Alexander Holevo

Probabilistic and Statistical Aspects of Quantum Theory

EDIZIONI
DELLA
NORMALE

1

QUADERNI

MONOGRAPHS

Alexander Holevo
Steklov Mathematical Institute
Moscow

Probabilistic and Statistical Aspects of Quantum Theory

Alexander Holevo

Probabilistic and Statistical Aspects of Quantum Theory

EDIZIONI
DELLA
NORMALE

ISBN: 978-88-7642-375-8

Contents

Foreword to the second English edition

In the present edition the *comments to chapters* in the main text and the References were extended and updated. During the past years the quantum (noncommutative) generalizations of probability, mathematical statistics and information theory were substantially developed. Theory of quantum stochastic processes was elaborated, unifying repeated and continuous measurement with the dynamics of open quantum systems. This material is reflected in the book [178] where the reader can find an extended bibliography. With the emergence of ideas of quantum computation the powerful impetus got the quantum information theory which was born more than half-century ago and shaped as an independent scientific discipline in the 1990-s. This progress enhanced, in particular, recent development of asymptotic methods of quantum estimation theory. An introduction to this circle of problems can be found in the books [152, 170, 179].

The present edition would not be possible without the enthusiasm and perseverance of Professor Vittorio Giovannetti and Professor Rosario Fazio to whom the author expresses his warm gratitude. The author is grateful to Dottoressa Luisa Ferrini, *Edizioni della Normale*, for her professional and most efficient assistance in preparation of the manuscript.

This book is dedicated to the memory of Academician Kamil A. Valiev who passed away in the summer 2010.

Alexander Holevo
Moscow, September 2010.

Foreword to the second Russian edition

When this book was first published in 1980 (the English edition appeared in 1982 in North Holland), the author addressed it to a broad audience of readers, both mathematicians and physicists having intention to make them acquainted with the new prospects and possibilities which emerge from the interaction of ideas of the mathematical statistics and the quantum theory. During the past period this approach became even more demanded. On one side, its advantages in the questions of foundations of quantum theory related to quantum measurements became more apparent and widely acknowledged. On the other hand, one should stress that these theoretical findings were not an end in itself: in modern high-precision physical experiments researchers become able to operate with single ions, atoms and photons which leads to potentially important applications, such as quantum communications, computation and cryptography. Of great importance is the question of extraction of maximal possible information from the state of a given quantum system. For example, in currently discussed proposals for quantum computing information is written in the states of elementary quantum memory cells, qubits, and then read off by means of quantum measurement. From the statistical point of view, measurement gives an estimate for the quantum state – as a whole, or for some of its parameters. In this way a new interest emerges to quantum estimation theory, the fundamentals of which are presented in this book.

One of essential consequences of penetration of the ideas of mathematical statistics into the theory of quantum measurement is the wide use of the mathematical notion of (non-orthogonal) resolution of the identity in the system Hilbert space (in the Western literature – POVM, positive operator-valued measures), describing the statistics of decision procedures. During the time passed resolutions of the identity became a standard tool both in mathematical and in physical literature on quantum measurements. All this, in my opinion, justifies publication of the second Russian edition of the book, moreover as the first one became a rare book.

The present edition includes Supplement discussing in some detail the problem of hidden variables in quantum mechanics which continues to provoke a keen interest. Moreover, the edition is complemented with Comments reflecting new results and achievements.

Academician K. A. Valiev
Moscow, April 2003.

Preface

The mathematical language of modern quantum mechanics is operator theory. Operators play there a role similar to functions in classical mechanics, probability theory and statistics. However, while the use of functions in classical theories is founded on premises which seem intuitively quite clear, in quantum theory the situation with operators is different.

Historically the "matrix mechanics" of Heisenberg and the "wave mechanics" of Schrödinger which gave rise to the contemporary form of quantum theory, originated from ingenious attempts to fit mathematical objects able to reflect some unusual (from the macroscopic point of view) features of microparticle behavior – in particular, a peculiar combination of continuous and discrete properties. The "probabilistic interpretation" developed later by Born and others elucidated the meaning of operator formalism by postulating rules connecting mathematical objects with observable quantities. However a good deal of arbitrariness remained in these postulates and the most convincing argument for quantum-theoretical explanations was still the "striking" coincidence of theoretical predictions with experimental data. This state of affairs gave rise to numerous attempts, on one hand, to find classical alternatives to quantum theory which would give an equally satisfactory description of the experimental data, and on the other hand, to find out physical and philosophical arguments for justifying the inevitability of the new mechanics.

Notwithstanding the impressive philosophical achievements in this field there was and still is a need for the structural investigation of quantum theory from a more mathematical point of view aimed at elucidating the connections between the entities of the physical world and the elements of operator formalism. The present book is essentially in this line of research opened by the classical von Neumann's treatise "Mathematical Foundations of Quantum Mechanics". However it differs from most subsequent investigations by the strong emphasis on the statistical rather than "logical" essence of quantum theory; it gives an account of recent

progress in the statistical theory of quantum measurement, stimulated by the new applications of quantum mechanics, particularly in quantum optics.

The first three chapters give an introduction to the foundations of quantum mechanics, addressed to the reader interested in the structure of quantum theory and its relations with classical probability. In spite of the mathematical character of the presentation it is not "axiomatic". Its purpose is to display the origin of the basic elements of operator formalism resting, as far as possible, upon the classical probabilistic concepts.

The present revision is not an end in itself – it emerged from the solution of concrete problems concerning the quantum limitations to measurement accuracy, arising in applications. So far there has been no general approach to such kind of problems. The methods of mathematical statistics adapted for classical measurements required radial quantum modification. The last chapters of the book are devoted to the recently developed quantum estimation theory, which is an analog of the corresponding branch of mathematical statistics.

We now give a more detailed account of the contents of the book. In Chapter 1 the general concepts of state and measurement are introduced on the basis of statistical analysis of an experimental situation. From the very beginning this approach leads to a substantial generalization of the Dirac-von Neumann concept of an observable. Mathematically it is reflected by the occurrence of arbitrary resolutions of identity in place of orthogonal ones (spectral measures) and the repudiation of self-adjointness as an indispensable attribute of an observable. In this way nonorthogonal resolutions of identity like the "overcomplete" system of coherent states known in physics for rather a long time find their proper place in quantum phenomenology. The new concept of quantum measurement is central for the whole book.

The notion of statistical model exploited in Chapter 1 is quite general and may find applications different from quantum theory. It gives us a new insight into the still controversial "hidden variables" problem.

In Chapter 2 the elements of operator theory in Hilbert space are introduced to provide mathematical background for the subsequent material. As compared to standard presentations relatively much attention is paid to nonorthogonal resolutions of identity and related questions. A novel feature is also the introduction of the $\mathscr{L}^2$ spaces of observables associated with a quantum state and playing a role similar to the Hilbert space of random variables with finite second moment in probability theory. These $\mathscr{L}^2$ spaces give the framework for a calculus of unbounded operators.

Of fundamental importance to quantum theory are groups of symmetries. In Chapter 3 elementary quantum mechanics is considered from this

point of view. An important result of this discussion is the isolation of the notion of covariant measurement which ties physical quantities with certain classes of resolutions of identity in the underlying Hilbert space. In this way we construct quantum measurements canonically corresponding to such quantities as time, phase of harmonic oscillator, angle of rotation and joint measurement of coordinate and velocity. Allowing the broader concept of quantum measurement enables us to resolve old troubles of quantum theory connected with the non-existence of self-adjoint operators having the required covariance properties.

Chapter 4 is devoted to a more advanced study of covariant measurements and extreme quantum limits for the accuracy of estimation of physical parameters. The latter problem becomes important in view of the progress in experimental physics. We present a unified statistical approach to "non-standard" uncertainty relations of the "angle-angular momentum" type. They appear to be related to the quantum analog of the Hunt-Stein theorem in mathematical statistics. A general conclusion which can be drawn from Chapter 4 is that the requirements of covariance and optimality, *i.e.*, extremal quantum accuracy, determine the canonical measurement of a "shift" parameter, such as angle, coordinate, time, uniquely up to a "gauge" transformation.

An example of a situation where quantum limitations are important is provided by optical communication. As it is known, "quantum noise" distorting the signal in the optical range can be much more significant than the thermal background radiation. As in ordinary communication theory the problem of signal estimation arises, but now it requires a specifically quantum-theoretic formulation and solution.

Chapter 5 is devoted to the so-called Gaussian states which, in particular, describe radiation fields in optical communication theory. The presentation is intended to make maximal use of the remarkable parallel with the Gaussian probability distributions. An important role is played here by quantum characteristic functions.

In Chapter 6 the general inequalities for the measurement mean square errors are derived, which are quantum analogs of the well-known Cramer-Rao inequality in mathematical statistics. The best unbiased measurements of the mean-value parameters of a Gaussian state are described.

Needless to say, the present book cannot (and is not intended to) replace the standard textbooks on quantum mechanics. Most of the important topics, such as perturbation theory, are apparently out of its scope. Nor does it pretend to give a full account of quantum measurements. We have discussed only those problems which concern measurement statistics and do not require consideration of state changes after measurements.

The references to the relevant work on "open" quantum systems and quantum stochastic processes can be found in the comments.

The author's intention was to write a book accessible to a wide circle of readers, both mathematicians and physicists. As a result, the presentation, being in general mathematical, is rather informal and certainly not "the most economic" from a mathematical point of view. On the other hand, it neglects some subtleties concerning measurability etc. As a rule a rigorous treatment can be found in the special papers refered to. The necessary background for the whole material is knowledge of fundamentals of probability theory. Mathematically the most elementary is Chapter 1 which uses mainly finite-dimensional linear analysis. The functional analytic minimum is given in Section 2.1-2.6 of Chapter 2, and a mathematically educated reader may just glance over it. On the other hand, a reader familiar with quantum mechanics can omit the detailed discussion of such topics as harmonic oscillators and spin in Chapter 3, included to make the presentation self-contained, and concentrate on less familiar things.

The Dirac notation is used intensively throughout the book but with round brackets for the inner product as accepted in mathematical literature. The angle brackets, associated with the averaging symbol in statistical mechanics, are reserved for the different inner product defining the correlation of a pair of observables. To denote a quantum state as well as its density operator we use the letter S (not the usual ρ) allied to the notation P for the classical state (probability distribution).

The author's thanks are due to Prof. D. P. Želobenko and the late Prof. Yu. M. Shirokov who read the manuscript and made useful comments.

In translating the book the author took the opportunity of improving the presentation which concerned mainly Chapters 3, 4. Few references were added. The author is grateful to Prof. Yu. A. Rozanov and Prof. P. R. Krishnaiah for providing the opportunity of translating this book for *North-Holland Series in Statistics and Probability*.

Chapter 1
Statistical models

1.1. States and measurements

Any theoretical model ultimately relies upon experience – the framework for a model is constituted by the array of experimental data relevant to the study of the object or phenomenon. Let us consider a very schematic and general description of an experimental situation and try to trace back the emergence of the principal components of a theoretical model.

The fundamental reproducibility condition requires at least in principle the unrestricted possibility of repetition of an experiment. Considering a sequence of identical and independent realizations of some experimental situation one always sees that practically the data obtained are not identically the same but subject to random fluctuations, the magnitude of which depends on the nature of the experiment and of the object under investigation.

There exist large classes of phenomena, for example, planetary motion or constant electric currents, in which these random fluctuations can be both practically and theoretically ignored. The corresponding theories – classical celestial mechanics and circuit theory – proceed from the assumption that the parameters describing the object can be measured with arbitrary accuracy, or, ultimately, with absolute precision. In such cases the object is said to admit deterministic description. Such a description, however self-contained it seems to be, is usually only an approximation to reality, valid in so far as it agrees with the experience.

The fruitfulness of the deterministic point of view in the classical physics of the 18-19$^{\text{th}}$ centuries gave rise to the illusion of its universality. However, with the penetration of experimental physics into the atomic domain the inapplicability of the classical deterministic approach and the relevance of statistical concepts in this domain became more and more evident. The behavior of atomic and subatomic objects is essentially probabilistic; an ordinary way to extract information about them is to observe a large number of identical objects to obtain statistical data. The

interested reader can find about the experimental evidence for statistical description in microphysics, which is now generally accepted, in any contemporary tract on quantum physics.

The possibility of statistical description presumes the fulfilment of the following statistical postulate, incorporating the previous requirement of reproducibility: *the individual results in a sequence of identical, independent realizations of an experiment may vary, but the occurrence of one or another result in a long enough sequence of realizations can be characterized by a definite stable frequency*. Then, abstracting from the practical impossibility of performing an infinite sequence of realizations, one can adopt that the results of the experiment are theoretically described by the *probabilities* of various possible outcomes. More precisely, we must distinguish an individual realization of the experiment which results in some concrete outcome from the experiment as a collection of all its possible individual realizations. In this latter sense, the final results of the experiment are theoretically described by probability distributions. The deterministic dependence of the experimental results on the initial conditions is replaced by the statistical one: the function of the initial data is now the output probability distribution.

As an example consider a beam of identical independent particles which are scattered by an obstacle and then registered by a photographic plate, so that an individual particle hitting the plate causes a blackening of the emulsion at the place of the collision. Exposing a beam which consists of a large enough number of particles will result in a photographic picture giving the visual image of the probability density for the point at which an individual particle hits the plate. The natural light is the chaotic flow of an immense number of specific corpuscules – the photons. The well-known optical diffraction pictures present the images of the probability density of an individual photon scattered by an aperture.

Of course, the statistical description is by no means subject to atomic or subatomic phenomena. When investigating a system which consists of a large number of components (*e.g.*, a gas or a liquid) the experimenter has at his disposal only a very restricted set of parameters to vary (say, pressure, volume or temperature). An immense number of parameters, giving a detailed description for the behavior of subsystems of the system are out of control; their uncontrolled changes may substantially influence the results of measurements. A study of these fluctuations is essential for understanding the mechanisms of phenomena occurring in large systems. The statistics of observations is most important in problems of information transmission, where the fluctuations in the physical carriers of information are the source of various “noises” distorting the signal.

The statistical approach is often appropriate in biometrical research. In studying the effect of a medicine, a physician can take into account a limited number of parameters characterizing his patients such as age, blood group etc. However the effect of the treatment in each individual case will depend not only on these "integral" parameters, but also on a number of other internal factors which were not, or could not be taken into account. In such cases the dependence of the effect on the "input parameters" is not deterministic and often can be successfully described statistically.

These two examples show that the origin of fluctuations in results of measurement may be uncertainty in the values of some "hidden variables" which are beyond the control of the experimenter. The nature of randomness in atomic and subatomic phenomena is still not so clear, though the relevance of the statistical approach is confirmed here by more than half a century experience of applications of quantum theory. We shall not touch here the issues concerning the nature of randomness in microscopic phenomena, but we shall comment on some mathematical aspects of the relevant "problem of hidden variables" in the Supplement. The main attention we shall pay here to the consequences of the statistical postulate irrespective of the nature of the object under consideration. We shall see that already on this very general level the notions of the state and the measurement arise, which play a basic role, in particular, in quantum theory.

In any experiment one can distinguish the two main stages. At the first stage of *preparation* a definite experimental set-up is settled, some initial conditions or "input data" of the experiment are established. At the following stage of the experiment the "prepared" object is coupled to a measuring device, resulting in these or the other output data (Figure 1.1).

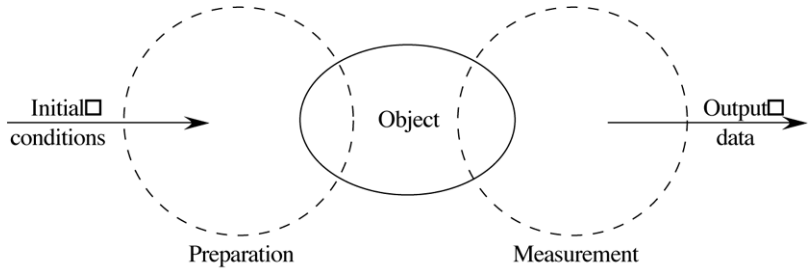

Figure 1.1.

Conventionally, one may conceive the object as a "black box" at the "input" of which one can impose some initial conditions $\widetilde{S}$. After the object has been definitely prepared, some measurement is performed, resulting in the output data u. These data may be of arbitrary nature; they may be

discrete if the measuring device registers the occurrence of some events, *e.g.*, the presence or absence of some particles; they may be represented by a scalar or vector quantity, if the measuring device has one or several scales; at last the result of a measurement may be a picture of a whole trajectory, as in a bubble chamber. To give a uniform treatment for all these possibilities we assume that the outcomes of measurement form a measurable space U with the σ-field of measurable subsets $\mathscr{A}(U)$. In the concrete cases we shall deal with, U will be usually a domain in the real n-dimensional space $\mathbb{R}^n$ with the Borel σ-field generated by open sets (or by multi-dimensional intervals). A measurable subset $B \subset U$ corresponds to the event: the result of the measurement u lies in B.

According to the statistical postulate, a result of an individual measurement can be considered as a realization of a random variable taking values in U. Let $\mu_{\widetilde{S}}(\mathrm{d}u)$ be the probability distribution of this random variable. The subscript $\widetilde{S}$ reflects the dependence of the statistics of the measurement upon the preparation procedure, *i.e.*, the initial conditions of the experiment, so that

$$\mu_{\widetilde{S}}(B) = \Pr\{u \in B|\widetilde{S}\}, \qquad B \in \mathscr{A}(U)$$

is the conditional probability of obtaining a result $u \in B$ under the initial condition $\widetilde{S}$. The map $\widetilde{S} \to \mu_{\widetilde{S}}(\mathrm{d}u)$ gives a complete statistical description for the results of the measurement. It should be stressed, however, that such a description does not contain indications either on a concrete mechanism of the measurement, or on its consequences for the object. From this point of view one should not distinguish between the measuring procedures giving the same statistics $\mu_{\widetilde{S}}$ under the same condition $\widetilde{S}$ however different their practical implementation may be. Thus the map $\widetilde{S} \to \mu_{\widetilde{S}}$ refers to this whole class of measuring procedures.

Similarly the initial conditions $\widetilde{S}_1$ and $\widetilde{S}_2$ are indiscernible from the point of view of the results of measurements if $\mu_{\widetilde{S}_1} = \mu_{\widetilde{S}_2}$ for any map $\widetilde{S} \to \mu_{\widetilde{S}}$ describing a measurement. We shall join the indiscernible preparation procedures $\widetilde{S}$ in the equivalence classes $S = [\widetilde{S}]$ which will be called *states*. Denote by $\mathfrak{S}$ the set of all possible states. Since the probability distribution $\mu_{\widetilde{S}}$ is the same for all $\widetilde{S}$ from the class S, it is a function of classes, $\mu_{\widetilde{S}} = \mu_S$. The map $S \to \mu_S$ transforming states $S \in \mathfrak{S}$ into the probability distributions on the space of outcomes U will be called a *measurement* (with values in U).

The set $\mathfrak{S}$ and the maps $S \to \mu_S$ enjoy an important structural property. Let $\{S_\alpha\}$ be a finite collection of states. Consider an infinite sequence of individual experiments in each of which the object is prepared in some of the states S_α, the occurrence of different values of α being

characterized by a probability distribution $\{p_\alpha\}$. This may reflect fluctuations in the values of some parameters in the preparation procedure. Let one and the same measurement be performed in each individual experiment. Then by the statistical postulate and the elementary properties of probabilities the occurrence of an outcome u will be described by the probability distribution $\mu(\mathrm{d}u) = \sum_\alpha p_\alpha \mu_{S_\alpha}(\mathrm{d}u)$. The situation described above can be considered as a special way of state preparation (mixing) when the value of the parameter α is not fixed but is chosen according to the prior distribution $\{p_\alpha\}$. Denoting such "mixed" state by

$$S = S(\{S_\alpha\}, \{p_\alpha\}) \tag{1.1.1}$$

we have for any measurement $S \to \mu_S$

$$\mu_S(\mathrm{d}u) = \sum_a p_\alpha \mu_{S_\alpha}(\mathrm{d}u). \tag{1.1.2}$$

Thus, we are led to adopt that for any finite set of states $\{S_\alpha\} \subset \mathfrak{S}$ and any probability distribution $\{p_\alpha\}$ there is uniquely defined "mixed" state $S(\{S_\alpha\}, \{p_\alpha\})$, which is characterized by (1.1.2). Then it turns out that the set of the states can be naturally identified with a convex set in a linear space, such that $S(\{S_\alpha\}, \{p_\alpha\}) = \sum_\alpha p_\alpha S_\alpha$. The exact formulation requires some knowledge in convexity presented in the next section.

1.2. Some facts about convex sets

Let $S_1, \ldots, S_n$ be the elements of a linear space $\mathscr{L}$, and $p_1, \ldots, p_n$ a set of real numbers satisfying

$$p_j \geq 0, \qquad j = 1, \ldots, n; \qquad \sum_{j=1}^{n} p_j = 1,$$

i.e., a finite probability distribution. Then the element

$$S = \sum_{j=1}^{n} p_j S_j$$

is called a *convex combination* of S_j with the coefficients (weights) $\{p_j\}$. The *convex hull* of a set $\mathscr{K} \subset \mathscr{L}$ is the collection of all convex combinations of all finite subsets $\{S_j\} \subset \mathscr{K}$. A set $\mathfrak{S}$ is called *convex* if it coincides with its convex hull, *i.e.*, if it contains convex combination of any finite subset of its elements. For two elements S_0, S_1, their convex combinations form the segment $[S_0, S_1]$:

$$\{S : S = p_0 S_0 + p_1 S_1;\ p_0, p_1 \geq 0,\ p_0 + p_1 = 1\}.$$

It is easy to see that the set $\mathfrak{S}$ is convex if and only if together with any two elements S_0, S_1, it contains the segment $[S_0, S_1]$.

An abstract set $\mathfrak{S}$ is called *mixture space* if there is a rule by which to any finite unordered collection $\{S_\alpha\}$ of elements of $\mathfrak{S}$ and any finite probability distribution $\{p_\alpha\}$ there corresponds a unique element $S(\{S_\alpha\}, \{p_\alpha\}) \in \mathfrak{S}$ called the mixture of the states S_α with weights p_α. It is assumed that mixing the collection consisting of copies of one and the same element S_0 gives again S_0, *i.e.*, if $S_\alpha \equiv S_0$ for all α, then $S(\{S_\alpha\}, \{p_\alpha\}) = S_0$. An example of mixture space is a convex set with the convex combination as the mixture.

Let F be a map from a mixture space $\mathfrak{S}$ into a linear space. The map is called *affine* if for any mixture $S(\{S_\alpha\}, \{p_\alpha\})$

$$F(S(\{S_\alpha\}, \{p_\alpha\})) = \sum_\alpha p_\alpha F(S_\alpha).$$

The set of affine maps is nonempty, since the map which sends any $S \in \mathfrak{S}$ into a constant vector b is affine. Clearly, the image of a convex set under an affine map is again convex. In linear space there is a close connection between affine and linear maps: namely, any affine map F of a convex set $\mathfrak{S} \subset \mathscr{L}$ has the form $F(T) = A(T) + b$, $T \in \mathfrak{S}$, where A is a linear map defined on $\mathscr{L}$. In particular, any affine functional (*i.e.*, map with values on the real line $\mathbb{R}$) is up to an additive constant a restriction to $\mathfrak{S}$ of a linear functional on $\mathscr{L}$.

A mixture space is called *separated* if for any two $S_1, S_2 \in \mathfrak{S}$ there is an affine functional φ on $\mathfrak{S}$ such that $\varphi(S_1) \neq \varphi(S_2)$.

An example of a separated mixture space is the set of states of Section 1.1. Indeed, for any measurement $S \to \mu_S$ and any subset $B \in \mathscr{A}(U)$ the functional $S \to \mu_S(B)$ is affine by (1.1.2). By construction, for any two states S_1 and S_2 there exists a measurement $S \to \mu_S$ such that $\mu_{S_1} \not\equiv \mu_{S_2}$, *i.e.*, $\mu_{s_1}(B) \neq \mu_{S_2}(B)$ for some B. The following simple statement shows that the set of states can be considered as a convex subset in a linear space, with the convex combinations as mixtures.

Proposition 1.2.1. *For any separated mixture space $\mathfrak{S}$ there is a one-to-one affine map of $\mathfrak{S}$ onto a convex subset of a linear space.*

Proof. Let $\mathfrak{A}(\mathfrak{S})$ be the linear space of all affine functional on $\mathfrak{S}$ and $\mathscr{L} = \mathfrak{A}(\mathfrak{S})'$ the dual space of all linear functionals on $\mathfrak{A}(\mathfrak{S})$. For each $S \in \mathfrak{S}$ introduce $\widehat{S} \in \mathscr{L}$, putting

$$\widehat{S}(\varphi) = \varphi(S), \qquad \varphi \in \mathscr{L}.$$

The map $S \to \widehat{S}$ is affine, since

$$\sum_j p_j \widehat{S}_j(\varphi) = \sum_j p_j \varphi(S_j) = \varphi(S(\{S_j\}, \{p_j\})) = \widehat{S}(\{S_j\}, \{p_j\})(\varphi),$$

and it is one-to-one since $\widehat{S}_1(\varphi) = \widehat{S}_2(\varphi)$ implies $\varphi(S_1) = \varphi(S_2)$ for all affine functionals φ. This proves the proposition. □

The most simple example of a convex set in an n-dimensional *simplex*, which is defined as a convex hull of $n+1$ points $S_0, \ldots, S_n$ in a space of dimension $\geq n$, such that vectors $S_0S_1, \ldots, S_0S_n$ are linearly independent. For $n = 1$ this is a segment, for $n = 2$ a triangle, for $n = 3$ a tetrahedron (Figure 1.2). The points $S_0, \ldots, S_n$ are the vertices of the simplex.

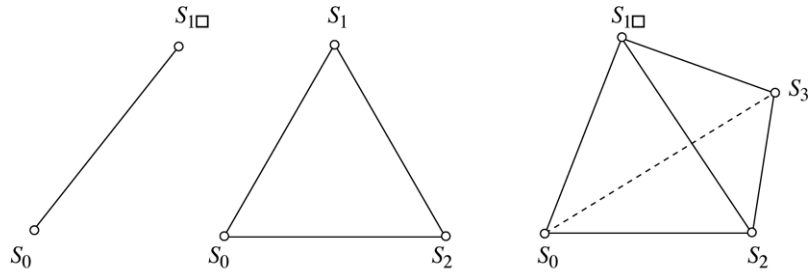

Figure 1.2.

The basic role in the theory of convex sets plays the notion of extreme point. The point S is an *extreme point* of a convex set $\mathfrak{S}$, if it is not an interior point of a segment, lying completely in $\mathfrak{S}$; that is, it cannot be represented in the form $S = p_0S_0 + p_1S_1$, where $p_0, p_1 > 0$, $p_0 + p_1 = 1$; $S_0, S_1 \in \mathfrak{S}$ and $S_0 \neq S_1$. For example, the extreme points of a simplex are its vertices. In a finite-dimensional space the following statement holds.

Theorem 1.2.2. *Any compact* (*i.e., bounded and closed*) *set* $\mathfrak{S}$ *coincides with the convex hull of the set of its extreme points.*

In general, there may be several ways to represent a point S of a convex set $\mathfrak{S}$ as a convex combination of extreme points. *The representation is unique for any point* $S \in \mathfrak{S}$ *if and only if* $\mathfrak{S}$ *is a simplex.*

Theorem 1.2.2, the definition and the characteristic property of a simplex can be generalized also to the infinite-dimensional case, but a careful treatment of these problems would require much more space; on the other hand, the finite-dimensional picture presented above will suffice for

understanding the following sections. Therefore we consider only one infinite-dimensional example, which we shall substantially need later.

Let $\mathfrak{P}(\Omega)$ be the collection of all probability distributions on a measurable space Ω. This set is convex, since any "mixture" of probability distributions $P_j(\mathrm{d}\omega)$ is again a probability distribution on Ω

$$P(A) = \sum_j p_j P_j(A), \qquad A \in \mathscr{A}(\Omega).$$

To any point $\omega \in \Omega$ there corresponds the *δ-distribution*, concentrated at ω,

$$\delta_\omega(A) = \begin{cases} 1, & A \ni \omega, \\ 0, & A \not\ni \omega. \end{cases}$$

We shall suppose, without loss of generality, that the σ-field $\mathscr{A}(\Omega)$ separates points of Ω, *i.e.*, for any two $\omega_1, \omega_2 \in \Omega$ there is $A \in \mathscr{A}(\Omega)$ such that $\omega_1 \in A$; $\omega_2 \notin A$. Then the correspondence $\omega \to \delta_\omega$ is one-to-one. The δ-distributions are precisely the extreme points of $\mathfrak{P}(\Omega)$. For any $P \in \mathfrak{P}(\Omega)$

$$P(A) = \int_\Omega \delta_\omega(A) P(\mathrm{d}\omega), \qquad A \in \mathscr{A}(\Omega). \tag{1.2.3}$$

This representation is a continual analog of a finite convex combination with respect to extreme points, the role of weights p_j is played by the probability distribution $P(\mathrm{d}\omega)$. The representation (1.2.3) is unique; so that the convex set $\mathfrak{P}(\Omega)$ has the property, characteristic of a simplex in the finite-dimensional case, and we shall keep this name for $\mathfrak{P}(\Omega)$.

If the space Ω consists of n points, $\Omega = \Omega_n$ ($\mathscr{A}(\Omega_n)$ being the Boolean field of subsets of Ω_n), the set

$$\mathfrak{P}_n = \left\{ P = [p_1, \ldots, p_n] : p_j \geq 0, \sum_{j=1}^n p_j = 1 \right\}$$

is clearly an (n–1)-dimensional simplex with the vertices [1,0,...,0], ..., [0, ..., 0, 1]. It will be convenient for us to represent P by the diagonal $n \times n$-matrix

$$P = \begin{bmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{bmatrix}.$$

Then the characteristic properties of P take the form

$$P \geq 0, \qquad \mathrm{Tr}\, P = 1, \tag{1.2.4}$$

where Tr denotes the *trace*, *i.e.*, the sum of diagonal elements of a matrix.

If X is a random variable on Ω_n with the values $x_1, \ldots, x_n$, then putting

$$X = \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{bmatrix}$$

we get that the expectation of X with respect to the probability distribution P is equal to

$$\sum_{j=1}^{n} p_j x_j = \operatorname{Tr} PX. \tag{1.2.5}$$

Consider the set $\mathfrak{O}_n$ of random variables, satisfying $0 \leq x_j \leq 1$, *i.e.*,

$$0 \leq X \leq I \tag{1.2.6}$$

where I is the unit $n \times n$-matrix. Then $\mathfrak{O}_n$ is a convex set – a unit hypercube, the extreme points being its vertices, represented by the matrices X for which x_j is either 0 or 1. Such matrices satisfy $X^2 = X$, so that the extreme points of $\mathfrak{O}_n$ are the idempotent (projection) matrices.

This elementary consideration leads naturally to the following construction which is of principal interest in connection with quantum theory. We can consider (complex) $n \times n$-matrices as operators acting in the n-dimensional unitary space $\mathscr{H}$ of column-vectors $\varphi = [\varphi_1]$, $\psi = [\psi_1], \ldots$ The inner product of φ and ψ is defined by: $(\varphi|\psi) = \sum \overline{\varphi}_j \psi_j$. We shall use Dirac's notation $|\varphi), |\psi), \ldots$ for column-vectors $\varphi, \psi, \ldots$ and $(\varphi|, (\psi|, \ldots$ for the Hermitean conjugated row-vectors $\varphi^*, \psi^*, \ldots$ The symbol for the inner product is then simply a graphic junction of symbols for the factors $(\varphi|$ and $|\psi)$. The "outer" product $|\psi)(\varphi|$ is then the $n \times n$-matrix with the components $[\psi_j \overline{\varphi}_k]$. If ψ is the unit vector, $(\psi|\psi) = 1$, then $S_\psi = |\psi)(\psi|$ is the matrix of the (orthogonal) projection on the vector ψ.

The finite-dimensional spectral theorem says that for any Hermitean $n \times n$-matrix X there is a complete orthonormal system of vectors $\{e_j;\ j = 1, \ldots, n\}$, in which X has the diagonal form

$$X = \sum_{j=1}^{n} \lambda_j |e_j)(e_j|, \tag{1.2.7}$$

where λ_j are the eigenvalues of X, which are real. It follows that e_j is an eigenvector of X corresponding to the eigenvalue λ_j.

In (1.2.7) the λ_j's are not necessarily different. Denote by $x_1, \ldots, x_m$ ($m \leq n$) the distinct eigenvalues of X numbered in the increasing order,

and by $E_k = \sum |e_j)(e_j|$ (the sum extends over all e_j belonging to λ_k) the matrix of projection onto the invariant subspace of X, corresponding to the eigenvalue λ_k. Then we have a different form of the spectral representation

$$X = \sum_{k=1}^{m} x_k E_k. \tag{1.2.8}$$

Consider the set $\mathfrak{S}_n$ of all Hermitean $n \times n$-matrices $S = [s_{jk}]$ satisfying the conditions

$$S \geq 0, \qquad \operatorname{Tr} S = 1, \tag{1.2.9}$$

which have the same form as (1.2.4). By the finite-dimensional spectral theorem

$$S = \sum_{j=1}^{n} \lambda_j S_{\psi_j}, \tag{1.2.10}$$

where λ_j are the eigenvalues of S, and $S_{\psi_j} = |\psi_j)(\psi_j|$ are the mutually orthogonal projections on the unit eigenvectors of S. The condition (1.2.7) implies that the eigenvalues of $S \in \mathfrak{S}_n$ constitute a probability distribution

$$\lambda_j \geq 0, \qquad \sum_{j=1}^{n} \lambda_j = 1.$$

In particular, $0 \leq \lambda_j \leq 1$ and (1.2.10) implies

$$S - S^2 = \sum_{j=1}^{n} \lambda_j (1 - \lambda_j) S_{\psi_j} \geq 0,$$

with the sign of equality attained if and only if $S = S_{\psi_k}$ for some ψ_k, *i.e.*, if S is a one-dimensional projection.

Proposition 1.2.3. *The set $\mathfrak{S}_n$ in convex, its extreme points being precisely one-dimensional projections.*

Proof. The first part of the statement follows from the fact that the conditions (1.2.9), defining $\mathfrak{S}_n$ sustain forming convex combinations. To prove that any one-dimensional projection S is an extreme point of $\mathfrak{S}_n$, assume that

$$\begin{aligned} &S = p_0 S_0 + p_1 S_1; \\ &p_0, p_1 > 1, \qquad p_0 + p_1 = 1. \end{aligned}$$

Taking square of this equality, subtracting and using the inequality $S \geq S^2$ for $S \in \mathfrak{S}_n$, we obtain

$$\begin{aligned} S - S^2 &= p_0(S_0 - S_0^2) + p_1(S_1 - S_1^2) + p_0 p_1 (S_0 - S_1)^2 \\ &\geq p_0 p_1 (S_0 - S_1)^2. \end{aligned}$$

Since S is a one-dimensional projection, then $S = S^2$, which implies that $S_0 = S_1$. Therefore S is an extreme point.

To prove the converse consider the spectral representation (1.2.10). Since $S_{\psi_j} \in \mathfrak{S}_n$ and $S_{\psi_j} \neq S_{\psi_k}$ for $j \neq k$, then for an extreme point S of $\mathfrak{S}_n$ the sum (1.2.10) can have only one nonzero term. Therefore $S = S_{\psi_j}$ for some ψ_j, which proves the proposition. □

The relation (1.2.10) is one of many possible representations of S as a convex combination of the extreme points.

We consider also the convex set $\mathfrak{X}_n$ of all Hermitean $n \times n$-matrices X, satisfying (1.2.6), and show that *the extreme points of this set are the* (*orthogonal*) *projections*, *i.e.*, (*Hermitean*) *matrices satisfying* $X^2 = X$. The proof of the statement that every projection is an extreme point is the same as in Proposition 1.2.3. To prove the converse, write the spectral decomposition of the matrix X in the form (1.2.8) with $0 \leq x_k \leq 1$ where x_k are the distinct eigenvalues of X, E_k is the projection onto the invariant subspace of X, corresponding to the eigenvalue x_k. Since $x_1 < \cdots < x_m$, then using in (1.2.8) the Abel transform and taking into account the equality $E_1 + \cdots + E_m = I$, we have

$$X = (1 - x_1) \cdot 0 + \sum_{k=1}^{m-1} (x_k - x_{k+1}) \cdot E'_k + x_m \cdot I,$$

where $E'_k = E_1 + \cdots + E_k$. Since the projections 0, I and E'_k belong to $\mathfrak{X}_n$ and the coefficients are nonnegative and sum to 1, then this is a convex combination of distinct projections. If X is an extreme point, then the sum can have only one nonzero term, and the matrix X must be a projection.

The difference between the sets $\mathfrak{P}_n$ and $\mathfrak{S}_n$ (correspondingly, between $\mathfrak{O}_n$ and $\mathfrak{X}_n$) is that in the latter case we consider all Hermitean matrices, satisfying (1.2.9) (correspondingly (1.2.6)), while in the former case only the diagonal matrices. We could consider as well a commuting family of matrices which can be simultaneously diagonalized. Therefore the latter case may be called the "noncommutative" analog to the former; defining the mean value by (1.2.5) one may treat $S \in \mathfrak{S}_n$ as a "noncommutative

probability distribution", and a Hermitean matrix X as a "noncommutative random variable". We shall see this connection to be deeper than a pure formal analogy.

We took the complex matrices because of their relevance to quantum theory, but the real matrices can be treated similarly.

At the end of this section we consider in some detail the structure of the convex set $\mathfrak{S}_n$ in the simplest "noncommutative" case, $n = 2$. Any matrix $S \in \mathfrak{S}_2$ can be represented as

$$S = \frac{1}{2}\begin{bmatrix} 1+\theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & 1-\theta_3 \end{bmatrix}, \tag{1.2.11}$$

where $\theta_1, \theta_2, \theta_3$ are the real numbers called the *Stokes parameters*. The condition $S \geq 0$ is equivalent to the inequality $\theta_1^2 + \theta_2^2 + \theta_3^2 \leq 1$. Thus, $\mathfrak{S}_2$ as a convex set can be represented by the unit ball in the three-dimensional real vector space; the extreme points are the matrices for which the vector $[\theta_1, \theta_2, \theta_3]$ lies on the sphere $\theta_1^2 + \theta_2^2 + \theta_3^2 = 1$.

If $n > 2$, then the set $\mathfrak{S}_n$ is a proper subset of the unit ball in the $(n^2 - 1)$-dimensional real vector space, and it cannot be represented so explicitly.

1.3. Definition of a statistical model

Motivated by consideration in Section 1.1, we define a *statistical model*[1] as a pair $(\mathfrak{S}, \mathfrak{M})$ where $\mathfrak{S}$ is a convex set and $\mathfrak{M}$ is a class of affine maps of $\mathfrak{S}$ into the collections of probability distributions on some measurable spaces U. The elements of $\mathfrak{S}$ are called *states*, and the elements of $\mathfrak{M}$ *measurements*. The problem of theoretical description of an object or a phenomenon satisfying the statistical postulate can then be described as a problem of construction of an appropriate statistical model. In more detail, the construction must first give a mathematical description of the set $\mathfrak{S}$ of theoretical states and the set $\mathfrak{M}$ of theoretical measurements and second, prescribe the rules for correspondence between the real procedures of preparation and measurement and the theoretical objects, *i.e.*, an injection of the experimental data into the statistical model.

The probability theory and statistics deal with the models in which the set of the states $\mathfrak{S}$ has the specifically simple structure. The statistical model of quantum theory is drastically different. We shall consider these models in detail in the following sections.

[1] The concept of statistical model will be considered in the Supplement in greater detail.

