x + uh) \overline{\psi(x)} \\
 &= \lim_{h \rightarrow 0} \frac{C}{\sqrt{p(x + uh) p(x)}} \sin \left(\frac{1}{h} \int_{x_1}^{x+uh} p(x) dx + \frac{\pi}{4} \right) \\
 &\times \sin \left(\frac{1}{h} \int_{x_1}^x p(x) dx + \frac{\pi}{4} \right) = \frac{C}{p(x)} \lim_{h \rightarrow 0} \left[\cos \left(\frac{1}{h} \int_x^{x+uh} p(x) dx \right) \right. \\
 &\quad \left. - \cos \left(\frac{1}{h} \int_{x_1}^{x+uh} p(x) dx + \frac{1}{h} \int_{x_1}^x p(x) dx + \frac{\pi}{2} \right) \right].
 \end{aligned}$$

We denote all normalization factors by C . The limit of the second term in the generalized function sense is zero, therefore,

$$F(x, u) = \frac{C}{p(x)} \cos(p(x)u).$$

Using (14.16), we find the distribution function for the limiting classical state:

$$\begin{aligned}
 \rho(q, p) &= \frac{C}{p(q)} \int_{-\infty}^{\infty} e^{-ipu} \cos(p(q)u) du \\
 &= \frac{C}{p(q)} \int_{-\infty}^{\infty} e^{-ipu} (e^{ip(q)u} + e^{-ip(q)u}) du.
 \end{aligned}$$

Finally, we get that

$$(14) \quad \rho(p, q) = \frac{C}{p(q)} [\delta(p - p(q)) + \delta(p + p(q))]$$

The state described by the function $\rho(q, p)$ has a very simple meaning. In this state the density of the coordinate distribution function is inversely proportional to the classical velocity of the particle, and the momentum of the particle at the point q can take the two values $\pm p(q)$ with equal probability. The formula (14) was obtained for the allowed region. In the same way it is not hard to verify that $\rho(q, p) = 0$ in the forbidden region.

§ 21. Three-dimensional problems in quantum mechanics. A three-dimensional free particle

The Schrödinger operator for a free particle in the coordinate representation has the form

$$(1) \quad H = -\frac{\hbar^2}{2m} \Delta.$$

The equation for the eigenfunctions (for $\hbar = 1$ and $m = 1/2$) is

$$(2) \quad -\Delta\psi = k^2\psi, \quad E = k^2 > 0,$$

and it has the solutions

$$(3) \quad \psi_{\mathbf{k}}(\mathbf{x}) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} e^{i\mathbf{k}\cdot\mathbf{x}}.$$

The normalization constant is chosen so that

$$\int_{\mathbf{R}^3} \psi_{\mathbf{k}}(\mathbf{x}) \overline{\psi_{\mathbf{k}'}(\mathbf{x})} d\mathbf{x} = \delta(\mathbf{k} - \mathbf{k}').$$

We see that the spectrum of H is positive and continuous, and has infinite multiplicity. To each direction of the vector \mathbf{k} there corresponds the eigenfunction (3) with the eigenvalue k^2 . Therefore, there are as many eigenfunctions as there are points on the unit sphere.

As in the one-dimensional case, the solution of the Cauchy problem

$$i \frac{d\psi}{dt} = H\psi, \quad \psi(0) = \varphi$$

for the nonstationary Schrödinger equation is most easily obtained in the momentum representation

$$i \frac{\partial\psi(\mathbf{p}, t)}{\partial t} = p^2\psi(\mathbf{p}, t), \quad \psi(\mathbf{p}, 0) = \varphi(\mathbf{p}).$$

It is obvious that

$$\psi(\mathbf{p}, t) = \varphi(\mathbf{p}) e^{-ip^2 t}.$$

Passing to the coordinate representation, we get that

$$\psi(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} \varphi(\mathbf{p}) e^{i(\mathbf{p}\cdot\mathbf{x} - p^2 t)} d\mathbf{p} = \int_{\mathbf{R}^3} \varphi(\mathbf{k}) \psi_{\mathbf{k}}(\mathbf{x}) e^{ik^2 t} dk.$$

Just as in the one-dimensional case, the functions $\psi(\mathbf{x}, t)$ or $\psi(\mathbf{p}, t)$ describe an infinite motion of the particle with a momentum distribution function independent of the time. Using the stationary phase method, we can show that a region in which there is a large probability of observing the particle moves with the classical velocity $\mathbf{v} = 2\mathbf{p}_0$ (we assume that the support of the function $\varphi(\mathbf{p})$ is concentrated in a neighborhood of the point \mathbf{p}_0). We have the estimate

$$|\psi(\mathbf{x}, t)| < \frac{C}{|t|^{3/2}},$$

where C is a constant. Finally, as in the one-dimensional case,

$$\lim_{t \rightarrow \infty} (\xi, \psi(t)) = 0$$

for any $\xi \in \mathcal{H}$.

§ 22. A three-dimensional particle in a potential field

In the coordinate representation, the Schrödinger operator for a particle in a potential field has the form

$$H = -\frac{\hbar^2}{2m} \Delta + V(\mathbf{x}).$$

The importance of the problem of the motion of a particle in a potential field is explained by the fact that the problem of the motion of two bodies reduces to it (as in classical mechanics). We show how this is done in quantum mechanics. Let us consider a system of two particles with masses m_1 and m_2 with mutual interaction described by the potential $V(\mathbf{x}_1 - \mathbf{x}_2)$. We write the Schrödinger operator of this system in the coordinate representation:

$$H = -\frac{\hbar^2}{2m_1} \Delta_1 - \frac{\hbar^2}{2m_2} \Delta_2 + V(\mathbf{x}_1 - \mathbf{x}_2).$$

Here Δ_1 and Δ_2 are the Laplace operators with respect to the coordinates of the first and second particles, respectively.

We introduce the new variables

$$\mathbf{X} = \frac{m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2}{m_1 + m_2}, \quad \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2;$$

\mathbf{X} is the coordinate of the center of inertia of the system, and \mathbf{x} gives the relative coordinates. Simple computations lead to the expression for H in the new variables:

$$H = -\frac{\hbar^2}{2M} \Delta_{\mathbf{X}} - \frac{\hbar^2}{2\mu} \Delta_{\mathbf{x}} + V(\mathbf{x}).$$

Here $M = m_1 + m_2$ is the total mass of the system, and $\mu = m_1 m_2 / (m_1 + m_2)$ is the so-called reduced mass. The first term in H can be interpreted as the kinetic energy operator of the center of inertia of the system, and the operator

$$H_1 = -\frac{\hbar^2}{2\mu} \Delta + V(\mathbf{x})$$

is the Schrödinger operator for the relative motion. In the equation

$$H\Psi = E\Psi,$$

the variables are separated, and the solutions of such an equation can be found in the form

$$\Psi(\mathbf{X}, \mathbf{x}) = \psi(\mathbf{X}) \psi_1(\mathbf{x}).$$

The functions $\psi(\mathbf{X})$ and $\psi_1(\mathbf{x})$ satisfy the equations ($\hbar = 1$)

$$\begin{aligned} -\frac{1}{2M} \Delta \psi(\mathbf{X}) &= \varepsilon \psi(\mathbf{X}), \\ -\frac{1}{2\mu} \Delta \psi_1(\mathbf{x}) + V(\mathbf{x}) \psi_1(\mathbf{x}) &= \varepsilon_1 \psi_1(\mathbf{x}), \end{aligned}$$

with $E = \varepsilon + \varepsilon_1$. The first of these equations has the solutions

$$\psi_{\mathbf{K}}(\mathbf{X}) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} e^{i\mathbf{K}\mathbf{X}}, \quad \frac{K^2}{2M} = \varepsilon.$$

The problem reduces to the solution of the second equation, which coincides in form with the Schrödinger equation for a particle with mass μ in the potential field $V(\mathbf{x})$. We note that the spectrum of the operator H is always continuous, since the spectrum of the operator $-\frac{1}{2M} \Delta$ is continuous.

The most important case of the problem of the motion of a particle in a potential field is the problem of motion in a central force field. In this case the potential $V(\mathbf{x}) = V(|\mathbf{x}|)$ depends only on $|\mathbf{x}| = r$. The problem of two particles reduces to the problem of a central field if the interaction potential depends only on the distance between the

particles. Before proceeding to this problem, we study properties of the angular momentum and some questions in the theory of representations of the rotation group. This will enable us to explicitly take into account the spherical symmetry of the problem.

§ 23. Angular momentum

In quantum mechanics the operators of the projections of the angular momentum are defined by the formulas

$$(1) \quad \begin{aligned} L_1 &= Q_2 P_3 - Q_3 P_2, \\ L_2 &= Q_3 P_1 - Q_1 P_3, \\ L_3 &= Q_1 P_2 - Q_2 P_1. \end{aligned}$$

We introduce one more observable, called the square of the angular momentum:

$$(2) \quad L^2 = L_1^2 + L_2^2 + L_3^2.$$

Let us find the commutation relations for the operators L_1 , L_2 , L_3 , and L^2 . Using the Heisenberg relations $[Q_j, P_k] = i\delta_{jk}$ and the properties of the commutators, we get that

$$\begin{aligned} [L_1, L_2] &= [Q_2 P_3 - Q_3 P_2, Q_3 P_1 - Q_1 P_3] \\ &= Q_2 P_1 [P_3, Q_3] + Q_1 P_2 [Q_3, P_3] \\ &= i(Q_1 P_2 - Q_2 P_1) \\ &= iL_3. \end{aligned}$$

Thus, the operators L_1 , L_2 , L_3 satisfy the following commutation relations:

$$(3) \quad \begin{aligned} [L_1, L_2] &= iL_3, \\ [L_2, L_3] &= iL_1, \\ [L_3, L_1] &= iL_2. \end{aligned}$$

It is not hard to verify that all the operators L_1 , L_2 , L_3 commute with L^2 :

$$(4) \quad [L_j, L^2] = 0, \quad j = 1, 2, 3.$$

Indeed,

$$\begin{aligned}[L_1, L^2] &= [L_1, L_1^2 + L_2^2 + L_3^2] \\&= [L_1, L_2^2] + [L_1, L_3^2] \\&= [L_1, L_2]L_2 + L_2[L_1, L_2] + [L_1, L_3]L_3 + L_3[L_1, L_3] \\&= iL_3L_2 + iL_2L_3 - iL_2L_3 - iL_3L_2 = 0.\end{aligned}$$

The properties of the commutators and the formulas (3) were used in the computation.

It follows from the commutation relations (3) and (4) that the projections of the angular momentum are not simultaneously measurable quantities. The square of the angular momentum and one of its projections can be simultaneously measured.

We write out the operators of the projections of the angular momentum in the coordinate representation:

$$\begin{aligned}L_1 &= i \left(x_3 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_3} \right), \\L_2 &= i \left(x_1 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_1} \right), \\L_3 &= i \left(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2} \right).\end{aligned}$$

It is not hard to see that the operators L_1 , L_2 , L_3 act only on the angle variables of the function $\psi(\mathbf{x})$. Indeed, if the function ψ depends only on r , then

$$\begin{aligned}L_3\psi(r) &= i \left(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2} \right) \psi(x_1^2 + x_2^2 + x_3^2) \\&= i\psi'(x_1^2 + x_2^2 + x_3^2)(x_2 \cdot 2x_1 - x_1 \cdot 2x_2) = 0.\end{aligned}$$

For what follows, it turns out to be useful to have the formula

$$(5) \quad \Delta = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) - \frac{1}{r^2} L^2,$$

which we also verify in Cartesian coordinates. To this end we compute the operator $-L^2$, using the notation x, y, z for the projections of \mathbf{x} :

$$\begin{aligned} -L^2 &= \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)^2 + \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right)^2 + \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right)^2 \\ &= (x^2 + y^2 + z^2) \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) - x^2 \frac{\partial^2}{\partial x^2} - y^2 \frac{\partial^2}{\partial y^2} - z^2 \frac{\partial^2}{\partial z^2} \\ &\quad - 2xy \frac{\partial^2}{\partial x \partial y} - 2yz \frac{\partial^2}{\partial y \partial z} - 2zx \frac{\partial^2}{\partial z \partial x} - 2x \frac{\partial}{\partial x} - 2y \frac{\partial}{\partial y} - 2z \frac{\partial}{\partial z}. \end{aligned}$$

Taking into account the relations

$$r \frac{\partial}{\partial r} = x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z},$$

$$\begin{aligned} \left(r \frac{\partial}{\partial r} \right)^2 &= x^2 \frac{\partial^2}{\partial x^2} + y^2 \frac{\partial^2}{\partial y^2} + z^2 \frac{\partial^2}{\partial z^2} + x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z} \\ &\quad + 2xy \frac{\partial^2}{\partial x \partial y} + 2yz \frac{\partial^2}{\partial y \partial z} + 2zx \frac{\partial^2}{\partial z \partial x}, \end{aligned}$$

we get that $-L^2 = r^2 \Delta - (r \frac{\partial}{\partial r})^2 - r \frac{\partial}{\partial r} = r^2 \Delta - \frac{\partial}{\partial r} (r^2 \frac{\partial}{\partial r})$.

§ 24. The rotation group

Denote by G the collection of all rotations of the space \mathbf{R}^3 about the origin of coordinates. Each rotation $g \in G$ generates a linear transformation

$$\mathbf{x}' = g\mathbf{x}$$

of three-dimensional space, where g now denotes the corresponding matrix $\|g_{ij}\|$, called a rotation matrix. It is well known that g is an orthogonal matrix and $\det g = 1$. The converse is also true: to each such matrix there corresponds a definite rotation. Therefore, in what follows we identify rotations with their matrices. The group of real orthogonal matrices of order three with determinant 1 is denoted by $SO(3)$ and called the rotation group

There are several ways to parametrize rotations. For example, in theoretical mechanics one often uses the Euler angles. For our purposes it is most convenient to specify a rotation by a vector $\mathbf{a} = n\alpha$ ($|n| = 1$) directed along the axis \mathbf{n} of rotation with length α equal to

the angle of rotation. Here it is assumed that the direction of rotation and the direction of the vector \mathbf{a} form a right-hand screw and that the angle α of rotation does not exceed π ($0 \leq \alpha \leq \pi$). If we start the vectors \mathbf{a} from a single point, then their endpoints fill a ball of radius π . Different rotations correspond to different interior points of this ball, and diametrically opposite points of the boundary correspond to the same rotations.

Thus, the rotation group is topologically equivalent to a ball with opposite points identified.

We shall establish a connection between the rotation group and the Lie algebra of skew-symmetric matrices of order three.

A matrix A is said to be skew-symmetric if the transposed matrix A' is $= -A$. An arbitrary skew-symmetric matrix is given by three parameters and has the form

$$A = \begin{pmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{pmatrix}.$$

The collection of skew-symmetric matrices becomes a Lie algebra if the commutator $[A, B]$ is taken as the Lie operation. This assertion follows from the properties of commutators and the equality $[A, B]' = -[A, B]$. The latter can be verified directly:

$$[A, B]' = (AB - BA)' = B'A' - A'B' = BA - AB = -[A, B].$$

It is convenient to choose the matrices

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

as the generators of the Lie algebra. Any skew-symmetric matrix can be represented in the form

$$A(\mathbf{a}) = a_1 A_1 + a_2 A_2 + a_3 A_3.$$

We find the commutation relations for the matrices A_1 , A_2 , and A_3 :

$$[A_1, A_2] = A_1 A_2 - A_2 A_1$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = A_3,$$

and the desired relations are

$$(1) \quad \begin{aligned} [A_1, A_2] &= A_3, \\ [A_2, A_3] &= A_1, \\ [A_3, A_1] &= A_2 \end{aligned}$$

A direct verification shows that

$$[A(\mathbf{a}), A(\mathbf{b})] = A(\mathbf{a} \times \mathbf{b}).$$

Let us consider the matrix $e^{\alpha A_3}$. To get an explicit expression for this matrix, we compute the integer powers of the matrix A_3 :

$$A_3^2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = -\tilde{I}_3, \quad A_3^3 = -A_3, \quad A_3^4 = \tilde{I}_3, \dots$$

Expanding $e^{\alpha A_3}$ in a series, we get that

$$\begin{aligned} e^{\alpha A_3} &= \sum_{n=0}^{\infty} \frac{\alpha^n A_3^n}{n!} = I + \tilde{I}_3 \sum_{k=1}^{\infty} \frac{(-1)^k \alpha^{2k}}{(2k)!} + A_3 \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \alpha^{2k-1}}{(2k-1)!} \\ &= I + \tilde{I}_3 (\cos \alpha - 1) + A_3 \sin \alpha, \end{aligned}$$

or

$$e^{\alpha A_3} = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We see that $e^{\alpha A_3}$ is a rotation about the third axis through the angle α ; that is, $g(0, 0, \alpha) = e^{\alpha A_3}$. It can be verified similarly that $g(\alpha, 0, 0) = e^{\alpha A_1}$ and $g(0, \alpha, 0) = e^{\alpha A_2}$. For rotations about the coordinate axes we shall use the abbreviated notation $g_i(\alpha)$, $i = 1, 2, 3$.

The matrices $\frac{\partial g(\mathbf{a})}{\partial \alpha_i} \Big|_{\mathbf{a}=0}$ are called infinitesimal generators of the rotation group. Using the formulas for rotations about the coordinate axes, we get that

$$\frac{\partial g(\mathbf{a})}{\partial a_1} \Big|_{\mathbf{a}=0} = \frac{\partial g(a_1, 0, 0)}{\partial a_1} \Big|_{a_1=0} = \frac{\partial e^{a_1 A_1}}{\partial a_1} \Big|_{a_1=0} = A_1.$$

Similarly,

$$\frac{\partial g(\mathbf{a})}{\partial a_2} \Big|_{\mathbf{a}=0} = A_2, \quad \frac{\partial g(\mathbf{a})}{\partial a_3} \Big|_{\mathbf{a}=0} = A_3.$$

We see that the skew-symmetric matrices A_1, A_2, A_3 are infinitesimal generators of the rotation group.

Now we can easily get a formula for an arbitrary rotation $g(\mathbf{n}\alpha)$. It is obvious that the product of the rotations about a single axis \mathbf{n} through angles α and β is the rotation through the angle $\alpha + \beta$ about the same axis:

$$g(\mathbf{n}\beta) g(\mathbf{n}\alpha) = g(\mathbf{n}(\alpha + \beta)).$$

Differentiating this identity with respect to β and setting $\beta = 0$, we get that

$$(n_1 A_1 + n_2 A_2 + n_3 A_3) g(\mathbf{n}\alpha) = \frac{d}{d\alpha} g(\mathbf{n}\alpha).$$

We have found a differential equation for $g(\mathbf{n}\alpha)$ which must be solved under the initial condition $g(\mathbf{n}0) = I$. The unique solution of this problem is

$$(2) \quad g(\mathbf{n}\alpha) = \exp[(n_1 A_1 + n_2 A_2 + n_3 A_3)\alpha].$$

§ 25. Representations of the rotation group

A representation of the rotation group G on a Hilbert space \mathcal{E} is defined to be a mapping W that carries elements g of G into bounded linear operators $W(g)$ on \mathcal{E} depending continuously on g and such that

- 1) $W(I) = I$,
- 2) $W(g_1 g_2) = W(g_1)W(g_2)$.

In the first condition the first I denotes the identity rotation, and the second denotes the identity operator on \mathcal{E} . It follows at once from the conditions 1) and 2) that

$$W(g^{-1}) = W^{-1}(g).$$

A representation is said to be unitary if the operators $W(g)$ are unitary; that is, $W^*(g) = W^{-1}(g)$.

In group theory it is proved that any representation of the rotation group is equivalent to some unitary representation, and therefore we can confine ourselves to the study of unitary representations.

We recall the two equivalent definitions of an irreducible representation:

- 1) A representation is said to be irreducible if there are no operators other than those of the form CI that commute with all the operators $W(g)$:
- 2) A representation is said to be irreducible if \mathcal{E} does not have a proper nontrivial subspace \mathcal{E}_0 invariant under the action of the operators $W(g)$.

It is easy to construct a representation of the rotation group on the state space $\mathcal{H} = L^2(\mathbf{R}^3)$. To do this we define the operators by the formula

$$(1) \quad W(g)\psi(\mathbf{x}) = \psi(g^{-1}\mathbf{x}).$$

The condition 1) of the definition is obvious, and the condition 2) is verified as follows:

$$W(g_1 g_2)\psi(\mathbf{x}) = \psi(g_2^{-1}g_1^{-1}\mathbf{x}) = W(g_1)\psi(g_2^{-1}\mathbf{x}) = W(g_1)W(g_2)\psi(\mathbf{x}).$$

The operators $W(g)$ are one-to-one mappings of \mathcal{H} onto itself, and therefore to verify that they are unitary it suffices to see that they preserve the norm of a vector ψ :

$$\|W(g)\psi\|^2 = \int_{\mathbf{R}^3} |\psi(g^{-1}\mathbf{x})|^2 d\mathbf{x} = \int_{\mathbf{R}^3} |\psi(\mathbf{x})|^2 d\mathbf{x} = \|\psi\|^2,$$

where we have used the equality $\det g = 1$. Let us show that

$$(2) \quad W(\mathbf{a}) = \exp[-i(L_1 a_1 + L_2 a_2 + L_3 a_3)].$$

Here we have denoted $W(g(\mathbf{a}))$ by $W(\mathbf{a})$, and the projections of the angular momentum by L_1, L_2, L_3 . To prove (2) we first consider a rotation about the third axis and show that

$$W_3(\alpha) = e^{-iL_3\alpha},$$

or, equivalently,

$$(3) \quad \psi(g_3^{-1}(\alpha)\mathbf{x}) = e^{-iL_3\alpha}\psi(\mathbf{x})$$

for any $\psi(\mathbf{x}) \in \mathcal{H}$. We verify the last equality by showing that the functions on the left- and right-hand sides satisfy one and the same equation with the same initial conditions.

The function

$$\psi(\mathbf{x}, \alpha) = e^{-iL_3\alpha}\psi(\mathbf{x})$$

obviously satisfies the equation

$$\frac{\partial \psi(\mathbf{x}, \alpha)}{\partial \alpha} = -iL_3\psi(\mathbf{x}, \alpha),$$

or, in more detail,

$$(4) \quad \frac{\partial \psi(\mathbf{x}, \alpha)}{\partial \alpha} = \left(x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2} \right) \psi(\mathbf{x}, \alpha).$$

The function on the left-hand side of (3), which we denote by $\psi_1(\mathbf{x}, \alpha)$, has the form

$$\psi_1(\mathbf{x}, \alpha) = \psi(x_1 \cos \alpha + x_2 \sin \alpha, -x_1 \sin \alpha + x_2 \cos \alpha, x_3).$$

We verify that this function satisfies the equation (4):

$$\begin{aligned} \frac{\partial \psi_1}{\partial \alpha} &= \frac{\partial \psi_1}{\partial x_1}(-x_1 \sin \alpha + x_2 \cos \alpha) + \frac{\partial \psi_1}{\partial x_2}(-x_1 \cos \alpha - x_2 \sin \alpha), \\ x_2 \frac{\partial \psi_1}{\partial x_1} &= x_2 \frac{\partial \psi}{\partial x_1} \cos \alpha - x_2 \frac{\partial \psi}{\partial x_2} \sin \alpha, \\ x_1 \frac{\partial \psi_1}{\partial x_2} &= x_1 \frac{\partial \psi}{\partial x_1} \sin \alpha + x_1 \frac{\partial \psi}{\partial x_2} \cos \alpha, \end{aligned}$$

from which (4) follows for the function $\psi_1(\mathbf{x}, \alpha)$. Finally, both functions satisfy the same initial condition

$$\psi(\mathbf{x}, 0) = \psi_1(\mathbf{x}, 0) = \psi(\mathbf{x}).$$

Thus, the operators of rotation about the coordinate axes have the form

$$(5) \quad W_j(\alpha) = e^{-iL_j\alpha}, \quad j = 1, 2, 3.$$

The infinitesimal operators of the representation are found at once from (5):

$$\frac{\partial W(\mathbf{a})}{\partial a_j} \Big|_{\mathbf{a}=0} = \frac{\partial W_j(\alpha)}{\partial \alpha} \Big|_{\alpha=0} = -iL_j.$$

Repeating the arguments that led us to the formula (24.2), we now get (2). We note that the operators $-iL_j$ have the same commutation relations as the matrices A_j , $j = 1, 2, 3$. Indeed, it follows from

$$[L_1, L_2] = iL_3$$

that

$$[-iL_1, -iL_2] = -iL_3.$$

In group theory it is proved that the commutation relations between the infinitesimal operators of a representation do not depend on the choice of the representation. Moreover, if some operators acting in a space \mathcal{E} satisfy the same commutation relations as the infinitesimal generators of a group, then they are the infinitesimal operators of some representation acting in \mathcal{E} .

In connection with the rotation group this means that if we find operators M_1 , M_2 , M_3 satisfying the relations $[M_1, M_2] = M_3$, $[M_2, M_3] = M_1$, $[M_3, M_1] = M_2$, then we can construct a representation W by the formula

$$W(\mathbf{a}) = \exp(a_1M_1 + a_2M_2 + a_3M_3).$$

§ 26. Spherically symmetric operators

An operator A is said to be spherically symmetric if it commutes with all the operators $W(g)$ in a representation of the rotation group.

Obviously, an operator A is spherically symmetric if $[A, L_i] = 0$ for $i = 1, 2, 3$.

Here are some examples of spherically symmetric operators.

1) The operator of multiplication by a function $f(r)$. Indeed, we have seen that the angular momentum operators act only on the angle variables. Therefore, $L_i f(r)\psi(\mathbf{x}) = f(r)L_i\psi(\mathbf{x})$ for any $\psi \in \mathcal{H}$.

2) The operator L^2 . The spherical symmetry of this operator follows from the relations $[L^2, L_i] = 0$ obtained earlier for $i = 1, 2, 3$

3) The kinetic energy operator $T = -(1/2m)\Delta$. The spherical symmetry of this operator is clear at once in the momentum representation, where it is the operator of multiplication by the function $p^2/(2m)$. In the momentum representation the angular momentum operators L_i have exactly the same form as in the coordinate representation.

4) The Schrödinger operator

$$H = -\frac{1}{2m}\Delta + V(r)$$

for a particle in a central field, as a sum of two spherically symmetric operators.

We direct attention to the fact that the very existence of spherically symmetric operators different from CI implies the reducibility of the representation of the rotation group constructed on the state space \mathcal{H} .

For the spectrum of the Schrödinger operator in a central field we now clarify some features connected with its spherical symmetry. Let ψ be an eigenvector corresponding to the eigenvalue E :

$$H\psi = E\psi.$$

Then

$$HW(g)\psi = W(g)H\psi = EW(g)\psi,$$

from which it is obvious that $W(g)\psi$ is also an eigenvector of H corresponding to the same eigenvalue.

We see that the eigenspace \mathcal{H}_E of H corresponding to the eigenvalue E is invariant under rotations (that is, under the action of the operators $W(g)$). The representation W of the rotation group on the space \mathcal{H} induces a representation W_E , $g \rightarrow W_E(g)$, on the subspace \mathcal{H}_E , where $W_E(g)$ is the restriction of $W(g)$ to \mathcal{H}_E (below we use the notation $W(g)$ also for the operators $W_E(g)$). There are two cases:

either the representation induced on \mathcal{H}_E is irreducible, or \mathcal{H}_E contains subspaces of smaller dimension that are invariant under $W(g)$, and then this representation will be equivalent to a direct sum of irreducible representations.²⁵

We see also that the multiplicity of an eigenvalue of a spherically symmetric Schrödinger operator is always no less than the dimension of some irreducible representation of the rotation group, and it coincides with this dimension in the first case mentioned.

In physics the appearance of multiple eigenvalues of energy is called degeneracy, and such energy levels are said to be degenerate. If on each of the eigenspaces the induced representation is irreducible, then one says that the operator H does not have accidental degeneracies. In this case the multiplicity of the spectrum is completely explained by the chosen symmetry of the problem. In the presence of accidental degeneracies there may exist a richer symmetry group of the Schrödinger equation. This is the case for the Schrödinger operator for the hydrogen atom, which, as we shall see, has accidental degeneracies with respect to the rotation group.

We remark that for a spherically symmetric operator H with pure point spectrum there are eigenvalues with arbitrarily large multiplicity. Indeed, in this case \mathcal{H} can be represented in the form

$$\mathcal{H} = \mathcal{H}_{E_1} \oplus \mathcal{H}_{E_2} \oplus \dots$$

On the other hand,

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \dots$$

where the \mathcal{H}_n are subspaces on which irreducible representations of the rotation group act. In studying such representations in § 30 we shall see that among the \mathcal{H}_n there are subspaces of arbitrarily large dimension. But for any \mathcal{H}_n at least one of the intersections $\mathcal{H}_n \cap \mathcal{H}_{E_k}$

²⁵Here we use a well-known theorem in group theory if $g \rightarrow W(g)$ is a unitary representation of the rotation group on a Hilbert space \mathcal{E} , then there exist finite-dimensional subspaces $\mathcal{E}_1, \mathcal{E}_2, \dots$ that are invariant with respect to $W(g)$ and such that the representation W is irreducible on each of them. They are pairwise orthogonal, and they have sum \mathcal{E} , that is, $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \dots$

is nonempty, and then it contains \mathcal{H}_n entirely, so that the eigenvalue E_k has multiplicity not less than the dimension of \mathcal{H}_n .²⁶

If the system does not have accidental degeneracies, then the eigenvalues of H can be classified with the help of irreducible representations of G in the sense that each eigenspace \mathcal{H}_E is also an eigenspace of the corresponding representation. Therefore, the problem of finding all the irreducible representations of the rotation group is important. We dwell on this question in the following sections.

In concluding this section we remark that such a relatively simple quantum mechanics problem as the problem of motion in a central field could be solved in general without invoking group theory. Our goal with this example is to show how group theory can be used in solving quantum mechanics problems. The microworld (atoms, molecules, crystal lattices) is very rich in diverse forms of symmetry. The theory of group representations makes it possible to take these symmetry properties into account explicitly from the very beginning, and it is often the case that one can obtain important results for very complex systems only by using an approach based on group theory.

§ 27. Representation of rotations by 2×2 unitary matrices

We shall construct a representation of the rotation group G acting in the space \mathbf{C}^2 . To this end we introduce three self-adjoint matrices

²⁶A very simple example of a spherically symmetric operator with pure point spectrum is the Schrödinger operator

$$H = \frac{P_1^2 + P_2^2 + P_3^2}{2} + \frac{\omega^2(Q_1^2 + Q_2^2 + Q_3^2)}{2},$$

or, in the coordinate representation,

$$H = -\frac{1}{2}\Delta + \frac{\omega^2}{2}r^2$$

The problem of separation of variables reduces to the one-dimensional case, and for the eigenvalues one obtains the formula

$$E_{n_1 n_2 n_3} = \left(n_1 + n_2 + n_3 + \frac{3}{2} \right), \quad n_1, n_2, n_3 = 0, 1, 2,$$

To each triple of numbers n_1, n_2, n_3 there corresponds an eigenvector $\psi_{n_1 n_2 n_3}$. It is clear that the multiplicity of the eigenvalues E grows unboundedly as $E \rightarrow \infty$.

with trace zero:

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

These are called the Pauli matrices. Let us compute the commutation relations for them:

$$[\sigma_1, \sigma_2] = \sigma_1 \sigma_2 - \sigma_2 \sigma_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = 2i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

that is, $[\sigma_1, \sigma_2] = 2i\sigma_3$; similarly,

$$[\sigma_2, \sigma_3] = 2i\sigma_1, \quad [\sigma_3, \sigma_1] = 2i\sigma_2.$$

It is not hard to see that the matrices $-i\sigma_j/2$, $j = 1, 2, 3$, have the same commutation relations as the infinitesimal generators A_j of the rotation group

$$\left[-\frac{i\sigma_1}{2}, -\frac{i\sigma_2}{2} \right] = -\frac{i\sigma_3}{2}.$$

Therefore, we can construct a representation $g \rightarrow U(g)$:

$$(1) \quad U(g) = \exp \left[-\frac{i}{2}(\sigma_1 a_1 + \sigma_2 a_2 + \sigma_3 a_3) \right].$$

It should be noted that this is not a representation in the usual sense of the word, since instead of

$$(2) \quad U(g_1) U(g_2) = U(g_1 g_2)$$

we have

$$(3) \quad U(g_1) U(g_2) = \sigma U(g_1 g_2),$$

where $\sigma = \pm 1$. It is not hard to see this in a simple example, computing the product $U(g_1)U(g_2)$ when $g_1 = g_2$ is the rotation through the angle π about the x_3 -axis:

$$U(g_1) U(g_2) = e^{-\frac{i}{2}\sigma_3\pi} e^{-\frac{i}{2}\sigma_3\pi} = e^{-i\sigma_3\pi} = \begin{pmatrix} e^{i\pi} & 0 \\ 0 & e^{i\pi} \end{pmatrix} = -I.$$

At the same time, $U(0, 0, 0) = I$ corresponds to the identity element of the group according to the formula (1). A mapping $g \rightarrow U(g)$ satisfying (3) with $|\sigma| = 1$ is called a projective representation with a multiplier. If we nevertheless want to preserve (2), then we shall have to consider that to each rotation there correspond two matrices

U differing in sign. In physics such representations are said to be two-to-one. These representations play as important a role in quantum mechanics as the usual representations. In what follows we shall not stress this distinction. We remark further that the appearance of such representations is explained by the fact that the rotation group is not simply connected.

We consider the properties of the matrices $U(g)$. It is obvious that they are unitary, since the σ_i are self-adjoint matrices. It is not hard to see that they have determinant equal to 1. Indeed, $U(g)$ has the form e^{iS} , where S is a self-adjoint matrix with trace zero. This matrix can always be reduced to diagonal form by a similarity transformation, and it takes the form $\begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix}$. Correspondingly, the diagonal form of the matrix will be $\begin{pmatrix} e^{i\lambda} & 0 \\ 0 & e^{-i\lambda} \end{pmatrix}$. The trace and determinant are invariant under a similarity transformation, so $\det U(g) = 1$.

Let us determine the general form of a unitary matrix with determinant equal to 1. The condition for unitarity is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} = \begin{pmatrix} a\bar{a} + b\bar{b} & a\bar{c} + b\bar{d} \\ c\bar{a} + d\bar{b} & c\bar{c} + d\bar{d} \end{pmatrix} = I.$$

From the equality $c\bar{a} + d\bar{b} = 0$, we have that $d = -c\bar{a}/b$, and from the conditions $\det U = 1$ and $a\bar{a} + b\bar{b} = 1$, we get that

$$ad - bc = -\frac{a\bar{a}c}{\bar{b}} - bc = -\frac{c}{\bar{b}}(a\bar{a} + b\bar{b}) = -\frac{c}{\bar{b}} = 1;$$

that is,

$$c = -\bar{b}, \quad d = a.$$

Thus, unitary matrices with determinant 1 have the form

$$U = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad |a|^2 + |b|^2 = 1.$$

The group of such matrices is denoted by $SU(2)$.

§ 28. Representation of the rotation group on a space of entire analytic functions of two complex variables

In this section we construct all irreducible representations of the rotation group. As the space of the representations we choose the Hilbert space \mathcal{D}_2 of functions $f(\xi, \eta)$ ($\xi \in \mathbf{C}, \eta \in \mathbf{C}$) of the form

$$f(\xi, \eta) = \sum_{n_1, n_2=0}^{\infty} C_{n_1, n_2} \frac{\xi^{n_1} \eta^{n_2}}{\sqrt{n_1! n_2!}}, \quad \sum_{n_1, n_2} |C_{n_1 n_2}|^2 < \infty.$$

with the scalar product

$$(f, g) = \frac{1}{\pi^2} \int f(\xi, \eta) \overline{g(\xi, \eta)} e^{-|\xi|^2 - |\eta|^2} d\mu(\xi) d\mu(\eta).$$

Just as in § 19, the functions $f_{n_1 n_2} = \xi^{n_1} \eta^{n_2} / \sqrt{n_1! n_2!}$ can be seen to form an orthonormal basis in this space: $(f_{n_1 n_2}, f_{n'_1 n'_2}) = \delta_{n_1 n'_1} \delta_{n_2 n'_2}$. Taking into account the connection between the groups $SO(3)$ and $SU(2)$, we can construct a representation of the group $SU(2)$. In what follows, it is convenient to represent $f(\xi, \eta)$ as $f(\zeta)$, where $\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbf{C}^2$. The representation $U \rightarrow W(U)$ is defined by the formula

$$(1) \quad W(U) f(\zeta) = f(U^{-1} \zeta).$$

We shall denote these operators also by $W(\mathbf{a})$ or $W(g)$, and the operators of rotation about the axes by $W_j(\alpha)$, $j = 1, 2, 3$.