In this chapter we shall often simplify our consideration by using only the measurements with a finite number of outcomes. In such a case the space U is finite and the probability distribution of the results of the measurement is described by a finite collection of real affine functionals $\{\mu_S(u); u \in U\}$ on $\mathfrak{S}$, satisfying

$$\mu_S(u) \geq 0, \qquad u \in U; \qquad \sum_{u \in U} \mu_S(u) = 1. \tag{1.3.12}$$

Here $\mu_S(u)$ is the probability of the result u when the state is S. For any $B \subset U$

$$\mu_S(B) = \sum_{u \in B} \mu_S(u).$$

Technically this case is much simpler than the continual one, being still sufficient to expose the essential features of the theory. In practice such measurements correspond to the procedures resulting in some classification of the data. Furthermore, one can easily imagine a finitely-valued approximation of a measurement with a continual space of the outcomes U by making a partition of U into a finite number of "small" pieces.

A two-valued measurement is called a *test*. Denoting one of the results of the test by 0, and the other by 1 we get that any test can be described by defining only one function on $\mathfrak{S}$, say $\mu_S(1)$, the probability of getting 1, since $\mu_S(0) = 1 - \mu_S(1)$. The probability $\mu_S(1)$ is an affine functional on $\mathfrak{S}$ satisfying $0 \leq \mu_S(1) \leq 1$.

Let $S \to \mu_S(\mathrm{d}u)$ be a measurement with an arbitrary space of results U. Then to any $B \in \mathscr{A}(U)$ there corresponds the test $S \to \{\mu_S(\overline{B}), \mu_S(B)\}$, the result of which is 0 if $u \in \overline{B}$ and is 1 if $u \in B$. ($\overline{B}$ denotes the complement of the set B). Thus, any measurement can be considered as a collection of tests (satisfying apparent compatibility conditions).

1.4. The classical statistical model

We have seen in Section 1.1 that the notion of state refers to the initial conditions of the experiment. Here we shall adopt that these conditions can be formally described by the points ω of some set Ω, which will be called *phase space*.

To take into account the possibility of variations in the initial data during the repetitions of an experiment, or uncertainties in some parameters in the preparation procedure we shall consider also the probability distributions on Ω. To do this we must accept that Ω is a measurable space with a σ-field $\mathscr{A}(\Omega)$; we assume that $\mathscr{A}(\Omega)$ separates the points of Ω.

Any probability distribution P on Ω will be called *classical state*. It should be interpreted as a statistical description of the preparation stage.

To any $\omega \in \Omega$ corresponds the *pure state* described by the δ-distribution $\delta_\omega(A)$, $A \in \mathscr{A}(\Omega)$. According to Section 1.2 the collection $\mathfrak{P}(\Omega)$ of all classical states is the convex set of the most simple structure, the simplex, and the pure states are its extreme points.

A measurement with values in U is described by an affine map

$$P \to \mu_P(\mathrm{d}u) \tag{1.4.13}$$

which transforms the set of classical states $\mathfrak{P}(\Omega)$ into the set of probability distribution $\mathfrak{P}(U)$. Denote by $M_\omega(\mathrm{d}u)$ the probability distribution of the given measurement with respect to a pure state δ_ω, so that $M_\omega(\mathrm{d}u) = \mu_{\delta_\omega}(\mathrm{d}u)$, and consider the mixture of the pure states

$$P(\mathrm{d}\omega) = \sum_\alpha p_\alpha \delta_{\omega_\alpha}(\mathrm{d}\omega).$$

Since (1.4.13) is affine, the probability distribution of the results with respect to this state will be given by

$$\mu_P(B) = \int M_\omega(B) P(\mathrm{d}\omega), \qquad B \in \mathscr{A}(U). \tag{1.4.14}$$

Under some additional assumptions this relation will be valid for any classical state P. We shall not discuss this question and simply restrict our attention to measurements $P \to \mu_P$, which have the form (1.4.14) where $M_\omega(\mathrm{d}u)$ is a *conditional probability distribution* on U[2]. While P describes the uncertainty in the initial conditions of the experiment, the probability distribution $M_\omega(\mathrm{d}u)$ characterizes the disturbance due to the measuring device. The relation (1.4.14) shows how these two sources of uncertainty enter into the overall measurement statistics. We shall denote by $\boldsymbol{M}$ both the conditional probability distribution $\{M_\omega(\mathrm{d}u)\}$ and the corresponding measurement (1.4.13).

The classical statistical model which we are going to define is based on the assumption of *complete observability*, according to which the values of any parameters of the object can be established with absolute precision. To give a precise formulation we introduce the following definition. The measurement $\boldsymbol{M} = \{M_\omega(\mathrm{d}u)\}$ is called *deterministic* if for any $\omega \in \Omega$ and $B \in \mathscr{A}(U)$ either $M_\omega(B) = 0$ or $M_\omega(B) = 1$. This means that if the object is in a pure state, then for any $B \in \mathscr{A}(U)$ the result of the

[2] This means that for any $\omega \in \Omega$, $M_\omega(\mathrm{d}u)$ is a probability distribution on U, and for any $B \in \mathscr{A}(U)$, $M_\omega(B)$ is a measurable function of ω.

measurement u is either in B or not in B with probability 1. This can be written in the following form

$$M_\omega(B)^2 = M_\omega(B), \qquad B \in \mathscr{A}(U). \tag{1.4.15}$$

The nature of this condition can be made clear by discussing finitely-valued measurements. Let $\boldsymbol{M} = \{M_\omega(u);\, u \in U\}$ be such a measurement, where $M_\omega(u)$ is the probability of the outcome u if the object is in the pure state δ_ω. These probabilities satisfy

$$M_\omega(u) \geq 0, \qquad \sum_u M_\omega(u) = 1; \qquad \omega \in \Omega. \tag{1.4.16}$$

If $\boldsymbol{M}$ is deterministic measurement, then $M_\omega(u)$ is equal to either 0 or 1. Introducing the *indicator* of a set $F \subset \Omega$ as the function $\mathbf{1}_F(\omega)$, which is equal to 1 on F and 0 outside F, we have $M_\omega(U) = \mathbf{1}_{\Omega_{(u)}}(\omega)$, where $\Omega_{(u)} = \{\omega : M_\omega(u) = 1\}$. It follows from (1.4.16) that the sets $\Omega_{(u)}$ for different values of u do not intersect, and the union of all $\Omega_{(u)}$ is equal to Ω; this is expressed by saying that the sets $\{\Omega_{(u)}\}$ form a *decomposition* of the set Ω. Therefore for any ω there is a unique $u = u(\omega)$ such that $M_\omega(u(\omega)) = 1$. For any $B \subset U$

$$M_\omega(B) = \sum_{u \in B} M_\omega(u) = \mathbf{1}_B(u(\omega)). \tag{1.4.17}$$

The function $\omega \to u(\omega)$ is a random variable on Ω with values in U; the relation (1.4.17) establishes the one-to-one correspondence between the deterministic measurements and the random variables with values in U. To make this connection more transparent consider a random variable $X(\omega)$ on Ω taking values in a finite subset $\{x\}$ of the real line $\mathbb{R}$. Let $\Omega_{(x)}$ be the subset of Ω on which $X(\omega)$ is equal to x, then

$$X(\omega) = \sum_X x \cdot \mathbf{1}_{\Omega_{(x)}}(\omega) = \sum_x x M_\omega(x). \tag{1.4.18}$$

Thus, to the random variable X there corresponds the unique deterministic measurement $\boldsymbol{M} = \{M_\omega(x)\}$ such that X takes a value x if and only if x is the result of the measurement $\boldsymbol{M}$.

The case of continuous random variables is technically much more involved but the conclusion is essentially the same: under some regularity conditions the relation (1.4.17) establishes the one-to-one correspondence between the random variables and the deterministic measurements.

We can now formalize the requirement of the complete observability by adopting the following definition. The *classical statistical model* is a

model $(\mathfrak{P}(\Omega), \mathfrak{M})$, where $\mathfrak{P}(\Omega)$ is the simplex of all probability distributions on the phase space Ω, and the class $\mathfrak{M}$ contains all the deterministic measurements.

Now consider the convex set $\mathfrak{M}(U)$ of all affine maps $P \to \mu_P$ with μ_P having the form (1.4.14).

Proposition 1.4.1. *The deterministic measurements are the extreme points of $\mathfrak{M}(U)$, and viceversa.*

Proof. Let $\boldsymbol{M} = \{M_\omega(B)\}$ be a deterministic measurement and let $\boldsymbol{M} = p_0\boldsymbol{M}^0 + p_1\boldsymbol{M}^1$; $p_0, p_1 > 0$; $p_0 + p_1 = 1$, *i.e.*,

$$M_\omega(B) = p_0 M^0_\omega(B) + p_1 M^1_\omega(B), \qquad B \in \mathscr{A}(U).$$

Taking the square of this equality and using the fact that $M_\omega(B) = M_\omega(B)^2$, we get after some algebra

$$p_0 M^0_\omega(B)[1 - M^0_\omega(B)] + p_1 M^1_\omega(B)[1 - M^1_\omega(B)] + p_0 p_1 [M^0_\omega(B) - M^1_\omega(B)]^2 = 0,$$

whence, using the inequality $M^0_\omega(B)[1 - M^0_\omega(B)] \geq 0$ and an analogous inequality for $\boldsymbol{M}^1$, we get $M^0_\omega(B) \equiv M^1_\omega(B)$. This means that $\boldsymbol{M}$ is an extreme point of $\mathfrak{M}(U)$.

Conversely, let $\boldsymbol{M} = \{M_\omega(B)\}$ be an extreme point of $\mathfrak{M}(U)$. Let B_1 be a fixed set from $\mathscr{A}(U)$, $B_2 = \overline{B}_1$, its complement. Put

$$M^\pm_\omega(B) = M_\omega(B) \pm [M_\omega(B_1)M_\omega(B \cap B_2) - M_\omega(B_2)M_\omega(B \cap B_1)]. \tag{1.4.19}$$

Then $M_\omega(B) = \frac{1}{2}M^+_\omega(B) + \frac{1}{2}M^-_\omega(B)$. We shall show that $\{M^\pm_\omega(B)\}$ are conditional probability distributions on U. For this it is sufficient to check that $M^\pm_\omega(B)$, $B \in \mathscr{A}(U)$, for any $\omega \in \Omega$ is a probability distribution. It is clear that $M^\pm_\omega(B)$ is a σ-additive function of $B \in \mathscr{A}(U)$, since all the terms in (1.4.19) are measures; moreover, $M^\pm_\omega(U) = M_\omega(U) = 1$. It remains to check that $M^\pm_\omega(B) \geq 0$, but this follows from the inequality

$$M^\pm_\omega(B) \geq M_\omega(B \cap B_1)[1 \mp M_\omega(B_2)] + M_\omega(B \cap B_2)[1 \pm M_\omega(B_1)] \geq 0.$$

Since $\boldsymbol{M}$ is an extreme point it follows that $M^\pm_\omega(B) = M_\omega(B)$, $B \in \mathscr{A}(U)$, *i.e.*,

$$M_\omega(B_1)M_\omega(B \cap B_2) = M_\omega(B_2)M_\omega(B \cap B_1).$$

Putting here $B = B_1$ and taking into account that $B_1 \cap B_2 = \emptyset$ we get $M_\omega(B_2)M_\omega(B_1) = 0$ or $M_\omega(B_1)[1 - M_\omega(B_1)] = 0$. Thus, $\{M_\omega(B)\}$ is a deterministic measurement. □

Let $\{\boldsymbol{M}^j\}$ be a finite collection of deterministic measurements from $\mathfrak{M}(U)$ and $\{p_j\}$ be a probability distribution. Then the convex combination

$$M_\omega(B) = \sum_j p_j M^j_\omega(B), \qquad B \in \mathscr{A}(U)$$

describes a *randomized measurement* in which a measurement $\boldsymbol{M}^j$ is performed with the probability p_j. Physically it can correspond to fluctuations in the measuring device. In the simplest case when both Ω and U are finite, the set $\mathfrak{M}(U)$ is a compact convex subset of a finite-dimensional space and Theorem 1.2.2 implies that any element of $\mathfrak{M}(U)$ can be regarded as a randomized measurement. Moreover, in the general case, under some natural assumptions about Ω and U it is shown that any conditional probability distribution $\{M_\omega(\mathrm{d}u)\}$ can be presented as a "continual convex combination" of the deterministic measurements

$$M_\omega(B) = \int M^\alpha_\omega(B) Q(\mathrm{d}\alpha), \qquad B \in \mathscr{A}(U).$$

This relation describes the randomized measurement, $Q(\mathrm{d}\alpha)$ being the randomizing distribution of the set of deterministic measurements. Thus, having in mind randomized procedures, we can include in the class of measurements $\mathfrak{M}$ of the classical statistical model all the affine maps $P \to \mu_P$ from $\mathfrak{P}(\Omega)$ to $\mathfrak{P}(U)$ defined by conditional probability distributions $\{M_\omega(\mathrm{d}u)\}$ according to (1.4.14) (for any fixed U).

At the end of this section we discuss the description of the tests in the classical statistical model. Any test is uniquely defined by the function $X(\omega) = M_\omega(1)$, $\omega \in \Omega$, which satisfies

$$0 \le X(\omega) \le 1. \tag{1.4.20}$$

The probability of the outcome 1 with respect to the classical state P is equal to

$$\int X(\omega) P(\mathrm{d}\omega).$$

For a deterministic test, either $X(\omega) = 0$ or $X(\omega) = 1$, so that $X(\omega) = 1_{\Omega_{(1)}}(\omega)$. Thus the deterministic test defines a dichotomy of the phase space: $\Omega = \Omega_{(1)} \cup \Omega_{(0)}$, $\Omega_{(1)} \cap \Omega_{(0)} = \emptyset$.

If the space Ω is finite, $\Omega = \Omega_n$, then the set of the classical tests (1.4.20) in the n-dimensional unit hypercube $\mathfrak{O}_n$, with the vertices as the extreme points (see Section 1.2). The probability of the outcome 1 for the test $\{X_\omega\}$ with respect to the state $\{P_\omega\}$ is equal to $\Sigma_\omega P_\omega X_\omega$.

1.5. Reduction of statistical model. Classical model with a restricted class of measurements

The assumption of complete observability underlying the classical statistical model is an idealization which is valid in so far as its consequences agree with the experience. An important fact we are going to discuss is that weakening in some way the requirement of complete observability will in general lead to statistical models which may be radically different from the classical one.

Consider a statistical model $(\mathfrak{S}, \mathfrak{M})$. The states $S_1, S_2 \in \mathfrak{S}$, $S_1 \neq S_2$ are called *indiscernible* if for any measurement $S \to \mu_S$ from the class $\mathfrak{M}$ the resulting probability distributions coincide: $\mu_{S_1} \equiv \mu_{S_2}$. There is no way to distinguish between such states S_1, S_2, basing on the results of the measurements, and from the point of view of an observer of the results, the states S_1 and S_2 are identical.

If we join the indiscernible states into the equivalence classes $[S]$ and put $\mu_{[S]} = \mu_S$, we obtain a new statistical model $(\mathfrak{S}', \mathfrak{M}')$, with $\mathfrak{S}'$ being the set of all equivalence classes $[S]$ and $\mathfrak{M}'$ the collection of affine maps of the form $[S] \to \mu_{[S]}$. This process will be called here *reduction*[3] of the initial model $(\mathfrak{S}, \mathfrak{M})$. The new model $(\mathfrak{S}', \mathfrak{M}')$ is separated in the sense that there are no indiscernible elements in $\mathfrak{S}'$. It was a separated model which was the ultimate product of the statistical analysis of experiments in Section 1.1.

Statistically the reduced model $(\mathfrak{S}', \mathfrak{M}')$ is completely equivalent to the initial one. However, as we shall see, the new set of states $\mathfrak{S}'$ can be very different from $\mathfrak{S}$. Moreover, by choosing an appropriate class of measurements $\mathfrak{M}$ we can reduce the classical simplex $\mathfrak{S} = \mathfrak{P}(\Omega)$ practically to the arbitrary convex set $\mathfrak{S}'$. In particular, uniqueness of decomposition into the extreme points need not hold for $\mathfrak{S}'$, as the following elementary example shows.

Consider the "object" with the phase space consisting of four points, $\Omega = \Omega_4$, so that a state is a vector $[P_1, \ldots, P_4]$ of the three-dimensional simplex $\mathfrak{P}_4$. For the measurements we shall take tests, *i.e.*, the vectors $[X_1, \ldots, X_4]$ from the four-dimensional hypercube $\mathfrak{O}_4$, satisfying in addition the equality

$$X_1 + X_2 = X_3 + X_4. \tag{1.5.21}$$

[3] This reduction of the state space of a statistical model (in the Supplement it is called "compression") should not be mixed with the notorious "quantum state reduction", i. e. the state change due to a measurement.

The states $P = \{P_j\}$ and $Q = \{Q_j\}$ are indiscernible by these tests if and only if (1.5.21) implies

$$\sum_j P_j X_j = \sum_j Q_j X_j,$$

i.e., the vector $[P_j - Q_j]$ is orthogonal to the hyperplane (1.5.21) in the four-dimensional space. The orthogonal complement to (1.5.21) is spanned by the vector $e = [1, 1, -1, -1]$, therefore the states $P = \{P_j\}$ and $Q = \{Q_j\}$ are indiscernible if and only if for some real t

$$\begin{aligned} P_1 &= Q_1 + t, & P_3 &= Q_3 - t, \\ P_2 &= Q_2 + t, & P_4 &= Q_4 - t. \end{aligned} \tag{1.5.22}$$

If $\mathfrak{P}_4$ is represented as a tetrahedron in the three-dimensional space (Figure 1.3), then P_j are the barocentric coordinates of a point in the tetrahedron, and the equations (1.5.22) define the set of parallel lines; a class of the indiscernible states will be presented by the segment of a line lying in the tetrahedron. The reduced set of states can therefore be identified with the projection of the tetrahedron along the direction of the lines onto an appropriate plane (Figure 1.3), which is a convex quadrangle.

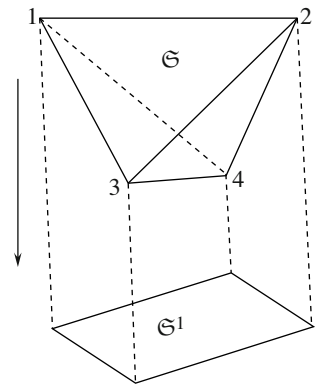

Figure 1.3.

This example shows that narrowing the set of possible tests may result in "sticking together" the states and originating new forms of convex states in which uniqueness of the decomposition into extreme points need not hold. Without entering into detail here, we mention that the restrictions on measurements may reflect the presence of some empirical symmetry relations. In the elementary example given above this role is played by the relation (1.5.21); in quantum mechanics, where the spatially-temporal description is essential, the symmetries with respect to the kinematical and dynamic transformations appear to be primarily important.

We now consider a description of the quantum-mechanical object – the spin $-\frac{1}{2}$ particle. As we shall see later, the states of this object in quantum theory are described by the 2×2-matrices of the form (1.2.11). The set of this matrices $\mathfrak{S}_2$ can be represented by the unit ball in the real three-dimensional space, as shown in Section 1.2. It is instructive to investigate what kind of a restriction can reduce a classical simplex to this convex set which in a sense is contrary to the simplex: the whole of its boundary consists of extreme points and the decomposition into them is highly non-unique.

We shall consider a schematic description of the *Stern-Gerlach experiment* which had led to the discovery of spin. The beam of silver atoms passes between the poles of the magnet creating the inhomogeneous magnetic field $\boldsymbol{B}$, increasing in the vertical direction. The particles which have passed through the field are collected by the plate $\mathscr{E}$. From the distribution of the substance accumulated by the plate $\mathscr{E}$ one can infer about the deviation of the particles under the action of the magnetic field $\boldsymbol{B}$ (Figure 1.4).

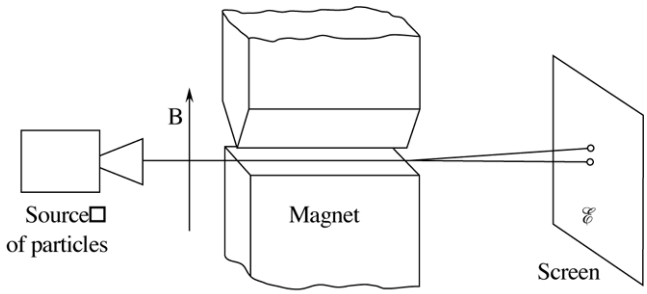

Figure 1.4.

While the classical theory predicted continuous scattering in all direction, *i.e.*, a more or less smooth distribution of the substance accumulated by the plate, the experiment showed the sharp splitting of the incoming beam into the two symmetrical beams in the vertical direction. By using other substances it was also possible to obtain splittings of the beam into more than two components. The spin of the particle is the integer of half-integral j such that $2j + 1$ is the number of the splitting components, so that in the Stern-Gerlach experiment $j = \frac{1}{2}$.

Placing instead of the plate $\mathscr{E}$ a screen with an aperture one obtains a filter which allows to pass, say, the upper outgoing beam and absorbs the lower one. The filtered beam, which is called polarized in the direction $\boldsymbol{B}$, does not split under the repeated action of the inhomogeneous field with the same direction $\boldsymbol{B}$ and thus it passes without absorption through the same filter which was used for its preparation. However it does split and

therefore is partially absorbed by the second filter with the other direction of the field $\boldsymbol{B}$. It is completely absorbed if the direction of $\boldsymbol{B}$ in the second filter is opposite to that of the initial one.

A schematic classical description of the filter is furnished by a unit vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$, giving the orientation of the filter, *i.e.*, the direction of the inhomogeneous magnetic field $\boldsymbol{B}$. All other parameters remain fixed and therefore can be omitted from the description. Consider the following experiment: the beam of a given intensity is first prepared by passing through the first filter $\boldsymbol{\theta}_{\text{in}}$ and then passing through the second filter $\boldsymbol{\theta}_{\text{out}}$ after which the intensity of the outgoing beam is measured. The ratio of this intensity to the half of the incoming intensity then gives the probability for a particle prepared by the filter $\boldsymbol{\theta}_{\text{in}}$ to pass through the filter $\boldsymbol{\theta}_{\text{out}}$ (It is supposed that the incoming beam is "chaotic", so that exactly one half of the incoming beam passes through the first filter.).

The "phase space" Ω is the set of all possible directions $\boldsymbol{\theta}_{\text{in}}$, *i.e.*, the unit sphere $\mathbb{S}^2$ in the real three-dimensional space. A classical state is a probability distribution $P(\mathrm{d}\theta)$ on $\mathbb{S}^2$, describing "partially polarized" beam. The pure states $\delta_{\boldsymbol{\theta}}$ corresponds to completely polarized beams and the uniform distribution to the "chaotic", unpolarized beam.

Denote by $\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\}$ the probability that the particle prepared by the filter $\boldsymbol{\theta}_{\text{in}}$ passes through the filter $\boldsymbol{\theta}_{\text{out}}$ (Figure 1.5). From the rotational symmetry it is natural to expect that this probability depends on the directions $\boldsymbol{\theta}_{\text{in}}$, $\boldsymbol{\theta}_{\text{out}}$ only through the angle φ between them, or through their inner product $t = \boldsymbol{\theta}_{\text{in}} \cdot \boldsymbol{\theta}_{\text{out}}$. Thus $\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\} = F(t)$, $-1 \le t \le 1$.

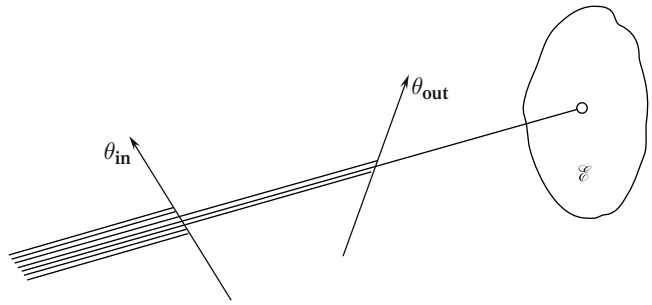

Figure 1.5.

If the directions $\boldsymbol{\theta}_{\text{in}}$ and $\boldsymbol{\theta}_{\text{out}}$ coincide, then $\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\} = 1$; if they are opposite, then $\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\} = 0$. Hence $F(1) = 1$, $F(0) = 0$. For any direction $\boldsymbol{\theta}_{\text{out}}$ the deviated particles go either in the direction $\boldsymbol{\theta}_{\text{out}}$ or in the opposite direction, whence $F(t) + F(-t) = 1$, or $\frac{1}{2} - F(t) = -\frac{1}{2} + F(-t)$. Thus $\frac{1}{2} - F(t)$ is an odd function of $t \in [-1, 1]$ taking values $\mp\frac{1}{2}$ at the ends of the interval.

The simplest continuous function satisfying these conditions is the linear function $F(t) = \frac{1}{2}(1+t)$, *i.e.*,

$$\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\} = \frac{1+\boldsymbol{\theta}_{\text{out}}\cdot\boldsymbol{\theta}_{\text{in}}}{2} = \cos^2\frac{\varphi}{2}. \tag{1.5.23}$$

We shall see later that it is this expression for the probability $\mathbf{Pr}\{\boldsymbol{\theta}_{\text{out}}|\boldsymbol{\theta}_{\text{in}}\}$ which is given by the quantum theory. This distribution is in agreement with experimental data.

If we agree that the result of the measurement is 1 if the particle, prepared by the first filter passes through the second filter $\boldsymbol{\theta}_{\text{out}}$, and 0 if it is absorbed, then this measurement can be considered as the test, defined by the function

$$X(\boldsymbol{\theta}) = \frac{1+\boldsymbol{\theta}_{\text{out}}\cdot\boldsymbol{\theta}}{2}; \qquad \boldsymbol{\theta}\in\mathbb{S}^2. \tag{1.5.24}$$

The probability of the result 1 when the prepared state is $P(\mathrm{d}\theta)$ is equal to

$$\mu_P(1) = \int \frac{1+\boldsymbol{\theta}_{\text{out}}\cdot\boldsymbol{\theta}}{2} P(\mathrm{d}\theta). \tag{1.5.25}$$

The set of the measurements $\mathfrak{M}$ for this model consists of all tests, defined by the functions of the form (1.5.24) with $\boldsymbol{\theta}_{\text{out}}\in\mathbb{S}^2$.

The two classical states P_1, P_2 are indiscernible if $\mu_{P_1}(1) = \mu_{P_2}(1)$ for all tests from $\mathfrak{M}$, *i.e.*,

$$\int \boldsymbol{\theta}_0\cdot\boldsymbol{\theta} P_1(\mathrm{d}\theta) = \int \boldsymbol{\theta}_0\cdot\boldsymbol{\theta} P_2(\mathrm{d}\theta), \qquad \theta_0\in\mathbb{S}^2,$$

or $\int \boldsymbol{\theta} P_1(\mathrm{d}\theta) = \int \boldsymbol{\theta} P_2(\mathrm{d}\theta)$. Thus the reduced states are in one-to-one affine correspondence

$$[P] \leftrightarrow \int \boldsymbol{\theta} P(\mathrm{d}\theta) = \left[\int \theta_1 P(\mathrm{d}\theta), \int \theta_2 P(\mathrm{d}\theta), \int \theta_3 P(\mathrm{d}\theta)\right]$$

with the three-dimensional vectors, representable in the form $\boldsymbol{\theta}_P = \int \boldsymbol{\theta} P(\mathrm{d}\theta)$, where $P(\mathrm{d}\theta)$ is a probability distribution of the sphere $\mathbb{S}^2$, *i.e.*, with the points of the unit ball. From (1.5.25)

$$\mu_{[P]}(1) = \frac{1+\boldsymbol{\theta}_{\text{out}}\cdot\boldsymbol{\theta}_P}{2}. \tag{1.5.26}$$

Now considering the components of vectors in the unit ball as Stokes parameters we can introduce the matrices

$$S = \frac{1}{2}\begin{bmatrix} 1 + \int \theta_3 P(\mathrm{d}\theta) & \int (\theta_1 - i\theta_2) P(\mathrm{d}\theta) \\ \int (\theta_1 + i\theta_2) P(\mathrm{d}\theta) & 1 - \int \theta_3 P(\mathrm{d}\theta) \end{bmatrix},$$

$$X = \frac{1}{2}\begin{bmatrix} 1 + \theta_3' & \theta_1' - i\theta_2' \\ \theta_1' + i\theta_2' & 1 - \theta_3' \end{bmatrix},$$

where $[\theta_1', \theta_2', \theta_3'] = \boldsymbol{\theta}_{\mathrm{out}}$. Then the reduced states are represented by matrices S from $\mathfrak{S}_2$, the tests by matrices X, and by (1.5.26) the probability of the result 1 for the test X and the state S is equal to

$$\mu_S(1) = \operatorname{Tr} SX = \int X(\boldsymbol{\theta}) P(\mathrm{d}\theta).$$

In fact, we have constructed a "hidden variables" model for the spin $-\frac{1}{2}$ particles. We shall return to this point in Section 1.7 after the introduction of the general quantum statistical model.

1.6. The statistical model of quantum mechanics

The main subject of our study will be the statistical model, in which states are described by complex Hermitean matrices S, satisfying

$$S \geq 0, \qquad \operatorname{Tr} S = 1,$$

and called *density matrices*. The set of all such matrices $\mathfrak{S}_n$ is a convex set, its extreme points being the one-dimensional projections $S_\psi = |\psi)(\psi|$, $(\psi|\psi) = 1$ (see Section 1.2). The corresponding states are called *pure*. In what follows we shall be mainly interested in the infinite-dimensional analog of the density matrix, but in this chapter we shall content ourselves with the finite-dimensional case to demonstrate the main features without entering into technical difficulties associated with infinite dimensionality.

Next we must describe the class of quantum-theoretic measurements. According to the general definition any measurement with values in the space U is described by an affine map of the set of states $\mathfrak{S}_n$ into the set of probability distributions on U. We shall assume that U is finite. In this case the structure of any such map is described by the following proposition.

Proposition 1.6.1. *The relation*

$$\mu_S(u) = \operatorname{Tr} SM_u, \qquad u \in U, \tag{1.6.27}$$

establishes the one-to-one correspondence between affine maps $S \to \mu_S$ of the set of density matrices $\mathfrak{S}_n$ into the set of probability distributions on U and the resolutions of identity $\{M_u; u \in U\}$, i.e., the collections of Hermitean matrices $\{M_u\}$, satisfying

$$M_u \geq 0, \qquad \sum_{u \in U} M_u = I. \tag{1.6.28}$$

Lemma 1.6.2. *Any affine functional $\mu(S)$ in $\mathfrak{S}_n$ has the form $\mu(S) = \operatorname{Tr} SM$, where M is a Hermitean matrix.*

Proof. The set of all density matrices spans the real linear space $\mathscr{L}$ of all Hermitean matrices. This means that any Hermitean matrix can be represented as $T = \sum_j t_j S_j$, where t_j are real numbers, S_j are density matrices (this follows, *e.g.*, from the spectral representation of T). Extend μ from $\mathfrak{S}_n$ to $\mathscr{L}$ putting

$$\mu(T) = \sum_j t_j \mu(S_j).$$

We must show that the sum in the right-hand side of this equality does not depend on the particular representation of T as a linear combination of density matrices, *i.e.*, that the equality $\sum_j t_j S_j = \sum_k t'_k S'_k$ implies

$$\sum_j t_j \mu(S_j) = \sum_k t'_k \mu(S'_k).$$

Transferring if necessary some terms into another side, we can make $t_j \geq 0$ and $t'_k \geq 0$, at least one of t_j and t'_k being strictly positive. Taking the trace of the resulting equality we get $\sum_j t_j = \sum_k t'_k = \tau > 0$, since $\operatorname{Tr} S_j = \operatorname{Tr} S'_k = 1$. Introduce the probability distributions $p_j = t_j/\tau$, $p'_k = t'_k/t$. It is sufficient to show that $\sum_j p_j S_j = \sum_k p'_k S'_k$ implies $\sum_j p_j \mu(S_j) = \sum_k p'_k \mu(S'_k)$, and this is a straightforward consequence of the affinity of μ on $\mathfrak{S}_n$.

By construction $\mu(T)$ is a real linear function on $\mathscr{L}$. Any such function of $T = [t_{jk}]$ evidently has the form

$$\mu(T) = \sum_{j,k} t_{jk} m_{jk} = \operatorname{Tr} TM,$$

where $M = [m_{jk}]$ with $m_{jk} = \overline{m}_{kj}$, so that $M = M^*$. □

Lemma 1.6.3. *Let X be a Hermitean matrix; then $X \geq 0$ if and only if* $\operatorname{Tr} SX \geq 0$ *for all* $S \in \mathfrak{S}_n$.

Proof. $X \geq 0$ means that $(\psi|X\psi) \geq 0$ for all (unit) vectors ψ, *i.e.*, $\operatorname{Tr} S_\psi X \geq 0$. By Proposition 1.2.3 it is equivalent to $\operatorname{Tr} SX \geq 0$, $S \in \mathfrak{S}_n$. □

Proof of Proposition 1.6.1. By Lemma 1.6.2 $\mu_S(u) = \operatorname{Tr} SM_u$, where $\{M_u\}$ is a collection of Hermitean matrices. Positivity of $\mu_S(u)$ and Lemma 1.6.3 imply that $M_u \geq 0$. Finally $\sum_u \mu_S(u) = \operatorname{Tr} S(\sum_u M_u) = 1$ for all $S \in \mathfrak{S}_n$ whence $\sum_u M_u = I$. Indeed, $\operatorname{Tr} S(\sum_u M_u - I) = 0$ for all $S \in \mathfrak{S}_n$ so that by Lemma 1.6.3 $\sum_u M_u - I \geq 0$ and $\sum_u M_u - I \leq 0$. This proves the proposition. □

The collection $\{M_u\}$ is formally analogous to the conditional probability distribution $\{M_\omega(u)\}$ which served to describe measurements in the classical statistical model. We have noticed the particular role played there by the deterministic measurements. The analog of the corresponding condition (1.4.15) in the quantum case has the form

$$M_u^2 = M_u, \qquad u \in U. \tag{1.6.29}$$

This means that M_u is a projection matrix. We now show that (1.6.29) implies

$$M_u M_v = 0, \qquad u \neq v, \tag{1.6.30}$$

i.e., M_u, M_v are mutually orthogonal projections.

Lemma 1.6.4. *Let A, B, C be Hermitean matrices satisfying* $0 \leq B \leq C$ *and* $CA = 0$. *Then* $BA = 0$.

Proof. $CA = 0$ implies $A^*CA = 0$, whence $0 = A^*CA \geq A^*BA \geq 0$ and $A^*BA = 0$. This can be written as $(\sqrt{B}A)^*(\sqrt{B}A) = 0$ where $\sqrt{B}$ is the positive square root of $B \geq 0$. It follows that $\sqrt{B}A = 0$ and $BA = 0$. □

To derive (1.6.30) we write (1.6.29) as $(I - M_u)M_u = 0$ and observe that by (1.6.28) $0 \leq M_v \leq I - M_u$ when $u \neq v$. It remains to apply Lemma 1.6.4 with $A = M_u$, $B = M_v$, $C = I - M_u$.

Thus, the formal quantum analog of classical deterministic measurements is the *orthogonal resolutions of identity* $\{E_u\}$:

$$E_u E_v = \delta_{uv} E_u, \qquad \sum_u E_u = I.$$

The corresponding quantum measurements are called *simple*. Whereas a classical deterministic measurement defines a decomposition of the phase space Ω into mutually disjoint sets $\Omega_{(u)}$, a simple quantum measurement furnishes a decomposition of the unitary vector space $\mathscr{H}$ into the direct sum of mutually orthogonal subspaces $\mathscr{H}_{(u)} = E_u(\mathscr{H})$. Let the outcomes of a simple measurement $\{E_x\}$ be real numbers $\{x\}$, then one can introduce the Hermitean operator

$$X = \sum_x x E_x. \tag{1.6.31}$$

This relation establishes one-to-one correspondence between simple measurements and Hermitean operators in $\mathscr{H}$ similar to the correspondence (1.4.18) between deterministic measurements and random variables in probability theory. Hermitean operators therefore play in quantum theory the same role as random variables in probability theory; they are also called (quantum) observables. The mean value of the measurement $\{E_x\}$ is expressed directly in terms of the corresponding observable,

$$\sum_x x \mu_S(x) = \sum_x x \operatorname{Tr} S E_x = \operatorname{Tr} S X.$$

In the standard presentations of quantum theory observables are the primary objects. Equivalently, one can start with the simple measurements described by orthogonal resolutions of identity. We have seen, however, that the consistent probabilistic treatment of an experimental situation leads to general resolutions of identity. Elimination of non-orthogonal resolutions and restriction to simple measurements should be based on some additional argument; a simple formal analogy with the probability theory is, apparently, not enough for this purpose. In the probability theory the exceptional role of the deterministic measurements is substantiated by the property expressed in Proposition 1.4.1, according to which the statistics of any measurement can be obtained from statistics of deterministic measurements by an appropriate randomization. However, it is important to realize that this fact has no analogy in quantum theory.

Denote by $\mathfrak{M}(U)$ the convex set of all resolutions of identity $\{M_u; u \in U\}$.

Proposition 1.6.5. *Any orthogonal resolution of identity $\{E_u; u \in U\}$ is an extreme point of $\mathfrak{M}(U)$. The converse is true only if $U = \{0, 1\}$; if U has more than two elements, then there exists an extreme point of $\mathfrak{M}(U)$, which is not an orthogonal resolution of identity.*

Proof. The first statement is proved in the same way as its classical analog in Proposition 1.4.1. If $U = \{0, 1\}$, then any resolution of identity has

the form $\{I-X, X\}$ where X satisfies $0 \leq X \leq I$. As we know from Section 1.2 this last convex set has projections as its extreme points, therefore if the measurement $\{I-X, X\}$ is an extreme point, then $I-X$ and X are both projections. This proves the first part of the second statement.

Let now $U = \{1, \ldots m\}$, $m > 2$. Consider first the case $n \equiv \dim \mathscr{H} = 2$. One may imagine that the density matrices $S \in \mathfrak{S}_2$ describe the states of the spin-$\frac{1}{2}$ particle (see Section 1.5). Then the state prepared by the filter with the direction $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ is described by the density matrix (1.2.11). In particular, the density matrix

$$\begin{aligned} S_\alpha &= \frac{1}{2}\begin{bmatrix} 1 & e^{-\mathrm{i}\alpha} \\ e^{\mathrm{i}\alpha} & 1 \end{bmatrix} = |\psi_\alpha)(\psi_\alpha|; \\ |\psi_\alpha) &= \frac{1}{\sqrt{2}}\begin{bmatrix} e^{-\mathrm{i}\alpha/2} \\ e^{\mathrm{i}\alpha/2} \end{bmatrix}, \end{aligned} \tag{1.6.32}$$

describes the state prepared by the filter with the direction $\boldsymbol{\theta}_{\text{in}} = [\cos\alpha, \sin\alpha, 0]$. Consider m directions, corresponding to $\alpha_u = 2\pi u/m$; $u = 1, \ldots, m$, which divide the coordinate plane θ_1, θ_2, into m equal angles (see Figure 1.6), so that

$$\sum_{u=1}^{m} \exp(\mathrm{i}\alpha_u) = 0. \tag{1.6.33}$$

Then the collection

$$M_u = \frac{2}{m}|\psi_u)(\psi_u| \qquad (\psi_u = \psi_{\alpha_u}), \tag{1.6.34}$$

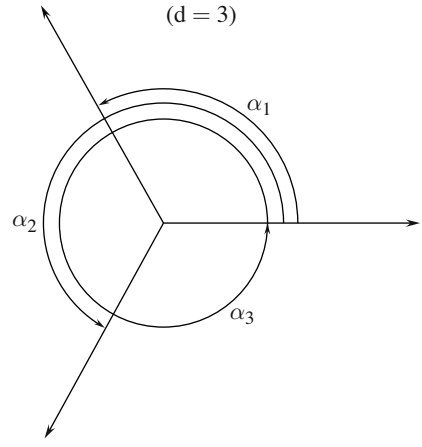

Figure 1.6.

constitutes a non-orthogonal resolution of identity. Indeed, $M_u \geq 0$ and via (1.6.33), (1.6.32), $\sum_u M_u = I$.

We shall show that for $m = 3$ the resolution of identity (1.6.34) is an extreme point. Indeed, let

$$M_u = p_0 M_u^0 + p_1 M_u^1; \qquad p_0, p_1 > 0,$$

then $M_u^0 \leq p_0^{-1} M_u$, $M_u^1 \leq p_1^{-1} M_u$, whence $M_u^j = \lambda_u^j |\psi_u)(\psi_u|$. The condition $\sum_u M_u^j = I$ together with (1.6.32) leads to equations

$$\sum_{u=1}^{3} \lambda_u^j \exp\left(\mathrm{i}\frac{2\pi u}{3}\right) = 0, \qquad \sum_{u=1}^{3} \lambda_u^j = 2.$$

The first means that λ_u^j are the lengths of the sides of a regular triangle, so that λ_u^j are all equal to each other; the second equality then implies $\lambda_u^j = \frac{2}{3}$. Thus $M_u^0 = M_u^1 = M_u$ and $\{M_u\}$ is an extreme point.

Now let $n \geq 2$ and $m \geq 3$. Decompose the n-dimensional space $\mathscr{H}_n$ into the direct orthogonal sum of a two-dimensional subspace $\mathscr{H}_2$ and its orthogonal complement $\mathscr{H}_{n-2}$ and denote by E the projection onto $\mathscr{H}_{n-2}$. Let $\{M_u; u = 1, 2, 3\}$ be the resolution of identity in $\mathscr{H}_2$ given by (1.6.34). Then the collection

$$\widetilde{M}_1 = M_1 \oplus 0, \qquad \widetilde{M}_2 = M_2 \oplus 0, \qquad \widetilde{M}_3 = M_3 \oplus E;$$
$$\widetilde{M}_u = 0, \qquad u \geq 3,$$

forms a non-orthogonal resolution of identity in $\mathscr{H}$, which is easily seen to be an extreme point of $\mathfrak{M}(U)$. □

Though the construction given above may seem somewhat artificial, these arguments do show that unlike classical statistics, in quantum theory there are no serious reasons to confine oneself to the orthogonal resolutions of identity. In fact, we shall see later that there is a variety of physical quantities, measurements of which are naturally described by the non-orthogonal resolutions of identity.

Based on the foregoing discussion we mean by the *statistical model, of quantum mechanics* the model in which states are described by the density matrices S, and measurements by all affine maps $S \to \mu_S$, transforming density matrices into probability distributions.

It is worthwhile to emphasize here again that statistical model is a mathematical object. The fact that an array of experimental data is satisfactorily described by the given model $(\mathfrak{S}, \mathfrak{M})$ means that there is an injection of the properly treated data into the model, *i.e.*, to any real preparation procedure one can ascribe a theoretical state $S \in \mathfrak{S}$ and to any real

measurement a theoretical measurement $M \in \mathfrak{M}$. Here we have given an abstract description of the quantum statistical model; in the following chapters the rules for correspondence between physical and mathematical objects will be elaborated. These rules (based mainly on the concepts of symmetry and covariance) allow to connect at least some theoretical states and measurements to their physical prototypes.

However in general there is no guarantee that any quantum-theoretic state or measurement can be physically implemented, even in principle. Possibility of such implementation requires special study in each particular case. The relevance and the usefulness of quantum theory are ultimately confirmed by the continuing success of its application.

Nevertheless the theoretical concepts of states and measurements reflect the essential features of real physical experiments; any general result obtained in the framework of quantum theory wittingly applies to "implementable" states and measurements in so far as quantum theory gives a correct model for reality. On the other hand, there results could not be obtained without the reference to the general concepts of state and measurement.

1.7. On the problem of "hidden variables"

The discussion in Section 1.5 shows that the reduction of a classical model can lead to convex sets of states, drastically different from the classical simplex. Here we shall demonstrate that in fact any convex set of states can be generated by the reduction of a classical model with an appropriately restricted class of measurements[4]. For simplicity we assume that the convex set is finite-dimensional, and consider only finitely-valued measurements, but the statement and the proof can be generalized to much more general cases.

Theorem 1.7.1. *Any separated statistical model* $(\mathfrak{S}, \mathfrak{M})$*, with* $\mathfrak{S}$ *being a compact subset of a finite-dimensional space,* $\mathfrak{M}$ *being a class of finitely-valued measurements, is a reduction of a classical model with a restricted class of measurements.*

Proof. We shall take for the phase space of the classical system the set Ω of all extreme points of $\mathfrak{S}$, which will be denoted by the letter ω. If $S \to \mu_S$ is a U-valued measurement from $\mathfrak{M}$, then the function $\mu_\omega(u)$;

[4] For a more detailed discussion see the Supplement.

$\omega \in \Omega$, $u \in U$, is a conditional probability distribution[5] on U. Thus to any measurement $\boldsymbol{M} : S \to \mu_S$ there corresponds the conditional probability distribution $\widetilde{\boldsymbol{M}} = \{\mu_\omega(u)\}$, the correspondence being one-to-one: Theorem 1.2.2 implies that the affine functionals coincident on the extreme points coincide on the whole convex set.

Consider the classical model with the set of states $\mathfrak{P}(\Omega)$ and the measurements described by the conditional probability distributions of the form $\widetilde{\boldsymbol{M}} = \{\mu_\omega(u)\}$, corresponding to all measurements $\boldsymbol{M}$ of the initial model. Denote by $\widetilde{\mathfrak{M}}$ the class of such classical measurements. We shall show that the reduction of the model $(\mathfrak{P}(\Omega), \widetilde{\mathfrak{M}})$ gives the model $(\mathfrak{S}, \mathfrak{M})$.

For any probability distribution P on Ω the vector-valued integral

$$\int_\Omega \omega P(\mathrm{d}\omega) \tag{1.7.35}$$

can be defined in a finite-dimensional space, containing the set $\mathfrak{S}$. If $P = \sum_j p_j \delta_{\omega_j}$, then

$$\int_\Omega \omega P(\mathrm{d}\omega) = \sum_j p_j \omega_j, \tag{1.7.36}$$

so that $\int \omega P(\mathrm{d}\omega)$ represents a point in $\mathfrak{S}$, *i.e.*, a state. In general the integral (1.7.35) is the limit of finite convex combinations of the form (1.7.36), and since $\mathfrak{S}$ is compact, the limit also belongs to $\mathfrak{S}$. The integral (1.7.35) is a continuous analog of a convex combination of pure states. Since any affine functional on a finite-dimensional space is evidently continuous then

$$\mu_S(u) = \int_\Omega \mu_\omega(u) P(\mathrm{d}\omega)$$

for any measurement $S \to \mu_S$ if $S = \int \omega P(\mathrm{d}\omega)$.

Let now P_1 and P_2 be two indiscernible classical states on Ω, so that

$$\int_\Omega \mu_\omega(u) P_1(\mathrm{d}\omega) = \int_\Omega \mu_\omega(u) P_2(\mathrm{d}\omega), \qquad u \in U,$$

[5] The requirement of measurability with respect to ω is satisfied since $\mu_\omega(u)$, $\omega \in \Omega$, is the restriction of the affine functional $\mu_S(u)$, $S \in \mathfrak{S}$ to the set of extreme points Ω, which is known to be a Borel set for any compact convex set $\mathfrak{S}$.

for all $\widetilde{\boldsymbol{M}} = \{\mu_\omega(u)\}$. By the foregoing argument this is equivalent to

$$\mu_{S_1}(u) = \mu_{S_2}(u), \qquad u \in U$$

where $S_j = \int \omega P_j(\mathrm{d}\omega)$; $j = 1, 2$. Since the initial statistical model is separated, it follows that

$$S_1 \equiv \int_\Omega \omega P_1(\mathrm{d}\omega) = \int_\Omega \omega P_2(\mathrm{d}\omega) \equiv S_2.$$

Thus, to a class of indiscernible classical states $[P]$ there corresponds the state $\int \omega P(\mathrm{d}\omega) \in \mathfrak{S}$ where P is arbitrary representative of the class $[P]$. This correspondence is apparently one-to-one; moreover, it maps affinely the set of all classes $[P]$ *onto* the set $\mathfrak{S}$, since for any extreme point ω of $\mathfrak{S}$ the class, containing the pure classical state δ_ω corresponds to the state ω. Indeed since by Theorem 1.2.2 any $S \in \mathfrak{S}$ can be represented as $S = \sum p_j \omega_j$, then $S = [\sum p_j \delta_{\omega_j}]$. Thus the reduction of the model $(\mathfrak{P}(\Omega), \widetilde{\mathfrak{M}})$ results in the set of states $\mathfrak{S}$ and the class of measurements $\mathfrak{M}$, so that

$$\int_\Omega \mu_\omega(u) P(\mathrm{d}\omega) = \mu_{[P]}(u), \tag{1.7.37}$$

for any state $S = [P]$ and measurement $S \to \mu_S$. □

The case of quantum statistical model is the most important and we consider the construction of Theorem 1.7.1 for this case in some detail. Let $\widehat{\Sigma}_n$ be the unit sphere in the complex n-dimensional space of column-vectors $|\psi)$,

$$\widehat{\Sigma}_n = \{\psi : (\psi|\psi) = 1\}.$$

The two vectors ψ, ψ' correspond to one and the same pure state $S_\psi = |\psi)(\psi|$ if $\psi = \lambda\psi'$ with $|\lambda| = 1$. Denote by Σ_n the set of the corresponding equivalence classes in $\widehat{\Sigma}_n$. The elements of Σ_n are in one-to-one correspondence with the pure states in $\mathfrak{S}_n$. The set Σ_n will play the role of the phase space for the classical model we are going to describe.

Let $P(\mathrm{d}\psi)$ be a probability distribution on Σ_n. Then the integral

$$S_P = \int_{\Sigma_n} |\psi)(\psi| P(\mathrm{d}\psi)$$

defines a density matrix $S_P \in \mathfrak{S}_n$. In fact, any density matrix is representable in such form. By (1.2.8), $S = \sum \lambda_j S_{\psi_j}$, so that $S = S_P$ with $P = \sum \lambda_j \delta_{\psi_j}$. Thus $P \to S_P$ is an affine map of the simplex $\mathfrak{P}(\Sigma_n)$ onto the set of quantum states $\mathfrak{S}_n$.

Let $S \to \mu_S(u)$, $u \in U$, be a quantum measurement. By Proposition 1.6.1 $\mu_S(u) = \operatorname{Tr} S M_u$, $u \in U$, where $\{M_u\}$ is a resolution of identity. Consider the conditional probability distribution on U:

$$M_\psi(u) \equiv \mu_{S_\psi}(u) = (\psi | M_u \psi). \tag{1.7.38}$$

Then for any measurement

$$\int_{\Sigma_n} M_\psi(u) P(\mathrm{d}\psi) = \mu_{S_P}(u), \qquad u \in U.$$

Thus the quantum statistical model is the reduction of the classical model with the phase space Σ_n and the restricted class of measurements described by transition probabilities of the form (1.7.38) with $\{M_u\}$ being an arbitrary resolution of identity.

Note that in the case $n = 2$ this construction gives the classical model for the spin-$\frac{1}{2}$ particle discussed in Section 1.5. In this case there exists a classical description for preparation of pure states in terms of the Stern-Gerlach filter. If $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3] \in \mathbb{S}^2$ is the direction of a filter, then the corresponding density matrix (1.2.11) is a one-dimensional projection $S_\psi = |\psi)(\psi|$, *i.e.*, a pure state. Therefore we can identify the set of pure states Σ_2 with the unit sphere $\mathbb{S}^2$ and any probability distribution on Σ_2 can be interpreted as a classical state in the experiment, described on the Figure 1.5.

Whether one can give such "physical" interpretation for the formal construction of the Theorem 1.7.1 in cases other than spin-$\frac{1}{2}$ particle depends on the possibility of interpreting an element $\psi \in \Sigma_n$ as "input data" for some classical preparation procedure. Without going into detail we mention that for particle with spin $j > \frac{1}{2}$ there are state vectors ψ which can hardly be interpreted in terms only of polarizing filter as it was done in Section 1.5 for $j = \frac{1}{2}$.

Nevertheless, Theorem 1.7.1 says that at least on the formal level any statistical model is equivalent to a "hidden variables" model. The role of the "variables" plays ω running over the "phase space" Ω, the attribute "hidden" reflecting the incomplete observability which is due to the restrictions onto classical measurements. This statement, of course, does not mean the possibility of reduction of quantum mechanics to a form of Newtonian classical mechanics but it sounds contradictory to the widely known thesis of the impossibility of "hidden variables" description in quantum theory. This thesis suggested by von Neumann also concerns only the statistical description of the results of quantum measurements. The original, rather controversial von Neumann's argument was later substantially modified and brought up to the statute of fairly

nontrivial theorems. In fact there is no contradiction at all. To prove a theorem one has first to formalize the notion of a hidden variables theory. The construction of Theorem 1.7.1 does not fulfil some additional requirements imposed in the existing "no-go" theorems (see the Supplement).

Roughly speaking, one usually requires that quantum observables should be necessarily described by classical random variables in a hidden variables theory, with preservation of some functional relations of logical nature. However the statistical analysis of experimental situations carried out in this chapter suggests that the tenet of observable can be replaced by the broader and more flexible concept of measurement. The conclusion of Theorem 1.7.1 is just a natural consequence of this approach, while the more severe logical restrictions on hidden variables theories seem to lack statistical motivation.

Anyhow, the introduction of the classical description is achieved in Theorem 1.7.1 at the price of drastic increase in the dimensionality of the set of states (from 3 to ∞ in the spin-$\frac{1}{2}$ case). Apparently it does not simplify the description of the object, introducing a lot of details which are not reflected in the measurement statistics. A concise and adequate description of all relevant statistical information is provided by the quantum theory.

The considerations of this chapter are of general nature and apply in fact to any situation in which the validity of the "statistical postulate" is ensured. So far quantum theory provides a unique example of a non-classical statistical model for a class of real objects and phenomena. Are there other fields in which nonclassical models may appear to be relevant? In this connection we wish to recall Bohr's observations concerning the features or "quantum-mechanical" behavior in the animate nature. Even without making any speculative hypothesis about possible mechanisms of these phenomena there is a general reason for expecting "non-classical" relationships in the animate world. As Bohr himself pointed out any observation of a living organism is limited in so far as it inevitably presumes an influence upon the organism. In quantum physics the "elementary" of the observed object is the property which does not allow to neglect the influence of measuring devices. The fundamental property of the living organism is its "wholeness" which excludes arbitrary intervention into the course of biological processes. The complete "classical" analysis of an organism is incompatible with maintaining it alive. A simple junction of admissible factors may occur to be inadmissible; the order of influences may also be important. This shows that a statistical model of a biological object, if it ever will be created, may well be a non-classical one, containing in its very structure the information about fundamental

limitations on observability of the object. Anyhow the classical model, which comes back to the Newtonian mechanics, is by no means a unique possibility in the statistical modelling of non-mechanical objects.

1.8. Comments

Sections 1.1-1.3. In view of the immensity of the literature on basic quantum physics we have to confine ourselves to the few sources which were used in writing this book. These are Dirac [30], Fock [37], Mandelstam [95], Bohm [17], Blokhintzev [13]. The last two books follow the historical development of quantum theory. For this see also the jubilee collections [86, 111] where different contemporary viewpoints are presented. A thorough discussion of an experimental situation with the emphasis on its two-staged character is given by Fock [38].

The description of a measurement in quantum mechanics can be realized with a different degree of detail. There are several levels of description, to each of them corresponds a definite mathematical object in the Hilbert space $\mathcal{H}$ of the system. Let us stress that the term "measurement" as it is used in the present book corresponds to the least detailed description which amounts only to the statistics of outcomes of a measurement procedure for arbitrarily prepared state. Theorem 2.1 of Chapter 2 shows that this is equivalent to defining some resolution of the identity in $\mathcal{H}$. In the modern literature this level of description is usually denoted as "generalized observable" or just "observable" (in the last case orthogonal resolutions of the identity correspond to "standard" or "sharp" observables). The most detailed description requires a dynamical picture of the measurement process, i. e. interaction of the system with a measuring device with subsequent observation of the outcomes, which, in particular, completely determines the statistics of outcomes. The reverse correspondence: "statistics – measurement process" is highly nonunique. Measurement of one and the same observable can be realized via different processes. For more detail see [178, Section 4.1.3]

The concept of statistical model originates largely from the recent attempts of foundation of quantum mechanics using the theory of partially-ordered vector spaces (see Ludwig [90], Hartkämper [49], Neumann [104]; Davies and Lewis [29], Davies [26–28]), though in fact it can be traced back to the earlier axiomatics of Mackey [92]. However, Mackey imposes further axioms aimed to get the "quantum-logic" description, excluding the appearance of non-orthogonal resolutions of identity. The definition of measurement as an affine map of states into the probability distributions was suggested by the author [56]. From the viewpoint adopted here the "state-space" approach to the foundations of quantum

theory (Gudder [46], Krause [82]) is the most relevant. The connection between this approach and the theory of partially-ordered vector spaces is clearly displayed in the paper of Ozawa [106].

Theorem 1.2.2 is due to Minkowski; it was extended to the case of a compact convex set in a general topological linear space by Krein and Milman. For the theory of convexity see Rockafellar [121] (finite-dimensional case), Valentine [134] and Alfsen [4].

For the spectral representation of Hermitean matrices and other topics from linear algebra see, *e.g.*, Halmos [47] and Maltsev [94].

Section 1.4. Since the publication of the basic treatise of Kolmogorov [78], a large amount of textbooks on probability theory have appeared. An accessible introduction to this theory, sufficient for our purposes, is given by Gnedenko [43]. In the classical mathematical statistics one speaks of "strategies" and "decision rules" instead of "measurements". Randomized strategies were introduced by Wald [140], the founder of statistical decision theory (see also Ferguson [34] and Chentzov [22]).

Section 1.5. An elementary quantum-mechanical treatment of the experiments with Stern-Gerlach filters is given in Feynmann's lectures [36]. Other hidden-variables models for the spin-$\frac{1}{2}$ particle were given by Kochen and Specker [77] and Bell [11].

Section 1.6. Fundamentals of the operator formalism of quantum mechanics were given in the classical treatise of Dirac [30]. The ideas of Dirac's approach underlie, in particular, the elementary course of Feynmann [35], based substantially on finite-dimensional spin models. It seems that Feynmann was one of the first to call the probabilists' attention to the interesting problems of connections between quantum mechanics and probability [35].

The first mathematically rigorous treatment of foundations of quantum mechanics, based on the theory of Hilbert space, was given by von Neumann [138]. The notion of density operator was developed in his book, motivated by initial findings of Landau and Weyl. Also the Dirac's concept of observable was made precise by using self-adjoint operators and the corresponding spectral theorem. This book was the starting point for several attempts to axiomatize quantum theory, *i.e.*, to construct a set of simple, physically meaningful postulates which would imply the Hilbert space formalism. Ideally, one would like something as efficient as Kolmogorov's axioms in probability theory. Mackey [92] formulated a set of axioms which leads to "propositional calculus" generalizing σ-fields of probability theory. The problem is then to characterize mathematically the "quantum logic" of projections in a Hilbert space. This problem was

discussed by several authors and was solved by Piron [110]. A detailed exposition of the quantum-logic approach to the foundations of quantum theory is given in the book of Jauch [71], the mathematical aspects of this approach are considered by Varadarajan [135]. Unfortunately, in contrast to Boolean σ-fields which constitute an inseparable part of the classical probability calculus, the quantum logics are rather of independent mathematical interest, the physical theory dealing directly with operators, not "propositions". Moreover, only Mackey's axioms 1-6 which are very close to the definition of a statistical model have an undeniable probabilistic meaning. Other axioms are introduced so as to get the conventional concept of observable. Being good for probability theory, this restricts the concept of quantum measurement to the orthogonal resolutions of identity.

The general resolutions of identity (they are also called positive operator-valued measures in distinction to projection-valued measures which are orthogonal resolutions of identity) where introduced in quantum measurement theory by Davies and Lewis [29] and Holevo [57]. Davies and Lewis came to this generalization by considering repeated quantum measurements. The book of Davies [27] contains an account of the corresponding results in which the state changes due to a measurement play an essential role. The exposition in Section 1.6 follows Holevo [56–60]. This statistical theory of measurement can be generalized to include both classical and quantum theories as extreme particular cases [64].

Section 1.7. An informal discussion of "no-go" results can be found in the survey by Wightman [143] (see also the Supplement). A detailed comparative physical consideration of various hidden-variables theories is contained in the book of Belinfante [10].

The relations between the atomic physics and the biological sciences are discussed in Bohr's philosophical papers [18] (see also Bohm [17]). The celebrated "complementarity principle" sheds light onto the nature of limitations in quantum-mechanical experiments, which however deserves further investigations.

Chapter 2
Mathematics of quantum theory

2.1. Operators in a Hilbert space

In the previous chapter the quantum statistical model was introduced in its simplest finite-dimensional form. However to describe many interesting and important properties of quantum-mechanical objects the infinite-dimensional generalization of this model is needed, where matrices are replaced by operators in a Hilbert space.

A Hilbert space is a complex linear space $\mathscr{H}$ of vectors $\varphi, \psi, \ldots$ with the inner product $(\varphi|\psi)$, which is complete with respect to the metric $\|\varphi - \psi\| = \sqrt{(\varphi - \psi|\varphi - \psi)}$. We shall deal only with separable spaces which have at most countable complete orthonormal system of vectors, *i.e.*, a orthonormal basis. For some reasons to be explained later it will be convenient to take $(\varphi|\psi)$ linear in ψ and conjugate-linear[1] in φ. A typical example is the space $\mathscr{L}^2(a, b)$ of the Lebesgue square-integrable functions on (a, b) with the inner product

$$(\varphi|\psi) = \int_a^b \overline{\varphi(x)}\psi(x)\,\mathrm{d}x.$$

A map $\varphi \to \hat{\varphi}$ of a Hilbert space $\mathscr{H}$ into a Hilbert space $\hat{\mathscr{H}}$ is called *isometric*, if

$$(\varphi|\psi) = (\hat{\varphi}|\hat{\psi}); \qquad \varphi, \psi \in \mathscr{H}.$$

Since it implies $\|\varphi\| \equiv \|\hat{\varphi}\|$, any isometric map is one-to-one. If there is a linear isometric map of $\mathscr{H}$ onto $\hat{\mathscr{H}}$, then $\mathscr{H}$ and $\hat{\mathscr{H}}$ are called *isomorphic*. For example, $\mathscr{L}^2(0, 2\pi)$ is isomorphic to the space l^2 of square-summable sequences $c = [c_k]$ with the inner product $(c|c') =$

[1] Conjugate-linearity means that coefficients in a linear combination of vectors transform into their complex conjugates.

$\Sigma_{k=-\infty}^{\infty} \bar{c}_k c'_k$. The corresponding map is given by

$$c_k = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) \mathrm{e}^{ikx} \, \mathrm{d}x; \qquad k = 0, \pm 1, \ldots,$$

the isometric nature of it is ensured by the Parseval relation. From the point of view of general theory difference between isomorphic Hilbert spaces is inessential, since any proposition about one of the isomorphic spaces can be in principle translated into the terms of another. However in fact the translation may be quite complicated; moreover, an appropriate choice of a concrete Hilbert space may substantially simplify description and study of a mathematical object. For example to study the operator of derivation in $\mathscr{L}^2(0, 2\pi)$ it may be convenient to pass to the space l^2 of the Fourier coefficients etc.

Our choice of the inner product, linear in the second argument, is related to the useful Dirac's notation for the vectors in a Hilbert space. This notation is widely used by physicists and we shall also adopt it.

By the fundamental *Riesz-Frechet lemma* any linear continuous (with respect to the norm $\| \cdot \|$) functional on $\mathscr{H}$ has the form $\varphi \to (\psi|\varphi)$ where ψ is a vector from $\mathscr{H}$ (and vice versa). Therefore the vector ψ can be considered not only as an element of $\mathscr{H}$, but also as an element of the dual space $\mathscr{H}^*$ of linear continuous functional on $\mathscr{H}$. Denote ψ, treated as an element of $\mathscr{H}$, by $|\psi)$, and treated as an element of $\mathscr{H}^*$, by $(\psi|$. Then $|\psi) \to (\psi|$ is an one-to-one conjugate-linear map of $\mathscr{H}$ onto $\mathscr{H}^*$. In the finite-dimensional case $|\psi)$ corresponds to the column-vector

$$\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \end{bmatrix}$$

and $(\psi|$ to the row-vector $[\overline{\psi}_1, \overline{\psi}_2, \ldots]$. As in the finite-dimensional case the symbol for inner product $(\varphi|\psi)$ is the graphic junction of the symbols $(\varphi|$ and $|\psi)$.

The main convenience of the Dirac's notation is the possibility of simple representation of operators in terms of an "outer product". Recall that in the finite-dimensional case the product of a column by a row of the same dimensionality gives a square matrix. We shall denote by $|\varphi_1)(\varphi_2|$ the operator which maps vector $|\psi)$ into the vector $|\varphi_1)(\varphi_2|\psi)$. Thus the action of $|\varphi_1)(\varphi_2|$ on $|\psi)$ is described by a simple graphic junction of the symbols. Operators of this form have *rank* 1, *i.e.*, they map onto a one-dimensional subspace. In particular, the (orthogonal) projection onto the unit vector ψ can be written as

$$S_\psi = |\psi)(\psi|. \tag{2.1.1}$$

Finite linear combinations (or sums, which is the same) of rank-1 operators

$$T = \sum_j |\varphi_j)(\psi_j| \tag{2.1.2}$$

describe *operators of finite rank*. Any finite collection of the finite-rank operators can be considered as acting in a finite-dimensional subspace $\tilde{\mathscr{H}} \subset \mathscr{H}$, therefore algebraic operations for such a collection reduce to those for matrices. The product of two finite-rank operators is given by

$$\left[\sum_j |\varphi_j)(\psi_j|\right] \cdot \left[\sum_k |\hat{\varphi}_k)(\hat{\psi}_k|\right] = \sum_{j,k} |\varphi_j)(\psi_j|\hat{\varphi}_k)(\hat{\psi}_k|. \tag{2.1.3}$$

In the space $\mathscr{L}^2(a, b)$ the finite-rank operators are the integral operators with degenerated kernels; the operator (2.1.2) is described by the kernel

$$T(x', x) = \sum_j \varphi_j(x')\overline{\psi_j(x)}. \tag{2.1.4}$$

Clearly, this class does not include very many operators of interest. One of the difficulties with the infinite-dimensional case is that an operator may be not defined (and not defineable) on the whole $\mathscr{H}$. An example is the operator of derivation in $\mathscr{L}^2(a, b)$. The most important is the class of bounded operators, which are naturally defined on the whole $\mathscr{H}$. The operator X is called *bounded* if

$$\|X\psi\| \leq c\|\psi\|$$

for some constant c and all $\psi \in \mathscr{H}$. Geometrically this means that X transforms norm bounded sets of $\mathscr{H}$ into the norm bounded sets. The least value of c, equal to

$$\|X\| = \sup_{\psi \neq 0} \frac{\|X\psi\|}{\|\psi\|}$$

is called the *norm* of the operator X.

To any bounded X there corresponds the sesquilinear (linear in ψ, conjugate-linear in φ) form on $\mathscr{H}$:

$$X(\varphi, \psi) = (\varphi|X\psi).$$

This relation establishes one-to-one correspondence between bounded operators in $\mathscr{H}$ and sesquilinear forms, continuous in $\varphi, \psi \in \mathscr{H}$. Consider the form $X^*(\varphi, \psi) = \overline{(\psi|X\varphi)}$. The corresponding operator is called *adjoint* to X and is denoted by X^*, so that

$$(X^*\varphi|\psi) = (\varphi|X\psi); \qquad \varphi, \psi \in \mathscr{H}. \tag{2.1.5}$$

The conjugation $X \to X^*$ is a conjugate-linear map, satisfying

$$(XY)^* = Y^*X^*$$

and preserving the norm

$$\|X^*\| = \|X\|. \tag{2.1.6}$$

Moreover,

$$X^{**} \equiv (X^*)^* = X.$$

A reader can check that for the finite-rank operators

$$\left[\sum_j |\varphi_j)(\psi_j|\right]^* = \sum_j |\psi_j)(\varphi_j|.$$

Let U be an isometric operator in $\mathscr{H}$, *i.e.*,

$$(U\varphi|U\psi) = (\varphi|\psi); \qquad \varphi, \psi \in \mathscr{H}.$$

Then U is bounded; by (2.1.5) the last relation can be written as

$$U^*U = I,$$

where I is the *unit operator*, mapping $|\psi)$ into $|\psi)$, $\psi \in \mathscr{H}$. An isometric operator mapping $\mathscr{H}$ onto $\mathscr{H}$ is called *unitary*. The condition of unitarity is

$$U^*U = UU^* = I.$$

A bounded operator X is called *Hermitean* if the corresponding form is Hermitean:

$$(X\varphi|\psi) \equiv \overline{(\psi|X\varphi)} = (\varphi|X\psi); \qquad \varphi, \psi \in \mathscr{H} \tag{2.1.7}$$

i.e., $X^* = X$. There is a polarization identity which linearly expresses a Hermitean form through its "diagonal" values for $\varphi = \psi$; therefore a Hermitean operator X is uniquely defined by the values $(\psi|X\psi)$, $\psi \in \mathscr{H}$. Thus to establish a linear relation between Hermitean forms or operators it is sufficient to check it only for the "diagonal" values, a device we shall often use. The norm of a Hermitean operator can be shown to be expressed through the "diagonal" values by

$$\|X\| = \sup_{\psi \neq 0} \frac{|(\psi|X\psi)|}{\|\psi\|^2}. \tag{2.1.8}$$

In finite-dimensional case this implies that the norm of a Hermitean operator is equal to the maximal absolute value of its eigenvalues. In general, $[-\|X\|, \|X\|]$ is the smallest closed interval, containing the spectrum of X, though the notion of spectrum in the infinite-dimensional case is more complicated (*cf.* Section 2.3).

Let $\mathscr{H}_1$ be a closed subspace of $\mathscr{H}$, then the decomposition into the *direct orthogonal sum* holds

$$\mathscr{H} = \mathscr{H}_1 \oplus \mathscr{H}_2,$$

where $\mathscr{H}_2 = \{\varphi : (\varphi|\psi) = 0, \psi \in \mathscr{H}_1\}$ is the *orthogonal complement* to $\mathscr{H}_1$. For any $\psi = \psi_1 \oplus \psi_2$ put $P\psi = \psi_1$. Then

$$P^2 = P, \qquad P^* = P.$$

Conversely, any operator in $\mathscr{H}$ satisfying these two conditions is the operator of (orthogonal) projection onto the subspace $\mathscr{H}_1 = \{\psi : P\psi = \psi\}$. We shall call such operators simply *projections*.

Let $\mathscr{H}_1$ be a finite-dimensional subspace; then the projection of ψ onto $\mathscr{H}_1$ can be written as

$$|\psi_1) = \sum_j |e_j)(e_j|\psi), \tag{2.1.9}$$

where $\{e_j\}$ is any complete orthonormal system in $\mathscr{H}_1$. Therefore the projection onto $\mathscr{H}_1$ is the finite-rank operator

$$P = \sum_j |e_j)(e_j|. \tag{2.1.10}$$

This relation holds also for infinite-dimensional closed subspace, however the question of convergences arises since an infinite sum is involved. The norm convergence is apparently not relevant here since $\||e_j)(e_j|\| = 1$. Two other types of convergence are useful. A sequence of operators $\{X_n\}$ *converges to X strongly* if $\lim_{n\to\infty} \|X_n\psi - X\psi\| = 0$ for any $\psi \in \mathscr{H}$; it *converges weakly* if $\lim_{n\to\infty}(\varphi|X_n\psi) = (\varphi|X\psi)$ for any φ, $\psi \in \mathscr{H}$. For Hermitean operators this is equivalent to

$$\lim_{n\to\infty} (\psi|X_n\psi) = (\psi|X\psi), \qquad \psi \in \mathscr{H}.$$

The connections between the types of convergence are the following

$$\begin{matrix}\text{norm}\\ \text{convergence}\end{matrix} \Rightarrow \begin{matrix}\text{strong}\\ \text{convergence}\end{matrix} \Rightarrow \begin{matrix}\text{weak}\\ \text{convergence.}\end{matrix}$$

Let $\{e_j\}$ be an orthonormal system in $\mathscr{H}$ and $\mathscr{H}_1$ the closed subspace spanned by $\{e_j\}$. Since for any ψ the vector series (2.1.9) converges in $\mathscr{H}$, the operator series (2.1.10) converges strongly and defines the projection operator onto $\mathscr{H}_1$. In particular, for any orthonormal basis in $\mathscr{H}$

$$I = \sum_j |e_j)(e_j|. \tag{2.1.11}$$

This is only an abbreviated form of the vector relation

$$|\psi) = \sum_j |e_j)(e_j|\psi), \tag{2.1.12}$$

expressing the *completeness* of the system $\{e_j\}$.

Using (2.1.11), we have for any bounded operator X

$$\begin{aligned} X &= \left[\sum_j |e_j)(e_j|\right] X \left[\sum_k |e_k)(e_k|\right] \\ &= \sum_j \sum_k |e_j)(e_j|Xe_k)(e_k|, \end{aligned} \tag{2.1.13}$$

where the series are strongly convergent. The numbers $(e_j|Xe_k)$ are called *matrix elements* of the operator X; this relation gives the decomposition of a bounded operator into the linear combination of operators of rank 1 – the *matrix units* $\{|e_j)(e_k|\}$. If X is a finite-rank operator, then for an appropriate system $\{e_j\}$ it has only a finite number of nonzero matrix elements.

A Hermitean operator X is called *positive* (denoted $X \geq 0$) if

$$(\psi|X\psi) \geq 0; \qquad \psi \in \mathscr{H}.$$

Clearly $X^*X \geq 0$ and $XX^* \geq 0$ for any X; $X \geq Y$ will mean that $X - Y \geq 0$.

The *trace* of a positive operator X is defined as

$$\operatorname{Tr} X = \sum_j (e_j|Xe_j), \tag{2.1.14}$$

where $\{e_j\}$ is an orthonormal basis in $\mathscr{H}$. The series consists of non-negative numbers; as in the finite-dimensional case, the sum does not depend on the choice of $\{e_j\}$, however it can be infinite. Thus, for $X \geq 0$, $0 \leq \operatorname{Tr} X \leq +\infty$.

If X is not positive, then the definition of trace by (2.1.14) may turn out to be incorrect; however as we shall see in Section 1.7 there are trace-class operators for which the trace is unambiguously defined by (2.1.14).

Meanwhile let us notice that this relation gives the correct definition for finite-rank operators. Indeed for any $\{e_j\}$

$$\sum_j (e_j|\varphi)(\psi|e_j) = (\psi|\varphi), \tag{2.1.15}$$

whence

$$\operatorname{Tr}|\varphi)(\psi| = (\psi|\varphi), \tag{2.1.16}$$

so that

$$\operatorname{Tr}\left[\sum_j |\varphi_j)(\psi_j|\right] = \sum_j (\psi_j|\varphi_j). \tag{2.1.17}$$

By (2.1.14), we obtain the relation for the trace of an integral operator in $\mathscr{L}^2(a, b)$ with degenerated kernel $T(x, y)$:

$$\operatorname{Tr} T = \int_a^b T(x, x)\,\mathrm{d}x.$$

Equalities (2.1.17), (2.1.13) imply the important relations

$$\operatorname{Tr} T^* = \overline{\operatorname{Tr} T}, \qquad \operatorname{Tr} TX = \operatorname{Tr} XT, \tag{2.1.18}$$

which will be generalized to a wider class of operators in Section 2.7.

2.2. Quantum states and measurements

A *density operator* is a positive Hermitean operator with unit trace:

$$S \geq 0, \qquad \operatorname{Tr} S = 1. \tag{2.2.19}$$

An example of a density operator is a one-dimensional projection (2.1.1). As we shall see in Section 2.7 any density operator has the spectral representation

$$S = \sum_j s_j |\psi_j)(\psi_j| \tag{2.2.20}$$

analogous to (1.2.10) for the finite-dimensional case. The series is convergent in the operator norm. From (2.2.19) it follows that the eigenvalues $\{s_j\}$ of a density operator satisfy

$$s_j \geq 0, \qquad \sum_j s_j = 1.$$

Denote by $\mathfrak{S}(\mathscr{H})$ the set of all density operators in $\mathscr{H}$; if $\{S_j\} \subset \mathfrak{S}(\mathscr{H})$ and $\{p_j\}$ is a finite probability distribution then the operator

$\sum p_j S_j$ satisfies (2.2.19). Thus $\mathfrak{S}(\mathscr{H})$ is a convex set. The density operators represent states in quantum mechanics. It follows from (2.2.20) that Proposition 1.2.3 can be generalized to the infinite-dimensional case, so that *the extreme points of* $\mathfrak{S}(\mathscr{H})$ *are the one-dimensional projections*. The corresponding states are called *pure*.

Now we pass to quantum measurements. Let U be a measurable space of outcomes of the measurement, *e.g.*, a finite set or a domain in $\mathbb{R}^n$ with the Borel σ-field. Following Section 1.6 we call a *quantum measurement* with values in U (U-measurement, for short) an affine map $S \to \mu_S(\mathrm{d}u)$ of the convex set of quantum states $\mathfrak{S}(\mathscr{H})$ into the set of all probability distributions on U. $\mu_S(\mathrm{d}u)$ is interpreted as the probability distribution of the results of the measurement in the state S. A generalization of Proposition 1.6.1 will describe quantum measurements in terms of *resolution of identity*; by this we mean a collection $\boldsymbol{M} = \{M(B);\, B \in \mathscr{A}(U)\}$ of Hermitean operators in $\mathscr{H}$, satisfying

(1) $M(\phi) = 0, \quad M(U) = I;$
(2) $M(B) \geq 0, \quad B \in \mathscr{A}(U);$
(3) for any at most countable decomposition $\{B_j\}$ of $B \in \mathscr{A}(U)$, $M(B) = \sum_j M(B_j)$ holds where the series is weakly convergent.

These conditions formally remind the definition of probability measure, and resolutions of identity are sometimes called probability operator-valued measures. If U is a finite set and $\{M_u;\, u \in U\}$ is a collection of Hermitean operators satisfying (1.6.28), *i.e.*, a finite resolution of identity in the sense of Section 1.6, then the relation

$$M(B) = \sum_{u \in B} M_u, \qquad B \subset U,$$

defines a operator-valued measure on the Boolean field of all subsets of U, *i.e.*, a resolution of identity in the sense of the definition given above, and conversely.

A particular but very important class constitute *orthogonal resolutions of identity*, satisfying the additional requirement

$$M(B_1)M(B_2) = 0 \qquad \text{if } B_1 \cap B_2 = \emptyset.$$

As in Section 1.6 one can show that this is equivalent to

$$M(B)^2 = M(B), \qquad B \in \mathscr{A}(U),$$

that is the measure $\{M(B)\}$ is projection-valued.
Now we formulate the generalization of Proposition 1.6.1, which will be proved in Section 2.7.