To get an expression for $W(g)$, we find the infinitesimal operators of the representation, which we denote by $-iM_j$, $j = 1, 2, 3$:

$$\begin{aligned} -iM_1 f(\zeta) &= \frac{\partial}{\partial a_1} W(\mathbf{a})|_{\mathbf{a}=0} f(\zeta) = \left. \frac{dW_1(\alpha)}{d\alpha} \right|_{\alpha=0} f(\zeta) \\ &= \frac{d}{d\alpha} f(U_1^{-1}(\alpha) \zeta)|_{\alpha=0} = \left. \frac{d}{d\alpha} f(e^{\frac{i}{2}\sigma_1 \alpha} \zeta) \right|_{\alpha=0} \\ &= \left. \frac{\partial f}{\partial \xi} \frac{d\xi(\alpha)}{d\alpha} \right|_{\alpha=0} + \left. \frac{\partial f}{\partial \eta} \frac{d\eta(\alpha)}{d\alpha} \right|_{\alpha=0}. \end{aligned}$$

Here we have used the definition (1) and have denoted by $\xi(\alpha)$ and $\eta(\alpha)$ the components of the vector $e^{\frac{i}{2}\sigma_1 \alpha} \zeta$. The last derivatives are

computed as follows:

$$\frac{d}{d\alpha} e^{i \frac{\sigma_1}{2} \alpha} \zeta \Big|_{\alpha=0} = \frac{i}{2} \sigma_1 e^{i \frac{1}{2} \sigma_1 \alpha} \zeta \Big|_{\alpha=0} = \frac{i}{2} \sigma_1 \zeta = \frac{i}{2} \begin{pmatrix} \eta \\ \xi \end{pmatrix},$$

and therefore

$$\frac{d\xi(\alpha)}{d\alpha} \Big|_{\alpha=0} = \frac{i}{2} \eta, \quad \frac{d\eta(\alpha)}{d\alpha} \Big|_{\alpha=0} = \frac{i}{2} \xi.$$

As a result we get that

$$-iM_1 f(\zeta) = \frac{i}{2} \left(\frac{\partial f}{\partial \xi} \eta + \frac{\partial f}{\partial \eta} \xi \right).$$

The operators M_2 and M_3 are found in exactly the same way. We write out the expressions for these operators:

$$(2) \quad \begin{aligned} M_1 &= -\frac{1}{2} \left(\eta \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \eta} \right), \\ M_2 &= -\frac{i}{2} \left(-\eta \frac{\partial}{\partial \xi} + \xi \frac{\partial}{\partial \eta} \right), \\ M_3 &= -\frac{1}{2} \left(\xi \frac{\partial}{\partial \xi} - \eta \frac{\partial}{\partial \eta} \right). \end{aligned}$$

It is easy to verify that the operators M_j have the same commutation relations as the angular momentum operators, and $-iM_j$, $j = 1, 2, 3$, have the same commutation relations as the matrices A_1, A_2, A_3 . For the operators $W(\mathbf{a})$ we get that

$$(3) \quad W(\mathbf{a}) = \exp[-i(M_1 a_1 + M_2 a_2 + M_3 a_3)].$$

The basic convenience of the representation space \mathcal{D}_2 is that it is very easily decomposed into a direct sum of invariant subspaces in which the irreducible representations act. Indeed, the invariance of some subspace with respect to the operators $W(\mathbf{a})$ is equivalent to the invariance with respect to the action of the operators M_1, M_2, M_3 . From the formulas (2) it is clear that the subspaces of homogeneous polynomials of degree $n = n_1 + n_2$ are such invariant subspaces. These subspaces have dimension $n + 1$, $n = 0, 1, 2, \dots$. It remains for us to show that such subspaces do not contain invariant subspaces of

smaller dimension. To do this we introduce the operators

$$(4) \quad \begin{aligned} M_+ &= M_1 + iM_2 = -\eta \frac{\partial}{\partial \xi}, \\ M_- &= M_1 - iM_2 = -\xi \frac{\partial}{\partial \eta} \end{aligned}$$

and consider how they act on the basis vectors $f_{n_1 n_2}$:

$$\begin{aligned} M_+ f_{n_1 n_2} &= -\eta \frac{\partial}{\partial \xi} \frac{\xi^{n_1} \eta^{n_2}}{\sqrt{n_1! n_2!}} = -n_1 \frac{\xi^{n_1-1} \eta^{n_2+1}}{\sqrt{n_1! n_2!}} \\ &= -\sqrt{n_1(n_2+1)} \frac{\xi^{n_1-1} \eta^{n_2+1}}{\sqrt{(n_1-1)!(n_2+1)!}}, \end{aligned}$$

that is,

$$(5) \quad \begin{aligned} M_+ f_{n_1 n_2} &= -\sqrt{n_1(n_2+1)} f_{n_1-1, n_2+1}, \\ M_- f_{n_1 n_2} &= -\sqrt{(n_1+1)n_2} f_{n_1+1, n_2-1}. \end{aligned}$$

It is obvious that

$$M_+ f_{0n} = 0, \quad M_- f_{n0} = 0.$$

From (5) it is clear that the subspaces of homogeneous polynomials do not contain invariant subspaces of smaller dimension.

Let us show that the basis vectors are eigenvectors of the operator M_3 and the operator $M^2 = M_1^2 + M_2^2 + M_3^2$. We have

$$M_3 f_{n_1 n_2} = -\frac{1}{2} \left(\xi \frac{\partial}{\partial \xi} - \eta \frac{\partial}{\partial \eta} \right) \frac{\xi^{n_1} \eta^{n_2}}{\sqrt{n_1! n_2!}} = -\frac{1}{2}(n_1 - n_2) \frac{\xi^{n_1} \eta^{n_2}}{\sqrt{n_1! n_2!}},$$

that is,

$$(6) \quad M_3 f_{n_1 n_2} = -\frac{1}{2}(n_1 - n_2) f_{n_1 n_2}.$$

For the operator M^2 we have the formula

$$M^2 = M_+ M_- + M_3^2 - M_3.$$

Indeed,

$$\begin{aligned} M_+ M_- &= (M_1 + iM_2)(M_1 - iM_2) = M_1^2 + M_2^2 - i(M_1 M_2 - M_2 M_1) \\ &= M^2 - M_3^2 + M_3. \end{aligned}$$

Next, we have

$$\begin{aligned} M^2 f_{n_1 n_2} &= (M_+ M_- + M_3^2 - M_3) f_{n_1 n_2} \\ &= \left(\sqrt{(n_1 + 1)n_2} \sqrt{n_2(n_1 + 1)} \right. \\ &\quad \left. + \frac{1}{4}(n_1 - n_2)^2 + \frac{1}{2}(n_1 - n_2) \right) f_{n_1 n_2}, \end{aligned}$$

so that

$$(7) \quad M^2 f_{n_1 n_2} = \left(\frac{1}{4}(n_1 + n_2)^2 + \frac{1}{2}(n_1 + n_2) \right) f_{n_1 n_2}.$$

It is convenient to rewrite all the relations obtained, replacing n_1, n_2 by j, m according to the formulas

$$\begin{aligned} J &= \frac{n_1 + n_2}{2}, \quad m = -\frac{1}{2}(n_1 - n_2), \quad j = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots, \\ &\quad m = -j, -j + 1, \dots, j, \end{aligned}$$

or

$$n_1 = j - m, \quad n_2 = j + m.$$

Then the formulas (5)–(7) take the form

$$(8) \quad M_+ f_{jm} = -\sqrt{(j-m)(j+m+1)} f_{j,m+1},$$

$$(9) \quad M_- f_{jm} = -\sqrt{(j-m+1)(j+m)} f_{j,m-1},$$

$$(10) \quad M_3 f_{jm} = m f_{jm},$$

$$(11) \quad M^2 f_{jm} = j(j+1) f_{jm}.$$

where f_{jm} denotes $f_{n_1 n_2} = f_{j-m, j+m}$.

The new indices j and m are convenient in that to each index j there corresponds a representation of dimension $2j + 1$, $j = 0, 1/2, 1, 3/2, \dots$. This representation is usually denoted by D_j , and j is called the representation index. The formulas (8)–(10) make it possible to easily construct the explicit form of the matrices M_1, M_2, M_3 for each D_j . Thus, we have constructed the finite-dimensional representations D_j of the rotation group for all dimensions.

§ 29. Uniqueness of the representations D_j

We prove that the irreducible representations D_j constructed are unique (up to equivalence). In the course of the proof we shall see

to what extent the spectrum of the angular momentum operators is determined by their commutation relations, and how an arbitrary representation of the rotation group is decomposed into irreducible representations.

Suppose that some irreducible representation is defined on an n -dimensional space \mathcal{E} , and denote by $-iJ_1, -iJ_2, -iJ_3$ the infinitesimal operators of this representation. They satisfy the commutation relations

$$\begin{aligned}[J_1, J_2] &= iJ_3, \\ [J_2, J_3] &= iJ_1, \\ [J_3, J_1] &= iJ_2.\end{aligned}$$

The equivalence of this representation to some representation D_j will be proved if we can prove that for a suitable choice of basis in \mathcal{E} the matrices J_k , $k = 1, 2, 3$, coincide with the matrices M_k . We note first of all that a representation of the rotation group is at the same time a representation of its subgroup of rotations about the x_3 -axis through angles α . This subgroup is Abelian, and therefore all its irreducible representations are one dimensional and have the form $e^{-im_k\alpha}$ (we do not assume anything about the numbers m_k , allowing the possibility of "multivalued" representations). This means that for a suitable choice of basis in \mathcal{E} the matrix of a rotation about the x_3 -axis has the form

$$e^{-iJ_3\alpha} = \begin{pmatrix} e^{im_1\alpha} & 0 & \dots & 0 \\ 0 & e^{-im_2\alpha} & \ddots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & e^{-im_n\alpha} \end{pmatrix}.$$

Thus, \mathcal{E} has a basis of eigenvectors of the operator J_3 :

$$(1) \quad J_3 e_m = m e_m.$$

Further, it follows from the irreducibility of the representation that the operator $J^2 = J_1^2 + J_2^2 + J_3^2$, which commutes with J_k for all $k = 1, 2, 3$, must be a multiple of the identity on the space \mathcal{E} . This is possible if all the basis vectors e_m are eigenvectors of J^2 corresponding to one and the same eigenvalue, which we denote by $j(j+1)$ (for the present this is simply a notation). The basis vectors will be denoted

by e_{jm} :

$$(2) \quad J^2 e_{jm} = j(j+1) e_{jm}.$$

We introduce the operators $J_{\pm} = J_1 \pm iJ_2$. It is easy to verify the formulas

$$(3) \quad [J^2, J_{\pm}] = 0,$$

$$(4) \quad [J_3, J_{\pm}] = \pm J_{\pm},$$

$$(5) \quad J^2 = J_{\pm} J_{\mp} + J_3^2 \mp J_3.$$

We find a bound for the possible values of $|m|$ for a given eigenvalue $j(j+1)$. To this end we take the scalar product with e_{jm} of the equality

$$(J_1^2 + J_2^2) e_{jm} = (j(j+1) - m^2) e_{jm}$$

and get that

$$((J_1^2 + J_2^2) e_{jm}, e_{jm}) = j(j+1) - m^2.$$

The left-hand side is nonnegative, therefore

$$(6) \quad |m| \leq \sqrt{j(j+1)}.$$

From the relations (3) and (4) it follows that

$$J^2 J_{\pm} e_{jm} = j(j+1) J_{\pm} e_{jm}, \quad J_3 J_{\pm} e_{jm} = (m \pm 1) J_{\pm} e_{jm}.$$

Therefore, the vectors $J_{\pm} e_{jm}$ (if they are nonzero) are eigenvectors of the operator J^2 with eigenvalue $j(j+1)$ and of the operator J_3 with eigenvalues $m \pm 1$. Thus, for an arbitrary basis element e_{jm} we can construct a chain of eigenvectors with the same eigenvalue of J^2 and with eigenvalues $m_1, m_1 + 1, \dots, m_2$ of J_3 . Here m_1 and m_2 denote the smallest and largest eigenvalues, respectively. The existence of m_1 and m_2 follows from the inequality (6): the chain of eigenvectors must break off in both directions.

Let us compute the norm of $J_{\pm} e_{jm}$, using the facts that $\|e_{jm}\| = 1$ and $J_{\pm}^* = J_{\mp}$.

$$\begin{aligned} \|J_{\pm} e_{jm}\|^2 &= (J_{\mp} J_{\pm} e_{jm}, e_{jm}) = ((J^2 - J_3^2 \mp J_3) e_{jm}, e_{jm}) \\ &= j(j+1) - m(m \pm 1) = (j \mp m)(j \pm m + 1). \end{aligned}$$

Therefore, we can write

$$(7) \quad J_{\pm} e_{jm} = -\sqrt{(j \mp m)(j \pm m + 1)} e_{j,m \pm 1}.$$

This formula enables us for any unit basis vector to construct new unit basis vectors $e_{j,m \pm 1}$ satisfying all the requirements. The minus sign in front of the square root is written for convenience.

We have not yet explained what values the numbers j and m can take. We start from the equalities

$$(8) \quad J_- e_{jm_1} = 0, \quad J_+ e_{jm_2} = 0$$

Multiplying these equalities by J_+ and by J_- and using (5), we get that

$$(J^2 - J_3^2 + J_3) e_{jm_1} = 0, \quad (J^2 - J_3^2 - J_3) e_{jm_2} = 0,$$

or

$$(9) \quad \begin{aligned} j(j+1) - m_1^2 + m_1 &= 0, \\ j(j+1) - m_2^2 - m_2 &= 0. \end{aligned}$$

From (9) we get at once that $(m_1 + m_2)(m_1 - m_2 - 1) = 0$. We take the solution $m_1 = -m_2$ of this equation, since $m_2 \geq m_1$. Further, $m_2 - m_1 = 2m_2$ is an integer or zero. Therefore, m_2 can take the values $0, 1/2, 1, 3/2, \dots$. Finally, we see from (9) that j can be taken to be m_2 . This means that the eigenvalues of J^2 have the form $j(j+1)$, where $j = 0, 1/2, 1, 3/2, \dots$, and for a given j the eigenvalues m of J_3 run through the $(2j+1)$ values $-j, -j+1, \dots, j-1, j$. The numbers j and m are simultaneously either integers or half-integers. We emphasize once more that these properties of the spectrum of the operators J^2 and J have been obtained using only the commutation relations.

To finish the proof it remains for us to see that the eigenvectors constructed form a basis in \mathcal{E} . This follows from the irreducibility of the representation. Indeed, the subspace \mathcal{E}' spanned by the vectors e_{jm} with $m = -j, -j+1, \dots, j$ will be invariant with respect to the operators J_k , $k = 1, 2, 3$, and therefore must coincide with \mathcal{E} , and the dimension of the representation is $n = 2j+1$. The formulas (1) and (7) show that for this choice of basis the matrices J_k coincide with the matrices M_k .

We note that at the same time we have constructed a way of decomposing an arbitrary representation into irreducibility representations. Suppose that a representation of the group of rotations $-iJ_k$ and its infinitesimal operators acts in some space \mathcal{E} . To distinguish the invariant subspaces, we must find the general solutions of the equations

$$(10) \quad J^2 e_{jm} = j(j+1) e_{jm}, \quad J_3 e_{jm} = m e_{jm}.$$

For given j and $m = -j, -j+1, \dots, j$ the vectors e_{jm} form a basis of an irreducible representation of dimension $2j+1$. The problem of finding the general solutions of (10) is solved most simply as follows. First a vector e_{jj} satisfying the equations

$$J_+ e_{jj} = 0, \quad J_3 e_{jj} = j e_{jj}$$

is found, and then the vectors e_{jm} are constructed using the formula

$$J_- e_{jm} = -\sqrt{(j+m)(j-m+1)} e_{j,m-1},$$

which makes it possible to find all the vectors e_{jm} from e_{jj} .

§ 30. Representations of the rotation group on the space $L^2(S^2)$. Spherical functions

In § 25 we constructed a representation of the rotation group on the state space $\mathcal{H} = L^2(\mathbf{R}^3)$ by the operators

$$W(\mathbf{a}) = \exp[-i(L_1 a_1 + L_2 a_2 + L_3 a_3)],$$

where L_1, L_2, L_3 are the angular momentum operators. We recall that these operators act only on the angle variables of a function $\psi(\mathbf{x}) \in L^2(\mathbf{R}^3)$, and therefore it is convenient to regard the space $L^2(\mathbf{R}^3)$ as $L^2(\mathbf{R}^+) \otimes L^2(S^2)$. Here $L^2(\mathbf{R}^+)$ is the space of square-integrable functions $f(r)$ with weight r^2 on \mathbf{R}^+ , and $L^2(S^2)$ is the space of square-integrable functions $\psi(\mathbf{n}) = \psi(\theta, \varphi)$ on the unit sphere. The scalar products in these spaces are given by the formulas

$$(f_1, f_2) = \int_0^\infty r^2 f_1(r) \overline{f_2(r)} dr, \quad (\psi_1, \psi_2) = \int_{S^2} \psi_1(\mathbf{n}) \psi_2(\mathbf{n}) d\mathbf{n},$$

where $d\mathbf{n} = \sin \theta d\theta d\varphi$ is the surface element of the unit sphere.

Somewhat cumbersome computations lead to the following form for the angular momentum operators in spherical coordinates:

$$\begin{aligned} L_1 &= i \left(\sin \varphi \frac{\partial}{\partial \theta} + \cot \theta \cos \varphi \frac{\partial}{\partial \varphi} \right), \\ L_2 &= -i \left(\cos \varphi \frac{\partial}{\partial \theta} - \cot \theta \sin \varphi \frac{\partial}{\partial \varphi} \right), \\ L_3 &= -i \frac{\partial}{\partial \varphi}. \end{aligned}$$

In these variables the operators $L_{\pm} = L_1 \pm iL_2$ have the form

$$L_+ = e^{i\varphi} \left(\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} \right), \quad L_- = e^{-i\varphi} \left(-\frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} \right).$$

The common eigenfunctions of the operators L^2 and L_3 will be denoted by $Y_{lm}(\mathbf{n})$ (the number j for L is usually denoted by l). In spherical coordinates the equations $L_+ Y_{ll} = 0$ and $L_3 Y_{ll} = l Y_{ll}$ have the form

$$(1) \quad \begin{aligned} -i \frac{\partial Y_{ll}}{\partial \varphi} &= l Y_{ll}, \\ \frac{\partial Y_{ll}}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \varphi} Y_{ll} &= 0. \end{aligned}$$

From the first equation in (1) we see that

$$Y_{ll}(\theta, \varphi) = e^{il\varphi} F_{ll}(\theta),$$

where l can take only the integer values $l = 0, 1, 2, \dots$. The second equation in (1) enables us to get an equation for $F_{ll}(\theta)$:

$$(2) \quad \frac{\partial F_{ll}(\theta)}{\partial \theta} - l \cot \theta F_{ll}(\theta) = 0.$$

Solving this equation, we get that

$$F_{ll}(\theta) = C \sin^l \theta.$$

We note that for each l there exists a single solution of the equation (2). Thus, we have found that

$$Y_{ll}(\theta, \varphi) = C \sin^l \theta e^{il\varphi},$$

where the constant C can be found from the normalization condition. The remaining functions $Y_{lm}(\mathbf{n})$ can be computed by the formula

$$Y_{l,m-1} = -\frac{1}{\sqrt{(l+m)(l-m+1)}} e^{-i\varphi} \left(-\frac{\partial}{\partial\theta} + i \cot\theta \frac{\partial}{\partial\varphi} \right) Y_{lm}.$$

We shall not go through the corresponding computations. The functions $Y_{lm}(\mathbf{n})$ are called the normalized spherical functions of order l . They can be expressed as

$$Y_{lm}(\theta, \varphi) = -\frac{1}{\sqrt{2\pi}} e^{im\varphi} P_l^m(\cos\theta),$$

where the functions

$$P_l^m(\mu) = \sqrt{\frac{(l+m)!}{(l-m)!}} \sqrt{\frac{2l+1}{2}} \frac{1}{2^l l!} (1-\mu^2)^{-\frac{m}{2}} \frac{d^{l-m}(\mu^2-1)^l}{d\mu^{l-m}}$$

are called the normalized associated Legendre polynomials.

Accordingly, a basis of an irreducible representation in the space $L^2(S^2)$ consists of the spherical functions $Y_{lm}(\mathbf{n})$ for a fixed l and $m = -l, -l+1, \dots, l$. The space $L^2(S^2)$ contains a subspace for an irreducible representation for each odd dimension $2l+1$ (l an integer). The theorem on decomposition of a representation of the rotation group into irreducible representations is equivalent in this case to the assertion that the spherical functions are complete in $L^2(S^2)$. Any function $\psi(\mathbf{n}) \in L^2(S^2)$ can be expanded in a convergent series

$$(3) \quad \varphi(\mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} Y_{lm}(\mathbf{n}).$$

We recall that irreducible representations of both even and odd dimensions acted in \mathcal{D}_2 . The space \mathcal{D}_2 can be represented as a direct sum $\mathcal{D}_2^+ \oplus \mathcal{D}_2^-$, where \mathcal{D}_2^+ and \mathcal{D}_2^- are orthogonal subspaces of even and odd functions, respectively. In \mathcal{D}_2^+ , as in $L^2(S^2)$, only representations of odd dimensions act. Any element $f(z) \in \mathcal{D}_2^+$ can be represented in the form

$$(4) \quad f(z) = \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{lm} \frac{\xi^{l-m} \eta^{l+m}}{\sqrt{(l-m)!(l+m)!}}.$$

It is clear that the one-to-one correspondence

$$f(z) \leftrightarrow \psi(\mathbf{n})$$

establishes an isomorphism between the spaces D_2^+ and $L^2(S^2)$ for which

$$\frac{\xi^{l-m} \eta^{l+m}}{\sqrt{(l-m)!(l+m)!}} \leftrightarrow Y_{lm}(\mathbf{n}),$$

$$M_k \leftrightarrow L_k, \quad k = 1, 2, 3.$$

In the conclusion of this section we consider a representation of the rotation group on the state space $L^2(\mathbf{R}^3) = L^2(S^2) \otimes L^2(\mathbf{R}^+)$. Let $\{f_n(r)\}$ be an arbitrary basis in $L^2(\mathbf{R}^+)$. Then $\{f_n(r)Y_{lm}(\mathbf{n})\}$ is a basis in the space $L^2(\mathbf{R}^3)$, and any function $\psi(\mathbf{x}) \in L^2(\mathbf{R}^3)$ can be expanded in a series:

$$\psi(\mathbf{x}) = \sum_n \sum_{l=0}^{\infty} \sum_{m=-l}^l C_{nlm} f_n(r) Y_{lm}(\theta, \varphi).$$

From this formula it is clear that $L^2(\mathbf{R}^3)$ can also be decomposed (in many ways) into subspaces in which the irreducible representations of the rotation group of order $(2l+1)$ act, and each representation D_l is encountered infinitely many times. Any of the invariant subspaces in which the irreducible representation D_l acts is a set of functions of the form $f(r) \sum_{m=-l}^l C_m Y_{lm}(\theta, \varphi)$, where $f(r) \in L^2(\mathbf{R}^+)$.

§ 31. The radial Schrödinger equation

Let us return to the problem of the motion of a particle in a central field. We shall look for solutions of the equation

$$-\frac{1}{2\mu} \Delta \psi + V(r) \psi = E\psi,$$

or, using the formula (23.5), the equation

$$(1) \quad -\frac{1}{2\mu r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) \psi + \frac{L^2}{2\mu r^2} \psi + V(r) \psi = E\psi.$$

We have seen that in the absence of accidental degeneracies the eigenspaces of the Schrödinger operator H must coincide with the subspaces of the irreducible representations D_l , while in the presence of accidental degeneracies they are direct sums of such subspaces. It is

clear that all the independent eigenfunctions of H can be constructed if we look for them in the form

$$(2) \quad \psi(r, \mathbf{n}) = R_l(r) Y_{lm}(\mathbf{n}).$$

These functions are already eigenfunctions of the operators L^2 and L_3 :

$$L^2\psi = l(l+1)\psi, \quad L_3\psi = m\psi,$$

and therefore they describe states of the particle with definite values of the square of the angular momentum and of its third projection.

Substitution of (2) in (1) gives us an equation for $R_l(r)$:

$$-\frac{1}{2\mu r^2} \frac{d}{dr} \left(r^2 \frac{dR_l(r)}{dr} \right) + \frac{l(l+1)}{2\mu r^2} R_l(r) + V(r) R_l(r) = E R_l(r).$$

We introduce a new unknown function by

$$R_l(r) = \frac{f_l(r)}{r},$$

and the equation for $f_l(r)$ takes the form

$$(3) \quad -\frac{1}{2\mu} \frac{d^2 f_l}{dr^2} + \frac{l(l+1)}{2\mu r^2} f_l + V(r) f_l = E f_l.$$

This equation is called the radial Schrödinger equation. We point out some of its features. First of all, the parameter m does not appear in the equation, which physically means that the energy of the particle does not depend on the projection of the angular momentum on the x_3 -axis. For each l its own radial equation is obtained. The spectrum of the radial equation is always simple (this can be proved), and therefore accidental degeneracies are possible if the equations (3) with different l have the same eigenvalues.

The radial equation coincides in form with the Schrödinger equation

$$-\frac{1}{2\mu} \frac{d^2 \psi}{dx^2} + V\psi = E\psi$$

for a one-dimensional particle if we introduce the so-called effective potential

$$(4) \quad V_{\text{eff}}(f) = V(r) + \frac{l(l+1)}{2\mu r^2}.$$

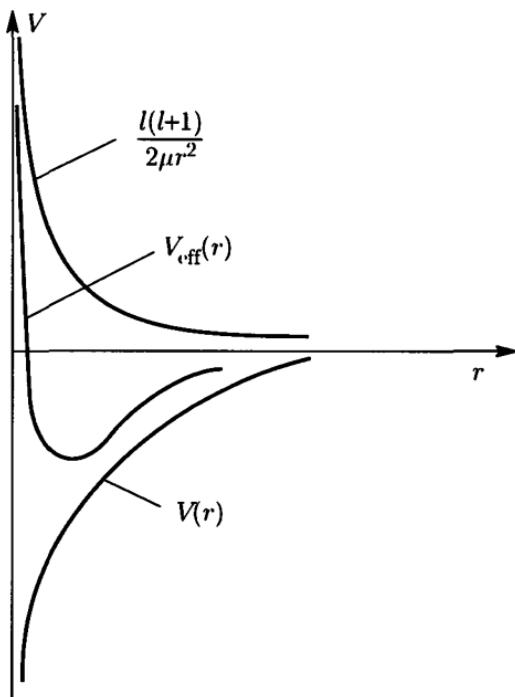


Figure 7

However, there is one essential difference. The function $\psi(x)$ is defined on \mathbf{R} while $f_l(r)$ is defined on \mathbf{R}^+ , and hence the radial equation is equivalent to the one-dimensional Schrödinger equation for the problem with the potential $V(x)$ under the condition that $V(x) = \infty$ for $x < 0$.

Figure 7 shows the graphs of the functions $V(r)$, $l(l + 1)/(2\mu r^2)$, and $V_{\text{eff}}(r)$, with $V(r)$ taken to be the Coulomb potential $-\alpha/r$ of attraction ($\alpha > 0$).

The expression $l(l + 1)/(2\mu r^2)$ can be interpreted as the potential of repulsion arising due to the centrifugal force. Therefore, this expression is usually called the centrifugal potential.

In quantum mechanics one has to solve problems with very diverse potentials $V(x)$. The most important of them seem to be the Coulomb potential $V(r) = \alpha/r$ describing the interaction of charged particles and the Yukawa potential $V(r) = g \frac{e^{-\mu r}}{r}$ often used in nuclear physics.

One usually considers potentials that are less singular than $1/r^{2-\epsilon}$ as $r \rightarrow 0$ ($\epsilon > 0$). In dependence on their behavior as $r \rightarrow \infty$, decaying potentials ($V(r) \rightarrow 0$) are divided into the short-range potentials, which satisfy $V(r) = o(1/r^{2+\epsilon})$ for some $\epsilon > 0$, and the long-range potentials, which do not satisfy this condition. The Yukawa potential is a short-range potential, while the Coulomb potential is a long-range potential. The spectrum of the radial Schrödinger operator

$$H_l = -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) \quad \left(\mu = \frac{1}{2}\right)$$

is well known for a very broad class of potentials. In the case of an increasing potential with $V(r) \rightarrow \infty$ as $r \rightarrow \infty$, the spectrum is a pure point spectrum and is simple. In the case of a decaying potential the interval $0 < E < \infty$ is filled by the continuous spectrum, and the negative spectrum is discrete. For a short-range potential the positive spectrum is simple and continuous, while the negative spectrum consists of a finite number of eigenvalues.

We shall give some simple arguments enabling us to understand the basic features of the spectrum of H_l . To this end we consider how the solutions of the radial equation behave as $r \rightarrow 0$ and $r \rightarrow \infty$. If as $r \rightarrow \infty$ we ignore the term V_{eff} in the radial equation, then it reduces to

$$\frac{1}{2\mu} f_l'' + Ef_l = 0.$$

For $E > 0$ this equation has the two linearly independent solutions e^{-ikr} and e^{ikr} , where $k^2 = 2\mu E > 0$ and $k > 0$. For $E < 0$ the linearly independent solutions have the form $e^{-\kappa r}$ and $e^{\kappa r}$, where $\kappa^2 = -2\mu E > 0$ and $\kappa > 0$.

In the case $r \rightarrow 0$, we can hope to get the correct behavior of the solutions of the radial equation if in this equation we leave the most singular $\frac{l(l+1)}{2\mu r^2} f_l$ of the terms linear in f_l , so that

$$f_l'' - \frac{l(l+1)}{r^2} f_l = 0.$$

This equation has the two linearly independent solutions r^{-l} and r^{l+1} .

Let us now consider what conditions should reasonably be imposed on the solutions of the radial equation. We are interested in the solutions $\psi(\mathbf{x}) = \frac{f_l(r)}{r} Y_{lm}(\mathbf{n})$ of the Schrödinger equation. The functions $\psi(\mathbf{x})$ must be continuous in \mathbf{R}^3 and either square-integrable or bounded on the whole space. In the first case they are eigenfunctions in the usual sense of the word, and in the second case they can be used to describe the continuous spectrum.

The continuity of $\psi(\mathbf{x})$ implies the condition $f_l(0) = 0$. Therefore, only solutions of the radial equation that behave like Cr^{l+1} as $r \rightarrow 0$ are of interest. This condition determines $f_l(r)$ up to a numerical factor. Next, for $E < 0$ we must find a solution $f_l(r)$ that behaves like $Ce^{-\kappa r}$ as $r \rightarrow \infty$ (otherwise, the solution will be unbounded). For arbitrary negative E these conditions turn out to be incompatible. Those values of E for which we can construct a solution having the correct behavior at zero and at infinity are precisely the eigenvalues.

For any $E > 0$ the solution $f_l(r)$ is bounded, and thus it suffices that it have the correct behavior at zero. The spectrum is continuous for $E > 0$.

The eigenfunctions of the discrete spectrum describe a particle localized in a neighborhood of the center of force; that is, they correspond to a finite motion. The eigenfunctions of the continuous spectrum can be used to describe the states in which the motion of the particle is infinite.

The eigenfunctions of the discrete spectrum of the radial equation will be denoted by $f_{kl}(r)$, where k indexes the eigenvalues E_{kl} of the equation for a given l :

$$H_l f_{kl} = E_{kl} f_{kl}.$$

The eigenfunctions of the continuous spectrum corresponding to the energy E will be denoted by f_{El} :

$$H_l f_{El} = E f_{El}.$$

For a broad class of potentials it has been proved that the system $\{f_{kl}, f_{El}\}$ of functions is complete for each $l = 0, 1, 2, \dots$. This means

that an arbitrary function²⁷ $f(r) \in L^2(0, \infty)$ has a representation

$$f(r) = \sum_k C_k f_{kl}(r) + \int_0^\infty C(E) f_{El}(r) dE,$$

where

$$C_k = \int_0^\infty f(r) \overline{f_{kl}(r)} dr, \quad C(E) = \int_0^\infty f(r) \overline{f_{El}(r)} dr.$$

Let us return to the three-dimensional problem. The functions of the discrete spectrum have the form

$$\psi_{klm}(\mathbf{x}) = \frac{f_{kl}(r)}{r} Y_{lm}(\mathbf{n}),$$

and the multiplicity of the eigenvalue E_{kl} is equal to $2l + 1$ (in the absence of accidental degeneracies). For the eigenfunctions of the continuous spectrum we have

$$\psi_{Elm} = \frac{f_{El}(r)}{r} Y_{lm}(\mathbf{n}).$$

The multiplicity of the continuous spectrum is infinite, since for any $E > 0$ there are solutions of the radial equations for all l and, moreover, $m = -l, -l + 1, \dots, l$.

The parameters k , l , and m that determine the eigenfunctions of the point spectrum are called the radial, orbital, and magnetic quantum numbers, respectively. These names go back to the old Bohr–Sommerfeld quantum theory, in which a definite classical orbit (or several such orbits) corresponded to each admissible value of the energy. The numbers k and l determined the size and shape of the orbit, and the number m determined the orientation of the plane of the orbit in space. The number m plays an essential role in magnetic phenomena, which explains its name.

The completeness of the system $\{f_{kl}, f_{El}\}$ of functions in $L^2(0, \infty)$ implies the completeness of the system $\{\psi_{klm}(\mathbf{x}), \psi_{Elm}(\mathbf{x})\}$ in $L^2(\mathbf{R}^3)$. For brevity of notation we consider the case when the point spectrum of the operator H is absent. In this case an arbitrary function

²⁷ $L^2(0, \infty)$ denotes the space of square-integrable functions (without the weight r^2) on \mathbf{R}^+ . If $f \in L^2(0, \infty)$, then $R = f/r \in L^2(\mathbf{R}^+)$

$\psi(\mathbf{x}) \in L^2(\mathbf{R}^3)$ can be represented in the form

$$\psi(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \int_0^{\infty} C_{lm}(E) \psi_{Elm}(\mathbf{x}) dE.$$

It is clear that ψ is determined by the sequence $\{C_{lm}(E)\}$ of functions, and therefore we get the representation

$$\psi \leftrightarrow \{C_{lm}(E)\}$$

in the space of sequences of functions. In this space the scalar product is given by

$$(\psi_1, \psi_2) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \int_0^{\infty} C_{lm}^{(1)} \overline{C_{lm}^{(2)}} dE.$$

From the fact that $\psi_{Elm}(\mathbf{x})$ is an eigenfunction of the operators H , L^2 , and L_3 , it follows easily that these operators act as follows in the representation constructed:

$$(5) \quad \begin{aligned} HC_{lm}(E) &= EC_{lm}(E), \\ L^2 C_{lm}(E) &= l(l+1) C_{lm}(E), \\ L_3 C_{lm}(E) &= m C_{lm}(E). \end{aligned}$$

Therefore, this representation is an eigenrepresentation for the three commuting operators H , L^2 , and L_3 . (The equation (5) should not be confused with the equations for the eigenvectors.)

§ 32. The hydrogen atom. The alkali metal atoms

The hydrogen atom is a bound state of a positively charged nucleus with charge e and an electron with charge $-e$ ($e > 0$ is the absolute value of the charge of an electron). Therefore, the potential $V(r)$ has the form

$$V(r) = -\frac{e^2}{r}.$$

We consider the problem of motion in the field

$$V(r) = -\frac{Ze^2}{r}.$$

Such a potential corresponds to a hydrogen atom with $Z = 1$ and to the hydrogen-like ions He^+ , Li^{++} , ... with $Z = 2, 3, \dots$. In the coordinate representation the Schrödinger operator has the form

$$H = -\frac{1}{2\mu}\Delta - \frac{Ze^2}{r},$$

where $\mu = mM/(m + M)$ is the reduced mass, and m and M are the masses of the electron and nucleus, respectively. We shall solve the problem in the so-called atomic system of units in which $\hbar = 1$, $\mu = 1$, and $e^2 = 1$. Then the radial Schrödinger equation takes the form

$$-\frac{1}{2}f_l''(r) + \frac{l(l+1)}{2r^2}f_l - \frac{Z}{r}f_l = Ef_l.$$

We are interested in the discrete spectrum, so we consider the case $E < 0$. It is convenient to use the notation $-2E = \kappa^2$. Then

$$(1) \quad f_l'' + \frac{2Z}{r}f_l - \frac{l(l+1)}{r^2}f_l - \kappa^2f_l = 0.$$

The arguments in the previous section about the behavior of a solution as $r \rightarrow 0$ and $r \rightarrow \infty$ suggest that it is convenient to look for a solution in the form

$$(2) \quad f_l(r) = r^{l+1}e^{-\kappa r}\Lambda_l(r).$$

If we can find $\Lambda_l(r)$ representable by a convergent power series,

$$(3) \quad \Lambda_l(r) = \sum_{i=0}^{\infty} a_i r^i$$

with $a_0 \neq 0$ and such that $f_l(r)$ satisfies (1), then the correct behavior of $f_l(r)$ as $r \rightarrow 0$ will also be ensured. Of course, the behavior of $f_l(r)$ as $r \rightarrow \infty$ depends on the asymptotics of the function $\Lambda_l(r)$ as $r \rightarrow \infty$.

It is convenient to make the substitution of (2) into the equation (1) in two steps. Introducing the function g by

$$f = e^{-\kappa r}g,$$

$$f'' = e^{-\kappa r}(\kappa^2 g - 2\kappa g' + g''),$$

we have

$$g'' - 2\kappa g' + \frac{2Z}{r}g - \frac{l(l+1)}{r^2}g = 0.$$

Further, setting

$$\begin{aligned} g &= r^{l+1} \Lambda, \\ g' &= r^{l+1} \left(\frac{\Lambda(l+1)}{r} + \Lambda' \right), \\ g'' &= r^{l+1} \left(\frac{l(l+1)\Lambda}{r^2} + \frac{2(l+1)\Lambda'}{r} + \Lambda'' \right), \end{aligned}$$

we get that

$$\Lambda''_l + \left(\frac{2(l+1)}{r} - 2\kappa \right) \Lambda'_l + \left(\frac{2Z}{r} - \frac{2\kappa(l+1)}{r} \right) \Lambda_l = 0.$$

Let us look for a solution of this equation in the form of a series (3).

$$\sum_{i=0}^{\infty} a_i [i(i-1)r^{i-2} + 2(l+1)ir^{i-2} - 2i\kappa r^{i-1} + (2Z - 2\kappa l - 2\kappa)r^{i-1}] = 0.$$

We make the change of summation index $i \rightarrow i+1$ in the first two terms in the square brackets, and then

$$\sum_{i=0}^{\infty} r^{i-1} \{ a_{i+1} [(i+1)i + 2(i+1)(l+1)] - a_i [2\kappa(i+l+1) - 2Z] \} = 0.$$

Equating the coefficients of the powers of r , we get that

$$(4) \quad a_{i+1} = 2 \frac{\kappa(i+l+1) - Z}{(i+1)(i+2l+2)} a_i$$

From d'Alembert's criterion it is clear that the series converges for all r . We estimate the behavior of the series with the coefficients defined by (4) for large r . Of course, the asymptotic behavior as $r \rightarrow \infty$ is determined by the coefficients of the higher powers, but then

$$a_{i+1} \cong \frac{2\kappa}{i+1} a_i;$$

that is,

$$a_i \cong C \frac{(2\kappa)^i}{i!}$$

and

$$\Lambda_l \cong C e^{2\kappa l}.$$

Thus, for the solution f_l we get that

$$f_l \cong Cr^{l+1} e^{\kappa r}$$

as $r \rightarrow \infty$ (This argument could be made more precise, of course.)

We see that a solution of the radial equation with the correct behavior as $r \rightarrow 0$ grows exponentially as $r \rightarrow \infty$. However, it is clear from (4) that there are values of κ such that the series breaks off at some term. In this case the function Λ_l turns out to be a polynomial, and the solution $f_l(r)$ is square integrable. Let k denote the index of the highest nonzero coefficient, that is, $a_k \neq 0$ and $a_{k+1} = 0$ for $k = 1, 2, \dots$. It is obvious from (1) that this is possible if

$$\kappa = \kappa_{kl} = \frac{Z}{k+l+1}.$$

From the formula $-2E = \kappa^2$, we get that

$$(5) \quad E_{kl} = -\frac{Z^2}{2(k+l+1)^2}.$$

The parameter k is the radial quantum number introduced earlier. We see that the eigenvalues E_{kl} depend only on $n = k + l + 1$. This number is called the principal quantum number. Recalling that $k = 0, 1, 2, \dots$ and $l = 0, 1, 2, \dots$ we get that $n = 1, 2, 3, \dots$. Furthermore, for a given n the quantum number l can take the values $0, 1, 2, \dots, n - 1$.

Accordingly, we have obtained the following results. The eigenvalues E are given by the formula

$$(6) \quad E_n = -\frac{Z^2}{2n^2},$$

and the eigenfunctions have the form

$$(7) \quad \psi_{nlm} = r^l e^{-\kappa_n r} \Lambda_{nl}(r) Y_{lm}(\theta, \varphi),$$

where Λ_{nl} is a polynomial of degree $n - l - 1$ whose coefficients are found from the formula (4), with a_0 found from the normalization condition. We see that the number of eigenvalues is infinite and has accumulation point $E = 0$. It is not hard to determine the multiplicity of the eigenvalue E_n . To each E_n there correspond eigenfunctions ψ_{nlm}

differing by the quantum numbers l and m , with $l = 0, 1, 2, \dots, n - 1$ and $m = -l, -l + 1, \dots, l$. For the multiplicity q we have

$$q = \sum_{l=0}^{n-1} (2l + 1) = n^2.$$

The multiplicity of the eigenvalues for the Coulomb field turns out to be greater than in the general case of a central field: there is an additional degeneracy with respect to l . We have already mentioned that this "accidental" degeneracy is explained by the presence of a symmetry group richer than $SO(3)$ for the Schrödinger operator for the hydrogen atom.

Let us now consider what physical information is given to us by the solution of the Schrödinger equation for the physical atom. First of all, we have found the admissible values of the energy, which it is reasonable to give in the usual units. For this it suffices to multiply the expression (6) for E_n by the atomic unit of energy equal to

$$\frac{\mu e^4}{h^2} = 4.36 \cdot 10^{-11} \text{ erg} = 27.21 \text{ eV}.$$

We assume that $Z = 1$, that is, we consider the hydrogen atom, and then

$$(8) \quad E_n = -\frac{\mu e^4}{2n^2 h^2}.$$

For the energy of the ground state of the hydrogen atom ($n = 1$), we have

$$E_1 = -\frac{\mu e^4}{2h^2} = -13.6 \text{ eV}$$

The absolute value of this energy is called the ionization potential or the binding energy of the electron in the atom and is equal to the work that must be done to strip the electron from the atom.

The formula (8) enables us to compute the frequencies of the spectral lines for the hydrogen atom. Quantum electrodynamics confirms Bohr's hypothesis that the frequency of a spectral line is determined by the formula

$$\hbar \omega_{mn} = E_n - E_m, \quad E_n > E_m,$$

and moreover, there is absorption of a quantum of light if the atom passes from a state with less energy to a state with greater energy, and emission of a quantum for the reverse transition.²⁸

For the frequencies of the spectral lines we have the formula

$$(9) \quad \omega_{nm} = \frac{\mu e^4}{2h^3} \left(\frac{1}{n^2} - \frac{1}{m^2} \right), \quad n < m.$$

This is called Balmer's formula and was discovered by him empirically long before the creation of quantum mechanics.

We direct attention to the dependence of the frequencies ω_{mn} on the reduced mass μ . In nature there are two forms of hydrogen: the ordinary hydrogen H whose nucleus is a proton with mass $M = 1836 m$ (m is the mass of the electron), along with a small quantity of heavy hydrogen or deuterium D whose nucleus is twice as heavy as the proton. Using the formula $\mu = mM/(m + M)$, we easily compute that $\mu_D/\mu_H = 1.000272$, that is, the reduced masses are very close. Nevertheless, the precision of spectroscopic measurements (wavelengths can be measured with an accuracy of 7–8 significant figures) make it possible to accurately measure the ratio ω_D/ω_H for the corresponding lines. This ratio is also equal to 1.000272 (for some lines there can be a deviation in the last significant figure). In general the values computed theoretically according to the formula (9) coincide with the experimental values of the frequencies to five significant figures. The existing deviations, however, can be removed by taking into account relativistic corrections.

Along with the transitions between stationary states of the discrete spectrum there can be transitions from the discrete spectrum to the continuous spectrum and the reverse transitions; physically, they correspond to ionization and recombination processes (capture of an electron by the nucleus). In these cases a continuous spectrum of absorption or emission is observed.²⁹

²⁸An absorption spectrum (dark lines on a bright background) arises if a light beam with continuous spectrum passes through a medium containing atomic hydrogen. Absorption lines are observed in the spectra of stars. A line spectrum of emission will be observed, for example, if an electrical discharge takes place in a medium with atomic hydrogen. Under the action of impacts with electrons, the hydrogen atoms will then pass into excited states. The transitions to levels with less energy lead to the appearance of bright lines.

²⁹The word "spectrum" is used here in two senses—the spectrum of an operator and the admissible values of the frequency of electromagnetic radiation.

The spectral lines of hydrogen on spectrograms are grouped into series corresponding to a definite value of n in the formula (9) and $m = n+1, n+2, \dots$. The first few series have been given names: the Lyman series ($n = 1$), the Balmer series ($n = 2$), and the Paschen series ($n = 3$). The lines of the Lyman series lie in the ultraviolet part of the spectrum, the first four lines of the Balmer series lie in the visible part of the spectrum, and the lines of the Paschen series and subsequent series lie in the infrared part of the spectrum. Toward the end of each series the lines converge to the so-called limit of the series, after which the continuous spectrum begins.

In Figure 8 the horizontal lines represent the energy levels of a hydrogen atom, and the vertical segments represent the possible transitions between them. The region of the continuous spectrum is shaded.

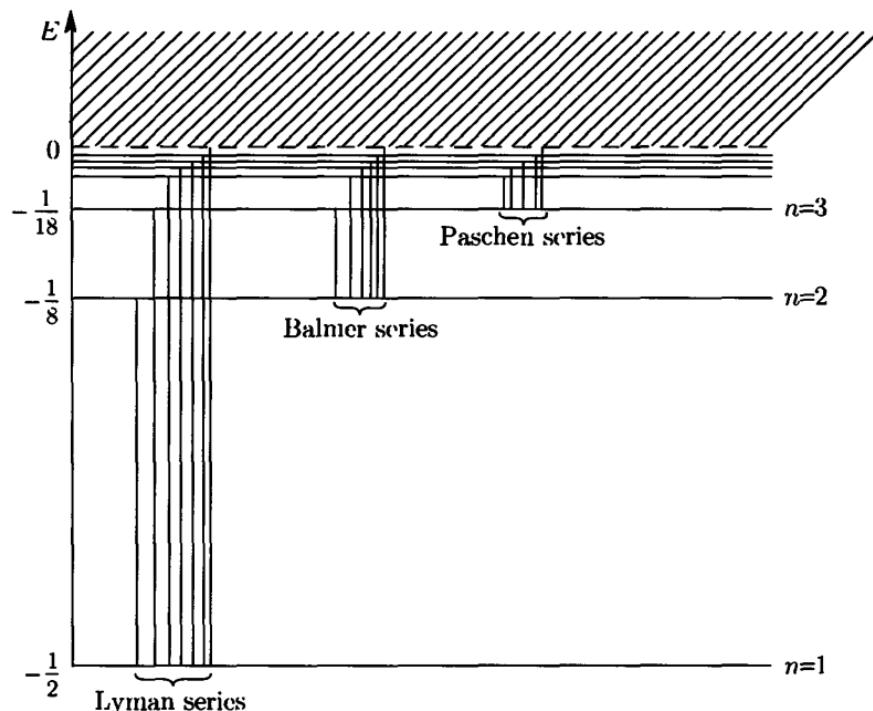


Figure 8



Figure 9

In Figure 9 we represent schematically the form of a spectral series, with the limit of the series represented by the dashed line.

The probabilities of transitions between states are important characteristics of atoms. The intensities of the spectral lines depend on the transition probabilities. The transitions that take place are spontaneous from an upper level to a lower level with emission of a quantum, forced (under the action of a light beam), or, finally, due to collisions with charged particles. Formulas for computing the probabilities of spontaneous and forced transitions are given by quantum electrodynamics, and transitions due to collisions are studied in the quantum theory of scattering. To compute all these characteristics, one must know the wave functions. Moreover, knowledge of the wave functions makes it possible to judge the size of an atom, the distribution of charge in the atom, and even the shape of the atom. We recall that $|\psi(\mathbf{x})|^2$ is the density of the coordinate distribution function. By the size of the atom we understand the size of the region in which $|\psi(\mathbf{x})|^2$ is not negligibly small. It is clear that the size of an atom is a conditional concept.

As an example let us consider the ground state of a hydrogen atom ($n = 1, l = 0, m = 0$). Since $Y_{00}(\mathbf{n}) = \text{const}$ and $\kappa_1 = 1$, we get by the formula (7) that

$$\psi_{100}(\mathbf{x}) = Ce^{-r}.$$

We find the constant C from the normalization condition

$$\int_{\mathbb{R}^3} |\psi|^2 d\mathbf{x} = |C|^2 4\pi \int_0^\infty e^{-2r} r^2 dr = |C|^2 \pi = 1,$$

from which $C = 1/\sqrt{\pi}$ and

$$\psi_{100}(\mathbf{x}) = \frac{1}{\sqrt{\pi}} e^{-r}.$$

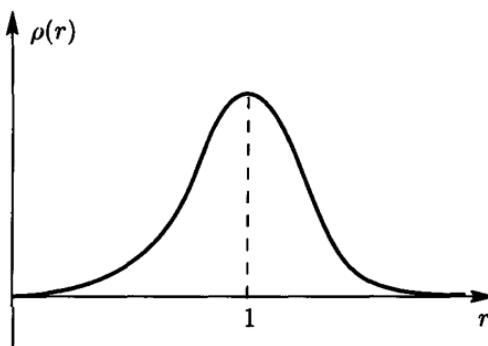


Figure 10

It is easy to see that

$$\rho(r) = 4\pi|\psi_{100}(r)|^2 r^2 = 4e^{-2r} r^2$$

is the density of the distribution function of the coordinate r . The graph of this function is pictured in Figure 10. The maximum of $\rho(r)$ is attained for $r_0 = 1$; that is, $r_0 = 1$ is the most probable distance of the electron from the nucleus. In the usual units, $r_0 = \hbar^2/\mu e^2 = 0.529 \cdot 10^{-8}$ cm. It is interesting to note that this number coincides with the radius of the first Bohr orbit. We see that the size of a hydrogen atom has order 10^{-8} cm.

By the density of the charge in an atom we understand the quantity $-e|\psi(\mathbf{x})|^2$, that is, we consider that because of its rapid motion about the nucleus, the electron is as if smeared over the volume of the atom, forming an electron cloud.

Finally, the form of the function (7) shows that for $l \neq 0$ the density of the coordinate distribution is not spherically symmetric. The dependence of this function on the angles lets us say something about the shape of the atom in various states.

In this same section we consider a simple model of alkali metal atoms, based on the assumption that their optical properties are explained by the motion of the valence electron in a certain central field $V(r)$. This potential $V(r)$ can be written as a sum of two terms,

$$V(r) = -\frac{Z}{r} + V_1(r),$$

where the first term describes the interaction of the electron with the nucleus, and $V_1(r)$ can be interpreted as the potential of the interaction of the electron with the negative charge of the other electrons distributed over the volume of the atom. The reasonableness of such a model for atoms of the alkali metals becomes clear only after we become acquainted with the properties of complex atoms and the Mendeleev periodic table.

We know very little about the potential $V(r)$, but nevertheless we can assert that

$$V(r) \cong -\frac{1}{r} \quad \text{for } r \rightarrow \infty$$

and

$$V(r) \cong -\frac{Z}{r} \quad \text{for } r \rightarrow 0.$$

The first condition follows from the obvious fact that when the valence electron is removed to infinity, it finds itself in the field of an ion with a single positive charge. The second condition follows from the continuity of the potential $V_1(r)$ of the volume distribution of charges.

As a model potential we choose

$$(10) \quad V(r) = -\frac{1}{r} - \frac{\alpha}{r^2}, \quad \alpha > 0.$$

Despite the fact that this potential has the correct behavior at infinity, its behavior at zero is different from that of the "true" potential. At the same time, the model potential correctly reflects the fact that upon approach to the nucleus the field becomes stronger than the Coulomb field $-1/r$. We assume that the parameter α is small (in what sense is indicated below). The numerical values of this parameter for the different alkali metal atoms is most reasonably chosen from a comparison of the results of computations of the energy levels with those found experimentally.

The radial equation for such a potential is very simple to solve. Indeed, it has the form

$$(11) \quad f_l'' + \frac{2}{r} f_l + \frac{2\alpha}{r^2} f_l - \frac{l(l+1)}{r^2} f_l - \varkappa^2 f_l = 0.$$

We introduce the number l' satisfying the equation

$$l'(l'+1) + 2\alpha - l(l+1) = 0$$

and the condition $\lim_{\alpha \rightarrow 0} l' = l$, so that $l' = -1/2 + \sqrt{(l+1/2)^2 - 2\alpha}$. The equation (11) can be rewritten in the form

$$f_l'' + \frac{2}{r} f_l - \frac{l'(l'+1)}{r^2} f_l - \kappa^2 f_l = 0;$$

that is, it coincides formally with the equation for the Coulomb field. All this can make sense only under the condition that $l(l+1) + 1/4 - 2\alpha > 0$. Otherwise we get complex values for l' .³⁰

If $\alpha < 1/8$, then the condition $l(l+1) + 1/4 - 2\alpha > 0$ holds for all l . One usually writes l' up to terms of order α^2 ; that is,

$$l' \cong l - \frac{\alpha}{l+1/2} = l - \sigma_l.$$

Using the formula (5) with $Z = 1$, we then get that

$$E_{kl} = -\frac{1}{2(k+l-\sigma_l+1)^2},$$

or, introducing the principal quantum number $n = k + l + 1$,

$$(12) \quad E_{nl} = -\frac{1}{2(n-\sigma_l)^2}.$$

It is clear from (12) that for the potential (9) the Coulomb degeneracy with respect to l is removed. The energy levels E_{nl} lie deeper than the levels E_n of the hydrogen atom, and the levels E_{nl} and E_n become close as n increases. The formula (12) describes the energy levels of the alkali metal atoms fairly well for an appropriate value of α . This formula was first obtained by Rydberg by analyzing experimental data. We remark that for the alkali metal atoms, as for hydrogen, the principal quantum number takes integer values, but the minimal value of n is not 1 but 2 for Li, 3 for Na, . . . , since the states with smaller principal quantum number are occupied by the electrons of the inner shells of the atom (this assertion will become clear after we become familiar with the structure of complex atoms).

In conclusion we note that this model illustrates a semi-empirical approach to the solution of complex quantum mechanics problems.

³⁰ It can be shown that for $2\alpha - l(l+1) > 1/4$ the radial Schrödinger operator

$$H_l = -\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1) - 2\alpha}{2r^2} - \frac{1}{r}$$

Such an approach consists in the following: instead of solving the problem in the exact formulation, one uses physical considerations to construct a simplified model of the system. The Schrödinger operator for the model problem usually depends on parameters which are just as difficult to find theoretically as it is to solve the problem in full scope. Therefore, the parameters are found by comparing the results of computations of the model problem with experimental data.

§ 33. Perturbation theory

In quantum mechanics there are relatively few interesting problems that admit the construction of exact solutions. Therefore, approximation methods play an important role. Approximation theories often turn out to be more valuable for understanding physical phenomena than exact numerical solutions of the corresponding equations. The main approximation methods in quantum mechanics are based on perturbation theory and the variational principle.

We describe the formulation of a problem in perturbation theory. Suppose that A is a given self-adjoint operator whose spectrum is known. It is required to find the spectrum of the operator $B = A + C$ under the condition that the self-adjoint operator C is small in some sense. We do not specify what is meant by smallness of C , since we are considering only a formal scheme in perturbation theory. Giving a rigorous basis for such a scheme requires the solution of some complex mathematical problems.

We shall analyze the case when A has a pure point spectrum, and we begin with the problem of perturbation of a simple eigenvalue. Let us consider the one-parameter family of operators

$$(1) \quad A_\varepsilon = A + \varepsilon C.$$

It is clear that $A_0 = A$ and $A_1 = B$. We know the eigenvectors ψ_n and eigenvalues λ_n of A , which satisfy the equation

$$(2) \quad A\psi_n = \lambda_n\psi_n.$$

It is assumed that the spectrum of A is simple; that is, to each λ_n there corresponds one eigenvector ψ_n .

The equation for the eigenvectors of the operator A_ε is

$$(3) \quad A_\varepsilon \psi_\varepsilon = \lambda_\varepsilon \psi_\varepsilon.$$

Our main assumption is that ψ_ε and λ_ε depend analytically on ε ; that is, they can be represented in the form

$$(4) \quad \lambda_\varepsilon = \lambda^{(0)} + \varepsilon \lambda^{(1)} + \varepsilon^2 \lambda^{(2)} + \dots,$$

$$(5) \quad \psi_\varepsilon = \psi^{(0)} + \varepsilon \psi^{(1)} + \varepsilon^2 \psi^{(2)} + \dots.$$

Substituting (4) and (5) in (3),

$$(A + \varepsilon C)(\psi^{(0)} + \varepsilon \lambda^{(1)} + \dots) = (\lambda^{(0)} + \varepsilon \lambda^{(1)} + \dots)(\psi^{(0)} + \varepsilon \psi^{(1)} + \dots),$$

and equating the coefficients of like powers of ε , we get the system of equations

$$A\psi^{(0)} = \lambda^{(0)}\psi^{(0)},$$

$$A\psi^{(1)} + C\psi^{(0)} = \lambda^{(0)}\psi^{(1)} + \lambda^{(1)}\psi^{(0)},$$

.....

$$A\psi^{(k)} + C\psi^{(k-1)} = \lambda^{(0)}\psi^{(k)} + \lambda^{(1)}\psi^{(k-1)} + \dots + \lambda^{(k)}\psi^{(0)}.$$

.....

which are more conveniently rewritten as

$$A\psi^{(0)} = \lambda^{(0)}\psi^{(0)},$$

$$(A - \lambda^{(0)})\psi^{(1)} = (\lambda^{(1)} - C)\psi^{(0)}.$$

$$(6) \quad (A - \lambda^{(0)})\psi^{(2)} = (\lambda^{(1)} - C)\psi^{(1)} + \lambda^{(2)}\psi^{(0)},$$

.....

$$(A - \lambda^{(0)})\psi^{(k)} = (\lambda^{(1)} - C)\psi^{(k-1)} + \dots + \lambda^{(k)}\psi^{(0)},$$

.....

From the first equation in (6) it follows³¹ that $\psi^{(0)}$ is an eigenvector of A , and from the assumption about simplicity of the spectrum we have

$$\lambda^{(0)} = \lambda_n, \quad \psi^{(0)} = \psi_n.$$

³¹Below we should equip the eigenvectors ψ_ε , the eigenvalues λ_ε and $\lambda^{(k)}$, and $\psi^{(k)}$ with the subscript n , but for brevity of notation we do not do that

Before turning to the subsequent equations in (6) we choose a normalization condition for the vector ψ_ε . It turns out that the condition

$$(7) \quad (\psi_\varepsilon, \psi^{(0)}) = 1$$

is most convenient. We assume that $\psi^{(0)}$ is normalized in the usual way: $\|\psi^{(0)}\| = 1$, and thus the condition (7) is equivalent to the conditions

$$(8) \quad (\psi^{(1)}, \psi^{(0)}) = 0, \dots, (\psi^{(k)}, \psi^{(0)}) = 0, \dots.$$

Thus, we can look for the corrections $\psi^{(1)}, \dots, \psi^{(k)}, \dots$ in the subspace orthogonal to the vector $\psi^{(0)} = \psi_n$.

Let us now consider the second equation in (6). This is an equation of the second kind with a self-adjoint operator A , and $\lambda^{(0)}$ is an eigenvalue of A . This equation has a solution if and only if the right-hand side is orthogonal to the vector $\psi^{(0)}$. It follows at once from the condition

$$(\psi^{(0)}, (\lambda^{(1)} - C) \psi^{(0)}) = 0$$

that

$$\lambda^{(1)} = (C\psi^{(0)}, \psi^{(0)}),$$

or, in more detail,

$$(9) \quad \lambda_n^{(1)} = (C\psi_n, \psi_n).$$

The formula (9) has a very simple physical interpretation. The first-order correction to the eigenvalue λ_n coincides with the mean value of the perturbation C in the unperturbed state ψ_n .

We consider what the second equation in (6) gives for the vector $\psi^{(1)}$. It might seem that we should write

$$(10) \quad \psi^{(1)} = (A - \lambda^{(0)}I)^{-1}(\lambda^{(1)} - C)\psi^{(0)}.$$

However, this formula needs to be made more precise. To understand why this is so, we consider more closely the operator $(A - \lambda I)^{-1}$, which is called the resolvent of A . The operator A can be written in the form

$$A = \sum_m \lambda_m P_m,$$

where P_m is the projection on the eigenvector ψ_m ; that is, $P_m\varphi = (\varphi, \psi_m)\psi_m$. Then for the operator $(A - \lambda I)^{-1}$ we have

$$(11) \quad (A - \lambda I)^{-1} = \sum_m \frac{P_m}{\lambda_m - \lambda}.$$

From (11) it is clear that the resolvent loses meaning for $\lambda = \lambda_n$, that is, precisely for the values of λ of interest to us. However, we recall that the right-hand side of the second equation in (6) is orthogonal to $\psi^{(0)} = \psi_n$ and $(\psi^{(1)}, \psi^{(0)}) = 0$. So in fact we need not the operator $(A - \lambda I)^{-1}$ but the operator $(A - \lambda I)^{-1}P$ acting in the subspace orthogonal to ψ_n , where P is the projection $I - P_n$ onto that subspace.

The operator $(A - \lambda I)^{-1}P$ can be represented in the form

$$(12) \quad (A - \lambda I)^{-1}P = \sum_{m \neq n} \frac{P_m}{\lambda_m - \lambda},$$

which preserves its meaning for $\lambda = \lambda_n$. Instead of (10) we must write

$$(13) \quad \psi^{(1)} = (A - \lambda^{(0)}I)^{-1}P(\lambda^{(1)} - C)\psi^{(0)}.$$

This expression can be transformed as follows:

$$\begin{aligned} & (A - \lambda_n I)^{-1}P[(C\psi_n, \psi_n)\psi_n - C\psi_n] \\ &= (A - \lambda_n I)^{-1}P(P_n - I)C\psi_n = -(A - \lambda_n I)^{-1}PC\psi_n, \end{aligned}$$

and therefore

$$(14) \quad \psi_n^{(1)} = -(A - \lambda_n I)^{-1}PC\psi_n$$

Using (12), we get that

$$(15) \quad \psi_n^{(1)} = \sum_{m \neq n} \frac{(C\psi_n, \psi_m)}{\lambda_n - \lambda_m} \psi_m.$$

Let us consider corrections of subsequent orders. From the orthogonality to $\psi^{(0)}$ of the right-hand side of the third equation in (6), we get at once that

$$(16) \quad \lambda^{(2)} = (C\psi^{(1)}, \psi^{(0)}).$$

Using the form of $\psi^{(1)}$, we find an explicit formula for the second correction to the eigenvalue λ_n :

$$(17) \quad \lambda_n^{(2)} = \sum_{m \neq n} \frac{|(C\psi_n, \psi_m)|^2}{\lambda_n - \lambda_m}.$$

We shall not present the detailed computations for $\psi^{(2)}$ and the subsequent corrections $\psi^{(k)}$ and $\lambda^{(k)}$, but note only that they can be found by the formulas $\lambda^{(k)} = (C\psi^{(k-1)}, \psi^{(0)})$ and $\psi^{(k)} = (A - \lambda^{(0)}I)^{-1}P \times$ [the right-hand side of the corresponding equation in (6)].

We now discuss the theory of perturbation of a multiple eigenvalue, confining ourselves to construction of the first-approximation correction $\lambda^{(1)}$. Let $\lambda_n = \lambda$ (we omit the index n) be an eigenvalue of A of multiplicity q :

$$A\psi_i = \lambda\psi_i, \quad i = 1, 2, \dots, q.$$

Denote by \mathcal{H}_λ the eigenspace of A corresponding to the eigenvalue λ , and by Q the projection on this subspace.

We turn again to the system of equations (5). As earlier, it follows from the first equation that $\lambda^{(0)} = \lambda$. As for the vectors $\psi^{(0)}$, we can only assert that $\psi^{(0)} \in \mathcal{H}_\lambda$. We now show that additional restrictions are imposed on the vectors $\psi^{(0)}$, and therefore they do not coincide with the eigenvectors ψ_i in the general case. Indeed, the second equation in (6) has solutions if its right-hand side is orthogonal to the subspace \mathcal{H}_λ ; that is,

$$Q(\lambda^{(1)} - C)\psi^{(0)} = 0.$$

Taking into account that $Q\psi^{(0)} = \psi^{(0)}$, we can rewrite the last equation in the form

$$(18) \quad QCQ\psi^{(0)} = \lambda^{(1)}\psi^{(0)}.$$

We see that the $\psi^{(0)}$ are eigenvectors of the q -dimensional operator QCQ , and the $\lambda^{(1)}$ are eigenvalues of it. In practice the problem reduces to the diagonalization of a matrix of order q . Indeed, substituting $\psi^{(0)} = \sum_{i=1}^q a_i\psi_i$ in (18) and using the fact that

$Q\varphi = \sum_{j=1}^q (\varphi, \psi_j) \psi_j$, we get that

$$\sum_j \sum_i a_i (C\psi_i, \psi_j) \psi_j = \lambda^{(1)} \sum_j a_j \psi_j.$$

so that

$$\sum_i C_{ji} a_i = \lambda^{(1)} a_j,$$

where $C_{ji} = (C\psi_i, \psi_j)$. The matrix $\|C_{ij}\|$ is self-adjoint, and thus can always be reduced to diagonal form. Denote the eigenvalues of this matrix by $\lambda_j^{(1)}$, $j = 1, 2, \dots, q$. To the multiple eigenvalue λ of the unperturbed operator A there correspond q eigenvalues of the operator $B = A + C$, which in the first approximation of perturbation theory have the form $\lambda + \lambda_j^{(1)}$, $j = 1, 2, \dots, q$. One usually says that the perturbation removes the degeneracy. Of course, the removal of the degeneracy can turn out to be incomplete if there are duplicates among the numbers $\lambda_j^{(1)}$, that is, if the operator QCQ has multiple eigenvalues.

Example. We consider a system with the Schrödinger operator

$$(19) \quad H = -\frac{1}{2} \Delta - \frac{1}{r} - \alpha B L_3.$$

Such a Schrödinger operator describes a hydrogen atom located in a constant homogeneous magnetic field with induction vector directed along the third axis.³²

As the unperturbed operator it is reasonable to take the operator

$$H_0 = -\frac{1}{2} \Delta - \frac{1}{r},$$

³²In electrodynamics the magnetic moment of a particle with charge e is defined to be the vector

$$\mathbf{M} = \frac{e}{2c} \mathbf{x} \times \mathbf{v} = \frac{e}{2c\mu} \mathbf{x} \times \mathbf{p} = \frac{e}{2c\mu} \mathbf{l}$$

Here μ is the mass of the particle, \mathbf{v} is its velocity, \mathbf{l} is its angular momentum, and c is the speed of light. The Hamiltonian function of the particle in a constant homogeneous magnetic field \mathbf{B} contains the additional term $-\mathbf{MB}$. For a hydrogen atom in a magnetic field directed along the third axis, the Hamiltonian function has the form

$$H(q, p) = \frac{p^2}{2\mu} - \frac{e^2}{r} - \frac{e}{2c\mu} Bl_3$$

The corresponding Schrödinger operator in atomic units coincides with (19) for $\alpha = 1/2c$.

that is, the Schrödinger operator for the hydrogen atom, and to regard

$$\Delta H = -\alpha B L_3$$

as the perturbation. From the physical point of view ΔH is small, since the magnetic force acting on an electron of an atom in attainable magnetic fields is less by several orders of magnitude than the Coulomb force of attraction to the nucleus. We recall that the eigenfunctions $\psi_{nlm}(\mathbf{x})$ of the operator H_0 are also eigenfunctions of the operator L_3 :

$$L_3 \psi_{nlm} = m \psi_{nlm}.$$

The matrix of the perturbation ΔH is at once diagonal, and its diagonal elements are equal to $-\alpha mB$. Therefore, for the energy of the hydrogen atom in a magnetic field we have the formula³³

$$(20) \quad E_{nm} = -\frac{1}{2n^2} - \alpha B m.$$

We see that the magnetic field removes the degeneracy with respect to the magnetic quantum number m , but leaves the degeneracy with respect to l that is characteristic for the Coulomb field.

The phenomenon consisting in a splitting of the energy levels of atoms in a magnetic field and in a corresponding splitting of their spectral lines is called the Zeeman effect.

It is interesting to look at this phenomenon from the point of view of group theory. The degeneracy with respect to m is explained by the spherical symmetry of the Schrödinger operator. A magnetic field directed along the x_3 -axis disturbs this symmetry. The symmetry group of the Schrödinger operator of the atom in the magnetic field is the group of rotations about the third axis. This group is Abelian, and all its irreducible representations are one dimensional. Therefore, the presence of such a symmetry group does not cause degeneracy; any degeneracy will be accidental.

³³This is not the most satisfactory example, because the functions ψ_{nlm} are exact eigenfunctions of H with eigenvalues (20)

§ 34. The variational principle

We consider the functional

$$(1) \quad E = \frac{(H\psi, \psi)}{(\psi, \psi)}, \quad \psi \in \mathcal{H}.$$

This functional has a simple physical meaning: E is the mean value of the energy of the system in the state given by the vector $\psi/\|\psi\|$. If $\psi = \psi_n$, where ψ_n is the eigenvector of H corresponding to the eigenvalue E_n , then $E = E_n$. Let us compute the variation of the functional (1):

$$\begin{aligned} \delta E &= \frac{(H\delta\psi, \psi) + (H\psi, \delta\psi)}{(\psi, \psi)} - \frac{(H\psi, \psi)[(\delta\psi, \psi) + (\psi, \delta\psi)]}{(\psi, \psi)^2} \\ &= 2 \operatorname{Re} \frac{((H - E)\psi, \delta\psi)}{(\psi, \psi)}. \end{aligned}$$

It is easy to see that the condition

$$(2) \quad \delta E = 0$$

for the functional E to be stationary is equivalent to the Schrödinger equation

$$(3) \quad H\psi = E\psi.$$

Indeed, (3) implies (2). To get the converse it suffices to consider $\delta\psi_1 = i\delta\psi$ along with the variation $\delta\psi$. Then it follows from the condition (2) that

$$\frac{((H - E)\psi, \delta\psi)}{(\psi, \psi)} = 0,$$

and we have (3), since $\delta\psi$ is arbitrary.