Theorem 2.2.1. *Let $S \to \mu_S$ be a U-measurement. Then there exists a unique resolution of identity $M = \{M(B);\ B \in \mathscr{A}(U)\}$ in $\mathscr{H}$ such that for any state S*

$$\mu_S(B) = \operatorname{Tr} SM(B), \qquad B \in \mathscr{A}(U). \tag{2.2.21}$$

Conversely, any resolution of identity defines an U-measurement by (2.2.21).

The relation (2.2.21) will sometimes be symbolically written in the form

$$\mu_S(\mathrm{d}u) = \operatorname{Tr} SM(\mathrm{d}u).$$

Notice that the right-hand side of (2.2.21) is the trace of a generally non-positive (and non-Hermitean) operator, which, however, is a trace-class as it will be shown in Section 2.7. Meanwhile we remark that for a pure state

$$\mu_{S_\psi}(B) = (\psi|M(B)\psi),$$

since $\operatorname{Tr}|\psi)(\psi|X = (\psi|X\psi)$ by (2.1.16).

Measurements described by orthogonal resolutions of identity will be called *simple*. Any simple measurement is an extreme point of the convex set $\mathfrak{M}(U)$ of all U-measurements, the proof being quite the same as in Proposition 1.6.5; the converse, of course, is not true.

The most important case is when the results of a measurement are real numbers. Then the simple measurements are described in terms of observables, *i.e.*, Hermitean operators in $\mathscr{H}$, which are quantum analogs of classical random variables. The following section is devoted to a more detailed study of this connection.

2.3. Spectral representation of bounded operators

In a finite-dimensional case any Hermitean operator X admits the spectral representation

$$X = \sum_k \lambda_k E_k, \tag{2.3.22}$$

where E_k are the projections onto the invariant subspaces corresponding to the distinct eigenvalues λ_k. The collection $\{E_k\}$ forms an orthogonal resolution of identity, so that

$$\sum_k E_k = I; \qquad E_j E_k = \delta_{jk} E_j. \tag{2.3.23}$$

An infinite-dimensional analog of (2.3.22) holds for compact Hermitean operators (see Section 2.7). However there are plenty of noncompact

Hermitean operators. In fact, a Hermitean operator need not have an eigenvector, as the following example shows.

Consider the operator Q of multiplication by the argument x in the space $\mathscr{L}^2(a, b)$ where (a, b) is a bounded interval. The equation

$$x\psi(x) = \xi\psi(x) \tag{2.3.24}$$

has the continual collection of formal solutions

$$\psi_\xi(x) \sim \delta(x - \xi); \qquad a < \xi < b, \tag{2.3.25}$$

which however do not correspond to any nonzero vectors in $\mathscr{H}$.

Nevertheless any Hermitean operator in a Hilbert space has a spectral representation in which a continuous analog of the sum (2.3.22) emerges. To explain the transition from sum to integral introduce the orthogonal resolution of identity on the real line $\mathbb{R}$, by putting

$$E(B) = \sum_{k:\lambda_k \in B} E_k; \qquad B \in \mathscr{A}(\mathbb{R}).$$

($\mathscr{A}(\mathbb{R})$ is the Borel σ-field of $\mathbb{R}$). Formally

$$E(\mathrm{d}\lambda) = \left[\sum_k \delta(\lambda - \lambda_k) E_k\right] \mathrm{d}\lambda. \tag{2.3.26}$$

Then the relation (2.3.22) can be rewritten as

$$X = \int \lambda E(\mathrm{d}\lambda). \tag{2.3.27}$$

Now let $E(\mathrm{d}\lambda)$ be an arbitrary orthogonal resolution of identity on $\mathbb{R}$. In this section we suppose that it is concentrated on a bounded subset Λ of $\mathbb{R}$, so that $E(\Lambda) = I$. Then for any $\psi \in \mathscr{H}$ the probability distribution $\mu_\psi(\mathrm{d}\lambda) = \operatorname{Tr} S_\psi E(\mathrm{d}\lambda) = (\psi|E(\mathrm{d}\lambda)\psi)$ will be concentrated on Λ and therefore the integral

$$\int \lambda(\psi|E(\mathrm{d}\lambda)\psi) = \int \lambda\mu_\psi(\mathrm{d}\lambda) \tag{2.3.28}$$

converges. This integral defines a continuous Hermitean form on $\mathscr{H}$, to which corresponds the Hermitean operator X such that

$$(\psi|X\psi) = \int \lambda(\psi|E(\mathrm{d}\lambda)\psi), \qquad \psi \in \mathscr{H}. \tag{2.3.29}$$

Thus (2.3.27) holds where the integral is weakly convergent in the sense that (2.3.29) holds. (In fact, one can show strong convergence).

Theorem 2.3.1 (Spectral theorem for bounded operators). *The relation* (2.3.27) *establishes one-to-one correspondence between the Hermitean operators X and the orthogonal resolutions of identity in $\mathscr{H}$ concentrated on bounded subsets of* $\mathbb{R}$.

The resolution of identity $E(\mathrm{d}\lambda)$ is also called the spectral measure of the operator X.

Return for a while to operators, corresponding to finite resolutions of identity (2.3.26). The relations (2.3.23) imply that for any polynomial $p(\lambda)$

$$p(X) = \sum_k p(\lambda_k) E_k = \int p(\lambda) E(\mathrm{d}\lambda).$$

In the general case, approximating integrals by sums, one shows that for any Hermitean X

$$p(X) = \int p(\lambda) E(\mathrm{d}\lambda).$$

Therefore one can define a Hermitean operator $f(X)$ where f is a bounded measurable function, by the relation

$$(\psi | f(X)\psi) = \int f(\lambda)(\psi | E(\mathrm{d}\lambda)\psi); \qquad \psi \in \mathscr{H}.$$

In particular, putting $f(x) = \mathbf{1}_B(x)$ we get the expression for the spectral projections of X

$$E(B) = \mathbf{1}_B(X); \qquad B \in \mathscr{A}(\mathbb{R}).$$

The equality

$$X^2 = \int \lambda^2 E(\mathrm{d}\lambda)$$

implies the important relation

$$\|X\psi\|^2 = \int \lambda^2 (\psi | E(\mathrm{d}\lambda)\psi).$$

The correspondence $f \to f(X)$ preserves algebraic relations and the ordering: $f(x) \geq 0, x \in \mathbb{R}$ implies $f(X) \geq 0$. Thus putting

$$|X| = \int |\lambda| E(\mathrm{d}\lambda), \tag{2.3.30}$$

we have $\pm X \leq |X|$ since $\pm x \leq |x|$ for all $x \in \mathbb{R}$.

As an example consider the operator Q in $\mathscr{L}^2(a,b)$. We have

$$Q = \int_a^b \xi\, E(\mathrm{d}\xi), \tag{2.3.31}$$

where $E(B) = \mathbf{1}_B(x)$. Indeed (2.3.31) means that

$$(\psi|Q\psi) = \int_a^b \xi\, \mu_\psi(\mathrm{d}\xi), \tag{2.3.32}$$

where

$$\begin{aligned}\mu_\psi(B) &= (\psi|E(B)\psi)\\ &= \int_a^b \mathbf{1}_B(x)|\psi(x)|^2\mathrm{d}x = \int_B |\psi(x)|^2\,\mathrm{d}x,\end{aligned}$$

so that $\mu_\psi(\mathrm{d}\xi) = |\psi(\xi)|^2\mathrm{d}\xi$ and (2.3.31) is obvious. Thus the collection of projections $\{\mathbf{1}_B(\cdot);\ B \in \mathscr{A}\ ((a,b))\}$ is the spectral measure of the operator of multiplication by x.

The operator Q is a typical example of an operator with "continuous spectrum". We have seen that it has no eigenvectors in $\mathscr{L}^2(a,b)$ though the equation (2.3.24) has the intuitively clear formal solutions (2.3.25). In this connection let us discuss Dirac's notation pertaining to operators with continuous spectrum. Following Dirac, denote by $|\xi)$ the formal eigenfunction (2.3.25). The family $\{|\xi);\ \xi \in (a,b)\}$ is "orthonormal" in the sense that

$$(\xi'|\xi) = \delta(\xi - \xi'),$$

and satisfies the formal completeness relation

$$\int_a^b |\xi)(\xi|\mathrm{d}\xi = I. \tag{2.3.33}$$

In fact (2.3.33) is just an abbreviated form of the equality

$$\int_a^b (\psi|\xi)(\xi|\psi)\mathrm{d}\xi = (\psi|\psi), \qquad \psi \in \mathscr{H},$$

which is meaningful if the symbol $(\xi|\psi)$ is interpreted as

$$(\xi|\psi) = \int_a^b \delta(x-\xi)\psi(x)\mathrm{d}x = \psi(\xi). \tag{2.3.34}$$

Thus the family $\{|\xi)\}$ is a formal continual analog of complete orthonormal system, the relation (2.3.33) being the analog of the com-

pleteness relation (2.3.32). The spectral representation of Q in Dirac's notation has the form

$$Q = \int_a^b \xi |\xi)(\xi| \mathrm{d}\xi \tag{2.3.35}$$

and looks like a direct continual analog of the spectral representation (1.2.7) for discrete spectrum. In fact (2.3.35) means that

$$(\psi|Q\psi) = \int_a^b \xi (\psi|\xi)(\xi|\psi)\mathrm{d}\xi, \qquad \psi \in \mathscr{H},$$

and this is equivalent to (2.3.32). Thus we obtain the relation between (2.3.31) and (2.3.35) by putting

$$E(\mathrm{d}\xi) = |\xi)(\xi|\mathrm{d}\xi.$$

Needless to say that this relation cannot be taken literally – the projection-valued measure $E(\mathrm{d}\xi)$ is not differentiable (has no operator density) with respect to Lebesgue measure $\mathrm{d}\xi$. However the scalar measures $\mu_\psi(\mathrm{d}\xi) = (\psi|E(\mathrm{d}\xi)\psi)$ are differentiable with respect to $\mathrm{d}\xi$ with

$$\mu_\psi(\mathrm{d}\xi) = |\psi(\xi)|^2 \mathrm{d}\xi = (\psi|\xi)(\xi|\psi)\mathrm{d}\xi. \tag{2.3.36}$$

The intuitive maintenance of Dirac's formalism gives certain advantages in presentation of quantum theory. It can be properly used in the framework within which it is equivalent to the spectral representation.

2.4. Spectral representation of unbounded operators

For quantum mechanics it is important to have a generalization of the spectral representation for unbounded operators. Let X be an operator defined on its domain $\mathscr{D}(X) \subseteq \mathscr{H}$; for unbounded operators domain considerations are essential since, with an immaterial exception, such an operator cannot be defined on the whole $\mathscr{H}$. The operator is called *symmetric* if the corresponding sesquilinear form is Hermitean on $\mathscr{D}(X)$, *i.e.*,

$$(X\varphi|\psi) = (\varphi|X\psi); \qquad \varphi, \psi \in \mathscr{D}(X). \tag{2.4.37}$$

The most important is the concept of a self-adjoint operator. Let the domain of X be dense in $\mathscr{H}$; denote by $\mathscr{D}(X^*)$ the subspace of vectors φ, for which $\psi \to (\varphi|X\psi)$ is a continuous functional of ψ. By the Riesz-Frechet lemma, $(\varphi|X\psi) = (\varphi^*|\psi)$, where φ^* is defined uniquely since ψ runs over the dense set $\mathscr{D}(X)$. Denote by X^* the operator $\varphi \to \varphi^*$ with the domain $\mathscr{D}(X^*)$. It is called the *adjoint* to X. Thus

$$(X^*\varphi|\psi) = (\varphi|X\psi); \qquad \psi \in \mathscr{D}(X), \quad \varphi \in \mathscr{D}(X^*).$$

In particular, X is symmetric if $X \subseteq X^*$, *i.e.*, $\mathscr{D}(X) \subseteq \mathscr{D}(X^*)$ and $X\psi = X^*\psi$ on $\mathscr{D}(X)$.

A densely defined operator X is called *self-adjoint* if $X = X^*$, *i.e.*, $\mathscr{D}(X^*) = \mathscr{D}(X)$ and (2.4.37) holds.

Sometimes symmetric operator X can be extended to a self-adjoint operator. If this can be done uniquely, then X is called *essentially self-adjoint*. A symmetric operator which does not have self-adjoint extensions is called *maximal*.

The following important examples illustrate these notions. Let Q be the (unbounded) operator of multiplication by x in $\mathscr{L}^2(\mathbb{R})$ with

$$\mathscr{D}(Q) = \left\{\psi : \int_{-\infty}^{\infty} x^2|\psi(x)|^2\mathrm{d}x < \infty\right\}.$$

For any $\varphi, \psi \in \mathscr{D}(Q)$ one has

$$(\varphi|Q\psi) = \int x\overline{\varphi(x)}\psi(x)\mathrm{d}x = (Q\varphi|\psi),$$

so that Q is symmetric. In fact it is self-adjoint; to prove it consider $\varphi \in \mathscr{D}(Q^*)$. Then $\int x\overline{\varphi}\psi\,\mathrm{d}x$ is a continuous functional of ψ and by the Riesz-Frechet lemma $\int x\overline{\varphi}\psi\,\mathrm{d}x = \int \overline{h}\psi\,\mathrm{d}x$ where $h \in \mathscr{L}^2(\mathbb{R})$. It follows that $h = x\varphi$ and $\int |x\varphi(x)|^2\mathrm{d}x < \infty$, so that $\mathscr{D}(Q^*) = \mathscr{D}(Q)$. The same operator considered on a smaller dense subspace $\mathscr{D} \subset \mathscr{D}(Q)$ is easily seen to be essentially self-adjoint.

Next consider the operator $P = \mathrm{i}^{-1}\,\mathrm{d}/\mathrm{d}x$ with the domain

$$\mathscr{D}(P) = \left\{\psi : \int_{-\infty}^{\infty} \left|\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right|^2 \mathrm{d}x < \infty\right\}.$$

The Fourier transform

$$\tilde{\psi}(\eta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\eta x}\psi(x)\,\mathrm{d}x$$

maps isometrically $\mathscr{L}^2(\mathbb{R})$ onto $\mathscr{L}^2(\mathbb{R})$, the operator P is transformed into the operator of multiplication by η with $\mathscr{D}(P)$ mapped onto the subspace $\{\tilde{\psi} : \int |\eta\hat{\psi}(\eta)|^2\,\mathrm{d}\eta < \infty\}$. From what is proved above for the multiplication operator it follows that P is also self-adjoint.

The next is an example of a symmetric operator which has no self-adjoint extension. Let $\mathscr{H} = \mathscr{L}^2(0,\infty)$ and consider the operator $P_+ = \mathrm{i}^{-1}\mathrm{d}/\mathrm{d}x$ with the domain

$$\mathscr{D}(P_+) = \left\{\psi : \psi(0) = 0, \int_0^{\infty} \left|\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right|^2 \mathrm{d}x < \infty\right\}. \qquad (2.4.38)$$

Then

$$\begin{aligned}(\varphi|P_+\psi) &= \mathrm{i}^{-1}\int_0^\infty \overline{\varphi(x)}\,\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\,\mathrm{d}x \\ &= \mathrm{i}^{-1}\overline{\varphi(0)}\psi(0) - \mathrm{i}^{-1}\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}x}\overline{\varphi(x)}\psi(x)\,\mathrm{d}x \\ &= (P_+^*\varphi|\psi), \qquad \varphi \in \mathscr{D}(P_+^*),\end{aligned} \tag{2.4.39}$$

where $P_+^* = \mathrm{i}^{-1}\mathrm{d}/\mathrm{d}x$ with

$$\mathscr{D}(P_+^*) = \left\{\psi : \int_0^\infty \left|\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right|^2 \mathrm{d}x < \infty\right\}.$$

Then $P_+ \subset P_+^*$, $P_+ \neq P_+^*$, since $\mathscr{D}(P_+) \neq \mathscr{D}(P_+^*)$. The operator P_+ is symmetric but P_+^* is not since if we try to compute $(P_+^*)^*$ by the formula analogous to (2.4.39), the term $\mathrm{i}^{-1}\overline{\varphi(0)}\psi(0)$ will be nonzero, yielding

$$(\varphi|P_+^*\psi) \neq (P_+^*\varphi|\psi); \qquad \varphi, \psi \in \mathscr{D}(P_+^*).$$

One can show that P_+ is maximal.

Now we pass to the spectral representation of self-adjoint operators. Consider the integral of the type (2.3.27), where $E(\mathrm{d}\lambda)$ is an arbitrary orthogonal resolution of identity on $\mathbb{R}$. One cannot expect that the integral in the right-hand side of (2.3.29) will converge for all $\psi \in \mathscr{H}$, but it certainly converges if ψ belongs to the subspace

$$\mathscr{D} = \left\{\psi : \int_{-\infty}^\infty \lambda^2(\psi|E(\mathrm{d}\lambda)\psi) < \infty\right\}. \tag{2.4.40}$$

One shows that $\mathscr{D}$ is dense in $\mathscr{H}$ and the Hermitean form (2.3.28) defines a self-adjoint operator X with $\mathscr{D}(X) = \mathscr{D}$, so that

$$(\psi|X\psi) = \int \lambda(\psi|E(\mathrm{d}\lambda)\psi), \qquad \psi \in \mathscr{D}, \tag{2.4.41}$$

$$\|X\psi\|^2 = \int \lambda^2(\psi|E(\mathrm{d}\lambda)\psi), \qquad \psi \in \mathscr{D}. \tag{2.4.42}$$

The last equality explains why it must be $\mathscr{D}(X) = \mathscr{D}$.

Theorem 2.4.1 (Spectral theorem for self-adjoint operators). *The relations* (2.4.41) *and* (2.4.40) *establish the one-to-one correspondence between self-adjoint operators and orthogonal resolutions of identity* (*spectral measures*) *in* $\mathscr{H}$.

Consider the self-adjoint operator Q of multiplication by x in $\mathscr{L}^2(\mathbb{R})$. The same argument as in Section 2.3 shows that its spectral measure is

$$E(B) = \mathbf{1}_B(x); \qquad B \in \mathscr{A}(\mathbb{R}), \tag{2.4.43}$$

or, formally $E(\mathrm{d}\xi) = |\xi)(\xi|\mathrm{d}\xi$, where $|\xi)$, $\xi \in \mathbb{R}$ are the formal eigenvectors of the operator of multiplication, so that the Dirac's notation

$$Q = \int_{-\infty}^{\infty} \xi |\xi)(\xi| \mathrm{d}\xi.$$

Consider now the operator $P = \mathrm{i}^{-1}\,\mathrm{d}/\mathrm{d}x$ in $\mathscr{L}^2(\mathbb{R})$ with the domain (2.4.39). This operator has the family of formal eigenfunctions

$$\frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}\eta x}; \qquad \eta \in \mathbb{R},$$

satisfying the conditions of orthonormality and completeness of the type (2.3.33). Denoting the formal eigenvectors by $|\eta)$; $\eta \in \mathbb{R}$ one may expect the spectral representation

$$P = \int_{-\infty}^{\infty} \eta |\eta)(\eta| \mathrm{d}\eta.$$

To give these considerations precise meaning, consider the Fourier transform $\tilde{\psi}(\eta) = (\eta|\psi)$. In Dirac's notations the Fourier transform corresponds to the change of the "orthonormal basis" $\{|\xi)\}$ to $\{|\eta)\}$,

$$(\eta|\psi) = \int (\eta|\xi)(\xi|\psi)\,\mathrm{d}\xi$$

since

$$(\eta|\xi) = \frac{1}{\sqrt{2\pi}} \int \mathrm{e}^{-\mathrm{i}\eta x} \delta(\xi - x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\mathrm{i}\eta\xi}.$$

Since the operator P is transformed into the operator of multiplication by η, the spectral measure $F(\mathrm{d}\eta)$ of the operator P is obtained by inverse Fourier transform of the spectral measure (2.4.43), namely

$$(\psi|F(B)\psi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_B \overline{\psi(x)} \psi(x') \mathrm{e}^{\mathrm{i}\eta(x-x')}\,\mathrm{d}\eta\,\mathrm{d}x\,\mathrm{d}x'. \tag{2.4.44}$$

Formally this can be written as

$$\begin{aligned}(\psi|F(\mathrm{d}\eta)\psi) &= \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{\psi(x)} \psi(x') \mathrm{e}^{\mathrm{i}\eta(x-x')} \mathrm{d}x\,\mathrm{d}x'\right] \mathrm{d}\eta \\ &= (\psi|\eta)(\eta|\psi)\,\mathrm{d}\eta,\end{aligned} \tag{2.4.45}$$

or $F(\mathrm{d}\eta) = |\eta)(\eta|\mathrm{d}\eta$, which is consistent with the formal spectral representation of P.

All that was said in Section 2.3 concerning the functions of Hermitean operators applies with obvious modifications to self-adjoint operators. Consider in particular the function

$$\mathrm{e}^{\mathrm{i}\theta X} = \int \mathrm{e}^{\mathrm{i}\theta\lambda} E(\mathrm{d}\lambda). \tag{2.4.46}$$

One checks that the family $\{V_\theta = \exp \mathrm{i}\theta X;\ \theta \in \mathbb{R}\}$ is a *group of unitary operators* (unitary group, for short), *i.e.*,

$$V_{\theta_1+\theta_2} = V_{\theta_1} V_{\theta_2}, \qquad V_\theta^* = V_{-\theta}, \qquad V_0 = I. \tag{2.4.47}$$

One can also show that the family $\{V_\theta\}$ is strongly continuous in θ, *i.e.*, for any $\psi \in \mathscr{H}$ the relation

$$\psi_\theta = \mathrm{e}^{\mathrm{i}\theta X}\psi; \qquad \theta \in \mathbb{R}, \tag{2.4.48}$$

defines a continuous curve in the Hilbert space $\mathscr{H}$. Moreover, *this curve is differentiable in* $\mathscr{H}$ *if and only if* $\psi \in \mathscr{D}(X)$; in this case $\{\psi_\theta\} \subset \mathscr{D}(X)$ and

$$\frac{\mathrm{d}\psi_\theta}{\mathrm{d}\theta} = \mathrm{i}X\psi_\theta. \tag{2.4.49}$$

The converse statement is known as Stone's theorem.

Theorem 2.4.2. *Any strongly* (*or weakly*) *continuous unitary group has the form* $V_\theta = \exp \mathrm{i}\theta X$ *where* X *is a uniquely defined self-adjoint operator.*

The operator X is called the *infinitesimal generator* of the group $\{V_\theta\}$. Let us calculate $\{\exp \mathrm{i}\theta P\}$ in $\mathscr{L}^2(\mathbb{R})$. Using (2.4.46) and (2.4.45) we obtain

$$\begin{aligned}(\psi|\mathrm{e}^{\mathrm{i}\theta P}\psi) &= \frac{1}{2\pi}\iiint \overline{\psi(x)}\psi(x')\mathrm{e}^{\mathrm{i}\theta\eta}\mathrm{e}^{\mathrm{i}\eta(x-x')}\mathrm{d}\eta\,\mathrm{d}x\,\mathrm{d}x' \\ &= (\psi|\psi_\theta),\end{aligned}$$

where $\psi_\theta(x) = \psi(x+\theta)$. Thus

$$\mathrm{e}^{\mathrm{i}\theta P}\psi(x) = \psi(x+\theta), \qquad \theta \in \mathbb{R}.$$

Now we pass to the spectral representation of symmetric but not necessarily self-adjoint operators.

Theorem 2.4.3. *For a densely defined symmetric operator X there exists in general non-unique resolution of identity (spectral measure) $M(\mathrm{d}\lambda)$, such that*

$$\begin{aligned} \mathscr{D}(x) &\subseteq \left\{\psi : \int \lambda^2 (\psi|M(\mathrm{d}\lambda)\psi) < \infty\right\}; \\ (\psi|X\psi) &= \int \lambda(\psi|M(\mathrm{d}\lambda)\psi), \qquad \psi \in \mathscr{D}(X); \\ \|X\psi\|^2 &= \int \lambda^2(\psi|M(\mathrm{d}\lambda)\psi), \qquad \psi \in \mathscr{D}(X). \end{aligned} \tag{2.4.50}$$

If X is maximal, and $\subseteq$ is changed to $=$ in the first relation, then $M(\mathrm{d}\lambda)$ is unique.

As an example we consider the operator P_+ in $\mathscr{L}^2(0,\infty)$, and show that its (unique) spectral measure can be represented as

$$M(\mathrm{d}\eta) = |\eta_+)(\eta_+|\mathrm{d}\eta, \tag{2.4.51}$$

where $|\eta_+)$ are the formal eigenfunctions $(1/\sqrt{2\pi})\mathrm{e}^{\mathrm{i}\eta x}$, $\eta \in \mathbb{R}$, which differ from $|\eta)$ only in that the argument x varies from 0 to ∞. The family $\{|\eta_+)\}$ obviously satisfies the formal completeness relation

$$\int_{-\infty}^{\infty} |\eta_+)(\eta_+|\mathrm{d}\eta = I, \tag{2.4.52}$$

but is not "orthogonal", since

$$\begin{aligned} (\eta'_+|\eta_+) &= \frac{1}{2\pi}\int_0^\infty \mathrm{e}^{\mathrm{i}(\eta-\eta')x}\mathrm{d}x \\ &= \frac{1}{2}\delta(\eta-\eta') + \frac{1}{2\pi\mathrm{i}}(\eta-\eta')^{-1}. \end{aligned} \tag{2.4.53}$$

Therefore (2.4.51) is not an orthogonal resolution of identity. Such families as $\{|\eta_+)\}$ are sometimes called "overcomplete".

The correct version of (2.4.51) can be given by the relation of the type (2.4.44):

$$(\psi|M(B)\psi) = \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_B \overline{\psi(x)}\psi(x')\mathrm{e}^{\mathrm{i}\eta(x-x')}\mathrm{d}x\mathrm{d}x'\mathrm{d}\eta. \tag{2.4.54}$$

Extending the functions defined on $(0,\infty)$ by making them equal to zero on $(-\infty, 0)$ we get a natural embedding of $\mathscr{L}^2(0,\infty)$ into $\mathscr{L}^2(\mathbb{R})$.

$M(\mathrm{d}\eta)$ is the restriction onto $\mathscr{L}^2(0,\infty)$ of the spectral measure $F(\mathrm{d}\eta)$ of the operator P, *i.e.*,

$$(\psi|M(\mathrm{d}\eta)\psi) = (\tilde{\psi}|F(\mathrm{d}\eta)\tilde{\psi}),$$

where $\tilde{\psi}$ is the extension of $\psi \in \mathscr{L}^2(0,\infty)$. If $\psi \in \mathscr{D}(P_+)$, then $\tilde{\psi} \in \mathscr{D}(P)$ with $\widetilde{P}_+\psi = P\tilde{\psi}$. Therefore

$$(\psi|P_+\psi) = (\tilde{\psi}|P\tilde{\psi}) = \int \eta(\tilde{\psi}|F(\mathrm{d}\eta)\tilde{\psi}) = \int \eta(\psi|M(\mathrm{d}\eta)\psi),$$
$$\|P_+\psi\|^2 = \|P\tilde{\psi}\|^2 = \int \eta^2(\tilde{\psi}|F(\mathrm{d}\eta)\tilde{\psi}) = \int \eta^2(\psi|M(\mathrm{d}\eta)\psi),$$

which shows that $M(\mathrm{d}\eta)$ satisfies the relations (2.4.50), characterizing the spectral measure of the operator P_+.

2.5. On realization of measurement

We have just seen that a nonorthogonal resolution of identity $\{M(B)\}$ in $\mathscr{H}$ may arise as a restriction of an orthogonal resolution of identity $\{E(B)\}$ in a larger Hilbert space $\tilde{\mathscr{H}}$:

$$M(B) = \widetilde{E}E(B)\widetilde{E}; \qquad B \in \mathscr{A}(U), \tag{2.5.55}$$

where $\widetilde{E}$ is the projection from $\tilde{\mathscr{H}}$ onto $\mathscr{H}$. It is easy to see that restricting in this way any resolution of identity in $\tilde{\mathscr{H}}$ one gets a resolution of identity in $\mathscr{H}$. The converse statement was proved by Naimark.

Theorem 2.5.1. *Any resolution of identity $\{M(B)\}$ in $\mathscr{H}$ can be dilated to an orthogonal resolution of identity in a larger Hilbert space $\tilde{\mathscr{H}}$, so that* (2.5.55) *will hold.*

We shall give the proof of this theorem only for finite resolutions of identity since it is technically simple and clearly indicates the idea of the proof in the general case.

Let $\mathscr{L}$ be a linear space; by a *pre-inner product* we mean a form $(\varphi|\psi)$; $\varphi, \psi \in \mathscr{L}$, possessing all properties of inner product except that $(\varphi|\varphi)$ may be equal to zero for a nonzero $\varphi \in \mathscr{L}$. There is a standard procedure of constructing a Hilbert space given a pre-inner product on $\mathscr{L}$: let $\mathscr{N} = \{\psi : (\psi|\psi) = 0\}$, then $(\cdot|\cdot)$ uniquely defines the inner product on the factor space $\mathscr{L}/\mathscr{N}$. The completion of $\mathscr{L}/\mathscr{N}$ with respect to this inner product is a Hilbert space which we call *completion of $\mathscr{L}$ with respect to the pre-inner product* $(\cdot|\cdot)$.

Proof of Theorem 2.5.1. Let $\boldsymbol{M} = \{M_u;\, u = 1, \ldots, m\}$ be a finite resolution of identity in $\mathscr{H}$. Let $\mathscr{L}$ be the direct sum of m copies of $\mathscr{H}$, so that the vectors of $\mathscr{L}$ are $\boldsymbol{\Psi} = [\psi_u;\, u = 1, \ldots, m]$. Then the form

$$(\boldsymbol{\Psi}|\boldsymbol{\Psi}')^{\sim} = \sum_{u=1}^{m} (\psi_u | M_u \psi'_u); \qquad \boldsymbol{\Psi}, \boldsymbol{\Psi}' \in \mathscr{L},$$

is clearly a pre-inner product, positivity following from the fact that $M_u \geq 0$. Let $\tilde{\mathscr{H}}$ be the completion of $\mathscr{L}$ with respect to $(\cdot|\cdot)^{\sim}$. The map $\varphi \to \boldsymbol{\Psi}_\varphi = [\varphi;\, u = 1, \ldots, m]$ from $\mathscr{H}$ into $\tilde{\mathscr{H}}$ is isometric since

$$(\boldsymbol{\Psi}_\varphi | \boldsymbol{\Psi}'_\varphi)^{\sim} = \sum_{u=1}^{m} (\varphi | M_u \varphi') = (\varphi|\varphi'); \qquad \varphi, \varphi' \in \mathscr{H}.$$

Identifying the image of $\mathscr{H}$ under this map with $\mathscr{H}$, we can assume that $\mathscr{H} \subset \tilde{\mathscr{H}}$. Define the operator E_v in $\mathscr{L}$ which cuts all the components of $\boldsymbol{\Psi} = [\psi_u]$ to zero except for ψ_v which remains unchanged. Then it is easy to see that operators $\{E_v;\, v = 1, \ldots, m\}$ give rise to an orthogonal resolution of identity in $\tilde{\mathscr{H}}$. For $\varphi \in \mathscr{H}$

$$(\varphi | M_u \varphi) = (\boldsymbol{\Psi}_\varphi | E_u \boldsymbol{\Psi}_\varphi)^{\sim}$$

which means that $\{E_u\}$ is a required dilation of $\{M_u\}$. The theorem is proved. □

Basing on Theorem 2.5.1 we shall show that any quantum measurement $\boldsymbol{M}$ in a sense reduces to a simple measurement $\boldsymbol{E}$ over a composite quantum system which consists of the initial object plus some additional quantum "degrees of freedom". For this we shall need the notion of tensor product of Hilbert spaces, which serves for the description of composite quantum systems.

Let $\mathscr{H}_1$, $\mathscr{H}_2$ be Hilbert spaces with the inner products $(\cdot|\cdot)_1$ and $(\cdot|\cdot)_2$. Consider the linear space $\mathscr{L}$ of formal linear combinations of the elements $\psi_1 \times \psi_2 \in \mathscr{H}_1 \times \mathscr{H}_2$. Introduce in $\mathscr{L}$ the pre-inner product $(\cdot|\cdot)$ putting

$$(\varphi_1 \times \varphi_2 | \psi_1 \times \psi_2) = (\varphi_1|\psi_1)_1 (\varphi_2|\psi_2)_2 \tag{2.5.56}$$

and extending it to $\mathscr{L}$ by linearity. The completion of $\mathscr{L}$ with respect to this pre-inner product is called the *tensor product* of Hilbert space $\mathscr{H}_1$, $\mathscr{H}_2$ and is denoted by $\mathscr{H}_1 \otimes \mathscr{H}_2$. The vector of $\mathscr{H}_1 \otimes \mathscr{H}_2$ corresponding to the equivalence class of the vector $\psi_1 \times \psi_2 \in \mathscr{L}$ is denoted by $\psi_1 \otimes \psi_2$. From (2.5.56) it follows that

$$(\varphi_1 \otimes \varphi_2 | \psi_1 \otimes \psi_2) = (\varphi_1|\psi_1)_1 (\varphi_2|\psi_2)_2. \tag{2.5.57}$$

A good illustration of this abstract definition is given by the following example. Let $\mathscr{H}_1 = \mathscr{L}^2(\mathbb{R}^n)$ be the space of the square-integrable functions $\psi_1(x)$, $x \in \mathbb{R}^n$, and $\mathscr{H}_2 = \mathscr{L}^2(\mathbb{R}^m)$ that of the functions $\psi_2(y)$, $y \in \mathbb{R}^m$. The $\mathscr{L}/\mathscr{N}$ consists of the finite sums of the form[2]

$$\psi(x, y) = \sum_j \psi_1^j(x)\psi_2^j(y)$$

with

$$(\varphi|\psi) = \iint \overline{\varphi(x, y)}\psi(x, y)\mathrm{d}^n x\, \mathrm{d}^m y,$$

and $\mathscr{H}_1 \otimes \mathscr{H}_2 = \mathscr{L}^2(\mathbb{R}^{n+m})$ is the space of all square-integrable functions $\varphi(x, y)$; $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ with the same inner product. The vector $\psi_1 \otimes \psi_2$ corresponds to the function $\psi_1(x)\psi_2(y)$.

Turning back to the general case, consider two orthonormal bases – $\{e_1^j\}$ in $\mathscr{H}_1$ and $\{e_2^k\}$ in $\mathscr{H}_2$. Then $\{e_1^j \otimes e_2^k\}$ forms an orthonormal basis in $\mathscr{H}_1 \otimes \mathscr{H}_2$. It follows easily that any vector $\varphi \in \mathscr{H}_1 \otimes \mathscr{H}_2$ can be represented in the form $\varphi = \sum_k \psi_1^k \otimes e_2^k$ where ψ_1^k are some uniquely defined vectors of $\mathscr{H}_1$, such that $\sum_k \|\psi_1^k\|_1^2 < \infty$. The subspaces $\mathscr{H}_1 \otimes e_2^k \subset \mathscr{H}_1 \otimes \mathscr{H}_2$ consisting of vectors of the form $\psi_1 \oplus e_2^k$, $\psi_1 \in \mathscr{H}_1$, are mutually orthogonal. Thus we have a decomposition of $\mathscr{H}_1 \otimes \mathscr{H}_2$ into the direct orthogonal sum of subspaces, each of which is isomorphic to $\mathscr{H}_1$:

$$\mathscr{H}_1 \otimes \mathscr{H}_2 = \sum_k \oplus[\mathscr{H}_1 \otimes e_2^k]. \tag{2.5.58}$$

The tensor product of operators $X_1 \otimes X_2$, where X_j is a bounded operator in $\mathscr{H}_j$, is defined by the relation

$$(X_1 \otimes X_2)(\psi_1 \otimes \psi_2) = X_1\psi_1 \otimes X_2\psi_2$$

on the vectors of the form $\psi_1 \otimes \psi_2$ and then uniquely extended to the whole $\mathscr{H}_1 \otimes \mathscr{H}_2$ by linearity and continuity.

Consider the decomposition (2.5.58). Then any bounded operator X in $\mathscr{H}_1 \otimes \mathscr{H}_2$ can be described by the block matrix $[X_{jk}]$ the elements of which are bounded operators in $\mathscr{H}_1$, as follows

$$X\left(\sum_k \psi_1^k \otimes e_2^k\right) = \sum_j \left(\sum_k X_{jk}\psi_1^k\right) \otimes e_2^j.$$

[2] More precisely, one should speak of the equivalence classes of the Lebesgue measurable functions.

In particular the operator $X_1 \otimes X_2$ is described by the matrix

$$[X_1(e_2^j|X_2e_2^k)_2].$$

Proposition 2.5.2. *For any measurement $\boldsymbol{M} = \{M(B)\}$ in $\mathscr{H}$ there are a Hilbert space $\mathscr{H}_0$, a pure state S_0 in $\mathscr{H}_0$ and a simple measurement $\boldsymbol{E} = \{E(B)\}$ in $\mathscr{H} \otimes \mathscr{H}_0$ such that*

$$\mu^{\boldsymbol{E}}_{S\otimes S_0}(B) = \mu^{\boldsymbol{M}}_S(B); \qquad B \in \mathscr{A}(U), \tag{2.5.59}$$

for any state S in $\mathscr{H}$. Conversely, any such triplet $(\mathscr{H}_0, S_0, \boldsymbol{E})$ gives rise to the unique measurement $\boldsymbol{M}$ in $\mathscr{H}$ satisfying (2.5.59).

Here $\mu^{\boldsymbol{E}}_{S\otimes S_0}$ (correspondingly $\mu^{\boldsymbol{M}}_S$) denotes the probability distribution of the measurement $\boldsymbol{E}$ (correspondingly $\boldsymbol{M}$) with respect to the state $S \otimes S_0(S)$.

Proof. Let $\boldsymbol{E}$ be the orthogonal resolution of identity in $\tilde{\mathscr{H}}$ the existence of which is asserted by Theorem 2.5.1. Extending if necessary $\tilde{\mathscr{H}}$ and $\boldsymbol{E}$ we can assume that $\tilde{\mathscr{H}} = \mathscr{H} \oplus \mathscr{H} \oplus \cdots$, $\mathscr{H}$ being embedded in $\tilde{\mathscr{H}}$ so that $\mathscr{H} = \mathscr{H} \oplus [0] \oplus [0] \oplus \cdots$. Then by the discussion preceding the formulation of the proposition, $\tilde{\mathscr{H}} = \mathscr{H} \otimes \mathscr{H}_0$ where $\mathscr{H}_0 = l^2$ is the space of square-summable sequences $[c_j : j = 1, 2, \ldots]$. Let $S_0 = S_\psi$ where $\psi = [1, 0, 0, \ldots]$; then operator $S \otimes S_0$ is described by the matrix

$$\begin{bmatrix} S & 0 & : \\ 0 & 0 & : \\ \cdots & \cdots & \end{bmatrix}$$

and it is clear that for any bounded X in $\mathscr{H} \otimes \mathscr{H}_0$ one has $\mathrm{Tr}(S \otimes S_0)X = \mathrm{Tr}\, S\widetilde{E}X\widetilde{E}$. Substituting $X = E(B)$ and using (2.5.55) we get (2.5.59).

To prove the converse note that the relation

$$\mu^{\boldsymbol{E}}_{S\otimes S_0}(B) = \mathrm{Tr}(S \otimes S_0)E(B); \qquad B \in \mathscr{A}(U),$$

defines an affine map $S \to \mu^{\boldsymbol{E}}_{S\otimes S_0}$ of the set of quantum states into the set of probability distributions on U. According to Theorem 2.2.1 there is a measurement $\boldsymbol{M}$ satisfying (2.5.59) and the proposition is proved. □

The proposition says that the measurements $\boldsymbol{E}$ and $\boldsymbol{M}$ are statistically equivalent in the sense that they have the same probability distributions for any state S.

We call the triple $(\mathscr{H}_0, S_0, \boldsymbol{E})$ a *realization* of measurement $\boldsymbol{M}$. A realization corresponds to a simple measurement over the system $\mathscr{H} \otimes \mathscr{H}_0$,

consisting of the initial system $\mathscr{H}$ and additional independent degrees of freedom $\mathscr{H}_0$ in the fixed state S_0. In classical statistics this would correspond to a randomized procedure in which the additional system can be regarded as a "roulette" generating random numbers according to the probability law S_0. The role of $\mathscr{H}_0$ in quantum case will be more clear when we shall consider an example (see Section 3.7).

2.6. Uncertainty relations and compatibility

Let $\boldsymbol{M} : S \to \mu_S(\mathrm{d}x)$ be a measurement with real values, to that $\mu_S(\mathrm{d}x)$ is a probability distribution on the real line $\mathbb{R}$. The most important characteristics of such distribution are the *mean value* and the *variance*

$$\begin{aligned} E_S\{\boldsymbol{M}\} &= \int x\mu_S(\mathrm{d}x), \\ D_S\{\boldsymbol{M}\} &= \int (x - E_S\{\boldsymbol{M}\})^2 \mu_S(\mathrm{d}x). \end{aligned} \tag{2.6.60}$$

The variance is apparently a measure of the mean square deviation of a random variable distributed according to $\mu_S(\mathrm{d}x)$, from its mean value. These quantities are well defined if, say, μ_S has a finite second moment. In this case we shall say that the measurement $\boldsymbol{M}$ has a finite second moment with respect to the state S.

By an *observable* in quantum mechanics one usually means a self-adjoint operator in $\mathscr{H}$; however having in mind Theorem 2.4.3 we shall extend the use of this term to arbitrary densely defined symmetric operator. By the spectral theorem to any observable X there corresponds a (generally, non-unique) measurement $M(\mathrm{d}x)$, such that $X = \int xM(\mathrm{d}x)$. As we shall see, the quantities (2.6.60) are the same for all spectral measures $M(\mathrm{d}x)$ of the operator X; for a pure state this follows directly from (2.4.41), (2.4.42) and (2.4.50). Therefore we can call (2.6.60), correspondingly, the mean $E_S(X)$ and the variance $D_S(X)$ of the observable X. Sometimes $E_S(X)$ will also be denoted $\overline{X}$ if it cannot cause a confusion. For a pure state $S = S_\psi$ with $\psi \in \mathscr{D}(X)$ we have[3]

$$\begin{aligned} E_{S_\psi}(X) &= (\psi | X\psi), \\ D_{S_\psi}(X) &= \|(X - E_{S_\psi}(X))\psi\|^2 = \|X\psi\|^2 - E_{S_\psi}(X)^2. \end{aligned} \tag{2.6.61}$$

[3] We shall sometimes drop the symbol I from the notation of an operator which is a multiple of identity, so that $X - E_S(X)$ means $X - E_S(X) \cdot I$ etc.

Consider a pair of observables X_1, X_2 and a state $S = S_\psi$ such that $\psi \in \mathscr{D}(X_1) \cap \mathscr{D}(X_2)$ assuming such a $\psi \neq 0$ exists. For any real c

$$\begin{aligned} 0 &\leq \|(X_1 - \overline{X}_1)\psi - \mathrm{i}c(X_2 - \overline{X}_2)\psi\|^2 \\ &= D_{S_\psi}(X_1) + 2c\,\mathrm{Im}(X_1\psi|X_2\psi) + c^2 D_{S_\psi}(X_2), \end{aligned} \tag{2.6.62}$$

whence

$$D_{S_\psi}(X_1) \cdot D_{S_\psi}(X_2) \geq |\,\mathrm{Im}(X_1\psi|X_2\psi)|^2, \tag{2.6.63}$$

the equality being achieved if and only if for some c

$$[(X_1 - \overline{X}_1) + \mathrm{i}c(X_2 - \overline{X}_2)]\psi = 0; \tag{2.6.64}$$

(we exclude here the case $D_{S_\psi}(X_2) = 0$ when $c = \infty$). If X_1, X_2 are bounded, then the right-hand side of (2.6.63) can be rewritten as $\frac{1}{4}|E_{S_\psi}(\mathrm{i}[X_1, X_2])|^2$, where

$$[X_1, X_2] = X_1X_2 - X_2X_1 \tag{2.6.65}$$

is the *commutator* of the operators X_1, X_2. The inequality (2.6.63) is called the *uncertainty relation*. In Section 2.9 we shall prove a rigorous and general version of this inequality for arbitrary states and measurements with finite second moments.

Sometimes it is asserted that the uncertainty relation sets a limitation to the accuracy of "joint measurement" of the observables X_1, X_2. In fact the statistical content of the uncertainty relation is different, and to show this we shall analyze the notion of "joint measurability".

In experimental practice it is common to measure several quantities during one individual experiment. An outcome of such experiment can be described by a collection of real numbers $x_1, \ldots, x_n$ taking values in some set Λ. Thus mathematically the statistics of a joint measurement must be described by a resolution of identity $M(\mathrm{d}x_1 \cdots \mathrm{d}x_n)$ on $\Lambda \subset \mathbb{R}^n$. It is customarily to speak of "simultaneous" measurements of several quantities. What is relevant for the notion of joint measurement is not the times at which the data $x_1, \ldots, x_n$ are obtained but the fact that all of them are obtained in a single individual experiment referred to a definite initial state S. Whether the measurement is in fact "simultaneous" or "repeated" will, of course, affect the measurement statistics but in any case it is described by an affine map $S \to \mu_S(\mathrm{d}x_1 \ldots \mathrm{d}x_n)$ and hence according to Theorem 2.2.1 by a resolution of identity $M(\mathrm{d}x_1 \cdots \mathrm{d}x_n)$.

We call two real-valued measurements, $M_j(\mathrm{d}x_j)$, $x_j \in \mathbb{R}$, $j = 1, 2$, *compatible* if there exists a measurement $M(\mathrm{d}x, \mathrm{d}x_2)$ on $\mathbb{R}_1 \times \mathbb{R}_2 = \mathbb{R}^2$ such that

$$M_1(\mathrm{d}x_1) = \int_{\mathbb{R}_2} M(\mathrm{d}x_1\,\mathrm{d}x_2), \qquad M_2(\mathrm{d}x_2) = \int_{\mathbb{R}_1} M(\mathrm{d}x_1\,\mathrm{d}x_2),$$

or, more precisely $M_1(B_1) = M(B_1 \times \mathbb{R}_2)$, $B_1 \in \mathscr{A}(\mathbb{R}_1)$, and analogously for $M_2(\mathrm{d}x_2)$. The measurement $M(\mathrm{d}x_1\,\mathrm{d}x_2)$ will be called a *joint measurement* for $M_j(\mathrm{d}x_j)$; $j = 1, 2$, and the latter will be called *marginal measurements* for $M(\mathrm{d}x_1\,\mathrm{d}x_2)$.

Proposition 2.6.1. *The simple measurements $E_j(\mathrm{d}x_j)$; $j = 1, 2$, are compatible if and only if*

$$[E_1(B_1), E_2(B_2)] = 0; \qquad B_j \in \mathscr{A}(\mathbb{R}_j); \quad j = 1, 2. \tag{2.6.66}$$

Proof. To prove sufficiency define the orthogonal resolution of identity $\boldsymbol{E}$ on $\mathbb{R}_1 \times \mathbb{R}_2$ putting

$$E(B_1 \times B_2) = E_1(B_1) \cdot E_2(B_2),$$

and extending $\boldsymbol{E}$ onto $\mathscr{A}(\mathbb{R}_1 \times \mathbb{R}_2)$ in a standard way. Then $\boldsymbol{E}_j$ are the marginal measurements for $\boldsymbol{E}$.

To prove necessity assume that $\boldsymbol{M}$ is a joint measurement for $\boldsymbol{E}_j$, $j = 1, 2$. Fix $B_j \in \mathscr{A}(\mathbb{R}_j)$ and let $\overline{B}_j$ denote the complement of B_j; $j = 1, 2$. Consider the following table

$$\begin{array}{ccccc}
E_1(B_1) & = & M(B_1 \times B_2) & + & M(B_1 \times \overline{B}_2) \\
+ & & + & & + \\
E_1(\overline{B}_1) & = & M(\overline{B}_1 \times B_2) & + & M(\overline{B}_1 \times \overline{B}_2) \\
\| & & \| & & \| \\
I & = & E_2(B_2) & + & E_2(\overline{B}_2),
\end{array}$$

where $E_1(B_1)E_1(\overline{B}_1) = E_2(B_2)E_2(\overline{B}_2) = 0$. Rewriting this in the form

$$\begin{aligned}
&[M(B_1 \times B_2) + M(B_1 \times \overline{B}_2)] \cdot [M(\overline{B}_1 \times B_2) + M(\overline{B}_1 \times \overline{B}_2)] \\
&= [M(B_1 \times B_2) + M(\overline{B}_1 \times B_2)] \cdot [M(B_1 \times \overline{B}_2) + M(\overline{B}_1 \times \overline{B}_2)] \\
&= 0
\end{aligned}$$

and applying several times the Hilbert space generalization of Lemma 1.6.4 we obtain that the product of any two operators $M(B_1 \times B_2)$, $M(B_1 \times \overline{B}_2)$, $M(\overline{B}_1 \times B_2)$, $M(\overline{B}_1 \times \overline{B}_2)$ is equal to zero. It follows that

$$E_1(B_1)E_2(B_2) = M(B_2 \times B_2)^2 = E_2(B_2)E_1(B_1)$$

and the proposition is proved. □

Now let X_j; $j = 1, 2$, be observables represented by self-adjoint operators and $\boldsymbol{E}_j$ be their spectral measures. The observables are called

compatible (or jointly measurable) if $\boldsymbol{E}_j$ are compatible. We see that this is the case if and only if the operators X_j *commute* in the sense that their spectral measures $\boldsymbol{E}_j$ satisfy the condition (2.6.66). One can show that this is equivalent to the condition

$$[\mathrm{e}^{\mathrm{i}\theta_1 X_1}, \mathrm{e}^{\mathrm{i}\theta_2 X_2}] = 0; \qquad \theta_1, \theta_2 \in \mathbb{R}, \tag{2.6.67}$$

on the unitary groups generated by X_1, X_2. If moreover X_1, X_2 are bounded, then this is equivalent to

$$[X_1, X_2] = 0.$$

These considerations extend to an arbitrary family of observables $X_1, \ldots, X_n$, represented by commuting self-adjoint operators. Let $E(\mathrm{d}x_1, \cdots \mathrm{d}x_n)$ be the orthogonal resolution of identity presenting the joint measurement of $X_1, \ldots, X_m$ ($E(\mathrm{d}x_1 \cdots \mathrm{d}x_n) = E_1(\mathrm{d}x_1) \cdot \ldots \cdot E_n(\mathrm{d}x_n)$, where $E_j(\mathrm{d}x_j)$ are the spectral measures for X_j). Then one can develop a functional calculus of $X_1, \ldots X_n$, putting

$$f(X_1, \ldots, X_n) = \int \ldots \int f(x_1, \ldots, x_n) E(\mathrm{d}x_1 \cdots \mathrm{d}x_n)$$

in a way analogous to the case of a single operator.

Returning to the uncertainty relation (2.6.63) we note that if the right-hand side of (2.6.63) is nonzero, then X_1 and X_2 cannot commute. Indeed, assuming X_1, X_2 commute we obtain from (2.6.67)

$$(\mathrm{e}^{-\mathrm{i}\theta_1 X_1}\psi | \mathrm{e}^{\mathrm{i}\theta_2 X_2}\psi) \equiv (\mathrm{e}^{\mathrm{i}\theta_1 X_1}\psi | \mathrm{e}^{-\mathrm{i}\theta_2 X_2}\psi).$$

Differentiating with respect to θ_1 and θ_2 at $\theta_1 = \theta_2 = 0$ we get by (2.4.49) $\mathrm{Im}(X_1\psi | X_2\psi) = 0$ contrary to the assumption. Thus X_1, X_2 are incompatible and by Proposition 2.6.1 no joint measurement exists for X_1 and X_2. Therefore it is senseless to say that (2.6.63) sets a limitation to the accuracy of joint measurement. To give the proper statistical interpretation one has to consider two large collections of copies of the object prepared in the same state S. Then, if the observable X_1 is measured in the first collection, and X_2 in the second, the product of variance of such independent measurements will satisfy (2.6.63). Otherwise assume that one of the variances, say, $D_S(X_1)$ is known *a priori* from the description of the preparation procedure; then the measured value $D_S(X_2)$ will again satisfy (2.6.63). The two interpretations are closely connected since evaluation of $D_S(X_1)$ by measuring X_1 in the first collection can be considered as a preliminary determination of a numerical characteristic of the state S.

As we shall see later, the uncertainty relation is indeed connected with a bound for accuracy of joint measurements but the connection is not so straightforward.

2.7. Trace-class operators and Hilbert-Schmidt operators[4]

Consider bounded operators which have the diagonal form in a fixed orthonormal basis $\{e_j\}$

$$X = \sum_j x_j |e_j)(e_j|, \tag{2.7.68}$$

where the series converges strongly. Any property of the sequence of eigenvalues $\{x_j\}$ corresponds to a property of the operator X. The norm of X is equal to

$$\|X\| = \sup_j |x_j|; \tag{2.7.69}$$

X is Hermitean if and only if x_j are all real, positivity of X corresponds to $x_j \geq 0$ etc.

To any classical space of sequences there corresponds a space of operators. Taking the Banach space c of all bounded sequences $\{x_j\}$ with the sup norm (2.7.69) we obtain all bounded diagonal operators with the operator norm. Restricting to a finite number of j we get the diagonal finite-rank operators. Note that the completion of the space of finite sequences with respect to the sup norm give only the proper subspace c_0 of c consisting of sequences $\{x_j\}$ tending to zero.

Other important space of sequences are l^1 and l^2, corresponding to the norms

$$\|X\|_1 = \sum_j |x_j| = \mathrm{Tr}\,|X|,$$

where $|X| = \sqrt{X^*X} = \sum_j |x_j||e_j)(e_j|$ for the operator X of the form (2.7.68), and

$$\|X\|_2 = \sqrt{\sum_j |x_j|^2} = \sqrt{\mathrm{Tr}\, X^*X}.$$

We are going to describe the noncommutative analogs of these spaces, which do not require X to be of diagonal form in a fixed basis $\{e_j\}$.

We shall proceed by completing the space $\mathfrak{F}(\mathscr{H})$ of all finite-rank operators with respect to corresponding norms. The completion of $\mathfrak{F}(\mathscr{H})$ with respect to the operator norm $\|\cdot\|$ is the space of *compact* (completely continuous) operators. We shall not discuss the properties of this important class; what is essential here is that the following analog of the finite-dimensional spectral theorem holds: *any compact Hermitean operator X has the spectral representation of the form* (2.7.68), *where* $\{e_j\}$ *is*

[4] The material of Sections 2.7-2.10 will be used essentially only in Chapters 5, 6.

the orthonormal basis of eigenvectors of X, $\{x_j\}$ are the eigenvalues of X. The sequence of eigenvalues of a compact Hermitean operator tends to zero.

Now we are going to define the operator analog of the space l^1. If X is Hermitean, then $|X|$ is defined by (2.3.30); for any bounded X put

$$|X| = \sqrt{X^*X};$$

where X^*X is (positive) Hermitean and therefore any continuous function of it is a well-defined Hermitean operator. Since $|X|^2 = X^*X$, then for all $\psi \in \mathscr{H}$

$$\||X|\psi\| = \|X\psi\|. \tag{2.7.70}$$

Observing that $T \in \mathfrak{F}(\mathscr{H})$ implies $|T| \in \mathfrak{F}(\mathscr{H})$ put

$$\|T_1\|_1 = \operatorname{Tr}|T|. \tag{2.7.71}$$

For any finite-rank operators T, X

$$|\operatorname{Tr} TX| \le \|T\|_1 \cdot \|X\|. \tag{2.7.72}$$

Indeed, since $|T|$ is a finite-rank Hermitean operator, there is an orthonormal basis $\{e_j\}$ of eigenvectors of $|T|$. From (2.7.70)

$$\|Te_j\| = \||T|e_j\| = (e_j||T|e_j),$$

so that

$$|\operatorname{Tr} TX| = \left|\sum_j (X^*e_j|Te_j)\right| \le \|X^*\| \cdot \sum_j \|Te_j\| = \|X\| \cdot \|T\|_1.$$

Putting in (2.7.72) $X = E$ where E is projection onto a finite-dimensional subspace containing all the vectors φ_j, ψ_j from the representation $T = \sum_j |\varphi_j)(\psi_j|$ we have $TE = T$, $\|E\| = 1$, so that

$$|\operatorname{Tr} T| \le \|T\|_1.$$

This shows that a natural domain of definition of trace would be the completion of $\mathfrak{F}(\mathscr{H})$ with respect to the norm $\|\cdot\|_1$. We shall only state the final result.

Theorem 2.7.1. *The relation* (2.7.71) *defines the norm on $\mathfrak{F}(\mathscr{H})$; the completion of $\mathfrak{F}(\mathscr{H})$ with respect to this norm is the Banach space $\mathfrak{T}^1(\mathscr{H})$ of trace-class operators satisfying $\|T\|_1 = \operatorname{Tr}|T| < \infty$. The unique continuous extension of the trace onto $\mathfrak{F}^1(\mathscr{H})$ is given by*

$$\operatorname{Tr} T = \sum_j (e_j|Te_j),$$

where the series converges to one and the same value for any basis $\{e_j\}$.

Any trace-class operator is compact. To show this notice that

$$\|T\| \leq \|T\|_1$$

for $T \in \mathfrak{F}(\mathscr{H})$ since by (2.7.70) and (2.1.8),

$$\|T\| = \||T|\| = \sup_{\varphi \neq 0} \frac{(\psi||T|\psi)}{(\psi|\psi)} \leq \operatorname{Tr}|T|.$$

Therefore the completion of $\mathfrak{F}(\mathscr{H})$ with respect to $\|\cdot\|_1$ is contained in the completion with respect to $\|\cdot\|$.

It follows that any Hermitean trace-class operator has the spectral representation

$$T = \sum_j t_j |e_j)(e_j|, \tag{2.7.73}$$

where the sum converges in the norm $\|\cdot\|_1$, since $\sum_j |t_j| = \operatorname{Tr}|T| < \infty$ and $\||e_j)(e_j|\|_1 = 1$. Also $\operatorname{Tr} T = \sum_j t_j$.

Putting

$$T_+ = \sum_{t_j > 0} t_j |e_j)(e_j|, \qquad T_- = -\sum_{t_j < 0} t_j |e_j)(e_j|,$$

we have

$$T = T_+ - T_-, \qquad |T| = T_+ + T_-, \tag{2.7.74}$$

so that

$$\|T\|_1 = \operatorname{Tr} T_+ + \operatorname{Tr} T_- = \|T_+\|_1 + \|T_-\|_1.$$

Any positive operator with finite trace is trace-class and therefore has the spectral representation (2.7.73) with $t_j \geq 0$. In particular (2.2.20) holds for any density operator.

It is well known that the space of continuous linear functional (the dual space) for l^1 is the space c. The following theorem is the noncommutative analog of this fact. We denote by $\mathfrak{B}(\mathscr{H})$ the Banach space of all bounded operators with the operator norm $\|\cdot\|$.

Theorem 2.7.2. *If T is trace-class, X is a bounded operator, then TX and XT are trace-class operators and* (2.1.18), (2.7.72) *hold for all such T and X. For any $X \in \mathfrak{B}(\mathscr{H})$ the correspondence*

$$T \to \operatorname{Tr} TX; \qquad T \in \mathfrak{T}^1(\mathscr{H}) \tag{2.7.75}$$

defines a continuous linear functional on $\mathfrak{T}^1(\mathscr{H})$ with the norm equal to $\|X\|$. Conversely any continuous linear function of $\mathfrak{T}^1(\mathscr{H})$ has this form.

Thus $[\mathfrak{T}^1(\mathscr{H})]^* = \mathfrak{B}(\mathscr{H})$. Supplying with the subscript h the corresponding real Banach spaces of Hermitean operators we have also $[\mathfrak{T}^1_{\rm h}(\mathscr{H})]^* = \mathfrak{B}_{\rm h}(\mathscr{H})$. The real linear functional (2.7.75) is positive, *i.e.*, $\operatorname{Tr} TX \geq 0$ for all $T \in \mathfrak{T}^1_{\rm h}(\mathscr{H})$, $T \geq 0$, if and only if $X \geq 0$. This can be proved as in Lemma 1.6.3. It follows that $T \geq 0$ and $X \leq Y$ imply $\operatorname{Tr} TX \leq \operatorname{Tr} TY$.

Proof of Theorem 2.2.1. Now we have tools for this proof. Let $S \to \mu_S$ be a measurement. Consider the real linear span of the set of states $\mathfrak{S}(\mathscr{H})$. Any linear combination $T = \sum t_j S_j$ of density operators is apparently a Hermitean trace-class operator. Conversely, let $T \in \mathfrak{T}^1_{\rm h}(\mathscr{H})$; then by (2.7.74)

$$T = t_+ S_+ - t_- S_-,$$

where $t_\pm = \operatorname{Tr} T_\pm$, $S_\pm = (t_\pm)^{-1} \cdot T_\pm$ are density operators. Thus the linear span of $\mathfrak{S}(\mathscr{H})$ is the space $\mathfrak{T}^1_{\rm h}(\mathscr{H})$ of Hermitean trace-class operators.

Fix a measurable set B and consider affine functional $S \to \mu_S(B)$ on $\mathfrak{S}(\mathscr{H})$. We construct a linear extension of it onto $\mathfrak{T}^1_{\rm h}(\mathscr{H})$ by putting

$$\mu(T) = \sum_j t_j \mu_{S_j}(B),$$

if $T = \sum_j t_j S_j$. Correctness of this extension can be verified as in finite-dimensional case (see Lemma 1.6.2). This functional is continuous since

$$\begin{aligned} |\mu(T)| &\leq \mu(T_+) + \mu(T_-) = t_+ \mu(S_+) + t_- \mu(S_-) \\ &\leq t_+ + t_- = \operatorname{Tr} |T| = \|T\|_1. \end{aligned}$$

Therefore by the second part of Theorem 2.7.2 there exists a bounded $M(B)$ such that $\mu(T) = \operatorname{Tr} TM(B)$, in particular $\mu_S(B) = \operatorname{Tr} SM(B)$. From $0 \leq \mu_S(B) \leq 1$, $S \in \mathfrak{S}(\mathscr{H})$, it follows that $0 \leq M(B) \leq I$, and from $\mu_S(\emptyset) = 0$, $\mu_S(U) = I$; $S \in \mathfrak{S}(\mathscr{H})$, if follows that $M(\emptyset) = 0$, $M(U) = I$. The proof of these facts is the same as in the finite-dimensional case and is based on the analog of Lemma 1.6.3.

To prove the σ-additivity (property (3) of resolution of identity in Section 2.2) note that for any S the probability distribution is σ-additive in B. Putting $S = S_\psi$ we have

$$(\psi | M(B) \psi) = \sum_j (\psi | M(B_j) \psi)$$

for any decomposition $\{B_j\}$ of B, and this means precisely the property (3).

Conversely let $\{M(B)\}$ be a resolution of identity in $\mathscr{H}$. Then the relation (2.2.21) defines a family of affine functionals on $\mathfrak{S}(\mathscr{H})$ and one needs only to check σ-additivity of μ_S, *i.e.*, to prove that property (3) implies

$$\operatorname{Tr} SM(B) = \sum_j \operatorname{Tr} SM(B_j) \tag{2.7.76}$$

for any density operator S. By the spectral representation (2.2.20) of S

$$\operatorname{Tr} SM(B) = \sum_j s_j(\psi_j|M(B)\psi_j), \tag{2.7.77}$$

where the right-hand side is the trace of $SM(B)$ calculated in the basis $\{\psi_j\}$ by the formula of Theorem 2.7.1. On the other hand by the σ-additivity of $\{M(B)\}$ we have

$$(\psi_j|M(B)\psi_j) = \sum_k (\psi_j|M(B_k)\psi_j).$$

Multiplying it by s_j, summing and changing the order of summation which is possible due to nonnegativity of summands, we obtain (2.7.76). The theorem is proved. □

We can now also establish a useful relation

$$\int f(x)\mu_S(\mathrm{d}x) = \operatorname{Tr} Sf(X), \tag{2.7.78}$$

where f is a bounded measurable function, X is a self-adjoint operator with the spectral measure $E(\mathrm{d}x)$, $\mu_S(\mathrm{d}x) = \operatorname{Tr} SE(\mathrm{d}x)$ and $f(X) = \int f(x)E(\mathrm{d}x)$. Indeed using (2.7.77) we get

$$\begin{aligned}\int f(x)\mu_S(\mathrm{d}x) &= \sum_j s_j \int f(x)(\psi_j|E(\mathrm{d}x)\psi_j) \\ &= \sum_j s_j(\psi_j|f(X)\psi_j) = \operatorname{Tr} Sf(X),\end{aligned}$$

due to boundedness of f.

We now turn to the noncommutative analog of the l^2 space. Introduce the inner product in $\mathfrak{F}(\mathscr{H})$

$$(T_1, T_2) = \operatorname{Tr} T_1^* T_2, \tag{2.7.79}$$

with the norm $\|T\|_2 = \sqrt{\operatorname{Tr} T^*T}$.

Theorem 2.7.3. *The completion of $\mathfrak{F}(\mathscr{H})$ with respect to the inner product* (2.7.79) *is the Hilbert space $\mathfrak{T}^2(\mathscr{H})$. The elements of this space, called Hilbert-Schmidt operators, are bounded operators T satisfying $\operatorname{Tr} T^*T \equiv \sum_j \|Te_j\|^2 < \infty$ for an orthonormal basis in $\mathscr{H}$. For any $T_1, T_2 \in \mathfrak{T}^2(\mathscr{H})$ the product $T_1 \cdot T_2$ is trace-class operator and the inner product of T_1 and T_2 is equal to $\operatorname{Tr} T_1^* T_2$. The noncommutative analog of the Cauchy inequality holds*

$$\|T_1 \cdot T_2\|_1 \le \|T_1\|_2 \cdot \|T_2\|_2. \tag{2.7.80}$$

The product of a bounded X and a Hilbert-Schmidt operator T (in any order) is again Hilbert-Schmidt with

$$\|TX\|_2 = \|XT\|_2 \le \|X\| \cdot \|T\|_2. \tag{2.7.81}$$

The spaces of operators introduced above are related by following diagram

$$\mathfrak{F}(\mathscr{H}) \subset \mathfrak{T}^1(\mathscr{H}) \subset \mathfrak{T}^2(\mathscr{H}) \subset (\text{compact operators}).$$

To prove the last inclusion it is sufficient to show that

$$\|T\| \le \|T\|_2,$$

since this implies that the completion of $\mathfrak{F}(\mathscr{H})$ with respect to $\|\cdot\|_2$ is contained in the completion with respect to $\|\cdot\|$. But

$$\|T\|^2 = \||T|\|^2 = \sup_{\psi\neq 0} \frac{\||T|\psi\|^2}{\|\psi\|^2} = \sup_{\psi\neq 0} \frac{\|T\psi\|^2}{\|\psi\|^2} \le \operatorname{Tr} T^*T$$

and the result follows. In particular, any Hermitean Hilbert-Schmidt operator has the spectral representation of the form (2.7.73) with

$$\|T\|_2 = \sqrt{\sum_j |t_j|^2} < \infty.$$

Furthermore $\|T\|_2 \le \|T\|_1$, since

$$\operatorname{Tr} T^*T = \operatorname{Tr}|T|^2 = \sum_j \tau_j^2 \le \left(\sum_j \tau_j\right)^2 = (\operatorname{Tr}|T|)^2.$$

It follows that any trace-class operator is Hilbert-Schmidt.

Finally let us consider Hilbert-Schmidt operators in $\mathscr{L}^2(a,b)$. Let for simplicity T be a Hermitean operator with the spectral representation

(2.7.73) satisfying $\sum_j t_j^2 < \infty$. Its eigenfunctions $\{e_j(x)\}$ form an orthonormal basis in $\mathscr{L}^2(a, b)$. Consider the kernel

$$T(x', x) = \sum_j t_j e_j(x')\overline{e_j(x)}.$$

Since $\int \bar{e}_j(x)e_k(x)\mathrm{d}x = \delta_{jk}$ and $\sum t_j^2 < \infty$ this series converges in $\mathscr{L}^2((a, b) \times (a, b))$ and defines a square-integrable function of two variables x, x', satisfying

$$\int_a^b \int_a^b |T(x, x')|^2 \mathrm{d}x\, \mathrm{d}x' = \sum_j t_j^2 \equiv \mathrm{Tr}\, T^2.$$

For any $\psi \in \mathscr{L}^2(a, b)$

$$T\psi(x') = \int_a^b T(x', x)\psi(x)\mathrm{d}x, \tag{2.7.82}$$

as one sees, *e.g.*, calculating the form $(\varphi|T\psi)$ in the basis $\{e_j\}$. If T is trace-class, the $\sum |t_j| < \infty$ and the function $T(x, x) = \sum t_j |e_j(x)|^2$ is integrable with

$$\int_a^b T(x, x)\mathrm{d}x = \mathrm{Tr}\, T. \tag{2.7.83}$$

In Dirac's notation the kernel $T(x', x)$ is expressed by the symbol $(x'|T|x)$ and may be regarded as a formal analog of the matrix elements for the continual "basis" $\{|x)\}$; using the completeness relation (2.3.33) one obtains formally

$$T = \iint |x')(x'|T|x)(x|\mathrm{d}x'\mathrm{d}x,$$

whence one gets Dirac's version of (2.7.82):

$$(x'|T\psi) = \int (x'|T|x)(x|\psi)\mathrm{d}x.$$

If T is Hilbert-Schmidt, then as we have seen the kernel $(x'|T|x)$ is a square-integrable function of x, x', and these manipulations have a straightforward mathematical substantiation.

2.8. $\mathscr{L}^2$ spaces associated with a quantum state

Many important quantum observables are represented by unbounded operators. Unboundedness is a source of serious technical difficulties in the noncommutative theory. For example, the definition of sum of random

variables presents no difficulties in probability theory, while the sum of incompatible quantum observables may not be well defined. That we have restricted only to pure states in the uncertainty relation (2.6.63) is also due to difficulties connected with unboundedness. We are going now to elaborate a technique which will allow us to operate freely enough with unbounded observables, in particular to calculate their first two moments with respect to a state defined by an arbitrary density operator S.

We shall introduce a noncommutative analog of the Hilbert space of random variables with finite second moment, which is an important concept in ordinary probability theory. This Hilbert space of "square-summable" observables turns out to be a convenient tool also in quantum theory; in particular there is no problem with the summation of such observables.

Let S be a fixed density operator and X, Y bounded operators in $\mathscr{H}$. Put

$$X \circ Y = \tfrac{1}{2}(XY + YX). \tag{2.8.84}$$

Assuming X, Y being Hermitean introduce the pre-inner product in $\mathfrak{B}_h(\mathscr{H})$

$$\langle Y, X\rangle_S = \operatorname{Tr} S(Y \circ X) \equiv \operatorname{Re}\operatorname{Tr} SYX, \tag{2.8.85}$$

with

$$\langle X, X\rangle_S = \operatorname{Tr} SX^2.$$

The completion of $\mathfrak{B}_{\mathrm{h}}(\mathscr{H})$ with respect to $\langle\cdot,\cdot\rangle_S$ is a real Hilbert space denoted by $\mathscr{L}^2_{\mathrm{h}}(S)$. The elements of $\mathscr{L}^2_{\mathrm{h}}(S)$ can be represented, generally, by unbounded operators in $\mathscr{H}$ in the following way.

A symmetric operator X is called *square-summable* with respect to the density operator S having the spectral representation (2.2.20) if $\sum_j s_j \|X\psi_j\|^2 < \infty$ (so that $\psi_j \in \mathscr{D}(X)$ if $s_j \neq 0$). Two such operators X_1, X_2 are *equivalent* if $X_1\psi_j = X_2\psi_j$ for $s_j \neq 0$. For square-summable X, Y put

$$\begin{aligned}\langle Y, X\rangle_S &= \sum_j s_j \frac{1}{2}[(Y\psi_j|X\psi_j) + (X\psi_j|Y\psi_j)] \\ &\equiv \operatorname{Re}\sum_j s_j(Y\psi_j|X\psi_j),\end{aligned} \tag{2.8.86}$$

the series being convergent by the Cauchy inequality. If X and Y are bounded, then they are square-summable and the sum (2.8.86) is equal to (2.8.85), since it is just the trace calculated in the basis $\{\psi_j\}$.

Theorem 2.8.1. *The elements of $\mathscr{L}^2_{\mathrm{h}}(S)$ can be naturally identified with the equivalence classes of square-summable operators with the inner*

product defined by (2.8.86)*; namely, if* $\{X_n\}$ *is a Cauchy sequence with respect to the pre-inner product* (2.8.85) *in* $\mathfrak{B}_{\mathrm{h}}(\mathscr{H})$*, then there is a square-summable* X *such that* $\lim_n \langle X_n - X, X_n - X\rangle = 0$*, and conversely any square-summable operator is a limit of a Cauchy sequence* $\{X_n\} \subset \mathfrak{B}(\mathscr{H})$.

The proof of this theorem is omitted. In what follows we shall denote by the same letter X both a square-summable operator and a corresponding element of $\mathscr{L}^2_{\mathrm{h}}(S)$.

Using the notion of Hilbert-Schmidt operator we can give an alternative description of square-summable operators. Consider the operator

$$\sqrt{S} = \sum_j \sqrt{s_j}|\psi_j)(\psi_j|,$$

which is apparently Hilbert-Schmidt since $\mathrm{Tr}(\sqrt{S})^2 < \infty$. Introducing the notation $\mathscr{R}(T)$ for the range of an operator T, we have

$$\mathscr{R}(\sqrt{S}) = \left\{ \psi : \psi = \sum_j \sqrt{s_j} c_j \psi_j, \sum_j |c_j|^2 < \infty \right\}.$$

Proposition 2.8.2. *Operator* X *with* $\mathscr{D}(X) \ni \psi_j$ *for* $s_j \neq 0$ *is square-summable if and only if* X *extends to* $\mathscr{R}(\sqrt{S})$ *so that* $X\sqrt{S}$ *is Hilbert-Schmidt. Moreover*

$$\begin{aligned} \langle Y, X\rangle_S &= \mathrm{Tr}\, \tfrac{1}{2}[(Y\sqrt{S})^*(X\sqrt{S}) + (X\sqrt{S})^*(Y\sqrt{S})] \\ &= \mathrm{Re}\, \mathrm{Tr}(Y\sqrt{S})^*(X\sqrt{S}). \end{aligned} \tag{2.8.87}$$

Proof. If X is a square-summable, then the extension is given by

$$X\left(\sum_j \sqrt{s_j} c_j \psi_j\right) = \sum_j \sqrt{s_j} c_j X\psi_j, \qquad \sum_j |c_j|^2 < \infty,$$

the series converging strongly by the square-summability. The operator $X\sqrt{S}$ is Hilbert-Schmidt since

$$\mathrm{Tr}(X\sqrt{S})^*(X\sqrt{S}) = \sum_j \|X\sqrt{S}\psi_j\|^2 = \sum_j s_j \|X\psi_j\|^2 < \infty.$$

The relation (2.8.87) follows from (2.8.86) in the same way. The converse statement is obvious. □

We now give one more useful formula for the inner product. Since by Theorem 2.7.3 a product of Hilbert-Schmidt operators is a trace-class operator, the expression

$$X \circ S \equiv \tfrac{1}{2}[(X\sqrt{S}) \cdot \sqrt{S} + \sqrt{S} \cdot (X\sqrt{S})^*]$$

defines a trace-class operator in $\mathscr{H}$. Using (2.1.18) we obtain

$$\langle Y, X\rangle_S = \mathrm{Tr}(X \circ S)Y, \tag{2.8.88}$$

for a bounded Y and $X \in \mathscr{L}_{\mathrm{h}}^2(S)$.