We point out one more important property of the functional E . For any vector $\psi \in \mathcal{H}$ we have $E \geq E_0$, where E_0 is the smallest eigenvalue, and equality holds only for $\psi = C\psi_0$. This is almost obvious, since the mean value of the energy is not less than the minimal possible value. We verify this formally for an operator H with simple pure point spectrum. Suppose that the eigenvalues are numbered in increasing order: $E_0 < E_1 < E_2 < \dots$. Substituting $\psi = \sum_{n=0}^{\infty} C_n \psi_n$

in (1), we get that

$$(4) \quad E - E_0 = \frac{\sum_n E_n |C_n|^2}{\sum_n |C_n|^2} - E_0 = \frac{\sum_n (E_n - E_0) |C_n|^2}{\sum_n |C_n|^2} \geq 0,$$

because $E_n - E_0 \geq 0$. Equality is attained in (4) if $C_n = 0$ for $n = 1, 2, \dots$. In this case $\psi = C_0 \psi_0$. It can be verified similarly that

$$(5) \quad \begin{aligned} E &\geq E_1 & \text{if } (\psi, \psi_0) = 0, \\ E &\geq E_2 & \text{if } (\psi, \psi_0) = 0, (\psi, \psi_1) = 0, \\ &\dots & \end{aligned}$$

The property $E \geq E_0$ makes the variational principle especially effective for computing the ground state of the system. Substituting an arbitrary vector $\psi \in \mathcal{H}$ in (1), we get an upper estimate for E_0 ; of the two values E' and E'' of the functional (1) the smaller one is closest to E_0 . The use of the properties (5) to estimate E_n encounters difficulties, since we do not know the eigenvectors $\psi_0, \dots, \psi_{n-1}$.

There is a second formulation of the variational principle asserting that the Schrödinger equation (3) is equivalent to the condition that the functional $(H\psi, \psi)$ be stationary for $(\psi, \psi) = 1$. Using Lagrange's method of undetermined multipliers, we can write the last condition in the form

$$(6) \quad \delta[(H\psi, \psi) - E(\psi, \psi)] = 0,$$

where E is a Lagrange multiplier. The equivalence of (6) and (3) is verified like the equivalence of (2) and (3).

Variational principles can be used in two ways to obtain approximate solutions of the equation (3). The first way is to look for an approximate wave function in the class of functions of a certain analytic form that depend on several parameters $\alpha_1, \dots, \alpha_k$. Then $E = E(\alpha_1, \dots, \alpha_k)$, and the parameters are found from the conditions

$$\frac{\partial E(\alpha_1, \dots, \alpha_k)}{\partial \alpha_i} = 0, \quad i = 1, 2, \dots, k$$

The second way is to construct, for a complex system (for example, a complex atom), an approximate eigenfunction $\psi(\mathbf{x}_1, \dots, \mathbf{x}_N)$ of H depending on several variables by using unknown functions of fewer variables (most often the approximation is represented as a product $\psi_1(\mathbf{x}_1)\psi_2(\mathbf{x}_2) \cdots \psi_N(\mathbf{x}_N)$ or as a linear combination of such

products). Equations for the functions ψ_1, \dots, ψ_N are found from the variational principle. We shall become familiar with this way when we study complex atoms.

Example. Let us use the variational principle for an approximate computation of the ground state of a helium atom. The Schrödinger operator for helium in atomic units has the form

$$H = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}}$$

As a test function we take³⁴

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \alpha) = e^{\alpha r_1} e^{\alpha r_2}.$$

Computations which we omit give a simple expression for the functional

$$E(\alpha) = \alpha^2 - \frac{27}{8}\alpha.$$

The minimum of this expression is attained for $\alpha = 27/16$, and the approximate theoretical value of the energy of the ground state is

$$E_0 = E(27/16) = -(27/16)^2 \cong -2.85.$$

The experimental value is $E_{0\text{exp}} = -2.90$. We see that such a simple computation leads to very good agreement with experiment. As would be expected, the theoretical value is greater than the experimental value.

We remark that $e^{-\alpha r}$ is an eigenfunction of the ground state of a particle in a Coulomb field $-\alpha/r$. Therefore, the approximate eigenfunction $e^{-27(r_1+r_2)/16}$ is an exact eigenfunction for the operator

$$H' = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{27}{16r_1} - \frac{27}{16r_2}.$$

The interaction between the electrons is taken into account in the approximate Schrödinger operator H' by replacing the charge $Z = 2$ of the nucleus by $Z' = 27/16$, and by the same token the screening of the nucleus by the electron charge is taken into account.

³⁴This choice of a test function can be explained by the fact that the function $e^{-2r_1-2r_2}$ is an exact eigenfunction of the operator $H - 1/r_{12}$. Indeed, if the term $1/r_{12}$ is removed from H , then by separation of variables the problem can be reduced to the problem of a hydrogen-like ion, and, as has been shown, the eigenfunction of the ground state of such an ion is $e^{-Z'}$, where Z is the charge of the nucleus.

In conclusion we remark that the computations of the helium atom used test functions with a huge number of parameters, and the accuracy attained was such that the existing deviations from experiment can be explained by relativistic corrections. Such an accurate solution of the problem of the ground state of the helium atom has had fundamental significance for quantum mechanics and confirms the validity of its equations for the three-body problem.

§ 35. Scattering theory. Physical formulation of the problem

We begin with a physical formulation of the scattering problem. Suppose that a beam of particles of type a obtained from an accelerator hits a target consisting of particles of type b . The scheme of such an experiment is represented in Figure 11

The particles a and b can be either elementary (for example, electrons, protons, neutrons) or compound (for example, atoms, molecules, atomic nuclei). The experimenter studies the physical characteristics of the particles going out from the target. If they differ from the corresponding characteristics of the impinging particles, then one can say that a particle a has been scattered.

One usually tries to choose the thickness of the target so that a sufficiently large number of particles a are scattered by the particles b , and at the same time so that the portion of the particles a experiencing multiple collisions is negligibly small. In this case to explain the experimental results it suffices to study the problem of scattering of a particle a by a particle b . This is a two-body problem

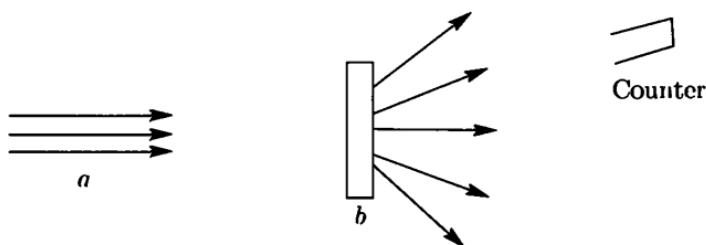


Figure 11

if a and b are elementary particles, and a many-body problem if a and b are compound particles. We recall that the two-body problem can be reduced to the problem of the motion of a particle in the field of a fixed force center by separating out the motion of the center of inertia. This is the simplest problem of scattering theory.

In scattering by a force center, a particle can change only its direction of motion due to the law of conservation of energy. In this case one speaks of elastic scattering. More complex processes are possible when compound particles collide. For example, when an electron collides with a hydrogen atom, elastic scattering is possible (the state of the atom does not change), scattering with excitation is possible (the electron transfers part of its energy to the atom, which then passes into a new state), and finally, ionization of the atom by the electron is possible. Each such process is called a scattering channel. Scattering of a particle by a force center is single-channel scattering, while scattering of compound particles is usually multichannel scattering. If, however, the colliding particles a and b are in the ground state and the energy of their relative motion is less than the energy of excitation, then the scattering is single channel.

The basic characteristic of the diverse scattering processes measured in experiments is their cross-section, which we define below. Some set of possible results of scattering is called a scattering process. The following are processes in this sense:

- 1) Elastic scattering into a solid angle element $d\Omega$ constructed about the direction \mathbf{n} ;
- 2) Elastic scattering at an arbitrary angle;
- 3) Scattering into a solid angle $d\Omega$ with excitation of the target from the i th level to the k th level;
- 4) Scattering with ionization of the target;
- 5) A process consisting in scattering in general taking place, and so on.

The probability N of some scattering process of a by b depends on a certain quantity characterizing the accuracy of using particles a to "shoot" a particle b . To introduce such a characterization of the state of an incoming particle a , we construct the plane passing

through the point where the scatterer b is located and perpendicular to the momentum of a . The probability dW of a crossing an area element dS of this plane is proportional to dS : that is, $dW = I dS$. It is clear that the probability N will be proportional to the quantity I computed at the point where the scatterer is located.³⁵

It now seems natural to define the cross-section σ by

$$\sigma = \frac{N}{l}.$$

For the five processes listed above, the cross-sections have the following names: 1) the differential cross-section of elastic scattering; 2) the total cross-section of elastic scattering. 3) the differential cross-section of excitation; 4) the ionization cross-section; 5) the total cross-section. The concept of cross-section becomes especially intuitive if we assume that there is complete determinism of the scattering results. In the case of such determinism, the scattering result would be determined by the point of the transverse cross-section of the beam through which the particle would pass in the absence of the scatterer. To the scattering process would then correspond some region in the plane of the transverse cross-section, and the cross-section would be equal to the area of this region.

There are two problems in scattering theory: from known interaction potentials between the particles determine the cross-sections of the various processes (the direct problem), and from the known cross-section obtain information about the interaction of the particles (the inverse problem).

In these lectures we confine ourselves to a study of the direct problem of scattering of a particle by a potential center, beginning with the simplest one-dimensional case.

§ 36. Scattering of a one-dimensional particle by a potential barrier

We structure the exposition of this problem according to the following plan. First we formulate the so-called stationary scattering problem.

³⁵Here the scatterer b is assumed to be distant, since I characterizes the state of a freely moving particle a .

For this we study the solutions of an equation $H\psi = E\psi$ of a certain special form. The physical meaning of such solutions will be explained later, after we have used them to construct solutions of the nonstationary (or time-dependent) Schrödinger equation $i\frac{\partial\psi}{\partial t} = H\psi$. For simplicity we employ a system of units in which $\hbar = 1$ and $m = 1/2$.

In the coordinate representation, the Schrödinger operator has the form

$$(1) \quad H = -\frac{d^2}{dx^2} + V(x).$$

We take the function $V(x)$ to be compactly supported ($V(x) = 0$ for $|x| > a$) and piecewise continuous.

Let us denote the regions $x < -a$, $x > a$, and $-a < x < a$ of the real axis by *I*, *II*, and *III*, respectively. The scattering problem is a problem of infinite motion of a particle. Such motion is possible for $E > 0$, and we know that for $E > 0$ the spectrum of the Schrödinger operator is continuous. From a mathematical point of view, scattering problems are problems about the continuous spectrum of the Schrödinger operator.

The stationary Schrödinger equation with the form

$$(2) \quad -\psi'' + V(x)\psi = k^2\psi, \quad E = k^2, \quad k > 0,$$

on the whole real axis simplifies in the regions *I* and *II*:

$$(3) \quad \psi'' + k^2\psi = 0.$$

The equation (3) has the two linearly independent solutions e^{ikx} and e^{-ikx} . The solutions of (2) on the whole axis can be constructed by sewing together the solutions in the regions *I* *III*. In doing this, we must use the conditions of continuity of the solutions and their first derivatives at the points $-a$ and a . This imposes four conditions on the six arbitrary constants appearing in the expressions for the general solutions in the regions *I* *III*. The fifth condition for these constants is the normalization condition. Therefore, we can construct

linearly independent solutions of (2) with the form³⁶

$$(4) \quad \begin{array}{lll} \text{I} & & \text{II} \\ \psi_1(k, x), & e^{ikx} + A(k) e^{-ikx}, & B(k) e^{ikx}. \\ \psi_2(k, x), & D(k) e^{-ikx}, & e^{-ikx} + C(k) e^{ikx}, \end{array}$$

in regions I and II. For example, in constructing ψ_1 we use the arbitrariness of one of the constants to get that the coefficient of e^{-ikx} is equal to zero in the region II. We then choose the coefficient of e^{ikx} to be equal to 1 in the region I, thereby determining the normalization of the function ψ_1 . The coefficients A and B are found from the conditions for sewing together, along with the constants m and n , where $m\varphi_1 + n\varphi_2$ is the general solution of (2) in the region III. It is not hard to see that the conditions for sewing together lead to a nonhomogeneous linear system of equations for A , B , m , and n with determinant that is nonzero if φ_1 and φ_2 are linearly independent. The solution ψ_2 is constructed similarly. The linear independence of ψ_1 and ψ_2 follows from the fact that the Wronskian of these solutions is not zero.

Let us determine the properties of the coefficients A , B , C , and D . For this we note that the Wronskian $W(\varphi_1, \varphi_2) = \varphi_1\varphi'_2 - \varphi'_1\varphi_2$ of the solutions φ_1 and φ_2 of (2) does not depend on x . Indeed, suppose that φ_1 and φ_2 satisfy (2). Then

$$\varphi''_1 - (V - k^2)\varphi_1 = 0, \quad \varphi''_2 - (V - k^2)\varphi_2 = 0.$$

Multiplying the first equation by φ_2 and the second by φ_1 and subtracting one from the other, we get that

$$\frac{d}{dx}(\varphi_1\varphi'_2 - \varphi'_1\varphi_2) = 0.$$

Using this property, we can equate the Wronskian for any pair of solutions in the regions I and II. Choosing (ψ_1, ψ_2) , $(\psi_1, \bar{\psi}_1)$, $(\psi_2, \bar{\psi}_2)$, and $(\psi_1, \bar{\psi}_2)$ as such pairs successively, and using the equalities $W(e^{-ikx}, e^{ikx}) = 2ik$ and $W(e^{\pm ikx}, e^{\pm ikx}) = 0$, we arrive at the

³⁶If the potential is not compactly supported but decreases sufficiently rapidly as $|x| \rightarrow \infty$, then these expressions for ψ_1 and ψ_2 in I and II should be regarded as the asymptotics of ψ_1 and ψ_2 as $x \rightarrow \pm\infty$. We shall not study this case.

following relations between the coefficients A , B , C , and D :

$$(5) \quad B = D,$$

$$(6) \quad |A|^2 + |B|^2 = 1,$$

$$(7) \quad |B|^2 + |C|^2 = 1,$$

$$(8) \quad A\bar{B} + B\bar{C} = 0.$$

(For example, $W(\psi_1, \bar{\psi}_1) = -2ik(1 - AA)$ in I and $W(\psi_1, \bar{\psi}_1) = -2ikBB$ in II. Equating these expressions, we get (6).)

The relations (5)–(8) show that the matrix S consisting of the coefficients A , $B = D$, and C , that is,

$$S = \begin{pmatrix} A & B \\ B & C \end{pmatrix},$$

is symmetric and unitary. This matrix is called the scattering matrix or simply the S -matrix. We shall see below that all the physically interesting results can be obtained if the S -matrix is known, and therefore the computation of its elements is the main problem in one-dimensional scattering theory.

Let us consider the question of normalization of the functions $\psi_1(k, x)$ and $\psi_2(k, x)$. We have the formulas

$$(9) \quad \int_{-\infty}^{\infty} \psi_1(k_1, x) \overline{\psi_2(k_2, x)} dx = 0, \quad k_1 > 0, \quad k_2 > 0,$$

$$(10) \quad \int_{-\infty}^{\infty} \psi_{1,2}(k_1, x) \overline{\psi_{1,2}(k_2, x)} dx = 2\pi\delta(k_1 - k_2),$$

that is, the same relations hold as for the functions e^{ikx} and e^{-ikx} , and the normalization does not depend on the form of the potential $V(x)$. The integrals in the formulas (9) and (10) are understood in the principal value sense. Let us verify (10) for the function ψ_1 . The remaining two relations are verified in the same way. Substitution of $\psi_1(k_1, x)$ and $\overline{\psi_1(k_2, x)}$ in (2) leads to the equalities

$$(11) \quad \psi''(k_1, x) + k_1^2 \psi(k_1, x) = V(x) \psi(k_1, x),$$

$$(12) \quad \overline{\psi''(k_2, x)} + k_2^2 \overline{\psi(k_2, x)} = V(x) \overline{\psi(k_2, x)}.$$

(For brevity we do not write the index 1 in the notation of the solution ψ_1 .) Multiplying (11) by $\overline{\psi(k_2, x)}$ and (12) by $\psi(k_1, x)$ and

subtracting the first from the second, we get that

$$\frac{d}{dx} W(\psi(k_1, x), \overline{\psi(k_2, x)}) = (k_1^2 - k_2^2) \psi(k_1, x) \overline{\psi(k_2, x)}.$$

Integrating this equality, we have

$$(13) \quad \int_{-N}^N \psi(k_1, x) \overline{\psi(k_2, x)} dx = \frac{1}{k_1^2 - k_2^2} W(\psi(k_1, x), \overline{\psi(k_2, x)}) \Big|_{-N}^N.$$

We are interested in the limit as $N \rightarrow \infty$ of the integral on the left-hand side of (13) in the sense of generalized functions. It is clear already from (13) that this limit depends only on the form of the solutions in the regions I and II (for the case of infinite potentials it depends only on the asymptotics of the solutions as $x \rightarrow \pm\infty$).

After simple computations we get that

$$\begin{aligned} & \int_{-N}^N \psi(k_1, x) \overline{\psi(k_2, x)} dx \\ &= -\frac{i}{k_1 - k_2} [(A(k_1) \overline{A(k_2)} + B(k_1) \overline{B(k_2)}) e^{i(k_1 - k_2)N} - e^{-i(k_1 - k_2)N}] \\ & \quad - \frac{i}{k_1 + k_2} [\overline{A(k_2)} e^{-i(k_1 + k_2)N} + \overline{A(k_1)} e^{i(k_1 + k_2)N}]. \end{aligned}$$

By the Riemann-Lebesgue theorem, the second term tends to zero (in the sense of generalized functions). The similar assertion is not true for the first term, since it is singular for $k_1 = k_2$. The singular part of this term does not change if we replace $\overline{A(k_2)}$ and $\overline{B(k_2)}$ by $\overline{A(k_1)}$ and $\overline{B(k_1)}$, respectively.³⁷ Using (6), we have

$$\int_{-N}^N \psi_1(k_1, x) \overline{\psi_1(k_2, x)} dx = \frac{2}{k_1 - k_2} \sin(k_1 - k_2) N + F(k_1, k_2, N),$$

where

$$\lim_{n \rightarrow \infty} F(k_1, k_2, N) = 0.$$

Finally, using the well-known formula

$$\lim_{N \rightarrow \infty} \frac{\sin Nx}{x} = \pi \delta(x),$$

³⁷For this it suffices that $A(k)$ and $B(k)$ be differentiable functions of k , which can be proved

we get that

$$\lim_{N \rightarrow \infty} \int_{-N}^N \psi_1(k_1, x) \overline{\psi_1(k_2, x)} dx = 2\pi\delta(k_1 - k_2),$$

which coincides with (10).

§ 37. Physical meaning of the solutions ψ_1 and ψ_2

To clarify the physical meaning of the solutions ψ_1 and ψ_2 , we use them to construct solutions of the nonstationary Schrödinger equation

$$i \frac{d\varphi}{dt} = H\varphi$$

For this equation we consider the solution

$$(1) \quad \varphi_1(x, t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty C(k) \psi_1(k, x) e^{-ik^2 t} dk$$

constructed from the function $\psi_1(k, x)$. The function $C(k)$ is assumed to be nonzero only in a small neighborhood of the point k_0 . In this case $\varphi_1(x, t)$ has the simplest physical meaning. Moreover, we assume that

$$\int_0^\infty |C(k)|^2 dk = 1.$$

and then it follows from (36.10) that

$$\int_{-\infty}^\infty |\varphi_1(x, t)|^2 dx = 1,$$

that is, the solution $\varphi_1(x, t)$ has the correct normalization. Using the concentration of the function $C(k)$ in a neighborhood of k_0 and the continuity of the functions $A(k)$ and $B(k)$, we can write approximate expressions³⁸ for the function $\varphi_1(x, t)$ in the regions *I* and *II*:

$$I: \varphi_1(x, t) \cong \varphi_+(x, t) + A(k_0) \varphi_-(x, t).$$

$$II. \varphi_1(x, t) \cong B(k_0) \varphi_+(x, t),$$

where

$$(2) \quad \varphi_\pm(x, t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty C(k) e^{\pm i k x - i k^2 t} dk.$$

³⁸These expressions can be regarded as having any desired degree of accuracy if the interval Δk inside which $C(k)$ is nonzero is sufficiently small. However, we cannot pass to the limit, replacing $C(k)$ by a δ -function, because we do not get a square-integrable solution of the Schrödinger equation

The functions $\varphi_{\pm}(x, t)$ are normalized solutions of the Schrödinger equation for a free particle, and were studied³⁹ in § 15. In the same section we constructed asymptotic expressions for these solutions as $t \rightarrow \pm\infty$:

$$\varphi_{\pm}(x, t) = \frac{1}{\sqrt{2|t|}} C\left(\pm \frac{x}{2t}\right) e^{i\chi} + O\left(\frac{1}{|t|}\right),$$

where χ is a real function whose form is not important for us. From this expression it is clear that for $|t| \rightarrow \infty$ the functions $\varphi_{+}(x, t)$ are nonzero only in a neighborhood of the points $x = \pm 2k_0 t$. Therefore, φ_{+} describes the state of a free particle moving from left to right with velocity⁴⁰ $v = 2k_0$, and φ_{-} describes a particle having the oppositely directed velocity (recall that $m = 1/2$)

It is now easy to see what properties the solution $\varphi_1(x, t)$ has as $t \rightarrow \pm\infty$. Suppose that $t \rightarrow -\infty$. Then in the regions I and II we have

$$I: \varphi_1(x, t) = \varphi_{-}(x, t),$$

$$II: \varphi_1(x, t) = 0,$$

since as $t \rightarrow -\infty$ we have $\varphi_{-}(x, t) = 0$ for $x < -a$ and $\varphi_{-}(x, t) = 0$ for $x > a$. Similarly, as $t \rightarrow +\infty$

$$I: \varphi_1(x, t) = A(k_0) \varphi_{-}(x, t),$$

$$II: \varphi_1(x, t) = B(k_0) \varphi_{+}(x, t).$$

We see that long before the scattering ($t \rightarrow -\infty$) the probability is 1 that the particle is to the left of the barrier and is moving toward the barrier with velocity $2k_0$.

Let us compute the probabilities W_I and W_{II} of observing the particle as $t \rightarrow +\infty$ in the regions I and II, respectively. We have

$$\begin{aligned} W_I &= \int_{-\infty}^{-a} |\varphi_1(x, t)|^2 dx = |A(k_0)|^2 \int_{-\infty}^a |\varphi_{-}(x, t)|^2 dx \\ &= |A(k_0)|^2 \int_{-\infty}^{\infty} |\varphi_{-}(x, t)|^2 dx = |A(k_0)|^2. \end{aligned}$$

³⁹There is an inessential difference in the notation. Here it is more convenient for us to take $k > 0$ and to write out the sign of the momentum in the exponential $e^{\pm ikx}$ explicitly. Moreover, the integration in (2) can be extended to the whole real axis, since $C(k) = 0$ for $k < 0$

⁴⁰More precisely, the region in which there is a nonzero of finding the particle moves with velocity $2k_0$

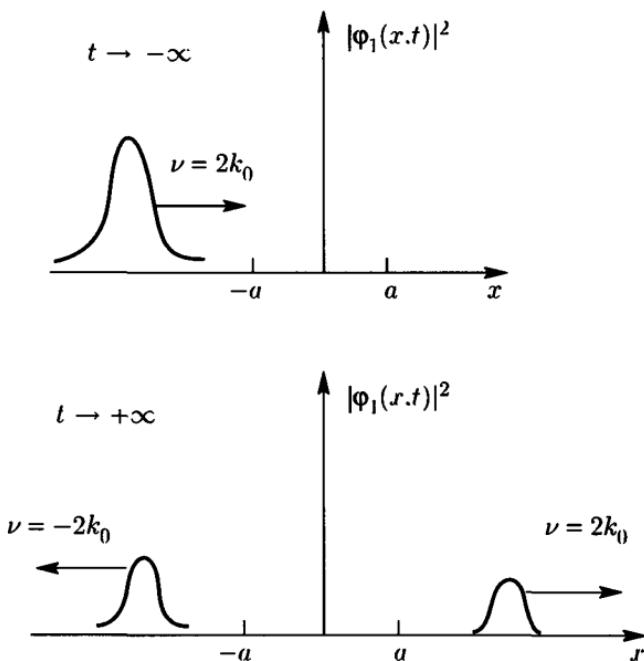


Figure 12

It is possible to replace the region I of integration by the whole real axis, because for $t \rightarrow +\infty$ we have $\varphi_-(x,t) \neq 0$ only in I . In exactly the same way we can verify that

$$W_{II} = |B(k_0)|^2.$$

The graphs of the function $|\varphi_1(x,t)|^2$ as a function of x for $t \rightarrow \pm\infty$ are shown in Figure 12.

Thus, the solution $\varphi_1(x,t)$ of the Schrödinger equation describes a particle that approaches the potential barrier with velocity $2k_0$ before the scattering and then with probability $|A(k_0)|^2$ it is reflected from the barrier or with probability $|B(k_0)|^2$ it passes through the potential barrier.⁴¹

We point out that the result does not depend on the form of the function $C(k)$. It is important only that the interval Δk in which

⁴¹In the one-dimensional problem the concept of cross-section loses its meaning. All information about the scattering is contained in the probabilities $|A|^2$ and $|B|^2$.

$C(k)$ is nonzero be small. Physically, this requirement is understandable if we want to experimentally determine the dependence of, say, the reflection coefficient $|A(k)|^2$ on k : we must use particles in a state with variance of k as small as possible (states with zero variance do not exist). In concrete computations of the reflection and transmission coefficients $|A(k)|^2$ and $|B(k)|^2$, it is not necessary to solve the nonstationary Schrödinger equation; it suffices to find the solution $\psi_1(k, x)$. We remark that the solution $\varphi_2(x, t)$, which can be constructed from the function $\psi_2(k, x)$, has the same meaning, except that the particle is approaching the barrier from the right.

Let us recall the properties of the scattering matrix S . The equality $B = D$ leads to equality of the probabilities of passing through the barrier in opposite directions and, as can be shown, is a consequence of the invariance of the Schrödinger equation with respect to time reversal. The equalities

$$|A|^2 + |B|^2 = 1, \quad |B|^2 + |C|^2 = 1$$

express the law of conservation of probability. Indeed, the normalization of the solutions $\varphi_1(x, t)$ and $\varphi_2(x, t)$ does not depend on the time, and as $t \rightarrow \infty$ we have

$$\lim_{t \rightarrow \infty} \int_{-\infty}^{\infty} |\varphi_1(x, t)|^2 dx = |A(k_0)|^2 + |B(k_0)|^2 = 1.$$

§ 38. Scattering by a rectangular barrier

We consider the concrete problem of scattering of a one-dimensional particle by a rectangular potential barrier. Let

$$V(x) = \begin{cases} V_0, & |x| \leq a, V_0 > 0, \\ 0, & |x| > a. \end{cases}$$

In this case the equation $H\psi = k^2\psi$ has the following form in the regions I–III:

$$(1) \quad \begin{aligned} \text{I and II: } & \psi'' + k^2\psi = 0, \\ \text{III: } & \psi'' + \alpha^2\psi = 0, \end{aligned}$$

where $\alpha = \sqrt{k^2 - V_0}$ (for definiteness we assume that $\alpha > 0$ for $k^2 - V_0 > 0$ and $-i\alpha > 0$ for $k^2 - V_0 < 0$).

We construct a solution $\psi_1(kx)$. In the regions I-III this solution has the form

$$(2) \quad I: e^{ikx} + Ae^{-ikx},$$

$$(3) \quad II: Be^{ikx}.$$

$$III: me^{i\alpha x} + ne^{-i\alpha x}.$$

The coefficients A , B , m , n are found from the conditions that ψ_1 and its derivative be continuous at the points a and $-a$:

$$e^{-ika} + Ae^{ika} = me^{-i\alpha a} + ne^{i\alpha a},$$

$$k(e^{-ika} - Ae^{ika}) = \alpha(me^{-i\alpha a} - ne^{i\alpha a}).$$

$$me^{i\alpha a} + ne^{-i\alpha a} = Be^{ika},$$

$$\alpha(me^{i\alpha a} - ne^{-i\alpha a}) = kB e^{ika}.$$

We write the expression for the coefficient B :

$$B = \left[\frac{(\alpha+k)^2}{4\alpha k} e^{2i(k-\alpha)a} - \frac{(\alpha-k)^2}{4\alpha k} e^{2i(k+\alpha)a} \right]^{-1}.$$

It is not hard to verify that $B \rightarrow 0$ under any of the conditions

$$1) k \rightarrow 0, \quad 2) V_0 \rightarrow \infty, \quad 3) a \rightarrow \infty, \quad k^2 < V_0.$$

On the other hand, $B \rightarrow 1$ under one of the conditions

$$4) k \rightarrow \infty, \quad 5) V_0 \rightarrow 0. \quad 6) a \rightarrow 0.$$

We see that in the limit cases 1)–5) the results obtained on the basis of quantum mechanics coincide with the classical results.⁴² Figure 13 shows the graph of the function $|B(k)|^2$. From this graph it is clear that for certain finite values of k the probability of transmission is $|B|^2 = 1$. It is interesting to note that the equations (1) together with the conditions (2) and (3) describe the passage of light waves through transparent plates. Here $|B|^2$ and $|A|^2$ are proportional to the intensities of the transmitted and reflected waves. In the case $|B|^2 = 1$, $|A|^2 = 0$ the reflected wave is absent. This phenomenon is used for coated optics.

⁴² According to classical mechanics a particle passes through the barrier with probability 1 when $k^2 > V_0$ and is reflected with probability 1 when $k^2 < V_0$.

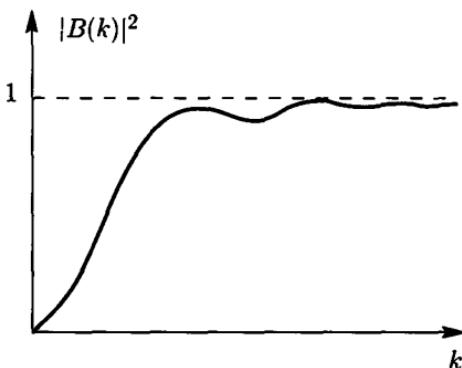


Figure 13

§ 39. Scattering by a potential center

For the problem of scattering by a potential center, the Schrödinger operator has the form

$$(1) \quad H = -\Delta + V(\mathbf{x}).$$

(We again assume that $m = 1/2$ and $\hbar = 1$.) In what follows we usually assume that the potential $V(\mathbf{x})$ is a compactly supported function ($V(\mathbf{x}) = 0$ for $|\mathbf{x}| > a$). The exposition is according to the same scheme as in the one-dimensional case, that is, we first consider some solutions of the stationary Schrödinger equation, and then we explain their physical meaning with the help of the nonstationary Schrödinger equation.

Our first problem is to formulate for the solution of the equation

$$-\Delta\psi(\mathbf{x}) + V(\mathbf{x})\psi(\mathbf{x}) = k^2\psi(\mathbf{x})$$

an asymptotic (as $r \rightarrow \infty$) condition that corresponds to the physical picture of scattering and which is an analogue of the conditions (36.4) for the one-dimensional problem. It is natural to expect that one term will correspond asymptotically to the particle impinging on the scattering center in a definite direction, while the second term will correspond to the scattered particle, which can have different directions of motion after the scattering and which goes away from the center. The analogues of the functions e^{ikx} and e^{-ikx} in the

three-dimensional case are the functions e^{ikr}/r and e^{-ikr}/r . Therefore, it is reasonable to assume that the terms $(e^{-ikr}/r)\delta(\mathbf{n} + \omega)$ and $(e^{ikr}/r)S(k, \mathbf{n}, \omega)$ correspond to the particle before and after scattering. We use the following notation: \mathbf{k} is the momentum of the impinging particle. $\omega = \mathbf{k}/k$. $\mathbf{n} = \mathbf{x}/r$, $\delta(\mathbf{n} - \mathbf{n}_0)$ is the δ -function on the unit sphere, defined by the equality

$$\int_{S_2} f(\mathbf{n}) \delta(\mathbf{n} - \mathbf{n}_0) d\mathbf{n} = f(\mathbf{n}_0), \quad d\mathbf{n} = \sin \theta d\theta d\varphi.$$

and $S(k, \mathbf{n}, \omega)$ is a certain function which will be shown to contain all information about the scattering process and which must be found in solving the problem. We shall see that $S(k, \mathbf{n}, \omega)$ is the kernel of a certain unitary operator S , which is called the scattering operator

We come to the following statement of the problem of scattering by a force center: it is required to find for the equation

$$(2) \quad -\Delta\psi(\mathbf{x}, \mathbf{k}) + V(\mathbf{x})\psi(\mathbf{x}, \mathbf{k}) = k^2\psi(\mathbf{x}, \mathbf{k})$$

a solution which has the asymptotic behavior

$$(3) \quad \psi(\mathbf{x}, \mathbf{k}) = \frac{e^{-ikr}}{r}\delta(\mathbf{n} + \omega) - \frac{e^{ikr}}{r}S(k, \mathbf{n}, \omega) + o\left(\frac{1}{r}\right)$$

as $r \rightarrow \infty$.

This formulation of the problem can be justified only with the help of the nonstationary formalism of scattering theory, and this will be done in the next section. The question of the existence of a solution of (2) with the condition (3) will also be discussed later. However, this question has a simple answer for the case $V(\mathbf{x}) = 0$. We show that the function

$$\psi_0(\mathbf{x}, \mathbf{k}) = \frac{k}{2\pi i}e^{i\mathbf{kx}} = \frac{k}{2\pi i}e^{ik\mathbf{n}\omega},$$

which is obviously a solution of (2) for $V(\mathbf{x}) = 0$, also satisfies the condition (3). and we find the form of the function $S(k, \mathbf{n}, \omega)$ for this case.

We are looking for the asymptotics of the function $\psi_0(\mathbf{x}, \mathbf{k})$ in the class of generalized functions, and therefore we must find an asymptotic expression for the integral

$$I = \int_{S_2} f(\mathbf{n}) \psi_0(r\mathbf{n}, k\omega) d\mathbf{n}$$

as $r \rightarrow \infty$, where $f(\mathbf{n})$ is a smooth function. Using the explicit form of $\psi_0(\mathbf{x}, \mathbf{k})$ and introducing a spherical coordinate system with polar axis directed along the vector ω , we have

$$\begin{aligned} I &= \frac{k}{2\pi i} \int_0^{2\pi} d\varphi \int_0^\pi f(\cos \theta, \varphi) e^{ikr \cos \theta} \sin \theta d\theta \\ &= \frac{k}{2\pi i} \int_0^{2\pi} d\varphi \int_{-1}^1 d\eta f(\eta, \varphi) e^{ikr\eta}. \end{aligned}$$

Integrating by parts, we get that

$$\begin{aligned} I &= \frac{k}{2\pi i} \int_0^{2\pi} d\varphi \left. \frac{1}{ikr} f(\eta, \varphi) e^{ikr\eta} \right|_{-1}^1 \\ &\quad - \frac{k}{2\pi i} \int_0^{2\pi} d\varphi \frac{1}{ikr} \int_{-1}^1 d\eta f'_\eta(\eta, \varphi) e^{ikr\eta}. \end{aligned}$$

Integrating by parts again, we easily see that the second term has order $O(1/r^2)$, and hence

$$\begin{aligned} I &= -\frac{1}{2\pi r} \int_0^{2\pi} d\varphi (e^{ik\eta} f(\omega) - e^{-ik\eta} f(-\omega)) + O\left(\frac{1}{r^2}\right) \\ &= \frac{e^{-ikr}}{r} f(-\omega) - \frac{e^{ikr}}{r} f(\omega) + O\left(\frac{1}{r^2}\right). \end{aligned}$$

Thus, we have shown that

$$(4) \quad \psi_0(\mathbf{x}, \mathbf{k}) = \frac{e^{-ikr}}{r} \delta(\mathbf{n} + \omega) - \frac{e^{ikr}}{r} \delta(\mathbf{n} - \omega) + O\left(\frac{1}{r^2}\right).$$

Comparing (4) with (3), we find the function S for a free particle:

$$S_0(k, \mathbf{n}, \omega) = \delta(\mathbf{n} - \omega).$$

Using (4), we can rewrite the asymptotic condition (3) in the form

$$(5) \quad \psi(\mathbf{x}, \mathbf{k}) = \frac{k}{2\pi i} \left(e^{ikr n \omega} + f(k, \mathbf{n}, \omega) \frac{e^{ikr}}{r} \right) + o\left(\frac{1}{r}\right).$$

where the function $f(k, \mathbf{n}, \omega)$ is connected with the function $S(k, \mathbf{n}, \omega)$ by the relation

$$(6) \quad S(k, \mathbf{n}, \omega) = \delta(\mathbf{n} - \omega) + \frac{ik}{2\pi} f(k, \mathbf{n}, \omega).$$

The function $f(k, \mathbf{n}, \omega)$ is called the scattering amplitude. Obviously, $f(k, \mathbf{n}, \omega) = 0$ for $V(\mathbf{x}) = 0$. We shall see that the scattering cross-section is most simply expressed in terms of this function.