The specific feature of the noncommutative case is the existence of an additional skew-symmetric form in $\mathscr{L}_{\mathrm{h}}^2(S)$ related to the commutator of operators. If X, Y are Hermitean, then $\mathrm{i}[Y, X]$ is also Hermitean. Therefore the relation

$$[Y, X]_S = \mathrm{i}\,\mathrm{Tr}\, S[Y, X] = 2\,\mathrm{Im}\,\mathrm{Tr}\, SXY \tag{2.8.89}$$

defines a real bilinear form on $\mathfrak{B}_{\mathrm{h}}(\mathscr{H})$. It can be extended to $\mathscr{L}_{\mathrm{h}}^2(S)$ by the equivalent relations

$$[Y, X]_S = 2\,\mathrm{Im}\sum_j s_j(X\psi_j|Y\psi_j) = 2\,\mathrm{Im}\,\mathrm{Tr}(X\sqrt{S})^*(Y\sqrt{S}),$$

analogous to (2.8.86), (2.8.87). If $X \in \mathscr{L}_{\mathrm{h}}^2(S)$, then the relation

$$[X, S] \equiv (X\sqrt{S}) \cdot \sqrt{S} - \sqrt{S} \cdot (X\sqrt{S})^*$$

determines a trace-class operator. Using again (2.1.18) we obtain for bounded Y

$$[Y, X]_S = \mathrm{i}\,\mathrm{Tr}[X, S] \cdot Y. \tag{2.8.90}$$

The form is *skew-symmetric*, *i.e.*,

$$[X, Y]_S = -[Y, X]_S; \qquad X, Y \in \mathscr{L}_{\mathrm{h}}^2(S).$$

If follows that

$$[X, X]_S = 0, \qquad X \in \mathscr{L}_{\mathrm{h}}^2(S). \tag{2.8.91}$$

From (2.8.90) with $Y = I$ and (2.1.18) we obtain also

$$[I, X]_S = 0, \qquad X \in \mathscr{L}_{\mathrm{h}}^2(S). \tag{2.8.92}$$

We shall also use the complexification of $\mathscr{L}_{\mathrm{h}}^2(S)$. Any bounded X can be uniquely presented in the form $X = X_1 + \mathrm{i}X_2$ where X_1, X_2 are Hermitean. Namely, one obtains

$$X_1 = \frac{1}{2}(X + X^*), \qquad X_2 = \frac{1}{2\mathrm{i}}(X - X^*).$$

Put

$$\langle Y, X\rangle_S = \operatorname{Tr} S(Y^* \circ X) = \operatorname{Tr}(S \circ X) \cdot Y^* \tag{2.8.93}$$

for $X, Y \in \mathfrak{B}(\mathscr{H})$. Then it is easy to see that

$$\langle X, X\rangle_S = \langle X_1, X_1\rangle_S + \langle X_2, X_2\rangle_S. \tag{2.8.94}$$

Denote by $\mathscr{L}^2(S)$ the completion of $\mathfrak{B}(\mathscr{H})$ with respect to the pre-inner product (2.8.93). Then any $X \in \mathscr{L}^2(S)$ still can be written in the form $X = X_1 + \mathrm{i}X_2$ with $X_1, X_2 \in \mathscr{L}^2_{\mathrm{h}}(S)$, so that (2.8.94) holds. This is expressed by saying that $\mathscr{L}^2(S)$ is the *complexification* of $\mathscr{L}^2_{\mathrm{h}}(S)$, and is written as

$$\mathscr{L}^2(S) = \mathscr{L}^2_{\mathrm{h}}(S) \oplus \mathrm{i}\mathscr{L}^2_{\mathrm{h}}(S).$$

The real bilinear skew-symmetric form (2.8.89) extends to the complex sesquilinear skew-Hermitean form, which is given on $\mathfrak{B}(\mathscr{H})$ by the relations

$$[Y, X]_S = \mathrm{i} \operatorname{Tr} S[Y^*, X] = \mathrm{i} \operatorname{Tr}[X, S] \cdot Y^*. \tag{2.8.95}$$

Other two useful complex pre-inner products on $\mathfrak{B}(\mathscr{H})$ are

$$\langle Y, X\rangle^+_S = \operatorname{Tr} SXY^*, \qquad \langle Y, X\rangle^-_S = \operatorname{Tr} SY^*X. \tag{2.8.96}$$

Since $\langle Y, X\rangle_S = \frac{1}{2}[\langle Y, X\rangle^+_S + \langle Y, X\rangle^-_S]$, then $\langle X, X\rangle^{\pm}_S \leq 2\langle X, X\rangle_S$. Therefore denoting by $\mathscr{L}^2_{\pm}(S)$ the completions of $\mathfrak{B}(\mathscr{H})$ with respect to the pre-inner products (2.8.96) we have $\mathscr{L}^2(S) \subseteq \mathscr{L}^2_{\pm}(S)$. Obviously

$$\begin{aligned} &\langle X, Y\rangle_S \pm \frac{\mathrm{i}}{2}[X, Y]_S = \langle X, Y\rangle^{\pm}_S; \\ &[X, Y]_S = \mathrm{i}(\langle X, Y\rangle^-_S - \langle X, Y\rangle^+_S); \qquad X, Y \in \mathscr{L}^2(S). \end{aligned} \tag{2.8.97}$$

Proposition 2.8.3. *The forms $\langle \cdot, \cdot\rangle_S$ and $[\cdot, \cdot]_S$ are related by the following equivalent inequalities:*

(1) $\langle X, X\rangle_S \geq \dfrac{\mathrm{i}}{2}[X, X]_S; \qquad X \in \mathscr{L}^2(S);$

(2) $\langle X, X\rangle_S \geq -\dfrac{\mathrm{i}}{2}[X, X]_S; \qquad X \in \mathscr{L}^2(S);$

(3) $\langle X_1, X_1\rangle_S + \langle X_2, X_2\rangle_S \geq [X_1, X_2]_S; \qquad X_1, X_2 \in \mathscr{L}^2_{\mathrm{h}}(S);$

(4) $\langle X_1, X_1\rangle_S \cdot \langle X_2, X_2\rangle_S \geq \frac{1}{4}[X_1, X_2]^2_S; \qquad X_1, X_2 \in \mathscr{L}^2_{\mathrm{h}}(S).$

Proof. The inequalities (1), (2) follow from

$$\langle X, X\rangle_S \pm \frac{\mathrm{i}}{2}[X, X]_S = \langle X, X\rangle^{\pm}_S \geq 0, \qquad X \in \mathscr{L}^2(S).$$

They are equivalent since $\langle X, X\rangle_S^+ = \langle X^*, X^*\rangle_S^-$, $X \in \mathfrak{B}(\mathscr{H})$. Let $X \in \mathscr{L}^2(S)$, then $X = X_1 + \mathrm{i}X_2$ with $X_j \in \mathscr{L}^2_{\mathrm{h}}(S)$ whence taking into account (2.8.91) and (2.8.96) we get

$$0 \le \langle X, X\rangle_S^- = \langle X_1, X_1\rangle_S + \langle X_2, X_2\rangle_S - [X_1, X_2]_S.$$

It follows that (1) is equivalent to (3). Inserting in (3) tX_1 in place of X_1 we obtain

$$t^2\langle X_1, X_1\rangle_S - t[X_1, X_2]_S + \langle X_2, X_2\rangle_S \ge 0, \qquad t \in \mathbb{R},$$

which is equivalent to (4).

It follows from (1), (2) that for any $X_1, \dots, X_n \in \mathscr{L}^2_{\mathrm{h}}(S)$

$$[\langle X_j, X_k\rangle_S] \ge \pm\frac{\mathrm{i}}{2}[[X_j, X_k]_S], \tag{2.8.98}$$

where on the left is a real symmetric $n \times n$-matrix, and on the right is $\pm\frac{1}{2}\mathrm{i}$ by a real skew-symmetric $(n \times n)$-matrix, both sides being considered as Hermitean $(n \times n)$-matrices.

Later we shall need the complex extension of inequality (4):

$$\langle X_1, X_1\rangle_S\langle X_2, X_2\rangle_S \ge \tfrac{1}{4}|[X_1, X_2]_S|^2; \qquad X_1, X_2 \in \mathscr{L}^2(S). \tag{2.8.99}$$

To prove it we first remark that (3) implies

$$\langle X_1, X_1\rangle_S + \langle X_2, X_2\rangle_S \ge \mathrm{Re}[X_1, X_2]_S \quad \text{for} \quad X_1, X_2 \in \mathscr{L}^2(S).$$

Then as in the proof of Proposition 2.8.3

$$\langle X_1, X_1\rangle_S\langle X_2, X_2\rangle_S \ge \tfrac{1}{4}(\mathrm{Re}[X_1X_2]_S)^2.$$

Replacing X_1 by λX_1 with $\lambda = [X_1, X_2]_S \cdot |[X_1, X_2]_S|^{-1}$ we get (2.8.99). □

2.9. Uncertainty relations for measurements with finite second moments

In probability theory the elements of $\mathscr{L}^2$ space are random variables with finite second moment. In the noncommutative case there is a correspondence between the elements of $\mathscr{L}^2_{\mathrm{h}}(S)$ and the real-valued measurements with finite second moment.

Let first X be an observable with finite second moment with respect to S which is represented by a densely defined symmetric operator. Then $X \in \mathscr{L}^2_{\mathrm{h}}(S)$ since the quantity

$$\sum_j s_j\|X\psi_j\|^2 = \sum_j s_j \int \lambda^2(\psi_j|M(\mathrm{d}\lambda)\psi_j) = \int \lambda^2\mu_S(\mathrm{d}\lambda)$$

is finite. We have used Theorem 2.4.3 in the first equality and positiveness and (2.7.77) in the second. In the same way we can prove

$$E_S(X) = \langle I, X\rangle_S, \tag{2.9.100}$$

$$D_S(X) = \langle X - E_S(X), X - E_S(X)\rangle_S. \tag{2.9.101}$$

If X is bounded, we get from (2.7.78)

$$E_S(X) = \operatorname{Tr} SX, \tag{2.9.102}$$

$$D_S(X) = \operatorname{Tr} S(X - E_S(X))^2, \tag{2.9.103}$$

which agrees with (2.9.100), (2.9.101).

Let X_1, X_2 be observables with finite second moment with respect to S. Applying the inequality (4) of Proposition 2.8.3 to $X_1 - E_S(X_1)$ and $X_2 - E_S(X_2)$ and taking into account (2.8.92) we obtain a generalization of the uncertainty relation (2.6.63) in the form

$$D_S(X_1) \cdot D_S(X_2) \geq \tfrac{1}{4}[X_1, X_2]_S^2. \tag{2.9.104}$$

Now let $\boldsymbol{M} = \{M(\mathrm{d}x)\}$ be an arbitrary real-valued measurement such that

$$\int x^2 \mu_S(\mathrm{d}x) < \infty$$

with $\mu_S(\mathrm{d}x) = \operatorname{Tr} SM(\mathrm{d}X)$. We shall define an integral

$$X_{\boldsymbol{M}} = \int x M(\mathrm{d}x),$$

converging in $\mathscr{L}_{\mathrm{h}}^2(S)$, so that the mean and the variance of the measurement $\boldsymbol{M}$ defined in (2.6.60) are evaluated through $X_{\boldsymbol{M}}$ by the relations

$$E_S\{\boldsymbol{M}\} = \langle I, X_{\boldsymbol{M}}\rangle, \tag{2.9.105}$$

$$D_S\{\boldsymbol{M}\} \geq \langle X_{\boldsymbol{M}} - E_S\{\boldsymbol{M}\},\ X_{\boldsymbol{M}} - E_S\{\boldsymbol{M}\}\rangle_S, \tag{2.9.106}$$

resembling (2.9.100), (2.9.101).

We proceed to define the integral of the type (2.9.104) first for simple real-valued function

$$f(x) = \sum_j f_j \mathbf{1}_{B_j}(x)$$

by the relation

$$\int f(x) M(\mathrm{d}x) = \sum_j f_j M(B_j).$$

Then the inequality holds:

$$\left[\int f(x)M(\mathrm{d}x)\right]^2 \le \int f(x)^2 M(\mathrm{d}x). \tag{2.9.107}$$

Indeed for the simple f this means

$$\left[\sum_j f_j M(B_j)\right]^2 \le \sum_j f_j^2 M(B_j),$$

and follows directly from

$$\sum_j \left[f_j - \sum_k f_k M(B_k)\right] M(B_j) \left[f_j - \sum_k f_k M(B_k)\right] \ge 0.$$

Multiplying (2.9.107) by S, taking trace and using (2.8.85) we get

$$\left\langle \int f(x)M(\mathrm{d}x), \int f(x)M(\mathrm{d}x)\right\rangle_S \le \int f(x)^2 \mu_S(\mathrm{d}x). \tag{2.9.108}$$

Now let f be square-integrable with respect to $\mu_S(\mathrm{d}x)$ and $\{f_n\}$ be a sequence of simple functions such that

$$\int (f_n(x) - f(x))^2 \mu_S(\mathrm{d}x) \to 0.$$

Applying (2.9.108) to the simple functions $f_n - f_m$ we see that

$$\left\{\int f_n(x)M(\mathrm{d}x)\right\}$$

is a Cauchy sequence of bounded operators in $\mathscr{L}^2_{\mathrm{h}}(S)$. Therefore it has the limit in $\mathscr{L}^2_{\mathrm{h}}(S)$ which we denote as $\int f(x)M(\mathrm{d}x)$. Moreover, the inequality (2.9.108) extends to any f, square-integrable with respect to $\mu_S(\mathrm{d}x)$.

If we proceed with complex-valued functions f and the inner products $\langle\cdot,\cdot\rangle_S^{\pm}$, the integral $\int f(x)M(\mathrm{d}x)$ will be an element of $\mathscr{L}^2_{\pm}(S)$ and the inequality (2.9.108) will change to

$$\left\langle \int f(x)\boldsymbol{M}(\mathrm{d}x), \int f(x)\boldsymbol{M}(\mathrm{d}x)\right\rangle_S^{\pm} \le \int |f(x)|^2 \mu_S(\mathrm{d}x). \tag{2.9.109}$$

Therefore we have proved

Proposition 2.9.1. *For any real-(complex)-valued* $f \in \mathscr{L}^2(\mu_S)$ *the integral* $\int f(x)M(\mathrm{d}x)$ *is defined as the limit in* $\mathscr{L}^2_{\mathrm{h}}(S)$ *(correspondingly in* $\mathscr{L}^2_{\pm}(S)$*) of the sequence* $\{\int f_n(x)M(\mathrm{d}x)\}$*, where* $\{f_n\}$ *is any sequence of simple functions converging to* f *in* $\mathscr{L}^2(\mu_S)$*. For any real* $f \in \mathscr{L}^2(\mu_S)$ *the inequality* (2.9.108) *holds*; *for any* $f \in \mathscr{L}^2(\mu_S)$ *the inequality* (2.9.109) *holds*.

To prove (2.9.106) we have only to put $f(x) = x - E_S\{\boldsymbol{M}\}$ in (2.9.108). The relation (2.9.105) follows from the more general relation

$$\int f(x)\mu_S(\mathrm{d}x) = \left\langle I, \int f(x)M(\mathrm{d}x)\right\rangle_S,$$

which is obvious for simple f's and extends to $f \in \mathscr{L}^2(\mu_s)$ by a standard limiting argument.

From (2.9.106) and Proposition 2.8.3(4) we obtain the most general form of the uncertainty relation

$$D_S\{\boldsymbol{M}_1\} \cdot D_S\{\boldsymbol{M}_2\} \geq \tfrac{1}{4}[X_{\boldsymbol{M}_1}, X_{\boldsymbol{M}_2}]^2,$$

which holds for any measurements $\boldsymbol{M}_1, \boldsymbol{M}_2$ with finite second moment. In contrast to (2.6.63) and (2.9.104) it applies also to joint measurements when $M_1(\mathrm{d}x_1)$ and $M_2(\mathrm{d}x_2)$ are marginal measurements with respect to some $M(\mathrm{d}x_1, \mathrm{d}x_2)$.

2.10. Matrix representation of square-summable operators. The commutation operator of a state

Elements of a $\mathscr{L}^2$ space can be naturally represented by infinite matrices. Assume first for simplicity that the state S is *exact*. This means that the density operator is nondegenerated, *i.e.*, has all eigenvalues $s_j > 0$. Then *the family of the matrix units*

$$E_{jk} = |\psi_j)(\psi_k|$$

where $\{\psi_j\}$ *is the orthonormal basis of eigenvectors of* S *in* $\mathscr{H}$ *form an orthogonal basis in* $\mathscr{L}^2_{\pm}(S)$, $\mathscr{L}^2(S)$ *with*

$$\langle E_{jk}, E_{jk}\rangle_S^+ = s_j, \qquad \langle E_{jk}, E_{jk}\rangle_S^- = s_k,$$

$$\langle E_{jk}, E_{jk}\rangle_S = \tfrac{1}{2}(s_j + s_k).$$

We shall prove it only for $\mathscr{L}^2(S)$. For any $X \in \mathscr{L}^2(S)$ we have

$$\langle E_{jk}, X\rangle_S = \tfrac{1}{2}(s_j + s_k)(\psi_j|X\psi_k)$$

by (2.8.86). It follows that $\langle E_{jk}, E_{j'k'}\rangle_S = \delta_{jj'} \cdot \delta_{kk'}$. The completeness of the system $\{E_{jk}\}$ follows from the fact that $\langle E_{jk}, X\rangle_S = 0$ for all j, k implies $X\psi_k = 0$ for all k, *i.e.*, $X = 0$ in $\mathscr{L}^2(S)$.

Therefore an arbitrary square-summable operator X is represented by the series

$$X = \sum_{jk} x_{jk}E_{jk} = \sum_{jk} |\psi_j)(\psi_j|X\psi_k)(\psi_k|, \tag{2.10.110}$$

where

$$x_{jk} = (\psi_j|X\psi_k) = \frac{\langle E_{jk}, X\rangle_S}{\langle E_{jk}, E_{jk}\rangle_S},$$

and the series converges in $\mathscr{L}^2(S)$. This is an extension of the matrix representation (2.1.13) to unbounded operators.

Consider now the real space $\mathscr{L}^2_{\mathrm{h}}(S)$. The representation (2.10.110) holds for $X \in \mathscr{L}^2_{\mathrm{h}}(s)$ but E_{jk} do not belong to $\mathscr{L}^2_{\mathrm{h}}(S)$ unless $j = k$ and x_{jk} are in general complex numbers. Introducing

$$C_{jk} = \frac{1}{2}(E_{jk} + E_{kj})^*, \qquad S_{jk} = \frac{1}{2\mathrm{i}}(E_{jk} - E_{jk}^*),$$

one checks that $\{C_{jk},\ j \le k;\ S_{jk},\ j < k\}$ form an orthogonal basis in $\mathscr{L}^2_{\mathrm{h}}(S)$ and for $X \in \mathscr{L}^2_{\mathrm{h}}(S)$

$$X = \sum_{j\le k} \alpha_{jk}C_{jk} + \sum_{j<k} \beta_{jk}S_{jk},$$

with real α_{jk}, β_{jk}.

Now let S be an arbitrary density operator. We shall denote by J_0 the set of indices j for which $s_j = 0$ and by J_1 the remaining set. Then S can be represented by the diagonal block matrix

$$\left[\begin{array}{ccc|c} s_1 & & 0 & \\ & \ddots & & \\ & & s_j & 0 \\ & & & \ddots \\ \hline 0 & & & \\ & & 0 & 0 \end{array}\right].$$

Consider first $\mathscr{L}^2_+(S)$. Since $\langle E_{jk}, E_{jk}\rangle_S^+ = 0$ for $j \in J_0$, then the orthogonal basis in $\mathscr{L}^2_+(S)$ consists of operators $\{E_{jk};\ j \in J_1, k \in J_0 \cup J_1\}$.

Therefore any element of $\mathscr{L}^2_+(S)$ can be represented by the block matrix of the form

$$X = \left[\begin{array}{c|c} X_{11} & X_{10} \\ \hline \sim & \sim \end{array}\right],$$

where the blocks X_{11}, X_{10} corresponding to the rows with the indices $j \in J_1$ are determined by X and the remaining blocks denoted by wavy lines are arbitrary (and can be made zero). Similarly, the elements of $\mathscr{L}^2_-(S)$ can be represented by the matrices of the form

$$X = \left[\begin{array}{c|c} X_{11} & \sim \\ \hline X_{01} & \sim \end{array}\right].$$

Turning to the matrix representation of the elements of $\mathscr{L}^2(S)$ we observe that $\langle E_{jk}, E_{jk}\rangle_S = 0$ if and only if $j, k \in J_0$, so that the basis in $\mathscr{L}^2(S)$ consists of $\{E_{jk};\, (j,k) \notin J_0 \times J_0\}$. Therefore the elements of $\mathscr{L}^2(S)$ are represented by matrices

$$X = \left[\begin{array}{c|c} X_{11} & X_{10} \\ \hline X_{01} & \sim \end{array}\right].$$

For $X \in \mathscr{L}^2_{\mathrm{h}}(S)$ the matrix is Hermitean so that $X^*_{11} = X_{11}$, $X^*_{10} = X_{01}$.

Especially simple is the case of the pure state $S = |\psi_1)(\psi_1|$. Then the elements of $\mathscr{L}^2_+(S)$ can be represented by infinite row vectors

$$X = \left[\begin{array}{c} x_{11}x_{12}\cdots \\ \hline \sim \end{array}\right], \qquad \sum_j |x_{1j}|^2 < \infty.$$

If we take the inessential part of the matrix equal to zero, then we have $X = |\psi_1)(\psi|$, $\psi \in \mathscr{H}$ so that $\mathscr{L}^2_+(S)$ is naturally isomorphic to the space $\mathscr{H}^*$ of continuous linear functionals on $\mathscr{H}$. Similarly the elements of $\mathscr{L}^2(S)$ are represented by

$$X = \left[\begin{array}{c|c} x_{11} & \\ x_{21} & \sim \\ \vdots & \end{array}\right]$$

i.e., we can take $X = |\psi)(\psi_1|$, $\psi \in \mathscr{H}$, so that $\mathscr{L}^2_-(S)$ naturally isomorphic to $\mathscr{H}$. For the elements of $\mathscr{L}^2(S)$ we have representation

$$X = \left[\begin{array}{c|c} x_{11}x_{12}\cdots & \\ x_{21} & \sim \\ \vdots & \end{array}\right], \tag{2.10.111}$$

i.e., $X = |\psi_1)(\varphi| + |\psi)(\psi_1|$, $\varphi, \psi \in \mathscr{H}$. Elements of $\mathscr{L}^2_{\mathrm{h}}(S)$ are represented by Hermitean matrices of the form (2.10.111), *i.e.*,

$$X = |\psi_1)(\psi| + |\psi)(\psi_1|, \qquad \psi \in \mathscr{H}. \tag{2.10.112}$$

We are now going to introduce an important notion of commutation operator of a state which will be very useful in the statistical theory of Chapters 5, 6. It follows from Proposition 2.8.3 that the form $[\cdot, \cdot]_S$ is continuous on $\mathscr{L}^2(S)$. Therefore it can be represented as

$$[Y, X]_S = \langle Y, \mathfrak{D} \cdot X\rangle_S, \tag{2.10.113}$$

where $\mathfrak{D}$ is a (complex)-linear bounded operator in $\mathscr{L}^2(S)$. It is called the *commutation operator of* S. Since $[\cdot, \cdot]_S$ is skew-Hermitean, the operator $\mathfrak{D}$ is skew-Hermitean, *i.e.* $\mathfrak{D}^* = -\mathfrak{D}$. The inequalities (1) and (2) of Proposition 2.8.3 then can be written as

$$1 \pm \frac{\mathrm{i}}{2}\mathfrak{D} \geq 0 \qquad \text{in} \quad \mathscr{L}^2(S). \tag{2.10.114}$$

It follows that

$$\left(1 + \frac{1}{4}\mathfrak{D}^2\right) = \left(1 + \frac{\mathrm{i}}{2}\mathfrak{D}\right)\left(1 - \frac{\mathrm{i}}{2}\mathfrak{D}\right) \geq 0.$$

Since the forms $\langle \cdot, \cdot\rangle_S$ and $[\cdot, \cdot]_S$ are real on the real subspace $\mathscr{L}^2_{\mathrm{h}}(S)$, this subspace is invariant under the operator $\mathfrak{D}$. Considered as an operator in $\mathscr{L}^2_{\mathrm{h}}(S)$ it is a (real)-linear bounded skew-symmetric operator satisfying $1 + \frac{1}{4}\mathfrak{D}^2 \geq 0$. The equation (2.8.92) implies

$$\mathfrak{D} \cdot I = 0. \tag{2.10.115}$$

To describe the action of the operator $\mathfrak{D}$ more explicitly we first observe that for bounded X, Y (2.10.113) implies via (2.8.88) and (2.8.90)

$$\mathrm{i}\,\mathrm{Tr}[X, S]Y^* = \mathrm{Tr}((\mathfrak{D} \cdot X) \circ S)Y^*.$$

If follows that $Z = \mathfrak{D} \cdot X \in \mathscr{L}^2(S)$ is the solution of the equation

$$Z \circ S = \mathrm{i}[X, S]. \tag{2.10.116}$$

It is convenient to solve this equation using the matrix representation (2.10.110). Applying both sides of (2.10.116) to ψ_k and taking the inner product with ψ_j we obtain

$$\tfrac{1}{2}(s_k + s_j)(\psi_j|Z\psi_k) = \mathrm{i}(s_k - s_j)(\psi_j|X\psi_k),$$

whence we get the matrix elements of $Z = \mathfrak{D} \cdot X$

$$(\psi_j | Z\psi_k) = \frac{2\mathrm{i}(s_k - s_j)}{(s_k + s_j)} (\psi_j | X\psi_k).$$

Therefore the action of $\mathfrak{D}$ is expressed by multiplication of matrix elements $x_{jk} = (\psi_j | X\psi_k)$ of X by $2\mathrm{i}(s_k - s_j)(s_k + s_j)^{-1}$:

$$\mathfrak{D}([x_{jk}]) = \left[\frac{2\mathrm{i}(s_k - s_j)}{(s_k + s_j)} x_{jk} \right]. \tag{2.10.117}$$

The basis $\{E_{jk}\}$ is thus the basis of eigenvectors of $\mathfrak{D}$ in $\mathscr{L}^2(S)$. Therefore for any function f the action of the operator $f(\mathfrak{D})$ is described by the relation

$$f(\mathfrak{D})([x_{jk}]) = \left[f\left(\frac{2\mathrm{i}(s_k - s_j)}{(s_k + s_j)} \right) x_{jk} \right].$$

In particular we shall need the following functions

$$\begin{aligned} \left(1 + \frac{\mathrm{i}}{2}\mathfrak{D}\right)([x_{jk}]) &= \left[\frac{2s_j}{s_k + s_j} x_{jk} \right], \\ \left(1 - \frac{\mathrm{i}}{2}\mathfrak{D}\right)([x_{jk}]) &= \left[\frac{2s_k}{s_k + s_j} x_{jk} \right], \\ \left(1 + \frac{1}{4}\mathfrak{D}^2\right)([x_{jk}]) &= \left[\frac{2s_k s_j}{(s_k + s_j)^2} x_{jk} \right]. \end{aligned} \tag{2.10.118}$$

From these relations follows

Proposition 2.10.1. *The state S is exact if and only if any of operators $1 \pm \frac{1}{2}\mathrm{i}\mathfrak{D}$, $1 + \frac{1}{4}\mathfrak{D}^2$ is nondegenerated.*

Assuming the condition of the proposition to hold we have the relation

$$\left(1 \pm \frac{\mathrm{i}}{2}\mathfrak{D}\right)^{-1} = \left(1 + \frac{1}{4}\mathfrak{D}^2\right)^{-1} \left(1 \mp \frac{\mathrm{i}}{2}\mathfrak{D}\right). \tag{2.10.119}$$

2.11. Comments

Section 2.1. For systematic exposition of the operator theory in a Hilbert space see Stone [129], Akhiezer and Glazman [2], Riesz and Sz-Nagy [119]. A modernized course of functional analysis adapted for applications in mathematical physics is given by Reed and Simon [118]. Rich complementary material can be found in Halmos [48]. In mathematical texts the outer product sometimes is denoted by a less expressive symbol $\bar{\varphi} \otimes \psi$.

Section 2.2. Introduced first in a somewhat implicit form by Carleman the general resolutions of identity (on $\mathbb{R}$) were extensively studied by Naimark [100, 101]. Theorem 2.2.1 is proved in [59].

Section 2.3-2.4. Proofs of the spectral theorems for Hermitean and self-adjoint operators can be found in the aforementioned textbooks. They are related to the names of Hilbert, von Neumann, Stone, Riesz and others. The spectral theorem for symmetric operators is due to Naimark [100] (see also Akhiezer and Glazman [2]). The formal Dirac's expansions over continuous systems of "eigenvectors" can be founded in the framework of rigged Hilbert spaces (see Gelfand and Vilenkin [40] and Bogoljubov, Logunov and Todorov [16]).

Section 2.5. Theorem 2.5.1 was proved by Naimark [101]. Our argument patterns more modern proofs (*cf.* Akhiezer and Glazman [2]). For tensor products of Hilbert space see *e.g.* Reed and Simon [118].

The concept of realization and its relations to Naimark's theorem and to randomized procedures of mathematical statistics were discussed in author's papers [57, 59, 64]. A different opportunity for emergence of nonorthogonal resolutions of identity (in physical language "overcomplete systems") is related to the so called indirect measurements (Mandelstam [95], Gordon and Louisell [44]). Under an indirect measurement the object $\mathscr{H}$ interacts with the "measuring apparatus" $\mathscr{H}_1$ which is initially in some state S_1; then a simple measurement over $\mathscr{H}_1$ is performed. One can readily see that the statistics of such indirect measurement is also described by the generally nonorthogonal resolution of identity in $\mathscr{H}$ (*cf.* Kraus [80]).

We do not touch the important question of a mechanism which transfers information from a microscopic level in the process of measurement. This should be described quantum-mechanically as an interaction between the object and the measuring apparatus. A fruitful point of view is to consider the apparatus as a "macroscopic system", *i.e.*, a system with a very large number of degrees of freedom. The interaction is then properly treated in the framework of quantum statistical mechanics and the phenomenological description of measurements can be recovered (see Daneri, Loinger and Prosperi [25]). One further step is to assume that the apparatus is a system with infinitely many degrees of freedom. Such systems are described by algebras of observables for which there exist the so called disjoint, *i.e.*, "macroscopically distinguishable" states. Interaction with the object drives the apparatus in one of the disjoint states labelled by the different values u of the result of the measurement. In the paper of Hepp [54] a number of interesting models of measurement

processes are considered which convincingly corroborate this point of view.

Section 2.6. The formal uncertainty relation $D(X)D(Y) \geq \frac{1}{4}\overline{[X, Y]^2}$ for arbitrary X, Y was established by Robertson [120], who generalized the Heisenberg uncertainty relation for position and momentum observables (see Section 3.3). Proposition 2.6.1 is proved in Davies' book [27]. The problem of compatibility and joint measurement has been discussed from various points of views by von Neumann [138], Urbanik [133], Varadarajan [135] and many others. For the functional calculus of several commuting self-adjoint operators see Riesz and Sz.-Nagy [119].

Section 2.7. Trace-class and Hilbert-Schmidt operators are considered in Schatten [122], Gelfand and Vilenkin [40], Reed and Simon [118].

Section 2.8. The $\mathscr{L}^2(S)$ spaces were introduced by Holevo [63,65] for a normal state of a von Neumann algebra where Theorem 2.8.1 was proved. The notion of square-summable operator appeared first in Holevo [59] and Kraus and Shröter [81].

Section 2.9. The rigorous uncertainty relation for a self-adjoint operator was obtained by Kraus and Schröter [81]. The construction of the integral $\int f(x)M(\mathrm{d}x)$ is given in Holevo [65].

Section 2.10. The matrix representation of unbounded operators meets well-known difficulties (see *e.g.*Akhiezer and Glazman [2]). The contents of this section show that if only "second moments" of the operator are concerned, then a satisfactory matrix representation exists. The commutation operator was introduced by Holevo [63, 65] for a normal state on a von Neumann algebra. There is a simple relation between the commutation operator $\mathfrak{D}$ and the modular operator Δ of the Tomita-Takesaki theory [132]; namely $\Delta = (1 + \frac{1}{2}\mathrm{i}\mathfrak{D})(1 - \frac{1}{2}\mathrm{i}\mathfrak{D})^{-1}$. For an exact S on $\mathfrak{B}(\mathscr{H})$ this follows from (2.10.118) and the fact that $\Delta \cdot X = SXS^{-1}$.

Chapter 3
Symmetry groups in quantum mechanics

3.1. Statistical model and Galilean relativity

The spatially-temporal structure which we are going to introduce is a distinctive feature in the description of mechanical objects. In classical mechanics the motion of point masses is described in a frame of reference which consists of a spatial Cartesian coordinate system and a clock. The distinguished class is formed by the so called inertial frames which are characterized by the property that a free point mass performs uniform rectilinear motion with respect to such a frame. It follows that the coordinates $(\boldsymbol{\xi}, \boldsymbol{\tau})$ and $(\boldsymbol{\xi}', \boldsymbol{\tau}')$ of a point mass in any two inertial frames of reference are related by the Galilei transformation

$$\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{x} + \boldsymbol{v}\boldsymbol{\tau}, \qquad \boldsymbol{\tau}' + \tau + t, \tag{3.1.1}$$

where $\boldsymbol{R}$ is an orthogonal matrix (rotation) which describes the orientation of the new spatial axes with respect to the old ones, $\boldsymbol{x}$ is the vector of the spatial shift of the origin, $\boldsymbol{v}$ is the relative velocity, and τ is the time shift showing the difference between the readings of the two clocks.

All transformations of the form (3.1.1) constitute the *Galilean group* which contains the subgroup of *kinematical transformations*

$$\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{x} + \boldsymbol{v}\boldsymbol{\tau}, \qquad \boldsymbol{\tau}' = \tau, \tag{3.1.2}$$

and the *Euclidean group* of spatial transformations

$$\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{x}.$$

The fundamental Galilean relativity principle says that the laws of mechanics are the same in all inertial frames or, mathematically, the equations of motion are invariant under all transformations of the form (3.1.1). Of course, this refers only to a free object; if the motion occurs in a field of a force, then the full Galilean relativity should be replaced by a restricted one which takes into account symmetry properties of the field.

For example, equations of motion in a time-independent field must be invariant with respect to all transformations

$$\boldsymbol{\xi}' = \boldsymbol{\xi} + \boldsymbol{v}\boldsymbol{\tau}, \qquad \boldsymbol{\tau}' = \boldsymbol{\tau} + t.$$

If the field is rotationally invariant, then the rotations $\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi}$ should be added etc.

Turning to quantum mechanics, we adopt that Galilean relativity holds also for microobjects. But we cannot take the formulation of the principle literally since the notion of point mass becomes meaningless. Given now are not coordinates $(\boldsymbol{\xi}, \boldsymbol{\tau})$ of point masses but the totality of statistical results of various measurements. Therefore an appropriate preliminary formulation of the Galilean relativity principle is: *the statistics of any experiment is the same in any inertial frame of reference*.

To give a more precise formulation consider a device which prepares a state S of the object. Then a measurement $\boldsymbol{M}$ is performed with the resulting probability distribution $\mu_S^{\boldsymbol{M}}$. Both the preparing and the measuring devices are macroscopic objects and we can associate with them the frame of reference $(\boldsymbol{\xi}, \boldsymbol{\tau})$ (Figure 3.1). Assume now that the frame $(\boldsymbol{\xi}, \boldsymbol{\tau})$ is transformed by means of a Galilean transformation g into the new frame $(\boldsymbol{\xi}', \boldsymbol{\tau}')$, the whole experimental set-up being the same in this new frame. Preparation of the state S with the subsequent change in the position of the preparing device can be considered as a new way of preparation which results in the new state gS; in the same way the measurement $\boldsymbol{M}$ accomplished after the transformation g can be considered as a new measurement $g\boldsymbol{M}$. Since the whole experimental set-up remains relatively unchanged we must have

$$\mu_{gS}^{g\boldsymbol{M}} = \mu_S^{\boldsymbol{M}}, \tag{3.1.3}$$

for all states S and measurements $\boldsymbol{M}$.

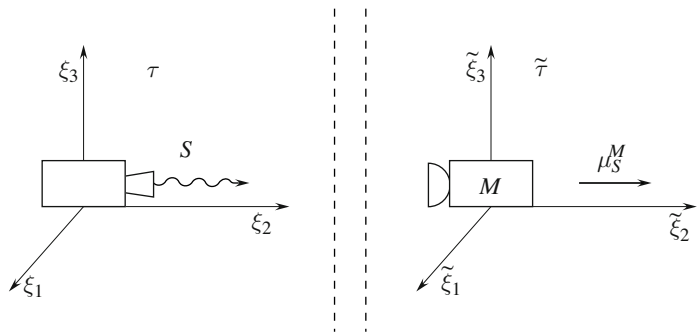

Figure 3.1.

Assuming full equivalence of all inertial frames of reference we must adopt that $S \to gS$ is a one-to-one map of the set of states $\mathfrak{S}$ onto $\mathfrak{S}$ (and correspondingly for $\boldsymbol{M} \to g\boldsymbol{M}$). We do not require for the present that $\mathfrak{S} = \mathfrak{S}(\mathscr{H})$; all that we need is that the statistical model $(\mathfrak{S}, \mathfrak{M})$ is separated. If follows then that the map $S \to gS$ is affine; let $S = \sum_j p_j S_j$, then

$$\mu_{gS}^{g\boldsymbol{M}} = \mu_S^{\boldsymbol{M}} = \sum_j p_j \mu_{S_j}^{\boldsymbol{M}} = \sum_j p_j \mu_{gS_j}^{g\boldsymbol{M}}.$$

Since the measurements $g\boldsymbol{M}$ constitute all $\mathfrak{M}$, $gS = \sum_j p_j (gS_j)$ what is asserted. An affine one-to-one map of the set of states $\mathfrak{S}$ onto $\mathfrak{S}$ is called an automorphism of $\mathfrak{S}$.

It is natural to assume that a successive application of transformations g_1, g_2 is equivalent to the transformation $g_1 g_2$, so that

$$\begin{aligned} g_1(g_2 S) &= (g_1 g_2) S, \\ g_1(g_2 \boldsymbol{M}) &= (g_1 g_2) \boldsymbol{M}. \end{aligned} \tag{3.1.4}$$

This means, first, that the group $G = \{g\}$ acts as a group of automorphisms of the set of states $\mathfrak{S}$ and, second, as a group of one-to-one transformations of the set of measurements $\mathfrak{M}$. Moreover the actions are connected by the requirement (3.1.3). This constitutes the mathematical formulation of the relativity principle.

By group G we have meant the Galilean group and the statistical model was one of quantum theory. However, the group could be any other symmetry group, *i.e.*, a group of transformations of an appropriate set. In particular, if the Galilean group is replaced by the Poincaré group the Galilean relativity cedes to Einstein relativity.

So far we have not used the specificity of the quantum statistical model. The following theorem going back to Wigner, reveals the structure of automorphisms of the set $\mathfrak{S}(\mathscr{H})$.

Theorem 3.1.1. *Any automorphism of the set of quantum states $\mathfrak{S}(\mathscr{H})$ has the form*

$$S \to VSV^*, \tag{3.1.5}$$

where V is a unitary or anti-unitary operator in the Hilbert space $\mathscr{H}$.

(An operator V from $\mathscr{H}$ onto $\mathscr{H}$ is called *anti-unitary* if it is conjugate-linear and satisfies $(V\varphi|V\psi) = (\overline{\varphi|\psi})$.)

It is important to note that the map (3.1.5) determines the operator V only up to a scalar factor of unit modulus: V can be multiplied by an $\omega \in \mathbb{C}$, $|\omega| = 1$, without changing the state VSV^*. Now let G act as

a group of automorphisms of $\mathfrak{S}(\mathscr{H})$. By Theorem 3.1.1 for any $g \in G$ there is a unitary or anti-unitary operator V_g in $\mathscr{H}$ such that $gS = V_g S V_g^*$. Moreover, by (3.1.4)

$$V_{g_2} V_{g_1} S V_{g_1}^* V_{g_2}^* = V_{g_2 g_1} S V_{g_2 g_1}^*; \qquad g_1, g_2 \in G,$$

for all $S \in \mathfrak{S}(\mathscr{H})$. It follows that there is a complex-valued function $\omega(g_1, g_2)$, with $|\omega(g_1, g_2)| \equiv 1$ such that

$$V_{g_2} V_{g_1} = \omega(g_2, g_2) V_{g_2 g_1}; \qquad g_1 g_2 \in G. \tag{3.1.6}$$

The groups we shall consider are continuous groups in the sense that they have a natural topological structure defining the notions of neighborhood, convergence etc. We shall assume that the map $g \to V_g$ is continuous[1] in the sense that $g' \to g$ implies $(\varphi|V_g\psi) \to (\varphi|V_g\psi)$ for all $\varphi, \psi \in \mathscr{H}$. We shall always put $V_e = I$, where e is the unit of the group G. Assume at last that the group G is connected, *i.e.*, any two elements of G can be connected by a continuous curve lying in G. Then all operators V_g are unitary – otherwise it would be possible to pass continuously from the unitary operator $V_e = I$ to an anti-unitary operator, which is clearly impossible.

A family of unitary operators $g \to V_g$; $g \in G$ in a Hilbert space $\mathscr{H}$, satisfying (3.1.6) is called *projective unitary representation* of G in $\mathscr{H}$. If $\omega \equiv 1$ it is called simply *unitary*. We shall consider only representations which are continuous in the aforementioned sense.

One of the principal problems of the theory of group representations is classification of all representations of a given group. Having such a classification for a symmetry group G one can describe all theoretically possible quantum objects with this type of symmetry. Of special importance are *irreducible* representations $g \to V_g$, characterized by the property that the only invariant closed subspaces of all operators $\{V_g; g \in G\}$ are [0] and $\mathscr{H}$. This means that the representation has no proper subrepresentations and in this sense is minimal. Under certain regularity assumptions any representation can be decomposed into a discrete or continuous sum (integral) of irreducible representations.

According to Wigner an irreducible representation of a symmetry group describes a "elementary system" with this type of symmetry. We may call it also "elementary object", but one should take care of the far-reaching

[1] It is not a serious restriction: there is a theorem that Borel measurability of all functions $(\varphi|V_{(\cdot)}\psi)$, $\varphi, \psi \in \mathscr{H}$, together with (3.1.6) implies continuity.

associations with "elementary particles" of quantum physics and remember that elementary system or object is essentially a mathematical concept. A complete answer to the question – what is an elementary particle – presumes a solution of the most profound problems of contemporary physics. Meanwhile the term "particle" is used loosely in different contexts, and we shall follow this custom calling by particle an elementary quantum object in the above-defined sense.

Equations for free quantum particles are in fact a suitable way for description of irreducible representations of the corresponding symmetry group. In particular, Galilean relativity implies full kinematical and dynamical description of a nonrelativistic quantum particle, including the basic Schrödinger equation. In this chapter we shall try to explain this point by elementary means, paying the main attention to the question – how symmetry properties allow one to relate physical parameters such as coordinate, velocity, time, angle etc. to certain quantum measurements, *i.e.*, resolutions of identity in the Hilbert space of the representation.

3.2. One-parameter shift groups and uncertainty relations

Assume that the apparatus preparing a quantum state is shifted along the axis ξ by the distance x, which is described by the Galilei transformation $\xi' = \xi - x$ (see the left side of Figure 3.2). If S was the basic state, the new state will be $S_x = V_x S V_x^*$ where $x \to V_x$, $x \in \mathbb{R}$, is a projective unitary representation of the additive group of the real line $\mathbb{R}$.

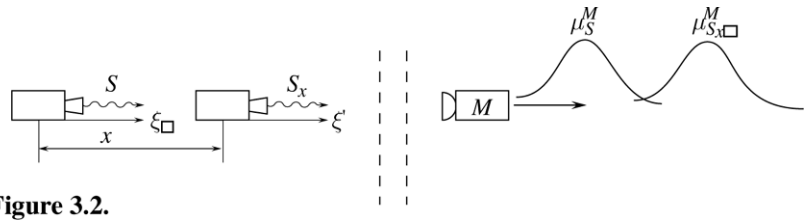

Figure 3.2.

Similarly, consider the transformations of the reference frame where the position of the apparatus is not changed but the time count is shifted according to the relation $\tau' = \tau + t$. This means simply that the preparation procedure is the same but begins t units of time earlier as compared to the initial one. Then the new prepared state S_t will be related to the basic one by the formula $S_t = V_t S V_t^*$, where $t \to V_t$, is again a projective unitary representation[2] of $\mathbb{R}$.

[2] Thus S_t is similar to S_{-x} rather than to S_x. With this choice S_t gives the temporal evolution of the state S.

These are the two most important examples when representations of $\mathbb{R}$ arise. We shall consider them here from a general point of view, postponing the applications to spatial and temporal shifts to the corresponding sections. However the results of this section are best visualized by having in mind the case of spatial shifts. Let us recall that by representation we always mean a continuous representation.

Proposition 3.2.1. *Any projective unitary representation $\theta \to \widetilde{V}_\theta$, $\theta \in \mathbb{R}$, of the additive group of $\mathbb{R}$ reduces to a unitary one; i.e., there exists a unitary representation $\theta \to V_\theta$, $\theta \in \mathbb{R}$, such that $V_\theta = \alpha_\theta \widetilde{V}_\theta$, $|\alpha_\theta| \equiv 1$.*

The family $\{V_\theta\}$ satisfies

$$V_{\theta_1} V_{\theta_2} = V_{\theta_1+\theta_2}; \qquad \theta_1, \theta_2 \in \mathbb{R}$$

and therefore constitutes an one-parameter group of unitary operators. By Stone's theorem of Section 2.4,

$$V_\theta = \exp(\mathrm{i}\theta A), \qquad \theta \in \mathbb{R},$$

where A is a self-adjoint operator in $\mathscr{H}$. Therefore the action of the one-parameter automorphism group on quantum states is described by

$$S \to S_\theta = \mathrm{e}^{\mathrm{i}\theta A} S \mathrm{e}^{-\mathrm{i}\theta A}. \tag{3.2.7}$$

In what follows we shall restrict for simplicity to pure states. If $S = |\psi)(\psi|$ with $\psi \in \mathscr{D}(A)$, then $S_\theta = |\psi_\theta)(\psi_\theta|$ with $\psi_\theta = \exp(\mathrm{i}\theta A)\psi \in \mathscr{D}(A)$.

We shall supply with the subscript θ the quantities referring to the state S_θ, so that $E_\theta(X) = (\psi_\theta | X \psi_\theta)$ is the mean value of an observable X with respect to the state S_θ. Proceeding formally, we get

$$\frac{\mathrm{d}}{\mathrm{d}\theta} E_\theta(x) = \left(\frac{\mathrm{d}\psi_\theta}{\mathrm{d}\theta} \middle| X \psi_\theta \right) + \left(\psi_\theta | X \frac{\mathrm{d}\psi_\theta}{\mathrm{d}\theta} \right).$$

This is easily verified if X is bounded. Then by the equation (2.4.49)

$$\frac{\mathrm{d}}{\mathrm{d}\theta} E_\theta(X) = 2 \operatorname{Im}(A\psi_\theta | X \psi_\theta). \tag{3.2.8}$$

Using the uncertainty relation (2.6.63) we get an important Mandelstam-Tamm inequality

$$D_\theta(X) \cdot D_\theta(A) \geq \frac{1}{4} \left| \frac{\mathrm{d}}{\mathrm{d}\theta} E_\theta(X) \right|^2. \tag{3.2.9}$$

We shall obtain a rigorous version of this inequality for arbitrary states and measurements with finite second moments in Section 6.3.

The importance of this result lies in the fact that it gives a principal lower bound for accuracy of quantum measurements, which substantially generalizes the uncertainty relation. To explain this point we must formulate the *quantum estimation problem*. Assume that an apparatus prepares the basic state S which is completely known. Then the apparatus is transformed according to a change of the shift parameter θ. The new prepared state will be the state S_θ of (3.2.7). The actual value θ is supposed to be unknown and the problem is to estimate this value statistically by making a measurement over the object. The measurement must be $\mathbb{R}$-valued and for simplicity we restrict here to observables. Any observable X presents a *statistical estimate* of parameter θ. The quality of the estimate X can be measured by the mean-square deviation $E_\theta((X-\theta)^2) = D_\theta(X) + (E_\theta(X)-\theta)^2$. Then the inequality (3.2.9) sets the lower bound

$$E_\theta((X-\theta)^2) \geq b(\theta)^2 + \frac{[1+b'(\theta)]^2}{4D_\theta(A)}, \tag{3.2.10}$$

where $b(\theta) = E_\theta(X) - \theta$ is the *bias* of the estimate X.

This way of reasoning is characteristic for statistical thinking and is new only in the context of quantum theory. From this point of view, to any physical parameter corresponds a variety of measurements which differ at least by their accuracy. However the class of all estimates is too general – the inequality (3.2.10) holds for observables X which in fact may have nothing to do with the parameter θ. One has to require some properties for observables X to relate them with the parameter θ. Following statistical terminology we call an estimate X *unbiased* if $b(\theta) = 0$ or

$$E_\theta(X) = \theta, \qquad \theta \in \mathbb{R}. \tag{3.2.11}$$

Unbiasedness means that there is no systematic error in the measurement results. For unbiased estimates the inequality (3.2.10) takes the form

$$D_\theta(X) \geq [4D_\theta(A)]^{-1}. \tag{3.2.12}$$

Note that $D_\theta(A) \equiv D_S(A)$ since A commutes with the unitary group $\{\exp(\mathrm{i}\theta A)\}$. Thus for any unbiased estimate of the shift parameter θ the variance is bounded from below by the quantity inversely proportional to the uncertainty of the observable A in the basic state S. Later we shall apply (3.2.12) to estimation of spatial and temporal shifts and now we are going to discuss the connection of (3.2.12) with traditional forms of the uncertainty relation.

An observable represented by the self-adjoint operator B is called *canonically conjugate* to A if the corresponding unitary groups satisfy the *Weyl canonical commutation relation* (CCR)

$$e^{i\chi B}e^{i\theta A} = e^{i\theta\chi}e^{i\theta A}e^{i\chi B}; \qquad \theta, \chi \in \mathbb{R}. \tag{3.2.13}$$

Obviously this is equivalent for A to be canonically conjugate to $-B$. Differentiating with respect to θ and χ the identity

$$(e^{-i\chi B}\psi|e^{i\theta A}\psi) = e^{i\theta\chi}(e^{-i\theta A}\psi|e^{i\chi B}\psi),$$

which follows form (3.2.13), we get with the help of (2.4.49)

$$2\,\mathrm{Im}(A\psi|B\psi) = (\psi|\psi); \qquad \psi \in \mathscr{D}(A) \cap \mathscr{D}(B). \tag{3.2.14}$$

If A and B were bounded this would imply the *Heisenberg CCR*

$$[A, B] = iI. \tag{3.2.15}$$

However as we shall see both canonically conjugate observables are necessarily unbounded. Therefore (3.2.15) involves domain troubles, and one can expect (3.2.15) to hold only on a dense domain. Moreover (3.2.13) and (3.2.15) are not strictly equivalent since there are pairs of operators satisfying (3.2.15) on a dense domain which do not satisfy (3.2.13). The Weyl CCR refers to physically more primary objects which are unitary groups describing transformations of states.

Substituting (3.2.14) into (2.6.63) we get the *Heisenberg uncertainty relation*

$$D_S(A) \cdot D_S(B) \geq \tfrac{1}{4}. \tag{3.2.16}$$

Let us show that this follows also from (3.2.12). Then (3.2.12) can be regarded as a generalization of the Heisenberg uncertainty relation to situations where the canonically conjugate observable B may not exist.

We shall deduce (3.2.16) from (3.2.12) by showing that the canonically conjugate observable B (provided it exists) is, up to a constant, unbiased estimate of the shift parameter θ. Rewriting (3.2.13) in the form

$$e^{-i\theta A}e^{i\chi B}e^{i\theta A} = e^{i\chi(B+\theta)}, \tag{3.2.17}$$

and denoting $G(d\lambda)$ the spectral measure of B we get by (2.4.46)

$$\int e^{i\chi\lambda}(e^{i\theta A}\psi|G(d\lambda)e^{i\theta A}\psi) = \int e^{i\chi(\lambda+\theta)}(\psi|G(d\lambda)\psi)$$

for all $\lambda \in \mathbb{R}$ and all unit vectors $\psi \in \mathscr{H}$. Both sides of this equality are the Fourier transforms of probability distributions. By the uniqueness property of the Fourier transform this is equivalent to

$$(e^{i\theta A}\psi|G(\Lambda)e^{i\theta A}\psi) = (\psi|G(\Lambda_{-\theta})\psi), \tag{3.2.18}$$

for all Borel sets $\Lambda \subset \mathbb{R}$, where $\Delta_{-\theta} = \{\lambda - \theta : \lambda \in \Lambda\}$ is the shift of the set Λ by $-\theta$. Since it holds for all $\psi \in \mathscr{H}$, then

$$e^{-i\theta A}G(\Lambda)e^{i\theta A} = G(\Lambda_{-\theta}); \qquad \theta \in \mathbb{R}, \quad \Lambda \in \mathscr{A}(\mathbb{R}). \tag{3.2.19}$$

A resolution of identity satisfying this relation is called *covariant* with respect to the representation $\theta \to \exp(i\theta A)$ of $\mathbb{R}$. Appropriately generalized, the covariance property will appear to be very important in the statistical analysis of quantum systems. Denoting by $\mu_\theta^{\boldsymbol{G}}(d\lambda)$ the probability distribution of the measurement $\boldsymbol{G} = \{G(d\lambda)\}$ with respect to the state $S_\theta = |\psi_\theta)(\psi_\theta|$ we get from (3.2.18)

$$\mu_\theta^{\boldsymbol{G}}(\Lambda) = \mu_0^{\boldsymbol{G}}(\Lambda_{-\theta}); \qquad \theta \in \mathbb{R}, \quad \Lambda \in \mathscr{A}(\mathbb{R}). \tag{3.2.20}$$

This means that the transformation of the preparing apparatus by the shift θ is reflected by the corresponding shift in the measurement probability distribution. This is an appealing reason to associate a resolution of identity satisfying (3.2.19) with a measurement of the shift parameter θ.

The covariance property implies

$$E_\theta(B) = \int_{-\infty}^{\infty} \lambda \mu_\theta^G(d\lambda) = \int_{-\infty}^{\infty} (\lambda + \theta)\mu_0^G(d\lambda) = E_0(B) + \theta,$$

so that B is up to a constant an unbiased estimate of θ. Moreover *the variance $D_\theta(B)$ does not depend on θ* and is equal to $D_0(B) \equiv D_S(B)$. Substituting $X = B - E_0(B)$, which is strictly unbiased, into (3.2.12) yields (3.2.16) as it was asserted.

3.3. Kinematics of a quantum particle in one dimension

In the case of one-dimensional motion a frame of reference is described by the pair of variables (ξ, τ) where ξ is the spatial coordinate and τ is the temporal coordinate. Consider the Galilei transformation

$$\xi' = \xi - x - v\tau, \qquad \tau' = \tau. \tag{3.3.21}$$

It corresponds to the change in the position of the preparing apparatus when it is shifted along the basic axis ξ by the distance x and is moving with the velocity v relative to its initial position (see the left side of

Figure 3.3). Any such transformation is described by the pair of real parameters $g = (x, v)$ with the law of composition given by

$$(x_1, v_1)(x_2, v_2) = (x_1 + x_2, v_1 + v_2).$$

Therefore the group of kinematical transformations in one dimension is the additive group of the plane $\mathbb{R}^2$.

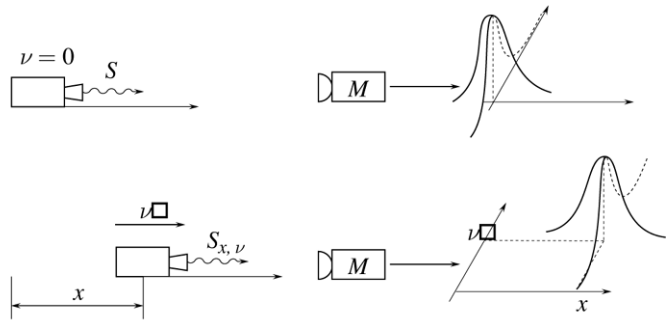

Figure 3.3.

According to the general scheme presented in Section 3.1 we shall look for projective unitary representations $(x, v) \to W_{x,v}$ of $\mathbb{R}^2$. First we show that choosing an appropriately multiplier of $X_{x,v}$ we can always reduce (3.1.6) to

$$W_{x_1,v_1} W_{x_2,v_2} = \exp\left[-\frac{\mathrm{i}\mu}{2}(x_1 v_2 - x_2 v_1)\right] W_{x_1+x_2,v_1+v_2}. \qquad (3.3.22)$$

This is sometimes called the *Weyl-Segal CCR*.

Put $V_x = W_{x,0}$, $U_v = W_{0,v}$. The automorphisms

$$S \to V_x S V_x^*, \qquad S \to U_v, S U_v^*$$

describe changes of states correspondingly under the spatial and the velocity shifts. By Proposition 3.2.1 we can assume that $\{V_x\}$ and $\{U_v\}$ form one-parameter groups of unitary operators:

$$V_{x_1} V_{x_2} = V_{x_1+x_2}, \qquad U_{v_1} V_{v_2} = U_{v_1+v_2}.$$

Consider the transformation which consists of the spatial shift by the distance x and change of the velocity from 0 to v. This can be accomplished in the two different ways

$$(x, v) = (x, 0)(0, v) = (0, v)(x, 0),$$

but the resulting state should not depend on the choice so that

$$S \to U_v V_x S V_x^* U_v^* = V_x U_v S U_v^* V_x^*$$

for any S. It follows that

$$U_v V_x = \mathrm{e}^{\mathrm{i}\eta(x,v)} V_x U_v, \tag{3.3.23}$$

where $\eta(\cdot,\cdot)$ is a real continuous function. For $x = 0$ or $v = 0$ (3.3.23) implies $1 \equiv \exp \mathrm{i}\eta(0,v) \equiv \exp \mathrm{i}\eta(x,0)$ so we can put $\eta(x,0) \equiv \eta(v,0) \equiv 0$, since $V_0 = U_0 = I$. Multiplying (3.3.23) by $U_{v'}$ we get

$$\eta(x, v + v') = \eta(x, v) + \eta(x, v') (\mathrm{mod}\, 2\pi).$$

The only continuous solution of this equation satisfying $\eta(x, 0) = 0$ is $\eta(x, v) = \eta(x) \cdot v$. In the same way $\eta(x + x', v) = \eta(x, v) + \eta(x', v) (\mathrm{mod}\, 2\pi)$ whence $\eta(x, v) = \mu x v$ with a real constant μ. We thus get

$$U_v V_x = \mathrm{e}^{\mathrm{i}\mu x v} V_x U_v. \tag{3.3.24}$$

Since $W_{x,v}$ is $V_x U_v$, up to an arbitrary factor of unit modulus, we can choose

$$W_{x,v} = \mathrm{e}^{\mathrm{i}\mu x v/2} V_x U_v. \tag{3.3.25}$$

Then $\{W_{x,v}\}$ is easily seen to satisfy (3.3.22). The operators $W_{x,v}$ are sometimes called *displacement operators*.

Assume now that the representation $(x, v) \to W_{x,v}$ is *irreducible*. Then if $\mu = 0$, (3.3.24) implies $[V_x, U_v] \equiv 0$, and one easily shows that the only possibility is the "trivial" one-dimensional representation $(x, v) \to \exp \mathrm{i}(\alpha x + \beta v)$ with $\alpha, \beta \in \mathbb{R}$. Thus, the following proposition holds.

Proposition 3.3.1. *Any projective unitary representation of the group of kinematical transformations* (3.3.21) *can be described by a family of unitary operators* $\{W_{x,v}\}$ *satisfying* (3.3.22). *If the representation is irreducible and* $\dim \mathscr{H} > 1$, *then* $\mu \neq 0$.

As we shall see later, the parameter μ corresponds to the *mass* of the object so that only the case $\mu > 0$ is physically relevant.

By Stone's theorem, in the Hilbert space of the representation $\mathscr{H}$ there are self-adjoint operators P and Q such that

$$V_x = \mathrm{e}^{-\mathrm{i}xP}, \qquad x \in \mathbb{R}, \tag{3.3.26}$$

$$U_v = \mathrm{e}^{\mathrm{i}\mu v Q}, \qquad v \in \mathbb{R}. \tag{3.3.27}$$

The relation (3.3.24) (which is equivalent to (3.3.22)) is just the Weyl CCR (3.2.13) for $A = -P$, $B = Q$ saying the Q and P *are canonically conjugate observables*.

The spectral measures of Q and P determined from the spectral representations

$$Q = \int \xi E(\mathrm{d}\xi), \qquad \mu^{-1} P = \int \eta F(\mathrm{d}\eta),$$

describe quantum measurements satisfying the covariance condition

$$\begin{aligned} V_x^* E(B) V_x &= E(B_{-x}), \qquad B \in \mathscr{A}(\mathbb{R}), \\ U_v^* F(B) U_v &= F(B_{-v}), \qquad B \in \mathscr{A}(\mathbb{R}), \end{aligned} \tag{3.3.28}$$

which follow from (3.2.19). We shall use the first of these relations to explain the kinematical meaning of the observable Q.

Consider the family of the states

$$S_x = \mathrm{e}^{-\mathrm{i}xP} S \mathrm{e}^{\mathrm{i}xP}, \qquad x \in \mathbb{R},$$

corresponding to different values of the *coordinate parameter* x. This parameter describes the position of the preparing apparatus and in this sense it reflects an information on the position of the microobject. If the measurement of the observable Q is performed, then by (3.2.20) applied to $A = -P$, $B = Q$, the resulting probability distributions satisfy

$$\mu_x^E(B) = \mu_0^E(B_{-x}); \qquad x \in \mathbb{R}, \quad B \in \mathscr{A}(\mathbb{R}).$$

This means that the spatial shift of the preparing apparatus is reflected by the corresponding shift of the resulting probability distribution (Figure 3.2). Therefore the observable Q and its spectral measure $E = \{E(\mathrm{d}\xi)\}$ are naturally associated with a measurement of coordinate parameter x.

The property which makes this association possible is the covariance property (3.3.28) with respect to the representation of the spatial shifts group. As we shall see in Chapter 4 there is a variety of covariant measurements: a parameter x can be measured in different ways with different accuracy. According to (3.2.12) the accuracy of measurement of the coordinate parameter x is bounded from below:

$$D_x(X) \geq [4 D_x(P)]^{-1}$$

for any unbiased estimate X of x. We shall show that there is a way to single out the observable Q as an optimal estimate of x. We shall call

Q the *canonical coordinate observable*, as is usually done in physical literature.

In the same way the observable $\mu^{-1}P$ can be associated with measuring the *relative velocity parameter* v in the family of states

$$S_v = \mathrm{e}^{\mathrm{i}\mu v Q} S \mathrm{e}^{-\mathrm{i}\mu v Q}, \qquad v \in \mathbb{R}.$$

Therefore we shall call $\mu^{-1}P$ *the canonical velocity observable*[3].

Since Q and P are canonically conjugate, they satisfy (3.2.14):

$$2\,\mathrm{Im}(Q\psi|P\psi) = (\psi|\psi), \qquad \psi \in \mathscr{D}(Q) \cap \mathscr{D}(P),$$

and the Heisenberg uncertainty relation

$$D_S(Q) \cdot D_S(P) \geq \tfrac{1}{4}. \tag{3.3.29}$$

In physics it is more convenient to use, instead of the velocity observable, the *momentum observable* which is defined as

$$p = m \cdot (\text{velocity observable}) = \hbar P,$$

where m is the "classical mass", $\hbar = m/\mu$ is the quantity proportional to Planck's constant. The meaning of these coefficients will be explained in Section 3.7. Redenoting $q \equiv Q$ we write (3.3.29) in the form

$$D_S(q) \cdot D_S(p) \geq \frac{\hbar^2}{4}, \tag{3.3.30}$$

which is usually formally derived from the Heisenberg CCR

$$[q, p] = \mathrm{i}\hbar. \tag{3.3.31}$$

3.4. Uniqueness theorem. The Schrödinger and the momentum representations

The CCR (3.3.22) is of fundamental importance in quantum theory. A concrete family of unitary operators $\{W_{x,v}\}$ in a Hilbert space $\mathscr{H}$ satisfying (3.3.22) is called a *representation of the CCR*. According to our convention we always mean continuous representations such that $(\varphi|W_{x,v}\psi)$ are continuous functions of (x, v). A complete description of such representations is given by the famous *uniqueness theorem* due to von Neumann and Stone. The proof of the theorem is postponed until Section 5.3.

[3] For the charged particle the velocity observable has a different form, but we shall not touch this question (see, *e.g.*, Jauch [71]).

Theorem 3.4.1 (Uniqueness). *Any two irreducible representations*

$$(x, v) \to W^{(j)}_{x,v}; \quad j = 1, 2,$$

of the CCR are unitarily equivalent: there is a linear isometric map U of $\mathscr{H}_2$ onto $\mathscr{H}_1$, where $\mathscr{H}_j$ is the space of j^{th} representation, such that

$$W^{(2)}_{x,v} = U^* W^{(1)}_{x,v} U; \qquad (x, v) \in \mathbb{R}^2.$$

Moreover any representation of the CCR is a discrete orthogonal sum of irreducible representations.

This implies that any representation of the CCR is unitarily equivalent to the representation of the form

$$(x, v) \to \begin{bmatrix} W^{(0)}_{x,v} & & 0 \\ & \ddots & \\ 0 & & W^{(0)}_{x,v} \end{bmatrix},$$

where $(x, v) \to W^{(0)}_{x,v}$ is a fixed irreducible representation. Thus the CCR gives an essentially unique description of the kinematics of quantum object with the given value of μ. To describe all representations it is sufficient to give only one irreducible representation.

Consider the space $\mathscr{H} = \mathscr{L}^2(\mathbb{R})$ of the complex square-integrable functions on the real line $\mathbb{R}$ and the family of unitary operators in $\mathscr{H}$ acting on $\psi \in \mathscr{H}$ by the relation

$$W_{x,v}\psi(\xi) = \exp\left[\mathrm{i}\mu v\left(\xi - \frac{x}{2}\right)\right]\psi(\xi - x). \tag{3.4.32}$$

One easily checks that $(x, v) \to W_{x,v}$ is a representation of the CCR. It is an irreducible representation. To explain it consider the one-parameter unitary groups $V_x = W_{x,0}$ and $U_v = W_{0,v}$ acting on ψ by the formulas

$$V_x\psi(\xi) = \psi(\xi - x), \tag{3.4.33}$$

$$U_v\psi(\xi) = \mathrm{e}^{\mathrm{i}\mu v\xi}\psi(\xi). \tag{3.4.34}$$

Assume that $\mathscr{L} \subset \mathscr{H}$ is an invariant closed subspace of $\{W_{x,v}; (x, v) \in \mathbb{R}^2\}$. Then it is invariant under (3.4.34). It follows that there is a measurable subset $B \subset \mathbb{R}$ such that $\mathscr{L} = \{\psi : \psi(\xi) = 0, \xi \in B\}$. But such a subspace can be invariant under all shifts (3.4.33) only if either B or the complement of B has zero Lebesgue measure. This corresponds to either $\mathscr{L} = \mathscr{H}$ or $\mathscr{L} = [0]$ so that the representation is irreducible.

The representation (3.4.32) of the CCR in $\mathscr{L}^2(\mathbb{R})$ is called the *Schrödinger representation*. From (3.4.34) and (3.3.27) we get the canonical observables in this representation

$$Q\psi(\xi) = \xi\psi(\xi), \qquad \psi \in \mathscr{D}(Q), \tag{3.4.35}$$

and from (3.4.33) and (3.3.26)

$$P\psi(\xi) = \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}\xi}\psi(\xi), \qquad \psi \in \mathscr{D}(P). \tag{3.4.36}$$

Apparently $\mathscr{D}(Q) \cap \mathscr{D}(P)$ contains the dense subspace $\mathscr{S}(\mathbb{R})$ of the infinitely differentiable functions which tend to zero together with all derivatives faster than any degree of ξ as $|\xi| \to \infty$.

Fix x and v and consider the family of unitary operators $\{W_{\theta x,\theta v}; \theta \in \mathbb{R}\}$. From (3.3.22) it follows that this is a one-parameter unitary group. From (3.4.32)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}W_{\theta x,\theta v}\psi(\xi)\bigg|_{\theta=0} &= \mu v\xi\psi(\xi) - x\mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}\xi}\psi(\xi)\\ &= (\mu vQ - xP)\psi(\xi)\end{aligned}$$

for ψ, say, from $\mathscr{S}(\mathbb{R})$. It follows that

$$W_{x,v} = \exp[\mathrm{i}(\mu vQ - xP)], \tag{3.4.37}$$

where $\mu vQ - xP$ means the self-adjoint extension of the sum defined, *e.g.*, on $\mathscr{S}(\mathbb{R})$.

Since by (3.4.35) Q is just the operator of multiplication by ξ in $\mathscr{L}^2(\mathbb{R})$, then by Section 2.4 Q is "diagonal" in the Schrödinger representation, *i.e.*, in Dirac's notations

$$Q = \int \xi|\xi)(\xi|\mathrm{d}\xi,$$

where $|\xi)$ can be regarded as "vector" of an "unphysical state" in which the object has a precisely determined coordinate ξ. The probability distribution of the coordinate observable $Q = q$ with respect to a state $S = |\psi)(\psi|$ is, according to (2.3.36), $|\psi(\xi)|^2\mathrm{d}\xi$. The more this distribution is concentrated at the point $\bar{q}$, *i.e.*, the smaller is $D_S(q)$, the more the quantum object resembles a classical point mass (a particle).

On the other hand, the momentum operator $p = \hbar P = \hbar\mathrm{i}^{-1}\mathrm{d}/\mathrm{d}\xi$ has the formal eigenfunctions $\exp(\mathrm{i}\xi\eta)$; $\eta \in \mathbb{R}$ (see Section 2.4), corresponding to "unphysical states" in which the object has strictly determined velocity η/μ and momentum $\hbar\eta$. From the Heisenberg uncertainty relation (3.3.30) it follows that the more definite is the momentum, *i.e.*, the

smaller $D_S(p)$, the less definite is the spatial localisation of the object. In the states with $D_S(p) \approx 0$ the object thus resembles a classical wave. Therefore, depending on preparation of the basic state, a quantum object may display features both of a classical wave and a classical particle.

Consider the Fourier transform

$$\tilde{\psi}(\eta) = \frac{1}{\sqrt{2\pi\hbar}} \int \mathrm{e}^{-\mathrm{i}\xi\eta/\hbar}\psi(\xi)\mathrm{d}\xi, \tag{3.4.38}$$

modified by the factor $\hbar^{-1}$, which maps isometrically $\mathscr{H} = \mathscr{L}^2(\mathbb{R})$ onto $\tilde{\mathscr{H}} = \mathscr{L}^2(\mathbb{R})$. Defining the operators $\widetilde{W}_{x,v}$ in $\tilde{\mathscr{H}}$ by $\widetilde{W}_{xv}\tilde{\psi} = \widetilde{W_{x,v}\psi}$ one calculates using (3.3.22) and (3.4.38)

$$\widetilde{W}_{x,v}\tilde{\psi}(\eta) = \exp\left[-\frac{\mathrm{i}x}{\hbar}\left(\eta - \frac{mv}{2}\right)\right]\tilde{\psi}(\eta - mv). \tag{3.4.39}$$

The family $(x, v) \to \widetilde{W}_{x,v}$ forms a different representation of the CCR in the space $\tilde{\mathscr{H}}$ which is called the *momentum representation*. The formula (3.4.38) describes the transform from the Schrödinger representation to the momentum representation and gives explicitly their unitary equivalence. In momentum representation

$$\tilde{p}\tilde{\psi}(\eta) \equiv \widetilde{p\psi(\eta)} = \eta\tilde{\psi}(\eta), \qquad \tilde{q}\,\tilde{\psi}(\eta) = \hbar\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}\eta}\tilde{\psi}(\eta),$$

so it "diagonalizes" the momentum observable.

3.5. Minimum-uncertainty states. The completeness relation

Pure states S_ψ for which the equality holds in the uncertainty relation (3.3.29) are called *minimum-uncertainty states*. According to (2.6.64) this is the case if and only if for a real c

$$[(Q - \overline{Q}) + \mathrm{i}c(P - \overline{P})]\psi = 0. \tag{3.5.40}$$

For each $c > 0$ this equation has a solution $\psi \in \mathscr{L}^2(\mathbb{R})$ which is unique up to a coefficient. Indeed in the Schrödinger representation (3.5.40) reads

$$\left[(\xi - \overline{Q}) + c\left(\frac{\mathrm{d}}{\mathrm{d}\xi} - \mathrm{i}\overline{P}\right)\right](\xi|\psi) = 0, \tag{3.5.41}$$

whence, using the normalization $\int |(\xi|\psi)|^2\mathrm{d}\xi = 1$,

$$(\xi|\psi) = \frac{k}{\sqrt[4]{\pi c}}\exp\left[\mathrm{i}\overline{P}\xi - \frac{(\xi - \overline{Q})^2}{2c}\right],$$

with $|k| = 1$, $c > 0$. Choosing $k = \exp(-\mathrm{i}\overline{Q}\,\overline{P}/2)$, putting $c = 2\sigma^2$ and denoting the resulting unit vector by $|\overline{P}, \overline{Q}; \sigma^2)$, we have

$$(\xi|\overline{P}, \overline{Q}; \sigma^2) = \frac{1}{\sqrt[4]{2\pi\sigma^2}} \exp\left[\mathrm{i}\overline{P}\left(\xi - \frac{\overline{Q}}{2}\right) - \frac{(\xi - \overline{Q})^2}{2\sigma^2}\right]. \quad (3.5.42)$$

The meaning of $\overline{P}$, $\overline{Q}$ and σ^2 is clear: $\overline{P}$ and $\overline{Q}$ are the mean values of P and Q in the state S_ψ and

$$\sigma^2 = D_S(Q) = [4D_S(P)]^{-1}.$$

The minimum-uncertainty states are sometimes called "wave packets"; under a special reparametrization they will appear also as "coherent states" in Section 3.10. Of special importance is the ground state with $\overline{P} = \overline{Q} = 0$ corresponding to the vector

$$(\xi|0, 0; \sigma^2) = \frac{1}{\sqrt[4]{2\pi\sigma^2}} \exp\left(-\frac{\xi^2}{4\sigma^2}\right). \quad (3.5.43)$$

The vector $|\overline{P}, \overline{Q}; \sigma^2)$ is obtained from it through the action of the displacement operator

$$|\overline{P}, \overline{Q}; \sigma^2) = W_{\overline{Q},\overline{P}/\mu}|0, 0; \sigma^2), \quad (3.5.44)$$

as follows from (3.4.32) and (3.5.42).

We shall show that for any fixed σ^2 the family of vectors $\{|\overline{P}, \overline{Q}; \sigma^2); (\overline{P}, \overline{Q}) \in \mathbb{R}^2\}$ satisfies the *completeness relation*

$$\iint |\overline{P}, \overline{Q}; \sigma^2)(\sigma^2; \overline{Q}, \overline{P}|\frac{\mathrm{d}\overline{P}\,\mathrm{d}\overline{Q}}{2\pi} = I. \quad (3.5.45)$$

In contrast to the formal completeness relations (2.3.33) and (2.4.52) this relation has strict mathematical meaning, the integral being understood in the sense of weak convergence, since $|\overline{P}, \overline{Q}; \sigma^2)$ are usual vectors of the Hilbert space. However these vectors are not orthogonal for different values of $\overline{P}$, $\overline{Q}$; moreover they are linearly dependent. This is expressed by saying that they form an overcomplete system.

The completeness relation (3.5.45) follows from the so called orthogonality relation for an irreducible representation of the CCR. Such relations have a very general nature (*cf.* Section 4.8), but we give an elementary proof of them for the particular case of interest.

Proposition 3.5.1. *Let $(x, v) \to W_{x,v}$ be an irreducible representation of the CCR in $\mathscr{H}$. Then the matrix elements $(\psi|W_{x,v}\varphi)$ are Lebesgue square-integrable functions of (x, v). If $\{e_j\}$ is an orthonormal basis in $\mathscr{H}$ then the functions*

$$\{\sqrt{\mu/2\pi}\,(e_j|W_{x,v}e_k)\}$$

form an orthonormal basis in the space $\mathscr{L}^2(\mathbb{R}^2)$ of complex square integrable functions of (x, v) so that the orthogonality relations hold:

$$\frac{\mu}{2\pi}\iint \overline{(e_j|W_{x,v}e_k)}(e_l|W_{x,v}e_m)\mathrm{d}x\,\mathrm{d}v = \delta_{jl}\delta_{km}. \tag{3.5.46}$$

Proof. By Theorem 3.4.1 we can deal with the Schrödinger representation. By (3.4.32)

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}(\varphi|W_{x,v}\psi) &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\mathrm{i}xy/2}\int \overline{\varphi(\xi)}\mathrm{e}^{\mathrm{i}y\xi}\psi(\xi - x)\mathrm{d}\xi \\ &= \frac{1}{2\pi}\mathrm{e}^{-\mathrm{i}xy/2}\iint \overline{\varphi(\xi)}\tilde{\psi}(\eta)\mathrm{e}^{\mathrm{i}\eta\xi}\mathrm{e}^{-\mathrm{i}(\eta x - y\xi)}\mathrm{d}\xi\,\mathrm{d}\eta,\end{aligned} \tag{3.5.47}$$

for all $\varphi, \psi \in \mathscr{H}$, where $\tilde{\psi}(\eta)$ is now the Fourier transform of $\psi(\xi)$ and $y = \mu v$. Since $\varphi, \tilde{\psi}$ are square-integrable, then $\overline{\varphi(\xi)}\ \tilde{\psi}(\eta)\exp(\mathrm{i}\eta\xi)$ is a square-integrable function of (ξ, η). Therefore the integral in (3.5.47) has the meaning as Fourier transform $\mathscr{F}$ of a square-integrable function and so the function $(\varphi|W_{x,v}\psi)$ is a square-integrable function of (x, v). If $\{e_j\}$ is an orthonormal basis in $\mathscr{L}^2(\mathbb{R})$, then the functions $\overline{e_j(\xi)}\tilde{e}_k(\eta)\exp(\mathrm{i}\eta\xi)$ form an orthonormal basis in $\mathscr{L}^2(\mathbb{R}^2)$. Since the Fourier transform $\mathscr{F}$ is isometric, then the functions $\mathscr{F}[\overline{e_j(\xi)}\tilde{e}_k(\eta)\exp(\mathrm{i}\eta\xi)]$ also form an orthonormal basis in $\mathscr{L}^2(\mathbb{R}^2)$. Therefore the functions

$$(2\pi)^{-1/2}(e_j|W_{x,y/\mu}e_k) = \exp(\mathrm{i}xy/2)\cdot\mathscr{F}[\overline{e_j(\xi)}\tilde{e}_k(\eta)\exp(\mathrm{i}\eta\xi)]$$

form an orthonormal basis in the space $\mathscr{L}^2(\mathbb{R}^2)$ of the variables $x,\ y = \mu v$, whence (3.5.46) follows. □

From (3.5.46) we get

$$\frac{\mu}{2\pi}\iint \overline{(\varphi_1|W_{x,v}\psi_1)}(\varphi_2|W_{x,v}\psi_2)\mathrm{d}x\,\mathrm{d}v = \overline{(\varphi_1|\varphi_2)}(\psi_1|\psi_2) \tag{3.5.48}$$

for arbitrary $\varphi_j,\ \psi_j \in \mathscr{H}\ ;\ j = 1, 2$. Putting $\psi_1 = \psi_2 = \psi$ with $(\psi|\psi) = 1$, we obtain

$$\frac{\mu}{2\pi}\iint (\varphi_2|W_{x,v}\psi)(\psi|W_{x,v}\varphi_1)\mathrm{d}x\,\mathrm{d}v = (\varphi_2|\varphi_2).$$

Since φ_1, φ_2 are arbitrary, this means that

$$\frac{\mu}{2\pi}\iint W_{x,v}|\psi)(\psi|W_{x,v}^{*}\mathrm{d}x\,\mathrm{d}v = I, \tag{3.5.49}$$

where the integral converges weakly. Thus for any unit vector ψ the family $\{W_{x,v}|\psi);\ (x, v) \in \mathbb{R}^2\}$ satisfies the completeness relation (3.5.49). Putting $|\psi) = |0, 0; \sigma^2)$ we get (3.5.45).

3.6. Joint measurements of coordinate and velocity

As follows from the uncertainty relation (3.3.29) the observables of coordinate Q and velocity $\mu^{-1}P$ are incompatible. This means that there is no measurement $M(\mathrm{d}x\,\mathrm{d}v)$ such that the measurements $E(\mathrm{d}x)$ and $F(\mathrm{d}v)$ described by the spectral measures of Q and $\mu^{-1}P$ are marginal with respect to $M(\mathrm{d}x\,\mathrm{d}v)$, *i.e.*,

$$E(\mathrm{d}x) = \int M(\mathrm{d}x, \mathrm{d}v), \qquad F(\mathrm{d}v) = \int M(\mathrm{d}x\,\mathrm{d}v).$$

However if one concludes that quantum theory unconditionally forbids joint measurements of coordinate and velocity, one meets serious difficulties. Experimentally velocity is often measured through the proportional quantity, the momentum P. A physicist considering classical mechanics as the limit of quantum theory as $\hbar \to 0$, encounters a troubling discontinuity at $\hbar = 0$; for all $\hbar \neq 0$, however small, there is no joint measurement for q, p, and for $\hbar = 0$ they are trivially jointly measurable. More sharply, if one properly considers classical mechanics only as approximation of the more basic quantum theory, one has to conclude the impossibility of joint measurements of coordinate and momentum for macroscopic objects as well. Indeed, consider a simplified "macroscopic object" consisting of an arbitrary large number N of identical quantum particles, the canonical observables q_j, p_j of which satisfy the Heisenberg CCR for N degrees of freedom

$$[q_j, p_k] = \mathrm{i}\hbar\delta_{jk}, \qquad [q_j, q_k] = [p_j, p_k] = 0. \tag{3.6.50}$$

Then the "macroscopic observable" – coordinate of the center of mass $q = N^{-1}\sum q_j$ and the total momentum $p = \sum p_j$ – satisfy the same Heisenberg CCR (3.3.31) as the "microscopic observables" q_j, p_j. Since $\hbar = 0$ (though it is extremely small if expressed in the classical units, $\hbar \approx 10^{-27}$ g cm^2 s^{-1}) one has to acknowledge the impossibility of joint measurements of the "macroscopic observables" q and p.