Instead of the asymptotic condition (5), one often uses

$$(7) \quad \psi(\mathbf{x}, \mathbf{k}) = e^{ik\mathbf{n}\omega} + f(\mathbf{k}, \mathbf{n}, \omega) \frac{e^{ikr}}{r} + o\left(\frac{1}{r}\right).$$

which leads to a more convenient normalization of the function $\psi(\mathbf{x}, \mathbf{k})$.

Let us now consider the properties of the function $S(k, \mathbf{n}, \omega)$. The following auxiliary assertion turns out to be convenient for the study of these properties.

Suppose that the functions $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$ satisfy the equation (2) and the asymptotic conditions

$$\begin{aligned} \psi_1(\mathbf{x}) &= A_1(\mathbf{n}) \frac{e^{-ikr}}{r} + B_1(\mathbf{n}) \frac{e^{ikr}}{r} + o\left(\frac{1}{r}\right), \\ \psi_2(\mathbf{x}) &= A_2(\mathbf{n}) \frac{e^{-ikr}}{r} + B_2(\mathbf{n}) \frac{e^{ikr}}{r} + o\left(\frac{1}{r}\right) \end{aligned}$$

as $r \rightarrow \infty$, which can be differentiated once with respect to r . Then

$$(8) \quad \int_{S_2} A_1(\mathbf{n}) B_2(\mathbf{n}) d\mathbf{n} = \int_{S_2} A_2(\mathbf{n}) B_1(\mathbf{n}) d\mathbf{n}$$

This assertion is easily proved with the help of Green's formula

$$(9) \quad \int_{\Omega} (\psi_1 \Delta \psi_2 - \psi_2 \Delta \psi_1) d\mathbf{x} = \int_{\partial\Omega} \left(\psi_1 \frac{\partial \psi_2}{\partial n} - \psi_2 \frac{\partial \psi_1}{\partial n} \right) d\Sigma.$$

We should choose the domain of integration Ω to be the ball of radius R and take the limit as $R \rightarrow \infty$, obtaining

$$\psi_1(\mathbf{x}) = \psi(\mathbf{x}, \mathbf{k}) = \frac{e^{-ikr}}{r} \delta(\mathbf{n} + \omega) - \frac{e^{ikr}}{r} S(k, \mathbf{n}, \omega) + o\left(\frac{1}{r}\right),$$

$$\psi_2(\mathbf{x}) = \psi(\mathbf{x}, \mathbf{k}') = \frac{e^{-ikr}}{r} \delta(\mathbf{n} + \omega') - \frac{e^{ikr}}{r} S(k, \mathbf{n}, \omega') + o\left(\frac{1}{r}\right),$$

$$\psi_3(\mathbf{x}) = \overline{\psi(\mathbf{x}, \mathbf{k}')} = -\frac{e^{-ikr}}{r} \overline{S(k, \mathbf{n}, \omega')} + \frac{e^{ikr}}{r} \delta(\mathbf{n} + \omega') + o\left(\frac{1}{r}\right),$$

where $\mathbf{k}' = k\omega'$. Applying the formula (8) to the functions ψ_1 and ψ_2 , we get that

$$S(k, -\omega, \omega') = S(k, -\omega', \omega),$$

or, replacing ω by $-\omega$,

$$(10) \quad S(k, \omega, \omega') = S(k, -\omega', -\omega).$$

This formula is an analogue of the equality $B = D$ for the one-dimensional scattering problem and expresses the fact that the values of S for the direct ($\omega' \rightarrow \omega$) and time-reversed ($-\omega \rightarrow -\omega'$) collision processes coincide. It can be shown that this property (just like the symmetry of the S -matrix in the one-dimensional case) is a consequence of the invariance of the Schrödinger equation with respect to time reversal.

Further, applying the formula (8) to the functions ψ_1 and ψ_2 , we get that

$$(11) \quad \int_{S_2} \overline{S(k, \mathbf{n}, \omega')} S(k, \mathbf{n}, \omega) d\mathbf{n} = \delta(\omega - \omega').$$

If we consider the function $S(k, \mathbf{n}, \omega)$ as the kernel of an integral operator

$$S\varphi(\omega) = \int_{S_2} S(k, \omega, \omega') \varphi(\omega') d\omega'$$

acting in $L^2(S_2)$, then (11) can be rewritten in the form

$$S^* S = I$$

In view of (10) it follows from (11) that

$$SS^* = I.$$

The last two relations imply that the operator S is unitary.

We see that for the problem of scattering by a potential center the operator S has the same properties as the S -matrix for the one-dimensional scattering problem. Since S is unitary, we have the important relation

$$(12) \quad \int_{S_2} |f(k, \mathbf{n}, \omega)|^2 d\mathbf{n} = \frac{4\pi}{k} \operatorname{Im} f(k, \omega, \omega),$$

which is called the optical theorem. Indeed, using (6) and (11), we get that

$$\begin{aligned} \int_{S_2} \left[\delta(\mathbf{n} - \omega') - \frac{ik}{2\pi} \overline{f(k, \mathbf{n}, \omega')} \right] \left[\delta(\mathbf{n} - \omega) + \frac{ik}{2\pi} f(k, \mathbf{n}, \omega) \right] d\mathbf{n} \\ = \delta(\omega - \omega') + \frac{ik}{2\pi} [f(k, \omega', \omega) - \overline{f(k, \omega, \omega')}] \\ + \frac{k^2}{4\pi^2} \int_{S_2} f(k, \mathbf{n}, \omega) f(k, \mathbf{n}, \omega') d\mathbf{n} = \delta(\omega - \omega'). \end{aligned}$$

Setting $\omega = \omega'$, we immediately arrive at the formula (12).

In § 42 we shall see that the integral on the left-hand side of (12) coincides with the total cross-section σ for scattering of a particle by a potential center. Therefore, (12) can be rewritten as

$$\sigma = \frac{4\pi}{k} \operatorname{Im} f(k, \omega, \omega).$$

This formula connects the total scattering cross-section with the imaginary part of the zero-angle scattering amplitude.

By using the unitarity of the operator S , it is easy to show that the functions $\psi(\mathbf{x}, \mathbf{k})$ satisfying the asymptotic condition (7) have the normalization

$$(13) \quad \int_{\mathbf{R}^3} \psi(\mathbf{x}, \mathbf{k}) \overline{\psi(\mathbf{x}, \mathbf{k}')} d\mathbf{x} = (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}');$$

that is, they are normalized just like the free-particle solutions $e^{i\mathbf{k}\mathbf{x}}$. To verify (13) we multiply the equalities

$$\Delta\psi(\mathbf{x}, \mathbf{k}) + k^2\psi(\mathbf{x}, \mathbf{k}) = V(\mathbf{x})\psi(\mathbf{x}, \mathbf{k}).$$

$$\Delta\overline{\psi(\mathbf{x}, \mathbf{k}')} + k'^2\overline{\psi(\mathbf{x}, \mathbf{k}'')} = V(\mathbf{x})\overline{\psi(\mathbf{x}, \mathbf{k}')}.$$

by $\overline{\psi(\mathbf{x}, \mathbf{k}')}$ and $\psi(\mathbf{x}, \mathbf{k})$, respectively. subtract the first from the second, and integrate over the ball of radius R . Then

$$\begin{aligned} \int_{\Omega_R} \psi(\mathbf{x}, \mathbf{k}) \psi(\mathbf{x}, \mathbf{k}') d\mathbf{x} \\ = \frac{1}{k^2 - k'^2} \int_{\Omega_R} [\psi(\mathbf{x}, \mathbf{k}) \Delta\overline{\psi(\mathbf{x}, \mathbf{k}')} - \overline{\psi(\mathbf{x}, \mathbf{k}'')} \Delta\psi(\mathbf{x}, \mathbf{k})] d\mathbf{x}. \end{aligned}$$

With the help of Green's formula, this equality can be rewritten in the form

$$\begin{aligned} & \int_{\Omega_R} \psi(\mathbf{x}, \mathbf{k}) \psi(\mathbf{x}, \mathbf{k}') d\mathbf{x} \\ &= \frac{1}{k^2 - k'^2} \\ & \times \int_S \left(\psi(\mathbf{x}, \mathbf{k}) \frac{\partial \overline{\psi(\mathbf{x}, \mathbf{k}')}}{\partial r} - \overline{\psi(\mathbf{x}, \mathbf{k}')} \frac{\partial \psi(\mathbf{x}, \mathbf{k})}{\partial r} \right) d\Sigma, \end{aligned}$$

where S_R is the sphere of radius R . The formula (13) is obtained from the last relation by passing to the limit as $R \rightarrow \infty$, and the integral on the right-hand side is computed with the help of the asymptotic expression for $\psi(\mathbf{x}, \mathbf{k})$. We omit the corresponding computations, since they repeat literally those which led us to the formulas (36.9) and (36.10).

§ 40. Motion of wave packets in a central force field

With the help of the function $\psi(\mathbf{x}, \mathbf{k})$, we construct for the nonstationary Schrödinger equation

$$i \frac{\partial \psi(\mathbf{x}, t)}{\partial t} = H \psi(\mathbf{x}, t)$$

a solution of the form

$$(1) \quad \psi(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} C(\mathbf{k}) \psi(\mathbf{x}, \mathbf{k}) e^{-ik^2 t} d\mathbf{k}.$$

If the function $C(\mathbf{k})$ satisfies the condition

$$(2) \quad \int_{\mathbf{R}^3} |C(\mathbf{k})|^2 d\mathbf{k} = 1,$$

then $\psi(\mathbf{x}, t)$ has the correct normalization

$$\int_{\mathbf{R}^3} |\psi(\mathbf{x}, t)|^2 d\mathbf{x} = 1$$

in view of (39.13).

As in the one-dimensional case, it is reasonable to consider a solution with the function $C(\mathbf{k})$ concentrated in a small neighborhood of the point $\mathbf{k}_0 = k_0 \omega_0$. By the Riemann Lebesgue theorem as $|t| \rightarrow \infty$, we have $\psi(\mathbf{x}, t) \rightarrow 0$, and $\int_{\Omega} |\psi(\mathbf{x}, t)|^2 d\mathbf{x} \rightarrow 0$ for any finite region Ω . Therefore, the function $\psi(\mathbf{x}, t)$ describes an infinite motion of a particle. We are interested in the behavior of this solution as $|t| \rightarrow \infty$ and $r \rightarrow \infty$, and we can replace $\psi(\mathbf{x}, \mathbf{k})$ in the formula (1) by its asymptotic expression

$$\psi(\mathbf{x}, \mathbf{k}) = \frac{2\pi i}{k} \left(\frac{e^{-ikr}}{r} \delta(\mathbf{n} + \boldsymbol{\omega}) - \frac{e^{ikr}}{r} S(k, \mathbf{n}, \boldsymbol{\omega}) \right) + o\left(\frac{1}{r}\right).$$

As $r \rightarrow \infty$, we have

$$\begin{aligned} \psi(\mathbf{x}, t) &\cong \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_0^\infty k^2 dk \int_{S_2} d\boldsymbol{\omega} C(k, \boldsymbol{\omega}) \frac{2\pi i}{k} \\ &\quad \times \left(\frac{e^{-ikr}}{r} \delta(\mathbf{n} + \boldsymbol{\omega}) - \frac{e^{ikr}}{r} S(k, \mathbf{n}, \boldsymbol{\omega}) \right) e^{-ik^2 t} \\ &= \psi_1(\mathbf{x}, t) + \psi_2(\mathbf{x}, t), \end{aligned}$$

where $\psi_1(\mathbf{x}, t)$ and $\psi_2(\mathbf{x}, t)$ correspond to the converging wave e^{-ikr}/r and the diverging wave e^{ikr}/r in the asymptotics.

The function $\psi_1(\mathbf{x}, t)$ can be rewritten in the form

$$(3) \quad \psi_1(\mathbf{x}, t) = \frac{i}{r\sqrt{2\pi}} \int_{-\infty}^{\infty} k C(k, -\mathbf{n}) e^{-ikr} e^{-ik^2 t} dk$$

(we consider integration over the whole real axis, setting $C(k, \mathbf{n}) = 0$ for $k < 0$). Computing the integral (3) by the stationary phase method, we get that

$$(4) \quad \psi_1(\mathbf{x}, t) = \left(\frac{1}{|2t|} \right)^{\frac{3}{2}} C\left(-\frac{r}{2t}, -\mathbf{n}\right) e^{i\chi} + O\left(\frac{1}{|t|^2}\right);$$

here χ , as always, denotes a real function that is not of interest to us.

To compute $\psi_2(\mathbf{x}, t)$ we use the formula (39.6) for the function S .

$$\begin{aligned}
 \psi_2(\mathbf{x}, t) &= -\frac{i}{r\sqrt{2\pi}} \int_{-\infty}^{\infty} k dk \int_{S_2} d\omega C(k, \omega) \\
 &\quad \times \left[\delta(\mathbf{n} - \omega) + \frac{ik}{2\pi} f(k, \mathbf{n}, \omega) \right] e^{ikr} e^{-ik^2 t} \\
 (5) \quad &= -\frac{i}{r\sqrt{2\pi}} \int_{-\infty}^{\infty} k dk \\
 &\quad \times \left[C(k, \mathbf{n}) + \frac{ik}{2\pi} C_1(k) f(k, \mathbf{n}, \omega_0) \right] e^{i(kr - k^2 t)}.
 \end{aligned}$$

Here we have introduced the notation $C_1(k) = \int_{S_2} C(k, \omega) d\omega$, and, taking into account that $C(k, \omega)$ is δ -shaped, we have replaced the function $f(k, \mathbf{n}, \omega)$ by its value at the point $\omega = \omega_0$. The integral in (5) can be computed by the stationary phase method:

$$\begin{aligned}
 \psi_2(\mathbf{x}, t) &= \left(\frac{1}{|2t|} \right)^{\frac{3}{2}} \\
 &\quad \times \left[C\left(\frac{r}{2t}, \mathbf{n}\right) + \frac{ir}{4\pi t} C_1\left(\frac{r}{2t}\right) f\left(\frac{r}{2t}, \mathbf{n}, \omega_0\right) \right] e^{ix_1} + O\left(\frac{1}{|t|^2}\right).
 \end{aligned}$$

Finally, since $C_1(k)$ is δ -shaped (it is concentrated in a neighborhood of k_0), we get that

$$\begin{aligned}
 (6) \quad \psi_2(\mathbf{x}, t) &= \left(\frac{1}{|2t|} \right)^{\frac{3}{2}} \\
 &\quad \times \left[C\left(\frac{r}{2t}, \mathbf{n}\right) + \frac{ik_0}{4\pi} C_1\left(\frac{r}{2t}\right) f(k_0, \mathbf{n}, \omega_0) \right] e^{ix_1} + O\left(\frac{1}{|t|^2}\right).
 \end{aligned}$$

It is clear from (4) and (6) that ψ_1 contributes to $\psi(\mathbf{x}, t)$ only as $t \rightarrow -\infty$, and ψ_2 only as $t \rightarrow +\infty$. For the density $|\psi(\mathbf{x}, t)|^2$ of the coordinate distribution function, we have

$$(7) \quad |\psi(\mathbf{x}, t)|^2 \cong \frac{1}{8|t|^3} \left| C\left(-\frac{r}{2t}, -\mathbf{n}\right) \right|^2, \quad t \rightarrow -\infty,$$

$$(8) \quad |\psi(\mathbf{x}, t)|^2 \cong \frac{1}{8|t|^3} \left[\left| C\left(\frac{r}{2t}, \mathbf{n}\right) \right|^2 - \frac{k_0}{\pi} \operatorname{Im} C_1\left(\frac{r}{2t}\right) \overline{C\left(\frac{r}{2t}, \mathbf{n}\right)} f(k_0, \mathbf{n}, \omega_0) + \frac{k_0^2}{4\pi^2} \left| C_1\left(\frac{r}{2t}\right) \right|^2 |f(k_0, \mathbf{n}, \omega_0)|^2 \right], \quad t \rightarrow +\infty.$$

It follows from (7) and (8) that the asymptotic expressions obtained for $\psi(\mathbf{x}, t)$ have the correct normalization as $t \rightarrow \pm\infty$. (This follows trivially for the case $t \rightarrow -\infty$ and is a consequence of (39.12) for the case $t \rightarrow +\infty$.)

Recalling that $C(k, \omega)$ is nonzero only in a small neighborhood of the point $k_0\omega_0$ and $C_1(k)$ is nonzero only in a small neighborhood of k_0 , we see that as $t \rightarrow -\infty$ the density $|\psi(\mathbf{x}, t)|^2$ of the coordinate distribution function is nonzero in a neighborhood of the point $r = -2k_0t$, $\mathbf{n} = -\omega_0$. As $t \rightarrow +\infty$ the density $|\psi(\mathbf{x}, t)|^2$ is nonzero interior to a thin spherical shell of radius $r = 2k_0t$. The angular probability distribution can be obtained⁴³ by integrating (8) with respect to the variable r with the weight r^2 . It is clear that the first two terms in (8) contribute to this distribution only in directions close to ω_0 . The angular distribution with respect to all the remaining directions is proportional to $|f(k_0, \mathbf{n}, \omega_0)|^2$.

We can now easily see how the motion takes place for a particle in the state described by the function $\psi(\mathbf{x}, t)$. Long before the scattering ($t \rightarrow -\infty$) the particle approaches the scattering center with velocity $2k_0$, moving in the direction ω_0 . After the scattering ($t \rightarrow +\infty$) it moves away from the scattering center with the same velocity, and it can be observed at any point of the spherical shell of radius $r = 2k_0t$ with angular probability distribution depending on $C(\mathbf{k})$ and $f(k_0, \mathbf{n}, \omega_0)$. In Figure 14 we have shaded the regions in which the probability of observing the particle is large as $t \rightarrow \pm\infty$. The regions

⁴³We do not write out the exact formulas for the angular probability distribution, since it is essentially dependent on the form of the function $C(\mathbf{k})$ and is therefore not a convenient characteristic of the scattering process (the function $C(\mathbf{k})$ corresponding to a concrete scattering experiment is never known). The cross-section is a suitable characteristic. As we shall see, the cross-section turns out to be insensitive to the form of $C(\mathbf{k})$: it is only important that this function be concentrated in a small neighborhood of k_0 . Physically, this requirement means that the momentum of the impinging particle must be almost specified.

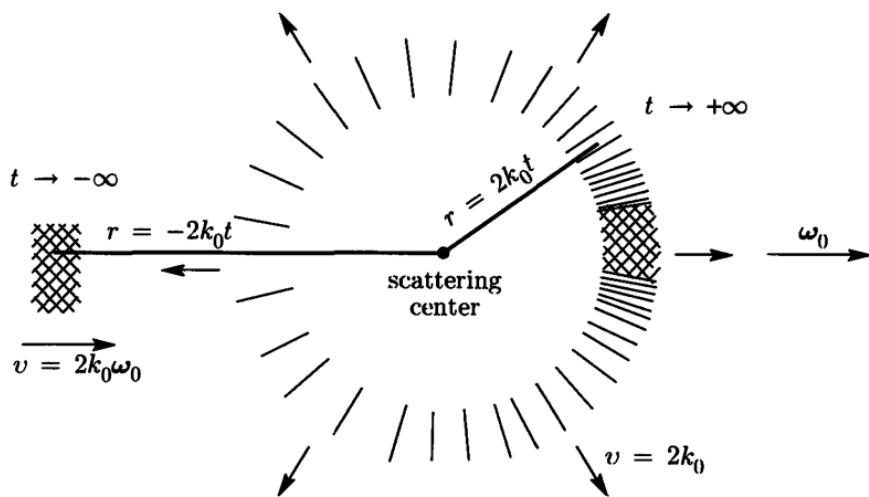


Figure 14

in which this probability is nonzero even in the absence of a scattering center is crosshatched.

The three terms in the formula (8) admit the following interpretation. The integral W_1 over the whole space of the sum of the first two terms is the probability that the particle goes past the force center without scattering. This probability is less than 1 because of the second term. The integral W_2 of the third term is the probability of scattering. We have already noted that the asymptotic expression (6) for $\psi(\mathbf{x}, t)$ has the correct normalization, and hence

$$W_1 + W_2 = 1.$$

We see that the solution $\psi(\mathbf{x}, t)$ of the Schrödinger equation constructed with the help of the function $\psi(\mathbf{x}, \mathbf{k})$ correctly describes the physical picture of the scattering. This justifies the choice of the asymptotic condition for $\psi(\mathbf{x}, \mathbf{k})$.

We mention some more features of the solution $\psi(\mathbf{x}, t)$. It is not hard to see that as $t \rightarrow -\infty$ this function has the same asymptotics as the solution

$$\varphi(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} C(\mathbf{k}) e^{i(\mathbf{kx} - k^2 t)} d\mathbf{k}$$

of the Schrödinger equation for a free particle. (Here $C(\mathbf{k})$ is the same function as in the integral (1).) Indeed, the diverging waves e^{ikr}/r do not contribute to the asymptotics as $t \rightarrow -\infty$, and the coefficients of e^{-ikr}/r in the asymptotic expression for the functions $e^{i\mathbf{k}\mathbf{x}}$ and $\psi(\mathbf{x}, \mathbf{k})$ coincide.

We show that also for $t \rightarrow +\infty$ the solution $\psi(\mathbf{x}, t)$ tends asymptotically to some solution of the Schrödinger equation for a free particle. Using the fact that for $t \rightarrow +\infty$ the converging waves e^{-ikr}/r do not contribute to the asymptotics, we get that

$$\begin{aligned} \psi(\mathbf{x}, t) &= \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} C(\mathbf{k}) \psi(\mathbf{x}, \mathbf{k}) e^{-ik^2 t} d\mathbf{k} \\ &\cong - \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_0^\infty k^2 dk \int_{S_2} d\omega C(k, \omega) \frac{2\pi i}{k} S(k, \mathbf{n}, \omega) \frac{e^{ikr}}{r} e^{-ik^2 t} \\ &= - \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_0^\infty k^2 dk \int_{S_2} d\omega \\ &\quad \times \int_{S_2} d\omega' \frac{2\pi i}{k} S(k, \omega', \omega) C(k, \omega) \delta(\omega' - \mathbf{n}) \frac{e^{ikr}}{r} e^{-ik^2 t} \\ &= - \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_0^\infty k^2 dk \int_{S_2} d\omega' \frac{2\pi i}{k} \tilde{C}(k, \omega') \delta(\omega' - \mathbf{n}) \frac{e^{ikr}}{r} e^{-ik^2 t} \\ &\cong \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} \tilde{C}(\mathbf{k}) e^{i(\mathbf{k}\mathbf{x} - k^2 t)} d\mathbf{k} = \tilde{\varphi}(\mathbf{x}, t). \end{aligned}$$

Here

$$(9) \quad \tilde{C}(\mathbf{k}) = \tilde{C}(k, \omega) = \int_{S_2} S(k, \omega, \omega') C(k, \omega') d\omega'.$$

We see that the solution $\psi(\mathbf{x}, t)$ of the Schrödinger equation tends asymptotically to the solution $\tilde{\varphi}(\mathbf{x}, t)$ for a free particle as $t \rightarrow \infty$. The function $\tilde{C}(\mathbf{k})$, which determines the final state of the free motion, is obtained from the function $C(\mathbf{k})$ giving the initial state as a result of the action of the operator S . The unitarity of S ensures the correct normalization of $\tilde{\varphi}(\mathbf{x}, t)$, since

$$\int_{\mathbf{R}^3} |\tilde{C}(\mathbf{k})|^2 d\mathbf{k} = 1$$

in view of the unitarity.

§ 41. The integral equation of scattering theory

Most approaches to the construction of the solutions $\psi(\mathbf{x}, \mathbf{k})$ and the scattering amplitude $f(k, \mathbf{n}, \omega)$ are based on the integral equation

$$(1) \quad \psi(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\mathbf{x}} - \frac{1}{4\pi} \int_{\mathbf{R}^3} \frac{e^{i\mathbf{k}|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} V(\mathbf{y}) \psi(\mathbf{y}, \mathbf{k}) d\mathbf{y},$$

which is often called the Lippmann–Schwinger equation.

We verify that a solution of this equation satisfies the equation (39.2) and the asymptotic condition (39.7).⁴⁴ Indeed, using the formulas

$$(\Delta + k^2) e^{i\mathbf{k}\mathbf{x}} = 0, \quad (\Delta + k^2) e^{i\mathbf{k}\mathbf{r}} / r = -4\pi\delta(\mathbf{x}).$$

we get that

$$(\Delta + k^2) \psi(\mathbf{x}, \mathbf{k}) = 0 + \int_{\mathbf{R}^3} \delta(\mathbf{x} - \mathbf{y}) V(\mathbf{y}) \psi(\mathbf{y}, \mathbf{k}) d\mathbf{y} = V(\mathbf{x}) \psi(\mathbf{x}, \mathbf{k}).$$

We verify the asymptotic condition for the case of a compactly supported potential ($V(\mathbf{x}) = 0$ for $r > a$). We have

$$\begin{aligned} \frac{1}{|\mathbf{x} - \mathbf{y}|} &= \frac{1}{r} + O\left(\frac{1}{r^2}\right), \quad r = |\mathbf{x}|, |\mathbf{y}| < a, \\ |\mathbf{x} - \mathbf{y}| &= r \sqrt{1 - \frac{2\mathbf{x}\mathbf{y}}{r^2} + \frac{\mathbf{y}^2}{r^2}} = r - \mathbf{n}\mathbf{y} + O\left(\frac{1}{r}\right), \\ e^{ik|\mathbf{x}-\mathbf{y}|} &= e^{i(kr - k\mathbf{n}\mathbf{y})} + O\left(\frac{1}{r}\right). \end{aligned}$$

Therefore, for the function $\psi(\mathbf{x}, \mathbf{k})$ we get that

$$\psi(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\mathbf{x}} - \frac{1}{4\pi} \frac{e^{ikr}}{r} \int_{\mathbf{R}^3} e^{-ik\mathbf{n}\mathbf{y}} V(\mathbf{y}) \psi(\mathbf{y}, \mathbf{k}) d\mathbf{y} + O\left(\frac{1}{r^2}\right).$$

Comparing the last formula with (39.7), we see that the solution of the integral equation has the correct asymptotics, and moreover, we get that

$$(2) \quad f(k, \mathbf{n}, \omega) = -\frac{1}{4\pi} \int_{\mathbf{R}^3} e^{-ik\mathbf{n}\mathbf{x}} V(\mathbf{x}) \psi(\mathbf{x}, k\omega) d\mathbf{x}.$$

⁴⁴Of course, it is also possible to verify the converse assertion

The formula (2), in which the scattering amplitude is expressed in terms of the solution of (1), often turns out to be useful for the approximate determination of $f(k, \mathbf{n}, \omega)$.

One of the approximation methods of scattering theory is based on the use of a series of iterations of the equation (42.1).

$$(3) \quad \psi(\mathbf{x}, \mathbf{x}) = \sum_{n=0}^{\infty} \psi^{(n)}(\mathbf{x}, \mathbf{k}),$$

where

$$\begin{aligned} \psi^{(0)}(\mathbf{x}, \mathbf{k}) &= e^{i\mathbf{k}\mathbf{x}}, \\ \psi^{(n+1)}(\mathbf{x}, \mathbf{k}) &= -\frac{1}{4\pi} \int_{\mathbf{R}^3} \frac{e^{i\mathbf{k}|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} V(\mathbf{y}) \psi^{(n)}(\mathbf{y}, \mathbf{k}) d\mathbf{y}. \end{aligned}$$

The series (3) is called the Born series, and substitution of (3) in (2) gives the Born series for the scattering amplitude. The Born series for the problem of scattering by a potential center has been thoroughly studied. It is known, for example, that it converges under the condition that

$$\max_{\mathbf{x}} \int_{\mathbf{R}^3} |V(\mathbf{y})| \frac{1}{|\mathbf{x}-\mathbf{y}|} d\mathbf{y} < 4\pi.$$

It is also known that under this condition on the potential the operator H does not have a discrete spectrum. If a discrete spectrum is present, then there are values of k for which the series (3) does not converge. At the same time, for sufficiently large k it does converge for a very broad class of potentials.

The simplest approximation for the scattering amplitude is obtained if in place of $\psi(\mathbf{x}, \mathbf{k})$ we substitute $\psi^{(0)}(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\mathbf{x}}$:

$$f_B(k, \mathbf{n}, \omega) = -\frac{1}{4\pi} \int_{\mathbf{R}^3} e^{-ik\mathbf{n}\mathbf{x}} V(\mathbf{x}) e^{ik\omega\mathbf{x}} d\mathbf{x}.$$

This approximation is called the Born approximation. The precise assertion is that for large k

$$f(k, \mathbf{n}, \omega) - f_B(k, \mathbf{n}, \omega) = o(1)$$

uniformly with respect to \mathbf{n} and ω .

The integral equation is used for mathematical investigation of the scattering problem. For a suitable choice of function space, this equation can be reduced to an equation of the second kind

$$\psi = \varphi + A\psi$$

with a compact operator. Therefore, the Fredholm theorems hold for the equation (1). For a broad class of potentials it has been shown that the corresponding homogeneous equation can have solutions only for imaginary values of the parameter k ($k_n = i\kappa_n$, $\kappa_n > 0$). By repeating the calculations carried out in the beginning of this section, it is easy to see that the solution $\psi_n(\mathbf{x})$ of the homogeneous equation satisfies the Schrödinger equation $[-\Delta + V(\mathbf{x})]\psi_n(\mathbf{x}) = -\kappa_n^2\psi_n(\mathbf{x})$ and has the asymptotic expression $\psi_n(\mathbf{x}) \cong f(\mathbf{n})e^{-\kappa_n r}/r$, where $f(\mathbf{n})$ is a function defined on the unit sphere. This means that the solutions of the homogeneous equation are eigenfunctions of the discrete spectrum of H . The completeness of the collection $\{\psi_n(\mathbf{x}), \psi(\mathbf{x}, \mathbf{k})\}$ of eigenfunctions of H was proved with the help of the integral equation (1).

§ 42. Derivation of a formula for the cross-section

The main characteristic of the process of scattering of a particle by a potential center is the differential cross-section. In accordance with the general definition of the cross-section, the differential cross-section is defined by the formula

$$(1) \quad d\sigma = \frac{dN}{I},$$

where dN is the probability of observing the particle after the scattering ($t \rightarrow \infty$) in the solid angle element $d\mathbf{n}$ about some direction \mathbf{n} , and I is the probability of a free particle crossing a unit area element oriented perpendicular to the motion of the particle. This probability is defined at the point where the scattering center is located. We stress that dN is a characteristic of a particle in the field of the scattering center, while I is a characteristic of a free particle.

Little is known about the state ω of a free particle in the beam. Indeed, it is known that the particle has momentum approximately

equal to $\mathbf{k}_0 = k_0 \boldsymbol{\omega}_0$, and the variance $(\Delta k)_{\text{avg}}^2 = (\Delta k_1)_{\text{avg}}^2 + (\Delta k_2)_{\text{avg}}^2 + (\Delta k_3)_{\text{avg}}^2$ of the momentum is known; moreover, for the scattering experiments to admit a simple interpretation, one tries to use beams of particles with a small variance of the momentum. Very little is usually known about the coordinates of a particle in the beam. But nevertheless it is known that at some moment of time (which can be taken to be $t = 0$), the particle is in a macroscopic region between the accelerator and the target. The transverse size of this region can be identified with the diameter of the beam. We remark that according to the Heisenberg uncertainty relation, the existence of such a region imposes lower bounds on the variance of the momentum.

In deriving the formula for the differential cross-section, we make two assumptions about the smallness of the square root Δk of the variance of the momentum:

- 1) $\Delta k \ll k_0$,
- 2) $|f(k, \mathbf{n}, \boldsymbol{\omega}) - f(k', \mathbf{n}, \boldsymbol{\omega}')| \ll |f(k, \mathbf{n}, \boldsymbol{\omega})|$ if $|k\boldsymbol{\omega} - k'\boldsymbol{\omega}'| < \Delta k$.

In principle these requirements can always be satisfied if $f(k, \mathbf{n}, \boldsymbol{\omega})$ is a continuous function of $\mathbf{k} = k\boldsymbol{\omega}$ (the continuity of f can be proved for a broad class of potentials). These conditions usually hold in practice. However, there are cases when a small change in the variable \mathbf{k} leads to a sharp change in the function $f(k, \mathbf{n}, \boldsymbol{\omega})$ (the case of narrow resonance). In the final analysis the quantity Δk is determined by the construction of the accelerator and in practice cannot be made arbitrarily small. Therefore, in some experiments the condition 2) may not hold. In this case one cannot use the formula obtained below for the cross-section.

We show that under the conditions 1) and 2) the differential cross-section depends on the state of the impinging particle only in terms of k_0 , and that

$$d\sigma = |f(k_0, \mathbf{n}, \boldsymbol{\omega}_0)|^2 dn.$$

The most general state for a free particle is a mixed state, given by a density matrix $M_0(t)$ which we represent as a convex combination of pure states:

$$M_0(t) = \sum_s \alpha_s P_{\varphi_s(t)}, \quad \sum_s \alpha_s = 1.$$

The pure state $P_{\varphi_s(t)}$ is determined by a wave function $\varphi_s(\mathbf{x}, t)$ with the form

$$\varphi_s(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} C_s(\mathbf{k}) e^{i(\mathbf{kx} - k^2 t)} d\mathbf{k}.$$

It can be asserted that an appreciable contribution is given only by the pure states in which the momentum variance does not exceed the momentum variance in the state $M_0(t)$. Therefore, we assume that all the functions $C_s(\mathbf{k})$ are nonzero only in a small neighborhood of the point $k_0 = k_0 \omega_0$, and that the diameter of this neighborhood does not exceed Δk . We do not have any information about the concrete form of the function⁴⁵ $C_s(\mathbf{k})$ nor about the weights α_s , and we do not make any assumptions about them.

We first compute the probability I . Let us choose a coordinate system with origin at the force center and with x_3 -axis directed along the vector ω_0 . For a pure state $\varphi_s(t)$ it is easiest to compute the probability $I^{(s)}$ by using the density vector

$$\mathbf{j}^{(s)}(\mathbf{x}, t) = i(\varphi_s \nabla \bar{\varphi}_s - \bar{\varphi}_s \nabla \varphi_s)$$

of the probability flow, and then

$$I^{(s)} = \int_{-\infty}^{\infty} j_3^{(s)}(\mathbf{x}, t) \Big|_{\mathbf{x}=0} dt.$$

Computing

$$\begin{aligned} j_3^{(s)}(\mathbf{x}, t) \Big|_{\mathbf{x}=0} &= i \left(\varphi_s \frac{\partial \bar{\varphi}_s}{\partial x_3} - \bar{\varphi}_s \frac{\partial \varphi_s}{\partial x_3} \right) \Big|_{\mathbf{x}=0} \\ &= \frac{1}{(2\pi)^3} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' (k_3 + k'_3) C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} e^{i(k'^2 - k^2)t}, \end{aligned}$$

⁴⁵ Corresponding to different functions $C_s(\mathbf{k})$ are both different forms and different arrangements of wave packets in the configuration space at some moment of time. For example, suppose that some function $C_s(\mathbf{k})$ gives a wave packet whose center of gravity $(\mathbf{x}(t))_{avg}$ moves along the x_3 -axis directed along the vector ω_0 and passes through the scattering center. Then corresponding to the function $C_l(\mathbf{k}) = e^{i\mathbf{k}\mathbf{u}_l} C_s(\mathbf{k})$ is the packet shifted by the vector \mathbf{u}_l in the configuration space ($e^{i\mathbf{k}\mathbf{u}}$ is the shift operator, written in the momentum representation). As a classical analogue (though not completely accurate) of the state $\varphi_s(\mathbf{x}, t)$ we can take the state of a particle moving on a rectilinear trajectory that passes through the scattering center. Then corresponding to the state $\varphi_l(\mathbf{x}, t)$ is a trajectory passing by the scattering center at a distance ρ equal to the projection of the vector \mathbf{u}_l on the plane of the cross-section of the beam. In classical scattering theory the distance ρ by which the particle would go by the center if there were no interaction is called the aiming parameter.

we get that

$$\begin{aligned}
 I^{(s)} &= \frac{1}{(2\pi)^3} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' \int_{-\infty}^{\infty} dt (k_3 + k'_3) C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} e^{i(k'^2 - k^2)t} \\
 &= \frac{1}{(2\pi)^2} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' (k_3 + k'_3) C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} \delta(k'^2 - k^2) \\
 &= \frac{1}{(2\pi)^2} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' \frac{k_3 + k'_3}{2k} C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} \delta(k' - k) \\
 &= \frac{1}{(2\pi)^2} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' C_s(\mathbf{k}) C_s(\mathbf{k}') d(k' - k).
 \end{aligned}$$

The computations use the equality

$$(2) \quad \delta(k^2 - k'^2) = \frac{1}{2k} (\delta(k - k') + \delta(k + k'))$$

and the condition $\Delta k \ll k_0$, which enables us to replace $(k_3 + k'_3)/2k$ by 1. The second term in (2) does not contribute to the integral, because $k > 0$ and $k' > 0$ in the region of integration.

It is obvious that for a mixed state $M_0(t)$

$$I = \sum_s \alpha_s \frac{1}{(2\pi)^2} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' C_s(\mathbf{k}) C_s(\mathbf{k}') \delta(k - k').$$

Let us now compute dN . In the computation we must use the density matrix $M(t)$ describing a state which long before the scattering ($t \rightarrow -\infty$) tends asymptotically to $M_0(t)$. The operator $M(t)$ can be written in the form

$$M(t) = \sum_s \alpha_s P_{\psi_s(t)},$$

where the pure states $P_{\psi_s(t)}$ are determined by the wave functions

$$(3) \quad \psi_s(t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} d\mathbf{k} C_s(\mathbf{k}) \psi(\mathbf{x}, \mathbf{k}) e^{-ik^2 t}.$$

The density of the coordinate distribution function in a pure state $P_{\psi_s(t)}$ is $|\psi_s(\mathbf{x}, t)|^2$, and for the probability $dN^{(s)}$ of observing the particle as $t \rightarrow \infty$ in a solid angle element $d\mathbf{n}$ in the state $P_{\psi_s(t)}$ we get

$$(4) \quad dN^{(s)} = \lim_{t \rightarrow +\infty} d\mathbf{n} \int_0^\infty r^2 dr |\psi_s(\mathbf{x}, t)|^2.$$

We know that as $t \rightarrow \infty$ the particle goes to infinity, and therefore when we substitute (3) into (4) we can replace $\psi(\mathbf{x}, \mathbf{k})$ by its asymptotic expression, obtaining

$$dN^{(s)} = \lim_{t \rightarrow +\infty} d\mathbf{n} \int_0^\infty r^2 dr \left| \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} d\mathbf{k} C_s(\mathbf{k}) \times \left(e^{i\mathbf{k}\mathbf{x}} + f(k, \mathbf{n}, \omega) \frac{e^{ikr}}{r} \right) e^{-ik^2 t} \right|^2.$$

The function $(2\pi)^{-3/2} \int_{\mathbf{R}^3} d\mathbf{k} C_s(\mathbf{k}) e^{i\mathbf{k}\mathbf{x} - ik^2 t}$ is a wave packet for a free particle and is nonzero only in that part of the configuration space where there is a finite probability of observing an unscattered particle of the beam, that is, inside a narrow cone constructed about the direction ω_0 . For all the remaining directions, only the diverging wave gives a contribution, and for such directions

(5)

$$dN^{(s)} = \lim_{t \rightarrow +\infty} \frac{d\mathbf{n}}{(2\pi)^3} \int_0^\infty dr \left| \int_{\mathbf{R}^3} C_s(\mathbf{k}) f(k, \mathbf{n}, \omega) e^{ikr - ik^2 t} d\mathbf{k} \right|^2.$$

Using the condition 2), we can replace the function $f(k, \mathbf{n}, \omega)$ under the integral sign by its value at the point $k = k_0$, $\omega = \omega_0$ and take it out from under the integral sign. Further, the integral $\int_{\mathbf{R}^3} C_s(\mathbf{k}) e^{ikr - ik^2 t} d\mathbf{k}$ becomes a one-dimensional wave packet after integration with respect to the angle variable of the vector \mathbf{k} , and as a function of r it is nonzero only for large positive r as $t \rightarrow \infty$. Therefore, the integration with respect to r in (5) can be extended over the whole real axis. Then

$$\begin{aligned} dN^{(s)} &= \lim_{t \rightarrow +\infty} \frac{d\mathbf{n}}{(2\pi)^3} |f(k_0, \mathbf{n}, \omega_0)|^2 \\ &\times \int_{-\infty}^\infty dr \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} e^{i(k-k')r} e^{-i(k^2 - k'^2)t} \\ &= d\mathbf{n} |f(k_0, \mathbf{n}, \omega_0)|^2 \frac{1}{(2\pi)^2} \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} \delta(k - k'). \end{aligned}$$

Finally, for the mixed state $M(t)$

$$dN = d\mathbf{n} |f(k_0, \mathbf{n}, \omega_0)|^2 \sum_s \alpha_s \frac{1}{(2\pi)^2} \times \int_{\mathbf{R}^3} d\mathbf{k} \int_{\mathbf{R}^3} d\mathbf{k}' C_s(\mathbf{k}) \overline{C_s(\mathbf{k}')} \delta(k - k'),$$

or

$$dN = d\mathbf{n} |f(k_0, \mathbf{n}, \omega_0)|^2 I,$$

from which we get at once that

$$(6) \quad d\sigma = |f(k_0, \mathbf{n}, \omega_0)|^2 d\mathbf{n}.$$

In conclusion we note that we have obtained (6) for all directions \mathbf{n} except for $\mathbf{n} = \omega_0$. The forward scattering cross-section cannot be defined by the formula (1), because the probability of observing the particle as $t \rightarrow \infty$ in a solid angle $d\mathbf{n}$ constructed about the direction $\mathbf{n} = \omega_0$ is not proportional to I (moreover, we cannot in general distinguish a particle that is scattered forward from an unscattered particle). When people speak of the forward scattering cross-section, they always mean the extrapolation to the zero angle of the cross-sections of scattering through small angles. With this understanding, the formula (6) is correct for the forward scattering cross-section.

For many problems a convenient characteristic of scattering turns out to be the total cross-section, defined by the formula⁴⁶

$$\sigma = \int_{S_2} |f(k_0, \mathbf{n}, \omega_0)|^2 d\mathbf{n}.$$

§ 43. Abstract scattering theory

In this section we shall use the following notation. The Schrödinger operator for a free particle is denoted by H_0 , and the Schrödinger operator for a particle in a field is denoted by H ,

$$H = H_0 + V,$$

where V is the potential energy operator. Any solution of the nonstationary Schrödinger equation for the particle in a field is denoted by

⁴⁶In classical mechanics the total cross-section σ is $= \infty$ if the potential is not compactly supported. A peculiarity of quantum mechanics is the finiteness of the cross-section σ for sufficiently rapidly decaying potentials.

$\psi(t)$. This vector is uniquely determined by its value at $t = 0$, which we denote by ψ : that is,

$$\psi(t) = e^{-iHt}\psi$$

(we recall that e^{-iHt} is the evolution operator).

Similarly, any solution of the Schrödinger equation for a free particle is denoted by $\varphi(t)$, and its value at $t = 0$ is denoted by φ :

$$\varphi(t) = e^{-iH_0 t}\varphi.$$

Different solutions $\varphi(t)$ can be supplied with additional indices. The wave functions corresponding to the vectors $\psi(t)$, ψ , $\varphi(t)$, φ are written as $\psi(\mathbf{x}, t)$, $\psi(\mathbf{x})$, $\varphi(\mathbf{x}, t)$, $\varphi(\mathbf{x})$ in the coordinate representation, and as $\psi(\mathbf{p}, t)$, $\psi(\mathbf{p})$, $\varphi(\mathbf{p}, t)$, $\varphi(\mathbf{p})$ in the momentum representation. Finally, the eigenvectors of the operator H (if there are any) will be denoted by χ_n : $H\chi_n = E_n\chi_n$.

For the Schrödinger equation we constructed a solution $\psi(\mathbf{x}, t) = e^{-iHt}\psi(\mathbf{x})$ in § 40 tending asymptotically to some solutions $\varphi_{\mp}(\mathbf{x}, t) = e^{-iH_0 t}\varphi_{\mp}(\mathbf{x})$ of the Schrödinger equation for a free particle as $t \rightarrow \mp\infty$, and therefore we can expect that

$$(1) \quad \lim_{t \rightarrow \mp\infty} \|e^{-iHt}\varphi - e^{-iH_0 t}\varphi_{\mp}\| = 0$$

for such a solution $\psi(t)$.

The physical picture of scattering can be represented as follows. Long before the scattering, the particle moves freely far from the scattering center, then it falls into the zone of action of the potential (scattering takes place), and finally the motion of the particle again becomes free over a sufficiently long period of time. Therefore, the following formulation of the nonstationary scattering problem seems natural.

1. For an arbitrary vector φ_- in the state space \mathcal{H} , construct a vector ψ such that (1) is valid as $t \rightarrow -\infty$.

2. For the vector ψ constructed, find a vector $\varphi_+ \in \mathcal{H}$ such that (1) is valid as $t \rightarrow +\infty$.

The vector $\psi(t) = e^{-iHt}\psi$ describes a state of the particle that coincides with $\varphi_-(t) = e^{-iH_0 t}\varphi_-$ in the distant past and becomes $\varphi_+(t) = e^{-iH_0 t}\varphi_+$ as $t \rightarrow +\infty$. Physics is interested in the connection

between the vectors φ_- and φ_+ . Therefore, the following can be added to the items 1 and 2 of the formulation.

3. Show that there exists a unitary operator S such that

$$\varphi_+ = S\varphi_-.$$

Let us begin with item 1. We pose the problem in a somewhat broader form and see whether it is possible, for arbitrary vectors $\varphi_-, \varphi_+ \in \mathcal{H}$, to construct vectors ψ such that (1) is valid as $t \rightarrow -\infty$ and $t \rightarrow +\infty$, respectively (of course, the vectors ψ constructed from φ_- and φ_+ do not have to coincide).

Rewriting (1) using the unitarity of the operator e^{-iHt} in the form

$$\lim_{t \rightarrow \pm\infty} \|\psi - e^{iHt} e^{-iH_0 t} \varphi_{\mp}\| = 0.$$

we see that the problem posed reduces to the existence of the strong limits

$$(2) \quad \lim_{t \rightarrow \pm\infty} e^{iHt} e^{-iH_0 t} = U_{\pm}.$$

If they exist, the operators U_{\pm} are called the wave operators. If U_- has been constructed, then the vector $\psi = U_- \varphi_-$ satisfies item 1 in the formulation of the problem.

We shall find a simple sufficient condition for the existence of the wave operators. Let us consider the operator

$$U(t) = e^{iHt} e^{-iH_0 t}$$

and compute its derivative⁴⁷

$$\frac{dU(t)}{dt} = ie^{iHt}(H - H_0)e^{-iH_0 t} = ie^{iHt} V e^{-iH_0 t}.$$

Obviously, $U(0) = I$, and hence

$$(3) \quad \begin{aligned} U(t) &= I + i \int_0^t e^{iHs} V e^{-iH_0 s} ds, \\ U_{\pm} &= I + i \int_0^{\pm\infty} e^{iHs} V e^{-iH_0 s} ds. \end{aligned}$$

The question of the existence of the operators U_{\pm} has been reduced to the question of the convergence of the integrals (3) at the upper

⁴⁷Note that the expression for the derivative takes into account that H and H_0 do not commute

limit A sufficient condition for the convergence of (3) is the existence of the integrals

$$(4) \quad \int_0^{\pm\infty} \|Ve^{-iH_0t}\varphi\| dt$$

for any $\varphi \in \mathcal{H}$ (we have taken into account that e^{iHt} is a unitary operator). Finally, the integrals (4) converge at the upper limits if for any $\varphi \in \mathcal{H}$

$$(5) \quad \|Ve^{-iH_0t}\varphi\| = o\left(\frac{1}{|t|^{1+\varepsilon}}\right), \quad \varepsilon > 0, \quad |t| \rightarrow \infty.$$

Let us see which potentials satisfy (5). In the coordinate representation we have an estimate⁴⁸

$$|\varphi(\mathbf{x}, t)| < \frac{C}{|t|^{\frac{3}{2}}},$$

uniform with respect to \mathbf{x} , for the wave function $\varphi(\mathbf{x}, t) = e^{-iH_0t}\varphi(\mathbf{x})$. Then

$$\|Ve^{-iH_0t}\varphi\|^2 = \int_{\mathbf{R}^3} |V(\mathbf{x}) \varphi(\mathbf{x}, t)|^2 d\mathbf{x} < \frac{C}{|t|^3} \int_{\mathbf{R}^3} |V(\mathbf{x})|^2 d\mathbf{x}.$$

We see that the condition (5) holds if the potential is square integrable. Of course, this condition is not necessary. The class of potentials for which U_{\pm} exist is broader, but there are potentials for which the wave operators cannot be constructed. An important example of such a potential is the Coulomb potential $V(r) = \alpha/r$. The reason for the absence of wave operators for this potential is that it decays too slowly at infinity. The solutions of the Schrödinger equation for the Coulomb potential do not tend to solutions for a free particle as $t \rightarrow \pm\infty$ (the particle “feels” the potential even at infinity). In connection with this, both the nonstationary and the stationary formulations of the scattering problem in a Coulomb field require serious modifications. The asymptotic form of the Coulomb functions $\psi(\mathbf{x}, \mathbf{k})$ is different from (39.7).

⁴⁸This estimate is obtained by the stationary phase method for the asymptotic computation of the integral $\varphi(\mathbf{x}, t) = (2\pi)^{-3/2} \int_{\mathbf{R}^3} \varphi(\mathbf{k}) e^{i(\mathbf{kx} - \mathbf{k}^2 t)} dk$. The stationary phase method can be used if $\varphi(\mathbf{k})$ satisfies certain smoothness conditions. But the set \mathcal{D} of smooth functions is dense in \mathcal{H} and the operators $U(t)$ and U_{\pm} are bounded, so it suffices to prove the existence of the strong limits of $U(t)$ on the set \mathcal{D} .

Let us return to our study of potentials for which the wave operators U_{\pm} exist, and discuss item 2. We can ask whether for any vector $\psi \in \mathcal{H}$ there are vectors φ_+ and φ_- such that (1) is valid as $t \rightarrow \pm\infty$.

It turns out that if H has eigenvectors χ_n , $\|\chi_n\| = 1$, then for them the equality (1) is false for all φ_+ and φ_- . Indeed, in this case one can easily compute that for any $\varphi \in \mathcal{H}$ with $\|\varphi\| = 1$

$$\begin{aligned} \lim_{t \rightarrow \pm\infty} \|e^{-iHt}\chi_n - e^{-iH_0 t}\varphi\| &= \lim_{t \rightarrow \pm\infty} \|e^{-iE_n t}\chi_n - e^{-iH_0 t}\varphi\| \\ &= \lim_{t \rightarrow \pm\infty} \sqrt{\|\chi_n\|^2 + \|\varphi\|^2 - 2 \operatorname{Re} e^{-iE_n t}(\chi_n, e^{-iH_0 t}\varphi)} = \sqrt{2}. \end{aligned}$$

We have taken into account that

$$(6) \quad \lim_{t \rightarrow \pm\infty} (\chi_n, e^{-iH_0 t}\varphi) = 0,$$

since the vector $e^{-iH_0 t}\varphi$ tends weakly to zero as $t \rightarrow \pm\infty$.

From physical considerations it is also easy to see why the state $e^{-iHt}\chi_n$ does not tend asymptotically to some state of free motion for $e^{-iH_0 t}\varphi$. The vector $e^{-iHt}\chi_n$ describes the state of a particle that is localized near the scattering center, while for any $\varphi \in \mathcal{H}$ the vector $e^{-iH_0 t}\varphi$ describes the state of a particle going to infinity as $t \rightarrow \pm\infty$.

It is clear that (1) fails also for any vector $\chi \in \mathcal{B}$, where \mathcal{B} is the subspace spanned by the eigenvectors of H . We shall call \mathcal{B} the subspace of bound states.

Accordingly, we see that for an arbitrary vector ψ , there does not in general exist a vector φ_+ satisfying (1) as $t \rightarrow +\infty$. To determine whether for a constructed vector $\psi = U_-\varphi_-$ it is possible to find a corresponding vector φ_+ , we need to study the properties of the wave operators.

Let \mathcal{R}_{\pm} denote the ranges of the operators U_{\pm} . We show that $\mathcal{R}_{\pm} \perp \mathcal{B}$.

It suffices to verify that the vectors $U_{\pm}\varphi$ are orthogonal to the eigenvectors χ_n for any $\varphi \in \mathcal{H}$. From (6) we get that

$$(U_{\pm}\varphi, \chi_n) = \lim_{t \rightarrow \pm\infty} (e^{iHt}e^{-iH_0 t}\varphi, \chi_n) = \lim_{t \rightarrow \pm\infty} e^{iE_n t}(e^{-iH_0 t}\varphi, \chi_n) = 0.$$

We state the following property without proof. It turns out that

$$(7) \quad \mathcal{R}_+ = \mathcal{R}_- = \mathcal{R}, \quad \mathcal{R} \ominus \mathcal{B} = \mathcal{H}.$$

The proof of this assertion in abstract scattering theory is very complex. The subspace \mathcal{R} which coincides with the ranges of the operators U_{\pm} is often called the subspace of scattering states.

We show further that the operators U_{\pm} are isometries. Indeed, from the strong convergence of the operator $U(t)$ as $t \rightarrow \pm\infty$ and from its unitarity it follows that

$$(U_{\pm}\varphi, U_{\pm}\varphi) = \lim_{t \rightarrow \pm\infty} (U(t)\varphi, U(t)\varphi) = (\varphi, \varphi);$$

that is, the operators U_{\pm} preserve the norm of a vector φ , and hence

$$(8) \quad U_{\pm}^* U_{\pm} = I.$$

The operators U_{\pm} are unitary only in the absence of eigenvectors of the discrete spectrum of H . In this case the operators U_{\pm} map \mathcal{H} bijectively onto itself, and then along with (8) we have the equality

$$(9) \quad U_{\pm} U_{\pm}^* = I.$$

If H has a discrete spectrum and $\psi \in \mathcal{R}$, then there are vectors φ_+ and φ_- such that

$$\psi = U_{\pm}\varphi_{\pm}.$$

Acting on this equality by U_{\pm}^* and using (8), we obtain

$$(10) \quad U_{\pm}^* \psi = \varphi_{\pm},$$

$$(11) \quad U_{\pm} U_{\pm}^* \psi = \psi, \quad \psi \in \mathcal{R}.$$

On the other hand, for $\chi \in \mathcal{B}$ and any $\varphi \in \mathcal{H}$,

$$(U_{\pm}^* \chi, \varphi) = (\chi, U_{\pm} \varphi) = 0,$$

and therefore

$$U_{\pm}^* \chi = 0, \quad U_{\pm} U_{\pm}^* \chi = 0, \quad \chi \in \mathcal{B}.$$

Any vector $\varphi \in \mathcal{H}$ can be represented as

$$\varphi = \chi + \psi, \quad \chi \in \mathcal{B}, \quad \psi \in \mathcal{R},$$

and

$$U_{\pm} U_{\pm}^* \varphi = \psi = (I - P) \varphi,$$

where P is the projection on the subspace \mathcal{B} of bound states. Thus, in the general case we have

$$(12) \quad U_{\pm}U_{\pm}^* = I - P$$

instead of (9), and therefore the operators U_{\pm} are not unitary in the presence of a discrete spectrum.

Finally, we show that

$$(13) \quad f(H)U_{\pm} = U_{\pm}f(H_0)$$

for any bounded function $f(s)$, $s \in \mathbf{R}$. Passing to the limit as $t \rightarrow \pm\infty$ in the equality

$$e^{iH\tau}e^{iHt}e^{-iH_0t}\varphi = e^{iH(t+\tau)}e^{-iH_0(t+\tau)}e^{iH_0\tau}\varphi, \quad \varphi \in \mathcal{H}, \quad \tau \in \mathbf{R},$$

we get that

$$e^{iH\tau}U_{\pm} = U_{\pm}e^{iH_0\tau},$$

which at once gives us (13).

Let us return to the nonstationary scattering problem formulated at the beginning of this section. If for some potential V the wave operators U_{\pm} exist and (7) holds, then the nonstationary scattering problem has a unique solution.

The vector ψ is found for any $\varphi_- \in \mathcal{H}$ from the formula

$$\psi = U_{-}\varphi_{-}.$$

This vector is in \mathcal{R} ; therefore by (10),

$$\varphi_{+} = U_{+}^*\psi$$

or

$$\varphi_{+} = U_{+}^*U_{-}\varphi_{-},$$

which can be rewritten in the form

$$(14) \quad \varphi_{+} = S\varphi_{-},$$

where

$$S = U_{+}^*U_{-}.$$

The unitarity of the scattering operator S follows from (8), (12), and the obvious equalities $PU_{\pm} = 0$. Indeed,

$$S^*S = U_{-}^*U_{+}U_{+}^*U_{-} = U_{-}^*(I - P)U_{-} = I,$$

and similarly

$$SS^* = I.$$

Further, it follows from (13) that

$$(15) \quad Sf(H_0) = f(H_0)S.$$

Let us write (15) in the momentum representation for $f(s) = s$:

$$S(\mathbf{k}, \mathbf{k}') k'^2 = k^2 S(\mathbf{k}, \mathbf{k}').$$

We see that the kernel of S can be written in the form

$$(16) \quad S(\mathbf{k}, \mathbf{k}') = \frac{2S(k, \omega, \omega')}{k} \delta(k^2 - k'^2), \quad \mathbf{k} = k\omega,$$

and the relation (14) takes the form

$$\begin{aligned} \varphi_+(k, \omega) &= \int_{S_2} d\omega' \int_0^\infty k'^2 dk' \frac{2S(k, \omega, \omega')}{k} \frac{\delta(k - k')}{2k} \varphi_-(k', \omega') \\ &= \int_{S_2} S(k, \omega, \omega') \varphi_-(k, \omega') d\omega'. \end{aligned}$$

Comparing this formula with (40.9), we get that the function $S(k, \omega, \omega')$ in (16) coincides with the function $S(k, \omega, \omega')$ in § 39. This connection between the S -operator of nonstationary scattering theory and the asymptotics of the wave functions of the stationary scattering problem can also be established in the framework of a rigorous theory, of course.

There is a simple connection between the wave operators U_\pm and the solutions $\psi(\mathbf{x}, \mathbf{k})$ introduced in § 39 of the Schrödinger equation. We recall that for the nonstationary Schrödinger equation the solution $\psi(\mathbf{x}, t)$ constructed from the function $\psi(\mathbf{x}, \mathbf{k})$ has the form

$$\psi(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} \varphi_-(\mathbf{k}) \psi(\mathbf{x}, \mathbf{k}) e^{-ik^2 t} d\mathbf{k}.$$

Here we have denoted the function $C(\mathbf{k})$ by $\varphi_-(\mathbf{k})$. We know that as $t \rightarrow -\infty$ the function $\psi(\mathbf{x}, t)$ tends asymptotically to

$$\varphi_-(\mathbf{x}, t) = \left(\frac{1}{2\pi} \right)^{\frac{3}{2}} \int_{\mathbf{R}^3} \varphi_-(\mathbf{k}) e^{i\mathbf{kx}} e^{-ik^2 t} d\mathbf{k}.$$

Setting $t = 0$ in these equalities, we get that

$$\begin{aligned}\psi(\mathbf{x}) &= \left(\frac{1}{2\pi}\right)^{\frac{3}{2}} \int_{\mathbf{R}^3} \psi(\mathbf{x}, \mathbf{k}) \varphi_-(\mathbf{k}) d\mathbf{k}, \\ \varphi_-(\mathbf{x}) &= \left(\frac{1}{2\pi}\right)^{\frac{3}{2}} \int_{\mathbf{R}^3} e^{i\mathbf{kx}} \varphi_-(\mathbf{k}) d\mathbf{k}.\end{aligned}$$

Comparing these formulas with $\varphi = U_- \varphi_-$, written in the momentum representation

$$\psi(\mathbf{k}) = \int_{\mathbf{R}^3} U_-(\mathbf{k}, \mathbf{k}') \varphi_-(\mathbf{k}') d\mathbf{k}',$$

we get that

$$U_-(\mathbf{k}, \mathbf{k}') = \frac{1}{(2\pi)^3} \int_{\mathbf{R}^3} e^{-i\mathbf{kx}} \psi(\mathbf{x}, \mathbf{k}') d\mathbf{x}.$$

It can also be shown that

$$U_+(\mathbf{k}, \mathbf{k}') = \frac{1}{(2\pi)^3} \int_{\mathbf{R}^3} e^{-i\mathbf{kx}} \overline{\psi(\mathbf{x}, -\mathbf{k}')} d\mathbf{x}.$$

The connection between the operators U_{\pm} and the eigenfunctions of the continuous spectrum of the operator H becomes especially clear if we write (13) with $f(s) = s$ together with (8) and (12) in the momentum representation

$$(17) \quad HU_{\pm}(\mathbf{p}, \mathbf{k}) = k^2 U_{\pm}(\mathbf{p}, \mathbf{k})$$

(H is the Schrödinger operator in the momentum representation),

$$(18) \quad \int_{\mathbf{R}^3} U_{\pm}(\mathbf{p}, \mathbf{k}_1) \overline{U_{\pm}(\mathbf{p}, \mathbf{k}_2)} d\mathbf{p} = \delta(\mathbf{k}_1 - \mathbf{k}_2),$$

$$(19) \quad \int_{\mathbf{R}^3} U_{\pm}(\mathbf{p}, \mathbf{k}) \overline{U_{\pm}(\mathbf{p}', \mathbf{k})} d\mathbf{k} + \sum_n \chi_n(\mathbf{p}) \overline{\chi_n(\mathbf{p}')} = \delta(\mathbf{p} - \mathbf{p}').$$

The formula (17) shows that the kernels of the operators U_{\pm} , as functions of \mathbf{p} , are eigenfunctions of the continuous spectrum corresponding to the eigenvalue k^2 . Then (18) is the condition of "orthonormality" of the eigenfunctions of $U_{\pm}(\mathbf{p}, \mathbf{k})$, and (19) is the condition of completeness of the systems $\{\chi_n(\mathbf{p}), U_{\pm}(\mathbf{p}, \mathbf{k})\}$.

§ 44. Properties of commuting operators

In this section we consider the properties of commuting operators, discuss once again the question of simultaneous measurement of observables, and introduce a concept important for quantum mechanics. the notion of a complete set of commuting observables.

To begin, let us consider two self-adjoint commuting operators A and B with pure point spectrum. We show that such operators have a common complete set of eigenvectors. Let

$$(1) \quad \begin{aligned} A\varphi_m^{(i)} &= \lambda_m \varphi_m^{(i)}, \quad i = 1, 2, \dots, \\ B\psi_n^{(j)} &= \mu_n \psi_n^{(j)}, \quad j = 1, 2, \dots, \end{aligned}$$

where the collections $\{\varphi_m^{(i)}\}$ and $\{\psi_n^{(j)}\}$ of vectors are complete in the state space \mathcal{H} .

Let \mathcal{H}_m denote the eigenspace of A corresponding to the eigenvalue λ_m , and let P_m be the projection onto this subspace. Let \mathcal{H}'_n and P'_n be the analogous objects for B . From the condition $AB = BA$ it follows that $P_m P'_n = P'_n P_m$. Let $P_{mn} = P_m P'_n$. Obviously, P_{mn} is the projection onto the subspace $\mathcal{H}_{mn} = \mathcal{H}_m \cap \mathcal{H}'_n$. If $\varphi \in \mathcal{H}_{mn}$, then it satisfies both the equations in (1). Further, $\mathcal{H}_{mn} \perp \mathcal{H}_{m'n'}$ if the index mn is different from the index $m'n'$. Let $\{\varphi_{mn}^{(k)}\}$, $k = 1, 2, \dots$, be a basis in the subspace \mathcal{H}_{mn} . To prove the completeness of the system $\{\varphi_{mn}^{(k)}\}$, $k, m, n = 1, 2, \dots$, of vectors in \mathcal{H} it suffices to verify that no nonzero vector φ in \mathcal{H} is orthogonal to all the subspaces \mathcal{H}_{mn} . It follows from the completeness of the set of eigenvectors of A that for any nonzero φ there is an index m such that $P_m \varphi \neq 0$. Similarly, there is an index n such that $P'_n P_m \varphi \neq 0$; that is, $P_{mn} \varphi \neq 0$ for any nonzero φ for some m and n .

Thus, we have shown that $\{\varphi_{mn}^{(k)}\}$ is a basis in \mathcal{H} consisting of common eigenvectors of the operators A and B .

The converse also holds. If two operators have a common complete set of eigenvectors, then they commute. Indeed, let $\{\varphi_{mn}^{(k)}\}$ be a common complete set of eigenvectors of A and B :

$$(2) \quad \begin{aligned} A\varphi_{mn}^{(k)} &= \lambda_m \varphi_{mn}^{(n)}, \\ B\varphi_{mn}^{(k)} &= \mu_n \varphi_{mn}^{(k)}, \quad k = 1, 2, \dots. \end{aligned}$$

Then $AB\varphi_{mn}^{(k)} = \lambda_m \mu_n \varphi_{mn}^{(k)}$ and $BA\varphi_{mn}^{(k)} = \lambda_m \mu_n \varphi_{mn}^{(k)}$. From the completeness of the collection it follows that $AB\varphi = BA\varphi$ for any $\varphi \in \mathcal{H}$.

The assertions proved are valid for an arbitrary number of pairwise commuting self-adjoint operators A_1, \dots, A_n, \dots with pure point spectrum and with certain stipulations for operators with a continuous spectrum.

We also have the following assertion. If the self-adjoint operators A_1, \dots, A_n commute, then there is a self-adjoint operator R such that all the operators A_i , $i = 1, 2, \dots, n$, are functions $A_i = F_i(R)$ of it.

It is quite easy to construct such an operator R for operators with pure point spectrum. Let us consider the case of two commuting operators A and B . They have a common complete system $\{\varphi_{mn}^{(k)}\}$ of eigenvectors which satisfy the equations (2). We define the operator R by the equalities

$$R\varphi_{mn}^{(k)} = r_{mn}^{(k)}\varphi_{mn}^{(k)},$$

where the $r_{mn}^{(k)}$ are distinct real numbers. Obviously, R is a self-adjoint operator with a simple pure point spectrum. We now introduce real functions $F(x)$ and $G(x)$ satisfying the conditions $\lambda_m = F(r_{mn}^{(k)})$ and $\mu_n = G(r_{mn}^{(k)})$. For x different from all the numbers $r_{mn}^{(k)}$, the values of $F(x)$ and $G(x)$ are not important. It follows at once from the definition of a function of an operator that $A = F(R)$ and $B = G(R)$. The general theorem for commuting operators A_1, \dots, A_n with arbitrary spectrum was proved by von Neumann.

We now discuss the physical consequences of the assertions formulated for commuting operators. We recall that the Heisenberg uncertainty relation does not impose any restrictions on the variance of the commuting observables, and in this sense we have called them simultaneously measurable observables. The concept of simultaneous measurability can now be made more precise. The last assertion shows that to measure the numerical values of commuting observables it suffices to measure the single observable R ; that is, it is possible in principle to find out the numerical values of all the observables A_1, \dots, A_n with a single measurement. The existence of a system of common eigenvectors implies the existence of an infinite set of states in which all these observables have definite numerical values. Finally,

it will follow from results in the next section that for any state it is possible to construct a common distribution function for the numerical values of simultaneously measurable observables.

We shall introduce the concept of a function of commuting operators A and B . For a real function $f(x, y)$ it is possible to construct a self-adjoint operator $f(A, B)$ by setting

$$f(A, B) \varphi_{mn}^{(k)} = f(\lambda_m, \mu_n) \varphi_{mn}^{(k)},$$

where the vectors $\varphi_{mn}^{(k)}$ satisfy (2). This definition agrees with the definition given earlier for a function of simultaneously measurable observables.

We have the following assertion. If the operator D commutes with any operator C commuting with all the operators A_1, \dots, A_n and $[A_i, A_j] = 0$, then D is a function of these operators. In the proof we again confine ourselves to the case of two operators with pure point spectrum.

Suppose that to each pair λ_m, μ_n of eigenvalues there corresponds a single eigenvector:

$$A\varphi_{mn} = \lambda_m \varphi_{mn}, \quad B\varphi_{mn} = \mu_n \varphi_{mn}.$$

In this case it suffices that D commute with just the operators A and B . Indeed, if $[D, A] = 0$ and $\varphi \in \mathcal{H}_m$, then also $D\varphi \in \mathcal{H}_m$, since $AD\varphi = DA\varphi = \lambda_m D\varphi$. Similarly, from the condition $[D, B] = 0$, we get that if $\varphi \in \mathcal{H}'_n$, then $D\varphi \in \mathcal{H}'_n$. Therefore, if $\varphi \in \mathcal{H}_{mn}$, then $D\varphi \in \mathcal{H}_{mn}$. By assumption, the spaces \mathcal{H}_{mn} are one dimensional, and $D\varphi_{mn}$ is only a numerical multiple, which can be different from φ_{mn} ; that is, $D\varphi_{mn} = \kappa_{mn} \varphi_{mn}$. Choosing the function $f(x, y)$ such that $\kappa_{mn} = f(\lambda_m, \mu_n)$, we see that $D = f(A, B)$.

Let us proceed to the more complicated case when the subspaces \mathcal{H}_{mn} are not one dimensional.

$$A\varphi_{mn}^{(k)} = \lambda_m \varphi_{mn}^{(k)}, \quad B\varphi_{mn}^{(k)} = \mu_n \varphi_{mn}^{(k)}, \quad k = 1, 2, \dots.$$

We consider the collection of eigenvectors corresponding to the pair λ_m, μ_n . For brevity we omit the index mn . The operators $C^{(j)}$

and $C^{(jl)}$ are introduced by specifying that

$$C^{(j)} \varphi^{(k)} = \delta_{jk} \varphi^{(k)}.$$

$$C^{(jl)} \varphi^{(k)} = \begin{cases} \varphi^{(j)}, & k = l, \\ \varphi^{(l)}, & k = j, \\ 0 & k \neq j, k \neq l, \end{cases}$$

$$C^{(j)} \varphi = C^{(jl)} \varphi = 0, \quad \varphi \in \mathcal{H}_{mn}^\perp.$$

where \mathcal{H}_{mn}^\perp is the orthogonal complement of the subspace \mathcal{H}_{mn} . It is easy to verify that all the operators $C^{(j)}$ and $C^{(jl)}$ commute with A and B . From the condition $[D, C^{(j)}] = 0$ we get that

$$D\varphi^{(j)} = DC^{(j)}\varphi^{(j)} = C^{(j)}D\varphi^{(j)},$$

from which it follows that $D\varphi^{(j)}$ is proportional to $\varphi^{(j)}$:

$$D\varphi^{(j)} = \kappa^{(j)}\varphi^{(j)}.$$

We show that all the numbers $\kappa^{(j)}$ coincide:

$$\begin{aligned} \kappa^{(j)}\varphi^{(j)} &= D\varphi^{(j)} = DC^{(jl)}\varphi^{(l)} = C^{(jl)}D\varphi^{(l)} \\ &= C^{(jl)}\kappa^{(l)}\varphi^{(l)} = \kappa^{(l)}\varphi^{(j)}. \end{aligned}$$

Thus, the vectors $\varphi_{mn}^{(k)}$ are eigenvectors of D , and the eigenvalues do not depend on the index k : $D\varphi_{mn}^{(k)} = \kappa_{mn}\varphi_{mn}^{(k)}$. Therefore, $D = f(A, B)$ as in the first case.

We point out that commutativity of D with just A and B would not be sufficient here. What is more, if the commutativity of D with A and B implies that $D = f(A, B)$, then we can assert that to each pair of eigenvalues λ_m and μ_n there corresponds a single eigenvector φ_{mn} . Indeed, if there are several such vectors, then we can always construct an operator that commutes with A and B but is not a function of them: for example, the operator $C^{(jl)}$.

We can now define the important concept of a complete system of commuting operators. A system of self-adjoint operators A_1, \dots, A_n is called a complete system of commuting operators if:

- 1) the operators A_i commute pairwise, that is, $[A_i, A_j] = 0$ for $i, j = 1, 2, \dots, n$;
- 2) none of the operators A_i is a function of the others;

3) any operator commuting with all the A_i is a function of these operators.

From the assertions proved above and the conditions 1) and 3) of the definition of a complete collection A_1, \dots, A_n , it follows that there is a common complete system of eigenvectors of all these operators,

$$A_i \varphi_{a_1, \dots, a_n} = a_i \varphi_{a_1, \dots, a_n}, \quad i = 1, 2, \dots, n,$$

and to each collection of eigenvalues a_1, \dots, a_n there corresponds a single eigenvector $\varphi_{a_1, \dots, a_n}$. The condition 2) implies that the last property is not enjoyed by a common complete system of eigenvectors for only a part of the operators A_1, \dots, A_n . In fact, 2) means that there are no “superfluous” operators among A_1, \dots, A_n .

If as a result of measurements it is known that the numerical values of a complete set of observables A_1, \dots, A_n in some state are certainly equal to a_1, \dots, a_n , then we can assert that this state is described by the vector $\varphi_{a_1, \dots, a_n}$.

In conclusion we note that if a set of pairwise commuting independent operators is not complete, then it can be supplemented—in many ways—to form a complete system.

§ 45. Representation of the state space with respect to a complete set of observables

Suppose that A_1, \dots, A_n is a complete set of operators with pure point spectrum. These operators have a common complete set of eigenvectors $\varphi_{a_1, \dots, a_n}$, and corresponding to each collection of eigenvalues is a single eigenvector $\varphi_{a_1, \dots, a_n}$. An arbitrary vector $\psi \in \mathcal{H}$ can be represented as a series

$$\psi = \sum_{a_1, \dots, a_n} \psi(a_1, \dots, a_n) \varphi_{a_1, \dots, a_n}, \quad \psi(a_1, \dots, a_n) = (\psi, \varphi_{a_1, \dots, a_n}).$$

This formula determines a one-to-one correspondence between the vectors ψ and the functions $\psi(a_1, \dots, a_n)$ defined on the spectrum of the operators A_1, \dots, A_n :

$$\psi \leftrightarrow \psi(a_1, \dots, a_n).$$

Obviously,

$$(\psi_1, \psi_2) = \sum_{a_1, \dots, a_n} \psi_1(a_1, \dots, a_n) \overline{\psi_2(a_1, \dots, a_n)},$$

$$A_i \psi \leftrightarrow a_i \psi(a_1, \dots, a_n);$$

that is, the representation constructed is an eigenrepresentation for all the operators A_1, \dots, A_n (the action of these operators reduces to multiplication by the corresponding variables).

The function $\psi(a_1, \dots, a_n)$ is called a wave function. To clarify its physical meaning we construct, as in the preceding section, an operator R such that $A_i = F_i(R)$ for $i = 1, 2, \dots, n$,

$$R \varphi_{a_1, \dots, a_n} = r_{a_1, \dots, a_n} \varphi_{a_1, \dots, a_n},$$

where r_{a_1, \dots, a_n} are distinct real numbers and $a_i = F_i(r_{a_1, \dots, a_n})$. We know that $|(\psi, \varphi_{a_1, \dots, a_n})|^2 = |\psi(a_1, \dots, a_n)|^2$ is the probability of obtaining the numerical value r_{a_1, \dots, a_n} for the observable R by a measurement. Therefore, $|\psi(a_1, \dots, a_n)|^2$ is the probability of obtaining the values a_1, \dots, a_n as a result of simultaneous measurement of the observables A_1, \dots, A_n .

All these results generalize to the case of a complete set of operators A_1, \dots, A_n with arbitrary spectrum. We formulate the theorem without proof.

Theorem. *Let A_1, \dots, A_n be a complete set of commuting operators. Then there exists a representation of the state space such that a vector $\psi \in \mathcal{H}$ is represented by a function $\psi(a_1, \dots, a_n)$ defined on some set \mathfrak{A} ($a = (a_1, \dots, a_n) \in \mathfrak{A}$). A measure $d\mu(a)$ is given on the set \mathfrak{A} , and a scalar product is defined by the formula*

$$(\psi_1, \psi_2) = \int_{\mathfrak{A}} \psi_1(a) \overline{\psi_2(a)} d\mu(a).$$

In this representation the operators A_1, \dots, A_n are the operators of multiplication by the corresponding variables

$$A_i \psi(a_1, \dots, a_n) = a_i \psi(a_1, \dots, a_n), \quad i = 1, 2, \dots, n.$$

The function $|\psi(a_1, \dots, a_n)|^2$ is the density of the common distribution function for the observables A_1, \dots, A_n with respect to the measure $d\mu(a)$.

Above we already had examples of complete sets of commuting operators and corresponding representations of the state space \mathcal{H} .

For a structureless particle the coordinate operators Q_1, Q_2, Q_3 form a complete set. The coordinate representation corresponds to this set. The momentum representation is constructed similarly from the complete set P_1, P_2, P_3 . A complete set is also formed by the operators H, L^2, L_3 , where H is the Schrödinger operator for a particle in a central field. The representation corresponding to this complete set was described in § 31. For a one-dimensional particle the Schrödinger operator H for the harmonic oscillator by itself represents a complete set. The corresponding representation was constructed in § 18.

§ 46. Spin

Up to this time we have assumed that the electron is a material point with mass m and charge $-e$; that is, it is a structureless particle whose state space \mathcal{H} can be realized, for example, as the space $L^2(\mathbf{R}^3)$ of square-integrable functions $\psi(\mathbf{x})$. On the basis of this representation of the electron, we calculated the energy levels of the hydrogen atom and obtained results that coincide with experimental results to a high degree of accuracy. Nevertheless, there are experiments showing that such a description of the electron is not complete.

We have already mentioned the experiments of Stern and Gerlach. These experiments showed that the projection on a certain direction of the magnetic moment of the hydrogen atom in the ground state can take two values. In § 34 we constructed the quantum observable “projection of the magnetic moment” of a structureless charged particle, and we saw that the third projection of the magnetic moment is proportional to the projection L_3 of the angular momentum. From the calculation of the hydrogen atom, we know that the numerical value of L_3 is zero for the ground state. Therefore, the magnetic moment of the atom in the ground state must also be zero. This contradiction can be explained if we assume that the electron itself has a magnetic moment and a mechanical moment whose projections on a certain direction can take two values

The electron's own angular momentum is called its *spin*, in contradistinction to the angular momentum associated with its motion in space, which is usually called its orbital angular momentum.

The existence of an observable, which can take two numerical values, leads to the necessity of considering that the electron can be in two different internal states independently of the state of its motion in space. This in turn leads to a doubling of the total number of states of the electron. For example, to each state of the electron in a hydrogen atom (without counting spin) there correspond two states that differ by the projection of the spin on some direction.

If we assume that there is no additional interaction connected with the spin, then the multiplicity of all the energy eigenvalues turns out to be twice that of a spinless particle. But if there are such interactions, then the degeneracies connected with the spin can disappear, and there can be a splitting of the energy levels. Experiments show that such a splitting does indeed take place.

In § 32 we described a model for an alkali metal atom, based on the assumption that the valence electron of the atom moves in a central field. This model is not bad for describing the configuration of the energy levels of alkali metal atoms, but neither this model itself nor any of its refinements can explain the observed splitting of the levels into two nearby levels when $l \neq 0$. The hypothesis of spin enables us to easily explain this splitting. In atomic physics there are many other phenomena which can be explained on the basis of this hypothesis. We shall see later that only the doubling of the number of electron states that is connected with spin makes it possible to explain the length of the periods in the periodic table.

We would like to stress that the hypothesis of spin for the electron is a hypothesis about the nature of a concrete elementary particle and does not concern the general principles of quantum mechanics. The apparatus of quantum mechanics turns out to be suitable for describing particles with spin.

Let us begin with the construction of the state space for an electron. Without spin taken into account, $L^2(\mathbf{R}^3)$ is the state space \mathcal{H} in the coordinate representation. The introduction of spin requires

us to extend the state space, since the number of states of a particle with spin is larger than for a spinless particle.

A doubling of the number of states without changing the physical content of the theory is easily attained by replacing the state space $\mathcal{H} = L^2(\mathbf{R}^3)$ by $\mathcal{H}_S = L^2(\mathbf{R}^3) \otimes \mathbf{C}^2$ and associating with each observable A acting in \mathcal{H} the observable $A \otimes I$ acting in \mathcal{H}_S .

We explain this in more detail. The elements of the state space of a particle with spin are pairs of functions

$$(1) \quad \Psi = \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix}.$$

The scalar product in the space \mathcal{H}_S is given by

$$(2) \quad (\Psi, \Phi) = \int_{\mathbf{R}^3} \psi_1(\mathbf{x}) \overline{\varphi_1(\mathbf{x})} d\mathbf{x} + \int_{\mathbf{R}^3} \psi_2(\mathbf{x}) \overline{\varphi_2(\mathbf{x})} d\mathbf{x}.$$

For the observables $A \otimes I$ acting in \mathcal{H}_S we use the same notation and terminology as for the observables A acting in \mathcal{H} . Thus, the coordinate operators Q_1, Q_2, Q_3 , the momentum projection operators P_1, P_2, P_3 , and so on, all act in \mathcal{H}_S . For example,

$$P_1 \Psi = P_1 \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\hbar}{i} \frac{\partial \psi_1(\mathbf{x})}{\partial x_1} \\ \frac{\hbar}{i} \frac{\partial \psi_2(\mathbf{x})}{\partial x_1} \end{pmatrix}.$$

To each pure state $\psi(\mathbf{x})$ in \mathcal{H} there now correspond the two orthogonal states

$$\Psi_1 = \begin{pmatrix} \psi(\mathbf{x}) \\ 0 \end{pmatrix}, \quad \Psi_2 = \begin{pmatrix} 0 \\ \psi(\mathbf{x}) \end{pmatrix}$$

in \mathcal{H}_S , or any linear combination of them. It is clear that the mean value of any observable A in a state ψ will be equal to the mean value of the observable $A \otimes I$ in the states Ψ_1 and Ψ_2 . Therefore, by introducing the space \mathcal{H}_S and confining ourselves to observables of the form $A \otimes I$, we have indeed attained a doubling of the number of states while preserving all the physical consequences of the theory.

However, along with the observables $A \otimes I$ there are also other observables acting in the space \mathcal{H}_S , for example, of the form $I \otimes S$, where S is a self-adjoint operator on \mathbf{C}^2 . Of course, there are observables acting in \mathcal{H}_S that cannot be represented in the form $A \otimes I$ or $I \otimes S$, for example, sums or products of such observables.

Let us consider observables of the type $I \otimes S$. First of all, it is clear that any such observable commutes with any observable $A \otimes I$ and is not a function of observables of the latter type. Therefore, on the space \mathcal{H}_S the observables Q_1, Q_2, Q_3 do not form a complete set of commuting observables, and we will have to supplement this set to obtain a complete set.

Any self-adjoint operator S on \mathbf{C}^2 is representable as a self-adjoint 2×2 matrix and can be expressed as a linear combination of four linearly independent matrices. It is convenient to take these matrices to be the identity matrix I and the Pauli matrices $\sigma_1, \sigma_2, \sigma_3$.

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

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

The properties of the Pauli matrices were discussed in § 27. The matrices $S_j = \sigma_j/2$, $j = 1, 2, 3$, have the same commutation relations as for the orbital angular momentum:

$$[S_1, S_2] = iS_3, \quad [S_2, S_3] = iS_1, \quad [S_3, S_1] = iS_2.$$

We have seen that the commutation relations are one of the most important properties of the angular momentum operators. Therefore, it is reasonable to identify the operators $I \otimes S_j$, $j = 1, 2, 3$, with the operators of the spin projections.⁴⁹ These operators will usually be denoted by S_j in what follows. The operators S_j have eigenvalues $\pm 1/2$, which are the admissible values of the projections of the spin on some direction.⁵⁰ The vectors $\Psi_1 = \begin{pmatrix} \psi(\mathbf{x}) \\ 0 \end{pmatrix}$ and $\Psi_2 = \begin{pmatrix} 0 \\ \psi(\mathbf{x}) \end{pmatrix}$ are eigenvectors of the operator S_3 with eigenvalues $+1/2$ and $-1/2$. Therefore, these vectors describe states with a definite value of the third projection of the spin.

⁴⁹Deeper considerations are based on the fact that for a system with spherically symmetric Schrödinger operator the operators $L_j + S_j$ are quantum integrals of motion. We shall discuss this question later.

⁵⁰We write all formulas in a system of units in which $\hbar = 1$. In the usual system of units, the operators of the spin projections have the form $S_j = (h/2)\sigma_j$, and the admissible numerical values of these projections are equal to $\pm h/2$. We remark that in passing to classical mechanics $h \rightarrow 0$, and correspondingly the projections of the spin tend to zero. Therefore, the spin is a specifically quantum observable.

For what follows, it is convenient for us to change the notation and write a vector $\Psi \in \mathcal{H}_S$ as a function $\Psi(\mathbf{x}, s_3)$, where $\mathbf{x} \in \mathbf{R}^3$ and s_3 takes the two values $+1/2$ and $-1/2$. Such notation is equivalent to (1) if we let

$$\Psi\left(\mathbf{x}, \frac{1}{2}\right) = \psi_1(\mathbf{x}), \quad \Psi\left(\mathbf{x}, -\frac{1}{2}\right) = \psi_2(\mathbf{x}).$$

From the equality

$$S_3 \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\psi_1(\mathbf{x}) \\ -\frac{1}{2}\psi_2(\mathbf{x}) \end{pmatrix}$$

it follows that

$$S_3 \Psi(\mathbf{x}, s_3) = s_3 \Psi(\mathbf{x}, s_3);$$

that is, the operator S_3 , like the operators Q_1 , Q_2 , Q_3 , is an operator of multiplication by a variable. We see that the representation constructed for the state space of a particle with spin is an eigenrepresentation for the operators Q_1 , Q_2 , Q_3 and S_3 , and these operators form a complete set of commuting operators acting in \mathcal{H}_S .

It is now easy to understand the physical meaning of the functions $\Psi(\mathbf{x}, s_3)$. According to the general interpretation, $|\Psi(\mathbf{x}, s_3)|^2$ is the density of the coordinate distribution function under the condition that the third projection of the spin has the value s_3 , and $\int_{\mathbf{R}^3} |\Psi(\mathbf{x}, s_3)|^2 d\mathbf{x}$ is the probability of obtaining the value s_3 as a result of measurement.

Along with the observables S_1 , S_2 , S_3 we can introduce the operator $S^2 = S_1^2 + S_2^2 + S_3^2$ of the square of the spin. Substituting $S_j = \frac{1}{2}\sigma_j$ in this expression and considering that $\sigma_j^2 = I$, we get that $S^2 = \frac{3}{4}I$. We see that any vector $\Psi \in \mathcal{H}_S$ is an “eigenvector” for the operator S^2 , with eigenvalue $3/4$. This eigenvalue can be written in the form⁵¹ $s(s+1)$, where $s = 1/2$. Therefore, one says that the spin of the electron is equal to $1/2$.

Let us construct a representation of the rotation group on the space \mathcal{H}_S . We recall that the representation of the rotations g act in the space $\mathcal{H} = L^2(\mathbf{R}^3)$ as the operators $W(g) = e^{-i(L_1 a_1 + L_2 a_2 + L_3 a_3)}$,

⁵¹In § 29 we showed that from the commutation relations for the angular momentum it follows that the eigenvalues of the operator of the square of the angular momentum have the form $j(j+1)$, where j is an integer or half-integer. For the spin operator this number is commonly denoted by s .

and in the space \mathbf{C}^2 as the operators $U(g) = e^{-i(S_1a_1 + S_2a_2 + S_3a_3)}$. The mapping $g \rightarrow W_S(g)$ with $W_S(g) = W(g) \otimes U(g)$ is a representation on the space \mathcal{H}_S , the product of the representations W and U .

An operator acting in \mathcal{H}_S is said to be spherically symmetric if it commutes with all the operators $W_S(g)$. If the Schrödinger operator H is spherically symmetric, then the operators $W_S(g)$ and the infinitesimal operators

$$\frac{\partial W_S}{\partial a_j} \Big|_{\mathbf{a}=0} = -i(L_j \otimes I + I \otimes S_j), \quad j = 1, 2, 3,$$

are integrals of motion. Therefore, for a system with spherically symmetric Schrödinger operator we have the law of conservation of the total angular momentum, whose projections are $J_j = L_j + S_j$.

We remark that in the general case of a spherically symmetric operator H , there are no separate laws of conservation of orbital angular momentum and spin angular momentum. However, if a spherically symmetric Schrödinger operator acting in \mathcal{H}_S commutes also with all the operators $W(g) \otimes I$, then it commutes with all the operators $I \otimes U(g)$, and there are laws of conservation for the observables L_j and S_j separately. An example of such a Schrödinger operator is the operator $H \otimes I$, where H is the Schrödinger operator for a particle in a central field.

§ 47. Spin of a system of two electrons

The space \mathbf{C}^2 introduced in the previous section is often called the spin space for the electron. For a system of two electrons the spin space is the space $\mathbf{C}^4 = \mathbf{C}^2 \otimes \mathbf{C}^2$. In \mathbf{C}^2 we choose a basis consisting of the eigenvectors

$$U_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad U_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

of the operator S_3 with eigenvalues $1/2$ and $-1/2$, respectively. As basis vectors in \mathbf{C}^4 we can take $U_+^{(1)}U_+^{(2)}$, $U_+^{(1)}U_-^{(2)}$, $U_+^{(1)}U_-^{(2)}$, and $U_-^{(1)}U_+^{(2)}$, where the indices (1) and (2) below indicate the spin subspaces of the electrons.

However, it turns out to be more convenient to use another orthonormal basis consisting of the vectors

$$\begin{aligned}W_1 &= U_{+}^{(1)}U_{+}^{(2)}, \\W_2 &= U_{-}^{(1)}U_{-}^{(2)}, \\W_3 &= \frac{1}{\sqrt{2}}(U_{+}^{(1)}U_{-}^{(2)} + U_{-}^{(1)}U_{+}^{(2)}), \\W_4 &= \frac{1}{\sqrt{2}}(U_{+}^{(1)}U_{-}^{(2)} - U_{-}^{(1)}U_{+}^{(2)})\end{aligned}$$

The convenience of the new basis consists in the fact that the vectors W_i , $i = 1, 2, 3, 4$, are eigenvectors of the operators

$$\begin{aligned}S_3 &= S_3^{(1)} + S_3^{(2)}, \\S^2 &= S_1^2 + S_2^2 + S_3^2.\end{aligned}$$

Here the operator S_3 is the third projection of the total spin of the two electrons,⁵² and S_1 and S_2 have the analogous meanings. The operator S^2 is the square of the total spin.

To verify our assertion about the vectors W_i , $i = 1, 2, 3, 4$, we find the results when the operators σ_1 , σ_2 , σ_3 are applied to the basis vectors U_+ and U_- :

$$\begin{aligned}\sigma_1 U_+ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = U_-, \\&\sigma_1 U_- = U_+, \\(1) \quad \sigma_2 U_+ &= iU_-, \\&\sigma_2 U_- = -iU_+, \\&\sigma_3 U_+ = U_+, \\&\sigma_3 U_- = -U_+.\end{aligned}$$

⁵²More precisely, $S_3 = \frac{1}{2}\sigma_3 \otimes I + \frac{1}{2}I \otimes \sigma_3$. The operators $\sigma_j \otimes I$ and $I \otimes \sigma_j$ will be denoted below by $\sigma_j^{(1)}$ and $\sigma_j^{(2)}$, respectively.

Using these formulas, we obtain

$$(2) \quad \begin{aligned} S_3 W_1 &= \frac{1}{2}(\sigma_1^{(1)} + \sigma_3^{(2)})U_+^{(1)}U_-^{(2)} = U_+^{(1)}U_+^{(2)} = 1W_1, \\ S_3 W_2 &= -1W_2, \\ S_3 W_3 &= 0W_3, \\ S_3 W_4 &= 0W_4. \end{aligned}$$

Along with (2) we have the formulas

$$(3) \quad \begin{aligned} S^2 W_j &= 2W_j, \quad j = 1, 2, 3, \\ S^2 W_4 &= 0W_4 \end{aligned}$$

We verify (3) for the vector W_3 :

$$\begin{aligned} S^2 W_3 &= (S_1^2 + S_2^2 + S_3^2)W_3 \\ &= \frac{1}{4}[(\sigma_1^{(1)} + \sigma_1^{(2)})^2 + (\sigma_2^{(1)} + \sigma_2^{(2)})^2 + (\sigma_3^{(1)} + \sigma_3^{(2)})^2]W_3 \\ &= \left[\frac{3}{2}I + \frac{1}{2}(\sigma_1^{(1)}\sigma_1^{(2)} + \sigma_2^{(1)}\sigma_2^{(2)} + \sigma_3^{(1)}\sigma_3^{(2)}) \right] \frac{1}{\sqrt{2}}(U_+^{(1)}U_-^{(2)} + U_-^{(1)}U_+^{(2)}) \\ &= \frac{3}{2}W_3 + \frac{1}{2}W_3 + \frac{1}{2}W_3 - \frac{1}{2}W_3 = 2W_3. \end{aligned}$$

Here we have used (1) and the equalities $\sigma_j^2 = I$, $j = 1, 2, 3$. Thus, the first three vectors W_1, W_2, W_3 describe states in which the square of the total spin of the system of two electrons is equal to 2. The number 2 can be written in the form $2 = S(S+1)$, where $S = 1$, and therefore the total spin in these states is equal to 1. In correspondence with the general properties of the angular momentum, the projection of the total spin takes the values ± 1 and 0 in these states. The vector W_4 describes a state with total spin equal to zero. One sometimes says that in the states W_1, W_2, W_3 the spins of the electrons are parallel, while in the state W_4 they are antiparallel.

We discuss the result obtained from the point of view of group theory. We know that the irreducible representation of the rotation group by the operators $U(g) = e^{-\frac{i}{2}(\sigma_1 a_1 + \sigma_2 a_2 + \sigma_3 a_3)}$ acts in the space \mathbf{C}^2 . It is clear that the mapping $g \rightarrow \tilde{U}(g) = U(g) \otimes U(g) = e^{-i(S_1 a_1 + S_2 a_2 + S_3 a_3)}$ is a representation of the rotation group on the space \mathbf{C}^4 . The representation \tilde{U} is the tensor product of the same two representations U , and it is reducible. According to results in

§ 29, the space \mathbf{C}^4 is representable as a direct sum of two subspaces that are invariant under the operators $\tilde{U}(g)$, and irreducible representations act in these subspaces. The first subspace is spanned by W_1, W_2, W_3 and the second by W_4 . Using the notation in § 29, we can write this result in the form

$$D_{\frac{1}{2}} \otimes D_{\frac{1}{2}} = D_0 \oplus D_1.$$

We remark that we have proved a particular case of the theorem on decomposition of a tensor product of irreducible representations of the rotation group. We state this theorem without proof:

If the irreducible representations D_{j_1} and D_{j_2} of the rotation group act in the spaces \mathcal{E}_1 and \mathcal{E}_2 , then the tensor product of the representations can be represented as the direct sum

$$D_{j_1} \otimes D_{j_2} = D_{|j_1 - j_2|} \oplus D_{|j_1 - j_2|+1} \oplus \cdots \oplus D_{j_1 + j_2}$$

of irreducible representations.

The last formula is called the Clebsch–Gordan decomposition. This decomposition is obtained by passing from the basis consisting of the vectors $e_{j_1 m_1} e_{j_2 m_2}$, $m_k = -j_k, -j_k + 1, \dots, j_k$, $k = 1, 2$, where the $e_{j_k m_k}$ are eigenvectors of the operators $(J^{(k)})^2$ and $J_3^{(k)}$, to the basis of vectors $e_{j_1 j_2 JM}$, $J = |j_1 - j_2|, |j_1 - j_2| + 1, \dots, j_1 + j_2$, $M = -J, -J + 1, \dots, J$. The vectors $e_{j_1 j_2 JM}$ are eigenvectors for the four commuting operators $(J^{(1)})^2, (J^{(2)})^2$,

$$J^2 = (J_1^{(1)} + J_1^{(2)})^2 + (J_2^{(1)} + J_2^{(2)})^2 + (J_3^{(1)} + J_3^{(2)})^2, \quad J_3 = J_3^{(1)} + J_3^{(2)}.$$

The vectors $e_{j_1 j_2 JM}$ are representable in the form

$$(4) \quad e_{j_1 j_2 JM} = \sum_{m_1, m_2} C_{j_1 j_2 JM, j_1 j_2 m_1 m_2} e_{j_1 m_1} e_{j_2 m_2},$$

where the summation indices m_k run through the values $-j_k, -j_k + 1, \dots, j_k$, $k = 1, 2$, and the coefficients C in the decomposition (4) are called the Clebsch–Gordan coefficients. We remark that in passing to the basis W_1, W_2, W_3, W_4 we found these coefficients for the case $j_1 = j_2 = 1/2$.

From the stated theorem it follows also that if in some state the angular momentum $J^{(1)}$ has the value j_1 and the angular momentum $J^{(2)}$ has the value j_2 , then the total angular momentum J can take

the values $|j_1 - j_2|, |j_1 + j_2| + 1, \dots, j_1 + j_2$. Both the orbital or spin angular momenta for different particles and the orbital and spin angular momentum for a single particle can combine in this way.

In concluding this section we mention a property of the basis elements W_i , $i = 1, 2, 3, 4$, that is important for what follows. Introducing the spin variables $s_3^{(1)}$ and $s_3^{(2)}$, each of which can take the two values $\pm 1/2$, we write the W_i as

$$(5) \quad \begin{aligned} W_1(s_3^{(1)}, s_3^{(2)}) &= U_+(s_3^{(1)}) U_+(s_3^{(2)}), \\ W_2(s_3^{(1)}, s_3^{(2)}) &= U_-(s_3^{(1)}) U_-(s_3^{(2)}), \\ W_3(s_3^{(1)}, s_3^{(2)}) &= \frac{1}{\sqrt{2}} [U_+(s_3^{(1)}) U_-(s_3^{(2)}) + U_-(s_3^{(1)}) U_+(s_3^{(2)})], \\ W_4(s_3^{(1)}, s_3^{(2)}) &= \frac{1}{\sqrt{2}} [U_+(s_3^{(1)}) U_-(s_3^{(2)}) - U_-(s_3^{(1)}) U_+(s_3^{(2)})]. \end{aligned}$$

where $U_+(\frac{1}{2}) = 1$, $U_+(-\frac{1}{2}) = 0$, $U_-(\frac{1}{2}) = 0$, $U_-(-\frac{1}{2}) = 1$.

It follows from (5) that the functions W_1 , W_2 , W_3 are symmetric functions of the spin variables; that is,

$$W_i(s_3^{(2)}, s_3^{(1)}) = W_i(s_3^{(1)}, s_3^{(2)}), \quad i = 1, 2, 3.$$

while W_4 is an antisymmetric function

$$W_4(s_3^{(2)}, s_3^{(1)}) = -W_4(s_3^{(1)}, s_3^{(2)}).$$

§ 48. Systems of many particles. The identity principle

So far we have studied mainly the behavior of a single quantum particle. In the coordinate representation, the state space of a spinless particle is the space $L^2(\mathbf{R}^3)$, and for a particle with spin $1/2$ it is $L^2(\mathbf{R}^3) \otimes \mathbf{C}^2$. A natural generalization of such spaces to the case of a system of n particles is the space

$$L^2(\mathbf{R}^{3n}) = L^2(\mathbf{R}^3) \otimes \cdots \otimes L^2(\mathbf{R}^3)$$

for spinless particles, and the space

$$L^2(\mathbf{R}^{3n}) \otimes \mathbf{C}^{2n} = L^2(\mathbf{R}^3) \otimes \mathbf{C}^2 \otimes \cdots \otimes L^2(\mathbf{R}^3) \otimes \mathbf{C}^2$$

for particles with spin $1/2$.

A comparison of theory with experiment shows, however, that this assumption about the state spaces of systems of n particles turns out to be valid only in the case when no two particles in the system are alike. In the presence of identical particles there are peculiarities in the behavior of quantum systems that can be explained on the basis of the so-called identity principle. We note at once that the identity principle is a new principle of quantum mechanics. It cannot be derived from the other fundamental tenets of quantum mechanics formulated earlier and must be postulated.

We use the single symbol \mathcal{H} to denote the spaces $L^2(\mathbf{R}^{3n})$ and $L^2(\mathbf{R}^{3n}) \otimes \mathbf{C}^{2n}$, and we write the elements of these spaces in the form $\Psi(\xi_1, \dots, \xi_n)$, where $\xi_i = \mathbf{x}^{(i)}$ for a spinless particle and $\xi_i = (\mathbf{x}^{(i)}, s_3^{(i)})$ for a particle with spin, $s_3^{(i)}$ being the spin variable of the i th particle. For a particle with spin $1/2$, the variables $s_3^{(i)}$ take the two values $\pm 1/2$.

The scalar product in the spaces \mathcal{H} can also be written in the uniform way

$$(\Psi, \Phi) = \int \Psi(\xi_1, \dots, \xi_n) \overline{\Phi(\xi_1, \dots, \xi_n)} d\xi_1 \dots d\xi_n,$$

where it is assumed that for a particle with spin the integration with respect to the variable ξ_i is integration with respect to the space variable $\mathbf{x}^{(i)}$ and summation with respect to the spin variable $s_3^{(i)}$.

Before passing to the statement of the identity principle, we consider the permutation group S_n , also called the symmetric group. The elements of this group are the permutations

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix},$$

and the identity element is the identity permutation

$$I = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}.$$

The product $\pi = \pi_2 \pi_1$ of two permutations is the permutation obtained as a result of the successive application of the permutations π_1 and π_2 .

It is easy to construct a representation of the group S_n on the space \mathcal{H} . We introduce the operators P_π by

$$P_\pi \Psi(\xi_1, \dots, \xi_n) = \Psi(\xi_{\pi(1)}, \dots, \xi_{\pi(n)}).$$

Obviously, P_π is a unitary operator, and the mapping $\pi \rightarrow P_\pi$ is a representation of the symmetric group on the space \mathcal{H} .

In \mathcal{H} we can immediately distinguish two invariant subspaces with respect to the operators P_π , namely, the subspace \mathcal{H}_S of symmetric functions, for which

$$P_\pi \Psi(\xi_1, \dots, \xi_n) = \Psi(\xi_1, \dots, \xi_n),$$

and the subspace \mathcal{H}_A of antisymmetric functions, for which

$$P_\pi \Psi(\xi_1, \dots, \xi_n) = (-1)^{[\pi]} \Psi(\xi_1, \dots, \xi_n),$$

where $[\pi]$ denotes the parity of the permutation π . It is clear that $\mathcal{H} = \mathcal{H}_A \perp \mathcal{H}_S$.

In the case of two particles, $\mathcal{H} = \mathcal{H}_A \oplus \mathcal{H}_S$. Indeed, any function $\Psi(\xi_1, \xi_2)$ can be written in the form

$$\begin{aligned} \Psi(\xi_1, \xi_2) &= \frac{\Psi(\xi_1, \xi_2) + \Psi(\xi_2, \xi_1)}{2} + \frac{\Psi(\xi_1, \xi_2) - \Psi(\xi_2, \xi_1)}{2} \\ &= \Psi_S(\xi_1, \xi_2) + \Psi_A(\xi_1, \xi_2), \quad \text{where } \Psi_S \in \mathcal{H}_S \text{ and } \Psi_A \in \mathcal{H}_A. \end{aligned}$$

In the case of a large number of particles there are also invariant subspaces more complicated than \mathcal{H}_S and \mathcal{H}_A , but these subspaces are not of interest. The identity principle asserts that the state space of a system of n identical particles is either the space \mathcal{H}_A or the space \mathcal{H}_S . The choice of one of these spaces as the state space depends only on the type of particles.

One says that particles whose states are described by symmetric functions are subject to Bose-Einstein statistics, while particles described by antisymmetric functions are subject to Fermi-Dirac statistics. The first particles are called bosons, and the second particles are called fermions.

It turns out that the spin of a particle determines the statistics to which it is subject. Particles with integer spin (including those without spin) are bosons, and those with half-integer spin are fermions. In nonrelativistic quantum mechanics there is no explanation for the

connection between spin and statistics. This connection is explained in part in relativistic quantum mechanics. Electrons, protons, and neutrons are fermions with spin 1/2, photons are bosons with spin 1, and mesons are bosons with spin zero. The statistics of identical compound particles (for example, atomic nuclei) is determined by the parity of the fermions in their structure, since a permutation of the same complex particles is equivalent to a permutation of several pairs of elementary particles. For example, deuterons, which are composed of a neutron and a proton, are bosons. We note that a deuteron has integer spin,⁵³ since the spins of a proton and a neutron are equal to 1/2.

We write the Schrödinger operator for a system of pairwise interacting particles in an external field:

$$H = - \sum_{i=1}^n \frac{1}{2m} \Delta_i + \sum_{i=1}^n V(\mathbf{x}^{(i)}) + \sum_{i < k}^n U(\mathbf{x}^{(i)} - \mathbf{x}^{(k)}).$$

The first term is the kinetic energy operator of the system of particles, the second term describes the interaction of the particles with the external field, and the third term is the interaction of the particles with each other. We point out that the masses of all the particles are the same, and the interaction potentials $V(\mathbf{x})$ and $U(\mathbf{x})$ do not depend on the numbers of the particles. As a consequence, the Schrödinger operator H commutes with all the operators P_π : $[H, P_\pi] = 0$.

§ 49. Symmetry of the coordinate wave functions of a system of two electrons. The helium atom

Electrons are fermions, and therefore the wave function for a system of two electrons must be antisymmetric:

$$\Psi(\xi_2, \xi_1) = -\Psi(\xi_1, \xi_2).$$

⁵³It is known from experiments that the spin of a deuteron is 1

We decompose the function $\Psi(\xi_1, \xi_2)$ with respect to the basis functions W_1, W_2, W_3, W_4 introduced in § 47:

$$(1) \quad \Psi(\mathbf{x}^{(1)} s_3^{(1)}, \mathbf{x}^{(2)} s_3^{(2)}) = \sum_{i=1}^4 \Psi_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) W_i(s_3^{(1)}, s_3^{(2)})$$

The first three terms in this sum correspond to the states with total spin 1, and the fourth term describes a state with total spin zero. The functions $\Psi_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), i = 1, 2, 3, 4$, introduced by the relation (1) are called the coordinate wave functions in contrast to $\Psi(\xi_1, \xi_2)$, which is called the total wave function. In § 47 we saw that the $W_i(s_3^{(1)}, s_3^{(2)})$ with $i = 1, 2, 3$ are symmetric with respect to a permutation of the spin variables, and $W_4(s_3^{(1)}, s_3^{(2)})$ is an antisymmetric function. Then it follows from the antisymmetry of the total wave function that

$$\begin{aligned} \Psi_i(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) &= -\Psi_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), \quad i = 1, 2, 3, \\ \Psi_4(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) &= \Psi_4(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), \end{aligned}$$

that is, the coordinate wave functions for the states with spin 1 are antisymmetric, and those for states with spin zero are symmetric.

Let us apply this result to the helium atom. The Schrödinger operator for a helium atom without considering spin interactions has the form⁵⁴

$$H = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}}.$$

If $\Psi(\xi_1, \xi_2)$ is a solution of the equation

$$H\Psi(\xi_1, \xi_2) = E\Psi(\xi_1, \xi_2),$$

then the coordinate functions $\Psi_i(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ also satisfy the Schrödinger equation with the same eigenvalue E . Therefore, the problem reduces to finding solutions of the equation

$$(2) \quad H\Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = E\Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$$

in subspaces of symmetric or antisymmetric functions. It is clear that there are fewer solutions of (2) in each such subspace than in the space $L^2(\mathbf{R}^6)$. The values of E for which (2) has a solution $\Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$

⁵⁴In a precise formulation of the problem, the Schrödinger operator contains terms depending on the spin, but it is possible to derive an expression for the spin interaction only in relativistic quantum mechanics. Moreover, for a helium atom these terms play the role of small corrections and are always taken into account according to perturbation theory.

in the subspace of antisymmetric functions correspond to states with spin 1, and the values of E for which there are symmetric solutions of (2) correspond to states with spin zero.

We see that the energy levels of a helium atom depend on the total spin even when we ignore the spin interactions in the Schrödinger operator. This dependence is a consequence of the identity principle and arises through the symmetry of the coordinate wave functions.

It can be shown that a symmetric coordinate wave function corresponds to the ground state of a helium atom: that is, the spin of the helium atom in the ground state is equal to zero.

It is interesting to note that transitions with emission or absorption of quanta between states with $S = 0$ and $S = 1$ turn out to have small probability. Therefore, the optical spectrum of helium is as if there are two sorts of helium, with $S = 0$ and $S = 1$. The first sort of helium is called parahelium, and the second sort is called orthohelium. Corresponding to each energy level of parahelium is the single spin state W_4 , and corresponding to an energy level of orthohelium are the three spin states W_1 , W_2 , W_3 . Therefore, the states of parahelium are called singlet states, and those of orthohelium are called triplet states. Taking into account the spin interactions leads to a splitting of the triplet energy levels into three nearby levels.⁵⁵

§ 50. Multi-electron atoms. One-electron approximation

The Schrödinger operator of a multi-electron atom or ion with charge Z of the nucleus has the form

$$H = -\frac{1}{2} \sum_{i=1}^n \Delta_i - \sum_{i=1}^n \frac{Z}{r_i} + \sum_{i < j}^n \frac{1}{r_{ij}}.$$

This operator is written neglecting the motion of the nucleus and the spin interactions. Neglecting the motion of the nucleus of a complex atom is completely valid, since the corrections arising when it is taken into account are less by several orders of magnitude than the errors

⁵⁵ It can be shown that such a splitting takes place if the total orbital angular momentum L of the two electrons is nonzero, and therefore it is not for all triplet energy levels that splitting is observed.

in contemporary methods of computation for complex atoms. The spin interactions are almost always taken into account according to perturbation theory, and their role will be discussed below.

The operator H corresponds to a neutral atom when $Z = n$, to a positive ion when $Z > n$, and to a negative ion when $Z < n$. The state space of a multi-electron atom or ion is the space \mathcal{H}_A of antisymmetric functions $\Psi(\xi_1, \dots, \xi_n)$. As in the already analyzed case of two particles, this space decomposes into a direct sum of spaces with a definite total spin. The corresponding coordinate functions satisfy certain symmetry conditions that are more complex than in the case of two electrons, and we do not describe them here. The problem of finding the spectrum of H acting in \mathcal{H}_A reduces to the problems of the spectrum of H acting in the subspaces \mathcal{H}_k of coordinate functions with certain symmetry conditions.

The spectrum of H acting in the subspaces \mathcal{H}_k has been thoroughly studied for $Z \geq n$. In this case the spectrum is continuous on an interval $-\mu \leq E < \infty$ with some generally positive μ and consists of an infinite series of negative eigenvalues accumulating at $-\mu$. The values of μ for different subspaces can be different. Further, the discrete spectrum for one of these subspaces can be superimposed on the continuous spectrum for another. Thus, on part of the negative semi-axis the spectrum of H is mixed, and eigenvalues lie on the continuous spectrum.

The case $Z < n$ has been investigated less thoroughly. Examples show that the discrete spectrum can be finite or it can be absent. The absence of a discrete spectrum means that such a negative ion does not have stable states. For example, it is known from experiments that there is a stable state of a negative ion of hydrogen ($Z = 1, n = 2$). Only the finiteness of the discrete spectrum for such a system has been proved rigorously.

The problem of constructing eigenfunctions of H for a multi-electron atom is extremely complex, since the equation $H\Psi = E\Psi$ does not admit separation of variables. This complexity is connected with the last term in H , which takes into account the interaction between the electrons. The attempt to take this term into account

according to perturbation theory leads to poor results, since the interaction between the electrons has the same order as the interaction of the electrons with the nucleus.

One simple method for approximately taking into account the interaction between the electrons is to replace the last term in H by the sum $\sum_{i=1}^n V(r_i)$, where the potential $V(r)$ can be interpreted as the potential of the interaction of an electron with the charge of the remaining electrons, distributed over the volume of the atom. The potential $V(r)$ is said to be self-consistent, since it depends on the state of the atom, which in turn depends on $V(r)$.

Replacing the last term in H by $\sum_{i=1}^n V(r_i)$ gives the approximate Schrödinger operator

$$H' = -\frac{1}{2} \sum_{i=1}^n \Delta_i + \sum_{i=1}^n U(r_i),$$

where $U(r) = -Z/r + V(r)$. A priori, we know about the potential $U(r)$ only that

$$U(r) \cong -\frac{Z}{r}, \quad r \rightarrow 0, \quad U(r) \cong -\frac{1}{r}, \quad r \rightarrow \infty.$$

In what follows we shall see that there are methods enabling us to find this potential. However, to understand the structure of complex atoms it suffices to assume that H can be fairly well approximated by the operator H' , and the difference $H' - H = W$ can be taken into account by the perturbation method.

The approximate Schrödinger equation

$$(1) \quad H' \Psi(\xi_1, \dots, \xi_n) = E \Psi(\xi_1, \dots, \xi_n)$$

admits separation of variables. Let us look for a solution of this equation in the form

$$(2) \quad \Psi(\xi_1, \dots, \xi_n) = \psi_1(\xi_1) \cdots \psi_n(\xi_n).$$

Substitution of (2) in (1) shows that if $\psi_1(\xi_1), \dots, \psi_n(\xi_n)$ are eigenfunctions of the Schrödinger operator for a particle in a central field $U(r)$,

$$(3) \quad -\frac{1}{2} \Delta \psi_k(\xi) + U(r) \psi_k(\xi) = E_k \psi_k(\xi),$$

then the function (2) is an eigenfunction of the operator H' corresponding to the eigenvalue $E = E_1 + \dots + E_n$. However, the function (2) does not satisfy the identity principle. The total wave function must be antisymmetric. We remark that if Ψ satisfies (1), then $P_\pi \Psi$ is also an eigenfunction of H' , with the same eigenvalue E . Therefore, as an eigenfunction satisfying the identity principle, we should take the antisymmetric combination of the functions $P_\pi \Psi$, that is,

$$(4) \quad \Psi_A(\xi_1, \dots, \xi_n) = C \begin{vmatrix} \psi_1(\xi_1) & \dots & \psi_1(\xi_n) \\ \dots & \dots & \dots \\ \psi_n(\xi_1) & \dots & \psi_n(\xi_n) \end{vmatrix},$$

where C is a constant found from the normalization condition. The function (4) is antisymmetric, because a permutation of the coordinates is equivalent to a permutation of the columns of the determinant.

The functions (2) and (4) can be interpreted as follows. The function (2) describes a state of a system of electrons in which the first electron is in the one-electron state ψ_1 , the second electron is in the state ψ_2 , and so on. The function (4) corresponds to a state of the system in which the n electrons fill n one-electron states, and it does not make sense to speak of which electron is in which state. We remark that the construction of the function (4) is not necessarily possible for an arbitrary solution (2). The determinant in (4) is nonzero only under the condition that no two of the functions ψ_1, \dots, ψ_n are the same. This result is called the Pauli principle, and it can be formulated as follows: no more than one electron can be in the same one-electron state.

The representation of the state of an individual electron is connected with the one-electron approximation. For an arbitrary state of the system the wave function cannot be represented in the form (2) or (4), and the concept of the state of a single electron becomes meaningless. There is a formulation of the Pauli principle that is not connected with the one-electron approximation, but in the general formulation the Pauli principle loses its clearness, and we shall not present this formulation. We remark that the Pauli principle is a consequence of the antisymmetry of the wave function and is correct only for fermions.

Let us consider the question of classifying the energy levels of a multi-electron atom. The exact Schrödinger operator for the atom can be written in the form

$$H = H' + W_C + W_S,$$

where

$$W_C = \sum_{i < j}^n \frac{1}{r_{ij}} - \sum_{i=1}^n V(r_i),$$

and W_S describes the spin interactions. We do not need the explicit form of the operator W_S . Calculations for atoms show that the corrections introduced by the operators W_C and W_S can be found with good accuracy by using perturbation theory, and for the atoms in the first half of the periodic table the main contribution is given by the corrections from W_C . Therefore, the possibility arises of a repeated application of perturbation theory; that is, as the unperturbed operator we can first take H' and regard the operator W_C as a perturbation, and then we can regard W_S as a perturbation of the operator $H' + W_C$.

The operator H' has a very rich collection of symmetries, and therefore its eigenvalues usually have fairly large multiplicity. Let us consider why the multiplicity arises. Without spin taken into account, the eigenfunctions of the equation (3) are classified by the three quantum numbers n, l, m . The spin states can be taken into account by introducing a spin quantum number m_s taking the two values $\pm 1/2$ according to the formula

$$U_{m_s}(s_3) = \begin{cases} U_+(s_3), & m_s = \frac{1}{2}, \\ U_-(s_3), & m_s = -\frac{1}{2}. \end{cases}$$

Then the eigenfunctions of (3) can be written in the form

$$(5) \quad \psi_{nlmm_s}(\mathbf{x}, s_3) = \psi_{nlm}(\mathbf{x}) U_{m_s}(s_3).$$

The eigenvalues of (3) depend only on the quantum numbers n and l . Therefore, the eigenvalue E of the operator H' depends on the set of quantum numbers n and l for all the electrons. This set of quantum numbers n and l is called the configuration of the atom. To express the configuration it is customary to assign the letters s, p, d, f, \dots to the values $l = 0, 1, 2, 3, \dots$. Then the one-electron state (5) with

$n = 1, l = 0$ is called the $1s$ -state, the state with $n = 2, l = 1$ is called the $2p$ -state, and so on. For the lithium atom, for example, possible configurations are $(1s)^2 2s$, $(1s)^2 2p$, $1s 2s 2p$, In the first of these configurations two electrons are in the $1s$ -state and one is in the $2s$ -state.

A collection of states with the same quantum numbers n and l is called a shell. The states of the electrons in a single shell differ by the quantum numbers m and m_s . For a shell with quantum number l the number of such states is equal to $2(2l + 1)$, since m takes the values $-l, -l + 1, \dots, l$ and $m_s = \pm 1/2$. A shell in which all $2(2l + 1)$ states are occupied is said to be filled. A completely determined set of functions (5) corresponds to a filled shell. For an unfilled shell a different choice of functions (5) is possible, differing by the numbers m and m_s . Therefore, the multiple eigenvalues of H' turn out to be those corresponding to configurations containing unfilled shells. For example, the multiplicity of an eigenvalue corresponding to the configuration $(1s)^2 (2p)^2$ is equal to $\binom{6}{2} = 15$, since there are six $2p$ -states, which are occupied by two electrons.

The corrections from the perturbation W_C can be found upon diagonalization of the matrix of this perturbation, which is constructed with the help of all possible functions (4) corresponding to the given configuration. However, such a diagonalization is usually not necessary. In the theory of atomic spectra it is proved that if one replaces the eigenfunctions (4) for a given configuration by linear combinations of them which are eigenfunctions of the square L^2 of the total orbital angular momentum operator, the square S^2 of the total spin angular momentum operator, and the operators L_3 and S_3 ,

$$L^2\Psi = L(L+1)\Psi,$$

$$S^2\Psi = S(S+1)\Psi,$$

$$L_3\Psi = M_L\Psi,$$

$$S_3\Psi = M_S\Psi,$$

then the perturbation matrix turns out to be diagonal with respect to the quantum numbers L, S, M_L, M_S , and its elements do not depend on the numbers M_L and M_S . Thus, a degenerate energy level corresponding to a configuration K splits into several levels corresponding

to the possible values of L and S . A collection of $(2M_L + 1)(2M_S + 1)$ states with a given configuration and given numbers L and S is called a *term*.⁵⁶

Taking the perturbation W_S into account can be done for each term separately. Here it turns out that the energy level corresponding to a term splits into several nearby levels. The collection of these levels is called a multiplet. It can be shown that the states corresponding to different levels of a multiplet differ by the quantum number J . This number characterizes the eigenvalues of the square of the total angular momentum, which is the sum of the total orbital angular momentum and the total spin angular momentum.

Finally, corresponding to each level of a multiplet are several states differing by the projection M_J of the total angular momentum. This degeneracy can be removed by putting the atom in a magnetic field.

We see that the classification of the energy levels of a complex atom (the configuration, L , S , J) corresponds to the hierarchy of summands H' , W_C , W_S of the Schrödinger operator.

§ 51. The self-consistent field equations

Though the approach in § 50 to the study of the spectrum of complex atoms makes it possible to understand the classification of energy levels, it is not convenient for practical calculations. Most efficient for our purpose is the self-consistent field method (the Hartree-Fock method), which is based on the use of the variational principle. In this section we discuss the main ideas of this method.

The Hartree-Fock method is also based on the one-electron approximation. The wave function of a complex atom is approximated

⁵⁶If there are more than two electrons in unfilled shells, then several terms with the same configuration and the same L and S can appear. In this case the matrix of the perturbation W_C is quasi-diagonal, and to compute the corrections of the first approximation, the matrix has to be diagonalized.

by a product $\Psi = \psi_1(\xi_1) \cdots \psi_n(\xi_n)$ of one-electron functions (the identity principle is not taken into account here), or by a determinant

$$\Psi = C \begin{vmatrix} \psi_1(\xi_1) & \dots & \psi_1(\xi_n) \\ \dots & \dots & \dots \\ \psi_n(\xi_1) & \dots & \psi_n(\xi_n) \end{vmatrix},$$

or by a linear combination of such determinants. From the condition of stationarity of the functional $(H\Psi, \Psi)$ under the additional condition $(\Psi, \Psi) = 1$, we obtain a system of integro-differential equations for the one-particle functions $\psi_1(\mathbf{x}), \dots, \psi_n(\mathbf{x})$.

Let us illustrate this approach with the example of the helium atom, for which the Schrödinger operator has the form

$$(1) \quad H = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}} = H_1 + H_2 + \frac{1}{r_{12}},$$

where

$$H_i = -\frac{1}{2}\Delta_i - \frac{2}{r_i}, \quad i = 1, 2.$$

We look for an approximate wave function in the form

$$\Psi(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \psi_1(\mathbf{x}^{(1)}) \psi_2(\mathbf{x}^{(2)}).$$

We remark that the condition

$$(\Psi, \Psi) = \int_{\mathbf{R}^6} |\psi_1(\mathbf{x}^{(1)}) \psi_2(\mathbf{x}^{(2)})|^2 d\mathbf{x}^{(1)} d\mathbf{x}^{(2)} = 1$$

can be replaced by the two conditions

$$(2) \quad \int_{\mathbf{R}^3} |\psi_1(\mathbf{x})|^2 d\mathbf{x} = 1, \quad \int_{\mathbf{R}^3} |\psi_2(\mathbf{x})|^2 d\mathbf{x} = 1,$$

which does not narrow down the class of functions to be varied. The functional $(H\Psi, \Psi)$ has the form

$$(3) \quad (H\Psi, \Psi) = \int_{\mathbf{R}^3} \bar{\psi}_1 H_1 \psi_1 d\mathbf{x}^{(1)} \int_{\mathbf{R}^3} |\psi_2|^2 d\mathbf{x}^{(2)} \\ + \int_{\mathbf{R}^3} \psi_2 H_2 \psi_2 d\mathbf{x}^{(2)} \int_{\mathbf{R}^3} |\psi_1|^2 d\mathbf{x}^{(1)} + \int_{\mathbf{R}^6} \frac{|\psi_1|^2 |\psi_2|^2}{r_{12}} d\mathbf{x}^{(1)} d\mathbf{x}^{(2)}.$$

Varying this functional with respect to the functions ψ_1 and ψ_2 under the conditions (2) and using the Lagrange method of undetermined multipliers, we have

$$(4) \quad \begin{aligned} H_1\psi_1 + \left(\int_{\mathbf{R}^3} \frac{|\psi_2|^2}{r_{12}} d\mathbf{x}^{(2)} \right) \psi_1 &= \varepsilon_1 \psi_1, \\ H_2\psi_2 + \left(\int_{\mathbf{R}^3} \frac{|\psi_1|^2}{r_{12}} d\mathbf{x}^{(1)} \right) \psi_2 &= \varepsilon_2 \psi_2. \end{aligned}$$

We have obtained a system of nonlinear integro-differential equations for the functions $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$. The equations (4) admit a very simple physical interpretation. For example, the first equation can be regarded as the Schrödinger equation for the first electron in the field of the nucleus and in the field created by the charge of the second electron. This charge is as if it were spread out over the volume of the atom with density $|\psi_2(\mathbf{x}^{(2)})|^2$ (recall that the charge of the electron is $e = 1$), and the integral in the second term in (4) is the potential of this volume distribution of charge. We note that this potential is unknown and will be found in solving the system (4). Moreover, in contrast to the model in the previous section, each electron is in its own potential field, which depends on the state of the other electron.

The equations (4) were first proposed by Hartree, who wrote them starting from the physical considerations given above. Fock established a connection between these equations and the variational principle, and he proposed a refinement of the Hartree method that takes into account the identity principle. The Fock equations turn out to be somewhat more complex, and some of the “potentials” (exchange potentials) in them no longer admit such a simple physical interpretation. These potentials arise because of symmetry properties of the wave function.

In practical calculations the Fock method is usually employed from the very start to find the one-electron functions $\psi_k(\xi)$ in the form of functions $\psi_{nlmm_s}(\xi) = \frac{R_{nl}(r)}{r} Y_{lm}(\mathbf{n}) U_{m_s}(s_3)$ of the central field.

The main steps of calculation consist in the following. An expression is first found for the wave function of a definite term in the

form of a linear combination of determinants. Then an expression is formed for the functional ($H\Psi, \Psi$). Finally, this functional is varied with respect to the radial functions $R_{nl}(r)$ (the techniques of all these operations have been worked out in detail). The result is a system of integro-differential equations for functions of one variable. The number of unknown functions in this system is equal to the number of shells in the particular configuration of the atom.

Comparison of the computational results with experimental data shows that the accuracy of computation of the energy levels of light atoms by the self-consistent field method is about 5%.

§ 52. Mendeleev's periodic system of the elements

The periodic law was discovered by Mendeleev in 1869 and is one of the most important laws of nature. Mendeleev based this system on the fact that if one arranges the elements in order of increasing atomic weight, then elements with closely related chemical and physical properties occur periodically.

At the time of discovery of the periodic law only 63 elements were known, the atomic weights of many elements had been determined incorrectly, and Mendeleev had to change them.⁵⁷ He left a series of cells in the table empty, considering that they corresponded to yet undiscovered elements. He predicted the properties of three such elements with astonishing accuracy. Finally, in several cases Mendeleev deviated from a strict arrangement of the elements in order of increasing atomic weight and introduced the concept of the atomic number Z .

The law discovered by Mendeleev was originally purely empirical. In his time there was no explanation for the periodicity of the properties of the elements; indeed, there could not be, since the electron was discovered by Thomson in 1897, and the atomic nucleus by Rutherford in 1910.

⁵⁷For example, the atomic weight of Cerium was reckoned to be 92, but Mendeleev assigned the value 138 to it (the modern value is 140)

An explanation of the periodic law in full scope is a complicated problem in quantum chemistry, but it is possible to understand the periodicity of the properties of the elements already in the framework of the simplified model of the atom described in § 50. We recall that the effective potential $V(r)$ for an electron in an atom is not the Coulomb potential, and the eigenvalues of the one-electron Schrödinger operator depend on the quantum numbers n and l . Computations show that for a typical effective potential of an atom, the eigenvalues E_{nl} increase in the order

$$1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, 5s, 4d, 5p, \dots$$

It is in this order that the electron shells of the atom fill up (in this section we consider only the ground states of atoms). However, this order is not strict, since each element has its own effective potential, and for some elements slight deviations are possible in the order in which the shells are filled.

To understand the principle according to which the elements are divided into periods, we consider the following feature of the d - and f -electrons which distinguishes them from the s - and p -electrons. Computations show that the density $|\psi(\mathbf{x})|^2$ of the coordinate distribution function for d - and f -electrons is concentrated in regions of smaller size than for s - and p -electrons with close energies. This means that on the average the d - and f -electrons are considerably closer to the nucleus than the s - and p -electrons. Therefore, the elements in which the d - and especially the f -shells are filling have similar chemical properties. (The chemical properties depend mainly on the states of the peripheral electrons of the atom. An explanation of this assertion is provided by the quantum theory of valence.)

The first element of each period of the Mendeleev table is an element for which an s -shell begins to fill. All these elements, with the exception of hydrogen, are alkali metals. The last element of each period is an element for which a p -shell has been completely filled (an exception is the first period, whose elements, hydrogen and helium, do not have p -electrons). The last elements of the periods are the noble gases. The configurations of atoms of noble gases consist of filled

shells. It can be shown that the corresponding states transform under rotations according to the irreducible representation D_0 , that is, are spherically symmetric. The same property is enjoyed by the singly charged positive ions of alkali metals, whose configurations coincide with the configurations of the preceding noble gas. This is the reason that the optical properties of alkali metals are well described by the model of a valence electron in a central field.

The sequence of electron shells is divided into periods as follows:

I	$1s.$	2 elements
II	$2s, 2p.$	8 elements
III	$3s, 3p.$	8 elements
IV	$4s, 3d, 4p,$	18 elements
V	$5s, 4d, 5p,$	18 elements
VI	$6s, 4f, 5d, 6p,$	32 elements
VII	$7s, 6d, 5f, \dots,$	

The number of a period is denoted by a roman numeral. The number of elements in the period is indicated to the right. This number is easy to determine if we recall that the number of one-electron states with quantum numbers n and l is equal to $2(2l+1)$. For example, the number of elements in the sixth period is $2 + 14 + 10 + 6 = 32$. The number of a period coincides with the principal quantum number of the s - and p -shells which are being filled.

The elements containing filled d - and f -shells (or not containing such shells) are called elements of the principal groups. The elements in which d - and f -shells are being filled are called elements of the intermediate groups. The elements of the intermediate groups have close chemical properties; this is especially true for the groups in which f -shells are filling (the shell $4f$ is filling in the group of rare-earth elements).

We remark that when d - and f -shells are filling, there is sometimes a deviation from the order indicated above. For example, the configurations of the ground states of the atoms V, Cr, and Mn are $(3d)^3(4s)^2$, $(3d)^54s$, and $(3d)^5(4s)^2$, respectively (we have not written out the configuration of the filled shells, which are the same for these three atoms: $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6$). It turns out that the configuration $(3d)^54s$ for the atom Cr is energetically more advantageous

than the configuration $(3d)^4(4s)^2$ which we would have assigned to chromium guided by our model. It is clear that this deviation cannot be explained in the framework of a model in which the effective potential is the same for all the electrons.

In conclusion we consider the variation in the binding energy⁵⁸ of an electron in an atom. This parameter is closely connected with the chemical properties of the elements, and characterizes the capacity of an atom to "give up" an electron when entering into a chemical combination.

We recall that for an atom the effective potential satisfies

$$U(r) \cong -1/r \text{ as } r \rightarrow \infty$$

and

$$U(r) \cong -Z/r \text{ as } r \rightarrow 0.$$

Therefore, it is natural to expect (and this is confirmed by computations) that with increasing Z the effective potential becomes stronger; that is, the corresponding one-electron energy levels for the next atom lie deeper than for the previous atom. Thus, the binding energy must increase in a series of elements corresponding to the filling of some shell. However, upon passage to the filling of the next shell, the binding energy can fall. An especially strong fall of the binding energy is observed at the beginning of each period, when a new s -shell begins to fill. Figure 15 shows experimental values of the binding energy for various elements. The atomic number Z of an element is plotted along the horizontal axis. From the figure it is clear that for each period the alkali metals have the minimal value of the binding energy and the noble gases have the maximal value. The chemical inertness of the noble gases is explained to a significant degree by the large value of the binding energy.

⁵⁸The binding energy is defined as the absolute value of the difference between the energy of the ground state of an atom and the energy of the ground state of the corresponding singly charged positive ion; in other words, the binding energy is equal to the minimal work required to remove an electron from the atom

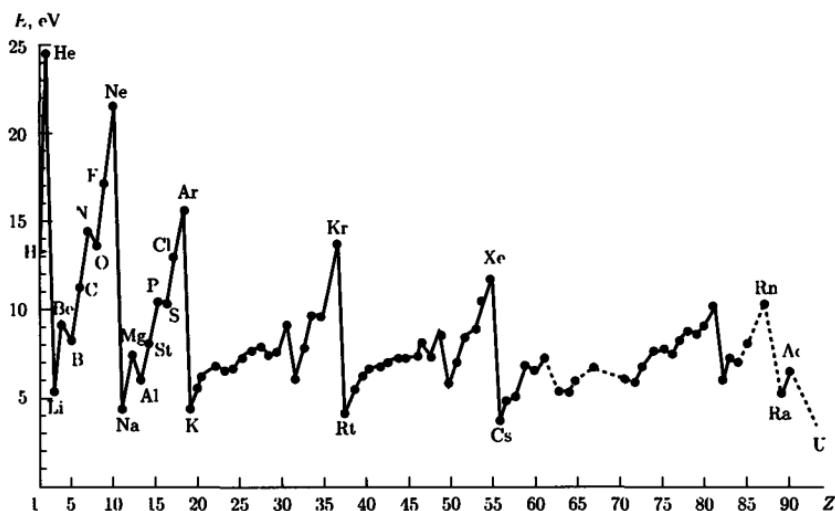


Figure 15

Appendix

Lagrangian Formulation of Classical Mechanics

For the convenience of the reader we present here, in a succinct form, the Lagrangian formulation of classical mechanics. For the detailed exposition of these standard facts we refer to the classic texts by R. Abraham and J. E. Marsden, *Foundation of Mechanics* (Addison-Wesley, Reading, MA, 1978; Amer. Math. Soc., Providence, RI, 2008), and V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, New York, 1989).

Classical mechanics describes systems of finitely many interacting particles. The position of a system in space is specified by the positions of its particles and defines a point in a smooth, finite-dimensional manifold M , the configuration space of the system. Coordinates $\mathbf{q} = (q_1, \dots, q_n)$ on M are called generalized coordinates of the system, and the dimension $n = \dim M$ is called the number of degrees of freedom. For the system of N interacting particles, considered in §1, the generalized coordinates are the Cartesian coordinates in \mathbb{R}^{3N} . The motion is described by the classical trajectory $\mathbf{q}(t) = (q_1(t), \dots, q_n(t))$ in the configuration space M , and the corresponding time derivatives $\dot{q}_i(t)$ are called generalized velocities and form a tangent vector $\dot{\mathbf{q}}(t) = (\dot{q}_1(t), \dots, \dot{q}_n(t))$ to the manifold M at the point $\mathbf{q}(t)$. The state of a classical system at any instant of time

is described by a point \mathbf{q} in M and by a tangent vector \mathbf{v} to M at this point.

It is a basic fact in classical mechanics that the motion of a closed system is completely described by the Lagrangian function L – a function $L(\mathbf{q}, \dot{\mathbf{q}})$ of the generalized coordinates and velocities – a smooth real-valued function on the tangent bundle TM to the manifold M . For mechanical systems describing the interacting particles considered in §1, $L = T - V$ – the difference between total kinetic energy and total potential energy of the system. The general principle of the least action was formulated by Hamilton (1834). It states that the classical trajectory which describes the motion between a point \mathbf{q}_0 and at time t_0 and a point \mathbf{q}_1 at time t_1 is the extremum of the action functional

$$(1) \quad S(\mathbf{q}(t)) = \int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}) dt$$

on the space of all smooth paths $\mathbf{q}(t)$ such that $\mathbf{q}(t_0) = \mathbf{q}_0$ and $\mathbf{q}(t_1) = \mathbf{q}_1$.

Finding extrema of the action functional (1) is a classical problem of the calculus of variations.¹ Namely, let $\delta\mathbf{q}(t) = (\delta q_1(t), \dots, \delta q_n(t))$ be the infinitesimal variation of the classical trajectory $\mathbf{q}(t)$ – a smooth function on the interval $[t_0, t_1]$ satisfying $\delta\mathbf{q}(t_0) = \delta\mathbf{q}(t_1) = 0$. Since $\mathbf{q}(t)$ is the extremum of the action functional S , its infinitesimal variation at $\mathbf{q}(t)$ vanishes: that is, for all $\delta\mathbf{q}(t)$

$$(2) \quad \delta S = S(\mathbf{q}(t) + \delta\mathbf{q}(t)) - S(\mathbf{q}(t)) = 0$$

up to the terms quadratic in $\delta\mathbf{q}(t)$. Keeping only linear terms in $\delta\mathbf{q}(t)$ and $\delta\dot{\mathbf{q}}(t)$ and using integration by parts, we can rewrite condition (2)

¹For the introduction to this mathematical discipline, we refer the reader to the classic textbook by I M Gelfand and S V Fomin. *Calculus of Variations* (Dover, New York, 2000)

as

$$\begin{aligned} 0 = \delta S &= \sum_{i=1}^n \int_{t_0}^{t_1} (L(\mathbf{q}(t) + \delta\mathbf{q}(t), \dot{\mathbf{q}}(t) + \delta\dot{\mathbf{q}}(t)) - L(\mathbf{q}(t), \dot{\mathbf{q}}(t)) dt \\ &= \sum_{i=1}^n \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) dt \\ &= \sum_{i=1}^n \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i dt + \sum_{i=1}^n \frac{\partial L}{\partial \dot{q}_i} \delta q_i \Big|_{t_0}^{t_1}. \end{aligned}$$

The second sum in the last line vanishes due to the property $\delta\mathbf{q}(t_0) = \delta\mathbf{q}(t_1) = 0$, and we obtain that

$$\sum_{i=1}^n \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i dt = 0$$

for arbitrary smooth functions δq_i on the interval $[t_0, t_1]$ which vanish at the endpoints. This implies that for each term in the sum, the integrand is identically zero, and we obtain that the classical trajectories — extrema of the action functional (1) — are described by the Euler–Lagrange equations

$$(3) \quad \frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \right) = 0.$$

When $L = T - V$ — the difference between the kinetic energy and the potential energy — the Euler–Lagrange equations reduce to Newton's equations. Thus for a particle in a force field with potential $V(\mathbf{x})$, considered in §1, we have

$$L = \frac{m}{2} \dot{\mathbf{x}}^2 - V(\mathbf{x}).$$

Denoting by

$$\mathbf{F} = \frac{\partial L}{\partial \mathbf{x}} = -\frac{\partial V}{\partial \mathbf{x}}$$

the force acting on the particle and using $\frac{\partial L}{\partial \dot{\mathbf{x}}} = m\dot{\mathbf{x}}$, we see that the Euler–Lagrange equations (3) give Newton's equations (1) in §1:

$$m\dot{\mathbf{v}} = \mathbf{F}, \quad \dot{\mathbf{x}} = \mathbf{v}.$$

In general,

$$(4) \quad \mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}})$$

are called generalized momenta, and introducing the generalized forces

$$\mathbf{F} = \frac{\partial L}{\partial \dot{\mathbf{q}}}$$

one can rewrite the Euler–Lagrange equations (3) in Newton’s form

$$\dot{\mathbf{p}} = \mathbf{F}.$$

The relation thus outlined between the Lagrangian formulation of classical mechanics and the Hamiltonian formulation, presented in §1, is given by the Legendre transform. Namely, suppose that instead of the coordinates $(\mathbf{q}, \dot{\mathbf{q}})$ one can use the coordinates (\mathbf{q}, \mathbf{p}) , where \mathbf{p} are the generalized momenta given by (4). By the implicit function theorem, locally this amounts to the condition that the corresponding Hessian matrix

$$\frac{\partial^2 L}{\partial \dot{\mathbf{q}} \partial \dot{\mathbf{q}}} = \left\{ \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \right\}_{i,j=1}^n$$

is non-degenerate. and such a Lagrangian function is called non-degenerate. The corresponding transformation $(\mathbf{q}, \dot{\mathbf{q}}) \mapsto (\mathbf{q}, \mathbf{p})$ is called the Legendre transformation.² When the Legendre transformation is globally defined and invertible, one can introduce the so-called Hamiltonian function $H(\mathbf{q}, \mathbf{p})$ by

$$H(\mathbf{q}, \mathbf{p}) = \mathbf{p}\dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}),$$

where $\dot{\mathbf{q}}$ is expressed in terms of \mathbf{q} and \mathbf{p} by using (4) and the implicit function theorem. Then the Euler–Lagrange equations (3) are equivalent to Hamilton’s equations

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}.$$

For the systems with the Lagrangian function $L = T - V$, derivation of these results is straightforward. Thus, for the Lagrangian function

$$L = \sum_{i=1}^N \frac{m_i}{2} \dot{q}_i^2 - \sum_{i < j}^N V_{ij}(\mathbf{x}_i - \mathbf{x}_j) - \sum_{i=1}^N V_i(\mathbf{x}_i)$$

the corresponding Hamiltonian function is given by formula (5) in §1, and Hamilton’s equations immediately follow from Newton’s equations.

²One needs to impose further conditions to ensure that the Legendre transformation is a diffeomorphism between tangent and cotangent bundles to M

Titles in This Series

- 47 **L. D. Faddeev and O. A. Yakubovskii**, Lectures on quantum mechanics for mathematics students, 2009
- 46 **Anatole Katok and Vaughn Climenhaga**, Lectures on surfaces: (Almost) everything you wanted to know about them, 2008
- 45 **Harold M. Edwards**, Higher arithmetic An algorithmic introduction to number theory, 2008
- 44 **Yitzhak Katznelson and Yonatan R. Katznelson**, A (terse) introduction to linear algebra, 2008
- 43 **Ilka Agricola and Thomas Friedrich**, Elementary geometry, 2008
- 42 **C. E. Silva**, Invitation to ergodic theory, 2007
- 41 **Gary L. Mullen and Carl Mummert**, Finite fields and applications, 2007
- 40 **Deguang Han, Keri Kornelson, David Larson, and Eric Weber**, Frames for undergraduates, 2007
- 39 **Alex Iosevich**, A view from the top Analysis, combinatorics and number theory, 2007
- 38 **B. Fristedt, N. Jain, and N. Krylov**, Filtering and prediction. A primer, 2007
- 37 **Svetlana Katok**, p -adic analysis compared with real, 2007
- 36 **Mara D. Neusel**, Invariant theory, 2007
- 35 **Jörg Bewersdorff**, Galois theory for beginners: A historical perspective, 2006
- 34 **Bruce C. Berndt**, Number theory in the spirit of Ramanujan, 2006
- 33 **Rekha R. Thomas**, Lectures in geometric combinatorics, 2006
- 32 **Sheldon Katz**, Enumerative geometry and string theory, 2006
- 31 **John McCleary**, A first course in topology Continuity and dimension, 2006
- 30 **Serge Tabachnikov**, Geometry and billiards, 2005
- 29 **Kristopher Tapp**, Matrix groups for undergraduates, 2005
- 28 **Emmanuel Lesigne**, Heads or tails An introduction to limit theorems in probability, 2005
- 27 **Reinhard Illner, C. Sean Bohun, Samantha McCollum, and Thea van Roode**, Mathematical modelling. A case studies approach, 2005
- 26 **Robert Hardt, Editor**, Six themes on variation, 2004
- 25 **S. V. Duzhin and B. D. Chebotarevsky**, Transformation groups for beginners, 2004
- 24 **Bruce M. Landman and Aaron Robertson**, Ramsey theory on the integers, 2004
- 23 **S. K. Lando**, Lectures on generating functions, 2003
- 22 **Andreas Arvanitoyeorgos**, An introduction to Lie groups and the geometry of homogeneous spaces, 2003