This is in apparent contradiction with the experimental evidence of classical mechanics. Moreover, there are experiments with quantum particles giving the data which can be interpreted as the joint measurements of coordinate and velocity. For example, from the trace of a charged particle in a bubble chamber it is possible to estimate both coordinate and momentum of the particle. In fact, even if only the momentum is measured, there is information about the localization of the particle: the experimenter knows at least that the particle is within the measuring apparatus. Adjoining this information to the results of the measurement one can interpret them as "joint measurement" of coordinate and momentum. Anyhow, joint measurements of coordinate and velocity actually do exist, so the problem is to find for them a proper quantum theoretical description.

This problem can be given a natural solution in the framework of the concept of measurement developed in Chapters 1-2. According to it, a joint measurement of a pair of parameters x and v must be described by a resolution of identity $M(\mathrm{d}x\,\mathrm{d}v)$ in $\mathscr{H}$, with the joint probability distribution of the results given by $\mu_S(\mathrm{d}x\,\mathrm{d}v) = \mathrm{Tr}\, SM(\mathrm{d}x\,\mathrm{d}v)$. To single out the resolutions of identity which actually correspond to joint measurements of the coordinate and the velocity, we shall make use of the covariance argument similar to one used in Section 3.3 for the explanation of the kinematical meaning of the observables Q and P.

Assume that a state S is prepared by an apparatus to which a frame of reference is related. If the apparatus is shifted to the distance x and moves with the velocity v relative to its basic position, then the prepared state will be

$$S_{x,v} = W_{x,v} S V_{x,v}^*.$$

The parameters x and v are thus the physical quantities which contain an information about the state of the microobject insofar as it is prepared by the apparatus the position of which is characterized by x and v. Let $M(\mathrm{d}x\,\mathrm{d}v)$ be a resolution of identity describing a joint measurement of x and v; it is then natural to require that the probability distribution of the measurement with respect to the "shifted" state $S_{x,v}$ be just the initial probability distribution shifted by the same vector (x, v), *i.e.*,

$$\mu_{S_{x,v}}(B) = \mu_S(B_{-x,-v}); \qquad B \in \mathscr{A}(\mathbb{R}^2),$$

where $B_{-x,-v} = \{(\xi - x, \eta - v) : (\xi, \eta) \in B\}$ (Figure 3.3).

This equality should hold for all S, whence

$$W_{x,v}^* M(B) W_{x,v} = M(B_{-x,-v}); \qquad B \in \mathscr{A}(\mathbb{R}^2). \tag{3.6.51}$$

The measurements $M(dx\,\mathrm{d}v)$ satisfying this condition will be called *covariant* with respect to the representation $(x, v) \to W_{x,v}$ of the group of

kinematical transformations. The condition (3.6.51) is analogous to the condition (3.2.19) for an one-parameter shift group.

In Chapter 4 we shall describe all resolutions of identity satisfying the condition (3.6.51); among them there are no orthogonal ones. This is another expression of the fact that there are no joint measurements of coordinate and velocity in the conventional sense. The generating example of the covariant measurement is given by

$$M(\mathrm{d}x\,\mathrm{d}v) = W_{x,v}|\psi)(\psi|W_{x,v}^*\frac{\mu\,\mathrm{d}x\,\mathrm{d}v}{2\pi}, \tag{3.6.52}$$

where ψ is a unit vector of $\mathscr{H}$. This means that for a Borel B we define

$$(\varphi|M(B)\varphi) = \iint_B (\varphi|W_{x,v}\psi)(\psi|W_{x,v}\varphi)\frac{\mu\,\mathrm{d}x\,\mathrm{d}v}{2\pi}, \qquad \varphi \in \mathscr{H},$$

the integral converging by Proposition 3.5.1. All properties of the resolution of identity are satisfied: that $M(B) \geq 0$ is obvious; the weak σ-additivity follows from the property of the integral, and the normalization $M(\mathbb{R}^2) = I$ is equivalent to the completeness relation (3.5.49). That (3.6.52) satisfies (3.6.51) follows from the CCR (3.3.22). The probability distribution of the measurement with respect to the state S is

$$\mu_S(\mathrm{d}x\,\mathrm{d}v) = (\psi|W_{x,v}^*SV_{x,v}\psi)\frac{\mu\,\mathrm{d}x\,\mathrm{d}v}{2\pi}. \tag{3.6.53}$$

To visualize this construction we give an idealized description of a procedure which can be considered as realization of the measurement (3.6.52) in the sense of Section 2.5. In addition to $\mathscr{H}$ with the canonical pair Q, P in $\mathscr{H}$ consider the identical space $\mathscr{H}_0$ and a canonical pair Q_0, P_0 in $\mathscr{H}_0$. In the tensor product $\mathscr{H}\otimes\mathscr{H}_0$ consider the operators

$$\widetilde{P} = P\otimes I_0 + I\otimes P_0, \qquad \widetilde{Q} = Q\otimes I_0 - I\otimes Q_0, \tag{3.6.54}$$

where I_0 is the unit operator in $\mathscr{H}_0$. These operators are infinitesimal generators of the unitary groups

$$\mathrm{e}^{\mathrm{i}\xi\widetilde{Q}} = \mathrm{e}^{\mathrm{i}\xi Q}\otimes\mathrm{e}^{-\mathrm{i}\xi Q_0}, \qquad \mathrm{e}^{\mathrm{i}\eta P}\otimes\mathrm{e}^{\mathrm{i}\eta P_0}, \tag{3.6.55}$$

which commute by the Weyl CCR (3.3.24), and therefore $\widetilde{P}$, $\widetilde{Q}$ are commuting self-adjoint operators in the sense of Section 2.6. It follows that $\widetilde{Q}$ and $\mu^{-1}\widetilde{P}$ are compatible and admit the joint measurement $E(\mathrm{d}x\,\mathrm{d}v)$.

For definiteness we may take the Schrödinger representation $Q = \xi_1$, $P = \mathrm{i}^{-1}\mathrm{d}/\mathrm{d}\xi_1$ in $\mathscr{H} = \mathscr{L}^2(\mathbb{R}_1)$, and $Q_0 = \xi_2$, $P_0 = \mathrm{i}^{-1}\mathrm{d}/\mathrm{d}\xi_2$ in $\mathscr{H}_0 =$

$\mathscr{L}^2(\mathbb{R}_2)$. Consider a "particle" with two degrees of freedom moving in the plane ξ_1, ξ_2. Then the Hilbert space of the particle will be $\mathscr{H} \otimes \mathscr{H}_0 = \mathscr{L}^2(\mathbb{R}^2)$. Introducing the new coordinate system ξ_1', ξ_2' rotated through the angle $-\pi/4$ relative to the initial one (Figure 3.4), we find the canonical observables corresponding to the new coordinate axes

$$P'_{\xi_2'} = \mathrm{i}^{-1}\frac{\partial}{\partial \xi_2'} = (\mathrm{i}\sqrt{2})^{-1}\left(\frac{\partial}{\partial \xi_1} + \frac{\partial}{\partial \xi_2}\right) = \sqrt{2}^{-1}(P_{\xi_1} + P_{\xi_2}),$$
$$Q_{\xi_1'} = \xi_1' = \sqrt{2}^{-1}(\xi_1 - \xi_2) = \sqrt{2}^{-1}(Q_{\xi_1} - Q_{\xi_2}),$$

so that $\widetilde{P} = \sqrt{2}P_{\xi_2'}$, $\widetilde{Q} = \sqrt{2}Q_{\xi_1'}$. The observables $P_{\xi_2'}$ and $Q_{\xi_1'}$ correspond to the mutually orthogonal axes and apparently are jointly measurable.

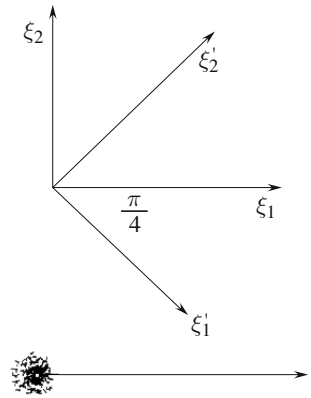

Figure 3.4.

Let the auxiliary degree of freedom ξ_2 be described by the state $S_0 = |\bar{\psi})(\bar{\psi}|$, where the vector $\bar{\psi}$ is given in the Schrödinger representation by the function $\overline{\psi(\xi_2)} = (\xi_2|\bar{\psi})$, which is complex conjugate to $\psi(\xi) = (\xi|\psi)$. Here ψ is the vector defining the covariant measurement by (3.6.52).

Proposition 3.6.1. *The triple* $(\mathscr{H}_0, S_0, \boldsymbol{E})$, *where* $\boldsymbol{E}$ *is the joint measurement of* $\widetilde{Q}$ *and* $\mu^{-1}\widetilde{P}$, *defined by* (3.6.54) *is a realization of the measurement* $\boldsymbol{M}$ *given by* (3.6.52) *in the sense that*

$$\mu^{\boldsymbol{E}}_{S\otimes S_0}(\mathrm{d}x\,\mathrm{d}v) = \mu^{\boldsymbol{M}}_S(\mathrm{d}x\,\mathrm{d}v)$$

for all states S *in* $\mathscr{H}$.

We postpone the proof of this statement until Section 5.3 where the necessary mathematical tools are developed.

Now the quantitative answer to the question – what is the best accuracy for a joint measurement of the coordinate x and the velocity v admitted by quantum theory – can be given. We shall restrict to covariant measurements (3.6.52) having finite second moments and zero mean values with respect to the state S. Then as at the end of Section 3.2 we can show that the measurements are unbiased

$$E_x\{\boldsymbol{M}\} \equiv \iint \hat{x}\mu^{\boldsymbol{M}}_{S_{x,v}}(\mathrm{d}\hat{x}\,\mathrm{d}\hat{v}) = x_0 + x,$$
$$E_v\{\boldsymbol{M}\} \equiv \iint \hat{v}\mu^{\boldsymbol{M}}_{S_{x,v}}(\mathrm{d}\hat{x}\,\mathrm{d}\hat{v}) = v_0 + v.$$

Moreover, the marginal variances

$$D_x\{\boldsymbol{M}\} \equiv \iint (\hat{x} - E_x\{\boldsymbol{M}\})^2\mu_{S_{x,v}}(\mathrm{d}\hat{x}\,\mathrm{d}\hat{v}),$$
$$D_v\{\boldsymbol{M}\} \equiv \iint (\hat{v} - E_v\{\boldsymbol{M}\})^2\mu_{S_{x,v}}(\mathrm{d}\hat{x}\,\mathrm{d}\hat{v})$$

are all the same for all values of x, v and so they are equal to the marginal variances with respect to the basic state S.

As the measure of accuracy of joint measurement we take

$$\mathscr{R}\{\boldsymbol{M}\} = g_x D_x\{\boldsymbol{M}\} + g_v D_v\{\boldsymbol{M}\}, \tag{3.6.56}$$

where g_x, g_v are positive constants, defining the relative scaling of the variances. By Proposition 3.6.1

$$D_x\{\boldsymbol{M}\} = D_S(Q) + D_{S_0}(Q_0), \qquad D_v\{M\} = \mu^{-2}[D_S(P) + D_{S_0}(P_0)].$$

By the inequality $a + b \geq 2\sqrt{ab}$ and the uncertainty relation (3.3.29)

$$g_x D_{S_0}(Q_0) + g_v\mu^{-2}D_{S_0} \geq 2\sqrt{g_x g_v\mu^{-2}D_{S_0}(Q_0)D_{S_0}(P_0)}$$
$$\geq \mu^{-1}\sqrt{g_x g_v}$$

with the equality achieved if and only if

$$D_{S_0}(Q_0) = \frac{1}{2\mu}\sqrt{\frac{g_v}{g_x}}, \qquad D_{S_0}(P_0) = \frac{\mu}{2}\sqrt{\frac{gx}{gv}}, \tag{3.6.57}$$

i.e., if and only if S_0 is the minimum-uncertainty state $|0, 0; \sigma^2)(\sigma^2; 0, 0|$ with σ^2 equal to $D_{S_0}(Q_0)$ from (3.6.57).

It follows that

$$\min_{\boldsymbol{M}} \mathscr{R}\{\boldsymbol{M}\} = g_x D_S(Q) + g_v D_S(\mu^{-1}P) + \mu^{-1}\sqrt{g_x g_v}, \tag{3.6.58}$$

with the minimum achieved for the unique optimal covariant measurement

$$M_*(\mathrm{d}x\,\mathrm{d}v) = |\mu v, x; \sigma^2)(\sigma^2; x, \mu v|\frac{\mu\,\mathrm{d}x\,\mathrm{d}v}{2\pi}, \tag{3.6.59}$$

with $\sigma^2 = (2\mu)^{-1}\sqrt{g_x/g_v}$. We call it the *canonical measurement*. Here we have used (3.6.52) and (3.5.44). In Section 4.8 we shall extend (3.6.58) to all covariant measurements of (x, v) with finite second moments. This result shows clearly the place of the minimum-uncertainty states in the problem of joint measurements of coordinate and velocity.

Putting $\mu = m/\hbar$, $p = \hbar P$, consider the classical limit with $\hbar \to 0$, $m = \text{const.}$, $D_S(p) = \text{const.}$. The optimal measurement is furnished by measuring the observables $q \otimes I_0 - I \otimes q_0$, $m^{-1}(p \otimes I_0 + I \otimes p_0)$, differing from $q, m^{-1}p$ by the terms $-q_0, m^{-1}p_0$, the variances of which by (3.6.57) are proportional to $\hbar$ and thus tend to zero. Moreover, by (3.6.58) the best quantum accuracy

$$\min_{\boldsymbol{M}} \mathscr{R}\{\boldsymbol{M}\} = g_x D_S(q) + g_v D_s(m^{-1}p) + \hbar m^{-1}\sqrt{g_x g_v}$$

tends to the classical expression $g_x D_S(q) + g_v D_S(m^{-1}p)$ for the joint measurement of coordinate and velocity. Thus in the classical limit the optimal quantum measurement passes smoothly into the classical measurement of the observables q and $m^{-1}p$ and the unphysical discontinuity is eliminated.

3.7. Dynamics of a quantum particle in one dimension

The purpose of this section is to show that the Galilean relativity determines not only quantum kinematics but also all possible dynamics, *i.e.*, temporal evolutions of quantum objects. In this approach the "correspondence principle", establishing a connection between some classical and quantum quantities, which was first introduced as an empirical rule, turns out to be a logical consequence of the Galilean covariance of the theories.

To include temporal evolution one needs to consider the full Galilean group of the transformations

$$\xi' = \xi + x + v\tau, \qquad \tau' = \tau + t.$$

Any such transformation is characterized by the three parameters (x, v, t), the law of composition of transformations being given by

$$(x_1, v_1, t_1)(x_2, v_2, t_2) = (x_1 + x_2 + v_1 t_2, v_1 + v_2, t_1 + t_2).$$

According to the general scheme of Section 3.1 one looks for irreducible projective unitary representations $(x, v, t) \to W_{x,v,t}$ of the Galilean group. As Bargmann and Wigner showed, the relation (3.1.6) for the Galilean group can be always reduced to the form

$$W_{x_1,v_1,t_1} W_{x_2,v_2,t_2} = \exp\left[-\frac{\mathrm{i}\mu}{2}(x_1 v_2 - x_2 v_1 + t_2 v_1 v_2)\right] \times W_{x_1+x_2+v_1 t_2, v_1+v_2, t_1+t_2}. \tag{3.7.60}$$

The restriction of this relation to the subgroup of the kinematical transformations $(x, v, 0)$ gives the CCR (3.3.22) for the operators $W_{x,v} \equiv W_{-x,-v,0}$. Since we already know the description of the representations of a kinematical group, we can use (3.7.60) to study the relation between kinematics and dynamics, *i.e.*, between $\{W_{x,v}\}$ and the one-parameter unitary group of time evolution $\{V_t\} \equiv \{W_{0,0,t}\}$. From (3.7.60) it follows that

$$V_t^* W_{x,v} V_t = W_{x-vt,v}.$$

Putting here $x = 0$ and $v = 0$ we get the two basic relations

$$V_t^* U_v V_t = W_{-vt,v}, \tag{3.7.61}$$

$$V_t^* V_x V_t = V_x. \tag{3.7.62}$$

According to Stone's theorem, $V_t = \exp(-\mathrm{i}tH)$, where H is a self-adjoint operator called the *Hamiltonian*. We shall give arguments showing that (3.7.61) and (3.7.62) determine the form of H, namely, (3.7.61) implies

$$H = \frac{P^2}{2\mu} + v(Q), \tag{3.7.63}$$

where $v(\cdot)$ is a real-valued function, and the additional restriction (3.7.62) leads to the unique, up to an additive constant, form of the Hamiltonian

$$H = \frac{P^2}{2\mu}. \tag{3.7.64}$$

Introduce the time-varying observables

$$Q(t) = V_t^* Q V_t, \qquad P(t) = V_t^* P V_t.$$

Then differentiating (3.7.61) with respect to v and using (3.4.34), (3.4.37) and (2.4.49) we get

$$Q(t) = Q + t\mu^{-1} P. \tag{3.7.65}$$

Similarly differentiating (3.7.62) with respect to x and using (3.4.33) we get

$$P(t) = P. \tag{3.7.66}$$

In what follows we shall proceed partly heuristically since the complete proofs would take too much place here. Differentiating (3.7.65) with respect to t we get

$$[H, Q] = \mu^{-1} P.$$

This is an inhomogeneous linear equation; its solution H is the sum of a particular solution H_0 and the general solution v of the corresponding homogeneous linear equation $[v, Q] = 0$. Since the representation is irreducible, then the general solution of this equation is $v = v(Q)$, as follows from consideration in the Schrödinger representation. To show that $H_0 = P^2/2\mu$ is a particular solution of the inhomogeneous equation we need the identities

$$\mathrm{i}[f(Q), P] = -f'(Q), \qquad \mathrm{i}[Q, f(P)] = -f'(P). \tag{3.7.67}$$

The first identity follows from the fact that in the Schrödinger representation $[f(x)\mathrm{d}/\mathrm{d}x - \mathrm{d}/\mathrm{d}xf(x)]\psi(x) = -f'(x)\psi(x)$, the second one is obtained similarly in the momentum representation. Putting in it $f(P) = P^2/2\mu$ we get $\mathrm{i}[Q, P^2/2\mu] = -P/\mu$ as required. Thus we have formally deduced (3.7.63) from (3.7.61). Differentiating (3.7.66) with respect to t we get $[H, P] = 0$, whence by (3.7.67) $v'(Q) = 0$ and (3.7.64) follows up to a constant.

Thus the full Galilean relativity determines essentially uniquely the Hamiltonian of a free quantum particle; if an external field is involved, then the spatial homogeneity requirement (3.7.62) should be omitted and the restricted Galilean relativity (3.7.61) gives the general form (3.7.63) of the Hamiltonian for a particle in the external field.

To explain the nature of the constant μ and other terms constituting the Hamiltonian consider a quantum state S_ψ for which the coordinate probability distribution $|\psi(\xi)|^2\mathrm{d}\xi$ is sharply peaked near the mean value $E(Q)$. Therefore the quantum object in the state S_ψ behaves as a "particle" localized near the point $E(Q)$. Supplying the expectations which refer to the time t with the corresponding index we obtain from (3.7.65) by averaging and differentiating

$$\frac{\mathrm{d}}{\mathrm{d}t} E_t(Q) = E_t(\mu^{-1} P). \tag{3.7.68}$$

This means that the classical velocity of the particle is $E_t(\mu^{-1}P)$ as is to be expected from the kinematical meaning of the observable $\mu^{-1}P$ (see

Section 3.3). Since the object displays itself as a classical particle, its "classical mass" m can be measured; the classical momentum which is defined as the product of the mass by the velocity is $m(\mathrm{d}/\mathrm{d}t)E_t(Q) = E_t(\hbar P)$, where $\hbar = m/\mu$. Therefore $p = \hbar P$ in the classical limit corresponds to the momentum observable. Differentiating formally the last equation we get

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}E_t(Q) = \hbar E_t(\mathrm{i}[H, P]).$$

Taking into account (3.7.67) and (3.7.63) we obtain $\mathrm{i}[H,P]=\mathrm{i}[v(Q),P]=-v'(Q)$, whence redenoting $Q = q$

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}E_t(q) = -E_t(\hbar v'(q)).$$

This is the Newton equation for the classical object with mass m and the potential energy $V(q) = \hbar v(q)$. Putting $E = \hbar H$ we can rewrite (3.7.63) in the form

$$E = \frac{p^2}{2m} + V(q),$$

completely corresponding to the classical expression representing the total energy as the sum of the kinetic energy $p^2/2m$ and the potential energy $V(q)$. Therefore $E = \hbar H$ corresponds to the *energy observable*.

The main concern of quantum mechanics is the temporal evolution of observed quantities such as probabilities or expectation values. An expectation value $E_t(X)$ referring to the time t can be written in the two equivalent forms $E_S(X(t)) = E_{S_t}(X)$, where $X(t) = V_t^* X V_t$ and $S_t = V_t S V_t^*$. The description in which states are kept fixed and observables are time-varying as in the conventional form of classical mechanics, is called the *Heisenberg picture*. The dual description corresponding to the second form of $E_t(X)$ is called the *Schrödinger picture* and is widely used in quantum mechanics.

If the initial state is pure, $S = |\psi)(\psi|$, then $S_t = |\psi_t)(\psi_t|$ is also pure with $\psi_t = V_t\psi$. Assuming $\psi \in \mathscr{D}(H)$ we have by (2.4.49)

$$\hbar\mathrm{i}\frac{\mathrm{d}\psi_t}{\mathrm{d}t} = E\psi_t; \qquad \psi_0 = \psi.$$

This is the general dynamical equation determining the temporal evolution of state vectors. Taking into account (3.7.63) we get in the Schrödinger representation

$$\mathrm{i}\frac{\partial\psi_t(x)}{\partial t} = -\frac{\hbar}{2m}\frac{\partial^2\psi_t(x)}{\partial x^2} + \frac{V(x)}{\hbar}\psi_t(x). \tag{3.7.69}$$

The ratio $\hbar = m/\mu$ essentially involved in this equation can be measured in experiments in which the object displays non-classical (wave) properties. The constant $\hbar$ turns out to be $\hbar = h/2\pi$ where h is the Planck's constant. Existence of the universal constant $\hbar$ means that there is a natural unit of mass; the constant $\hbar$ is just the coefficient relating it with the classical unit of mass. If the mass is measured in natural units, then $\hbar = 1$; we shall often use this convention.

3.8. Time observable. The "time-energy" uncertainty relation

The Stone-von Neumann uniqueness theorem implies that any pair of canonically conjugate observables is essentially the canonical Schrödinger pair (3.4.35), (3.4.36). In particular both operators are necessarily unbounded from above and below. From (3.7.63) it follows that if as usual the potential $v(\cdot)$ is bounded from below, the energy observable is also bounded from below. Therefore the energy observable cannot have a canonically conjugate observable represented by a self-adjoint operator. For the reason to be explained below this means the nonexistence of a self-adjoint operator representing time observable. However time measurements are quite common experimentally, and a theoretical representation for them in quantum mechanics should exist. We shall show a way to resolve the difficulty using the broader concept of measurement.

Consider the family of states

$$S_t = \mathrm{e}^{-\mathrm{i}tH} S \mathrm{e}^{\mathrm{i}tH}, \qquad t \in \mathbb{R}, \tag{3.8.70}$$

corresponding to different values of the time evolution parameter t. The parameter t describes the backward time shift in the state preparation. Motivated by the consideration of coordinate measurements in Section3.3 consider the covariance condition

$$V_t^* M(B) V_t = M(B_{-t}); \qquad t \in \mathbb{R}, \quad B \in \mathscr{A}(\mathbb{R}), \tag{3.8.71}$$

for a measurement $\boldsymbol{M} = \{M(\mathrm{d}t)\}$ with $V_t = \exp(-\mathrm{i}tH)$. Denoting by $\mu_t^{\boldsymbol{M}}$ the measurement probability distribution with respect to the state S_t we get

$$\mu_t^{\boldsymbol{M}}(B) = \mu_0^{\boldsymbol{M}}(B_{-t}); \qquad t \in \mathbb{R}, \quad B \in \mathscr{A}(\mathbb{R}).$$

This means that a time shift in the preparation procedure causes the same shift in the resulting probability distribution. For this reason any $\boldsymbol{M}$ satisfying (3.8.71) if it at all exists can be associated with a measurement of the time parameter t.

If the Hamiltonian H is bounded from below, then there are no simple measurements $\boldsymbol{M} = \{M(\mathrm{d}t)\}$ satisfying (3.8.71), for otherwise (3.8.71) would be just (3.2.19) for $A = -H$, $\boldsymbol{G} = \boldsymbol{M}$ and $-H$ would have the canonically conjugate observable $B = \int t M(\mathrm{d}t)$. This expresses appropriately the non-existence of a self-adjoint time observable. However there are measurements represented by non-orthogonal resolutions of identity which satisfy the covariance condition (3.8.71).

It is convenient to consider (3.8.71) in the "energy representation" in which the energy observable E is "diagonal". For definiteness we shall first consider the case of a free particle in one dimension. Since for a free particle $E = p^2/2m$, then in the momentum representation (3.4.38) the energy operator is multiplication by $\epsilon = \eta^2/2m$. We have

$$\begin{aligned}(\tilde{\psi}|\tilde{\psi}) &\equiv \int_{-\infty}^{\infty} |\tilde{\psi}(\eta)|^2 \mathrm{d}\eta \\ &= \sqrt{\frac{m}{2}}\left[\int_0^{\infty} |\tilde{\boldsymbol{\psi}}(\sqrt{2m\epsilon})|^2 \frac{\mathrm{d}\epsilon}{\sqrt{\epsilon}} + \int_0^{\infty} |\tilde{\boldsymbol{\psi}}(-\sqrt{2m\epsilon})|^2 \frac{\mathrm{d}\epsilon}{\sqrt{\epsilon}}\right] \\ &= \int_0^{\infty} |\boldsymbol{\psi}_{\boldsymbol{\epsilon}}|^2 \mathrm{d}\epsilon,\end{aligned} \tag{3.8.72}$$

where

$$\boldsymbol{\psi}_{\boldsymbol{\epsilon}} = \sqrt[4]{\frac{m}{2\epsilon}}\begin{bmatrix} \tilde{\psi}(\sqrt{2m\epsilon}) \\ \tilde{\psi}(-\sqrt{2m\epsilon}) \end{bmatrix}, \tag{3.8.73}$$

and $|\boldsymbol{\psi}_{\boldsymbol{\epsilon}}|^2 = \boldsymbol{\psi}_{\boldsymbol{\epsilon}}^* \boldsymbol{\psi}_{\boldsymbol{\epsilon}}$ is the squared norm of the two-dimensional vector $\boldsymbol{\psi}_{\boldsymbol{\epsilon}} \in \mathbb{C}^2$. The relation (3.8.73) describes the transformation from the momentum representation to the *energy representation* in which the energy operator is multiplication by ϵ and the time evolution group $\{V_t\}$ is just multiplication by $\{\exp(-\mathrm{i}t\boldsymbol{\epsilon}/\hbar)\}$. The space of energy representation is thus the space $\mathscr{L}^2_{\mathbb{K}}(0, \infty)$ of the functions with values in $\mathbb{K} = \mathbb{C}^2$ with the norm given by (3.8.72). One could write down the representation of the CCR in $\mathscr{L}^2_{\mathbb{K}}(0, \infty)$ but we shall not need it.

In $\mathscr{L}^2_{\mathbb{K}}(0, \infty)$ consider the operator

$$T = \mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}\epsilon}$$

with the domain

$$\mathscr{D}(T) = \left\{ \boldsymbol{\psi}_{\boldsymbol{\epsilon}} : \boldsymbol{\psi}_0 = 0, \int_0^{\infty} \left| \frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{\psi}_{\boldsymbol{\epsilon}} \right|^2 \mathrm{d}\epsilon < \infty \right\}.$$

Similar to the operator P_+ in $\mathscr{L}^2(0,\infty)$ (see Section 2.4) this is the maximal symmetric operator. Analogously to (2.4.54) the spectral measure of T is the nonorthogonal resolution of identity $\boldsymbol{M}(\mathrm{d}\tau)$ defined by

$$(\psi|M(\mathrm{d}\tau)\psi) = \left[\int_0^\infty \int_0^\infty \boldsymbol{\psi}^*_{\boldsymbol{\epsilon}}\boldsymbol{\psi}_{\boldsymbol{\epsilon}'}\mathrm{e}^{\mathrm{i}(\boldsymbol{\epsilon}'-\boldsymbol{\epsilon})\tau/\hbar}\mathrm{d}\epsilon\,\mathrm{d}\epsilon'\right]\frac{\mathrm{d}\tau}{2\pi\hbar} \qquad (3.8.74)$$

or formally in Dirac's notations

$$(\boldsymbol{\epsilon}|M(\mathrm{d}\tau)|\epsilon') = \mathrm{e}^{\mathrm{i}(\boldsymbol{\epsilon}'-\boldsymbol{\epsilon})\tau/\hbar}\frac{\mathrm{d}\tau}{2\pi\hbar}.$$

The covariance property (3.8.71) of $M(\mathrm{d}\tau)$ follows directly from this definition.

Let $\psi \in \mathscr{D}(T)$, then by (2.6.61) the mean and the variance of T in the state S_ψ are, correspondingly

$$E_{S_\psi}(T) = \hbar\mathrm{i}\int_0^\infty \boldsymbol{\psi}^*_{\boldsymbol{\epsilon}}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\boldsymbol{\psi}_{\boldsymbol{\epsilon}}\mathrm{d}\epsilon,$$
$$D_{S_\psi}(T) = \hbar^2\int_0^\infty \left|\frac{\mathrm{d}}{\mathrm{d}\epsilon}\boldsymbol{\psi}_{\boldsymbol{\epsilon}}\right|^2\mathrm{d}\epsilon - E_{S_\psi}(T)^2,$$

and $\mathscr{D}(T) = \{\psi : D_{S_\psi}(T) < \infty\}$. Going back to momentum representation by (3.8.72) we get

$$E_{S_\psi}(T) = m\hbar\mathrm{i}\int_{-\infty}^\infty \operatorname{sgn}\eta\frac{\tilde\psi(\eta)}{\sqrt{|\eta|}}\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\tilde\psi(\eta)}{\sqrt{|\eta|}}\mathrm{d}\eta,$$
$$D_{S_\psi}(T) = (m\hbar)^2\int_{-\infty}^\infty \left|\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\tilde\psi(\eta)}{\sqrt{|\eta|}}\right|^2\frac{\mathrm{d}\eta}{|\eta|} - E_{S_\psi}(T)^2.$$

It follows that

$$T = m\hbar\mathrm{i}\operatorname{sgn}\eta\frac{1}{\sqrt{|\eta|}}\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{1}{\sqrt{|\eta|}} = m\operatorname{sgn}p|p|^{-1/2}q|p|^{-1/2} = mp^{-1}\circ q,$$

$$\mathscr{D}(T) = \left\{\tilde\psi(\eta) : \int_{-\infty}^\infty \left|\frac{\mathrm{d}}{\mathrm{d}\eta}\frac{\tilde\psi(\eta)}{\sqrt{|\eta|}}\right|^2\frac{\mathrm{d}\eta}{|\eta|} < \infty\right\}.$$

In general, assume that the energy operator is the multiplication by the independent variable ϵ in the space $\mathscr{L}^2_{\mathbb{K}}(0,\infty)$ where $\mathbb{K}$ is a Hilbert space; this means that the representation space consists of $\mathbb{K}$-valued func-

tions $\boldsymbol{\psi}(\cdot)$ on $(0,\infty)$ satisfying $\int_0^\infty |\boldsymbol{\psi}_{\boldsymbol{\epsilon}}|^2 \mathrm{d}\epsilon < \infty$ where $|\cdot|$ is the norm in $\mathbb{K}$. Then the operator

$$\begin{aligned} T &= \hbar \mathrm{i} \frac{\mathrm{d}}{\mathrm{d}\epsilon}; \\ \mathscr{D}(T) &= \left\{ \boldsymbol{\psi}_{\boldsymbol{\epsilon}} : \boldsymbol{\psi}_0 = 0, \int_0^\infty |\boldsymbol{\psi}_{\boldsymbol{\epsilon}}|^2 \mathrm{d}\epsilon < \infty \right\} \end{aligned} \tag{3.8.75}$$

is maximal symmetric in $\mathscr{L}^2_{\mathbb{K}}(0,\infty)$. Its spectral measure satisfying the covariance condition (3.8.71) can be constructed analogously to (3.8.74). We call (3.8.75) the *canonical time observable* in the energy representation. Whether it is possible to give an expression for T in momentum or Schrödinger representation depends on the possibility of an explicit diagonalization procedure for the energy operator. For example, for the three-dimensional free particle one obtains $T = m \sum_{j=1}^3 \frac{p_j}{|p|^2} \circ q_j$.

Let us return to the problem of estimation of the time evolution parameter t in the family (3.8.70). Assume that we are interested in measuring a distinguished moment in the history of the quantum object, this moment having the property that shifting in time the preparation of the basic state leads to the same shift of the distinguished moment. Typical quantities of such kind are "arrival times" or "passage times" $\tilde{T}$ related to our time observable T as $\tilde{T} = \text{const.} - T$. Then the inequality (3.2.12) implies that the variance of any unbiased estimate X of such a time quantity is bounded from below by the quantity inversely proportional to the energy uncertainty $D_t(E) \equiv D_S(E)$

$$D_t(X) \geq [4D_t(H)]^{-1} = \hbar^2 [4D_t(E)]^{-1}. \tag{3.8.76}$$

As we have seen the set of unbiased estimates of t is not empty, since it contains the observable $T - E_S(T)$. Indeed, by Section 3.2 an estimate corresponding to a covariant measurement is unbiased up to a constant. Thus (3.8.76) implies the uncertainty relation of the type (3.3.30)

$$D_S(T) \cdot D_S(E) \geq \hbar^2/4.$$

Another interesting example of an unbiased estimate of the time evolution parameter of a free quantum particle of mass m is given by $\widehat{T} = (m/\bar{p}) \cdot q$, where the mean momentum $\bar{p} = E_S(p)$ is assumed to be non-zero. By (3.7.68)

$$\frac{\mathrm{d}}{\mathrm{d}t} E_t(\widehat{T}) = \frac{m}{\bar{p}} \frac{\mathrm{d}}{\mathrm{d}t} E_t(q) = \frac{1}{\bar{p}} E_t(p) = 1,$$

since the mean momentum of a free particle is conserved. To explain the meaning of this estimate let the basic state S be the minimum-uncertainty

state described in the momentum representation by the function

$$\tilde{\psi}(\eta) = (2\pi\sigma_p^2)^{-1/4} \exp\left[-\frac{(\eta-\bar{p})^2}{4\sigma_p^2}\right]. \tag{3.8.77}$$

Then physically (3.8.70) describes the "wave packet moving with the velocity $\bar{p}/m$". Apparently (3.8.77) does not belong to the domain of the time operator T (since in particular $\tilde{\psi}(0) \neq 0$) and therefore $D_S(T) = \infty$. However it does not mean that the "arrival time of the wave packet" cannot be measured with finite variance; one can use $\widehat{T}$ which corresponds to measuring t by measuring the coordinate of the "wave packet" and dividing it by the known velocity $\bar{p}/m$. The variance of this measurement is finite for (3.8.77) and grows as $t^2\sigma_p^2/\bar{p}^2$ with $t \to \infty$, while for a covariant measurement the variance would be constant.

Finally we want to show that there is a modified form of "canonical conjugateness" between the energy and time observables. Obviously, on a dense domain the Heisenberg type CCR $[T, H] = \mathrm{i}\hbar I$ holds. Moreover there is a kind of the Weyl CCR. Consider the energy shift operators in $\mathscr{L}^2_{\mathbb{K}}(0,\infty)$

$$P_e\boldsymbol{\psi}_{\boldsymbol{\epsilon}} = \begin{cases} \boldsymbol{\psi}_{\boldsymbol{\epsilon}-e}; & \epsilon \geq e, \\ 0; & \epsilon < e \end{cases}$$

(see Figure 3.5). Apparently $P_e^* P_e = I$, $P_e P_e^* \leq I$, $e > 0$. The family $\{P_e; e \geq 0\}$ constitutes a semigroup of isometric operators in $\mathscr{L}^2_{\mathbb{K}}(0,\infty)$. Using (3.8.74) one gets

$$P_e = \int \mathrm{e}^{\mathrm{i}e\tau/\hbar} M(\mathrm{d}\tau),$$

where $M(\mathrm{d}\tau)$ is the spectral measure of $T = \mathrm{i}\hbar\mathrm{d}/\mathrm{d}\epsilon$ and formally $P_e = \exp(\mathrm{i}eT/\hbar)$. The fact that P_e are not unitary is strictly connected to the non-self-adjointness of T. From the definition of P_e we obtain the relations

$$\begin{aligned} &V_t^* P_e V_t = \mathrm{e}^{\mathrm{i}et/\hbar} P_e \\ &P_e^* V_t P_e = \mathrm{e}^{-\mathrm{i}et/\hbar} V_t; \qquad t \in \mathbb{R}, \quad 0 \leq e, \end{aligned} \tag{3.8.78}$$

which are algebraically similar to (3.2.17).

Also there is the *time representation* in which the observable T is diagonal. Just as the momentum representation is obtained via the Fourier transform from the Schrödinger representation, the time representation

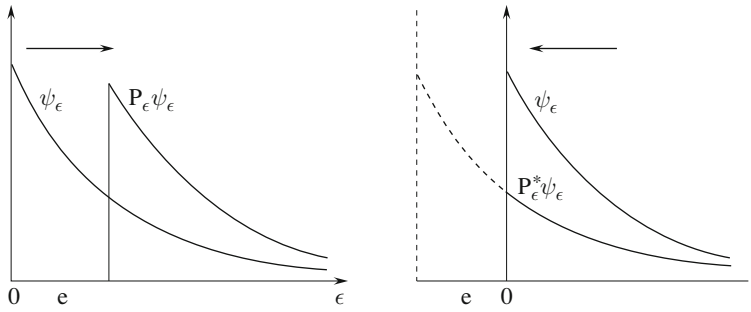

Figure 3.5.

can be defined through the Fourier transform of the energy representation

$$\tilde{\psi}(t) = \frac{1}{\sqrt{2\pi\hbar}} \int_0^\infty \mathrm{e}^{\mathrm{i}\epsilon t/\hbar} \psi_\epsilon \, \mathrm{d}\epsilon.$$

The Hilbert space which is obtained in this way from $\mathscr{L}^2(0, \infty)$ is called the Hardy class $\mathscr{H}^2$ for the half-plane. A reader may elaborate the time representation of the CCR for the case of a free particle.

3.9. Quantum oscillator and phase measurement

The relation (3.7.63) gives a formal expression for the Hamiltonian H; to define quantum dynamics, *i.e.*, the group of unitary operators $V_t = \exp(-\mathrm{i}tH)$, $t \in \mathbb{R}$, the operator H needs to be essentially self-adjoint. Proving this property for various potentials $V(\cdot)$ is one of the main mathematical problems of quantum mechanics. The other one is spectral analysis of the Hamiltonian. These problems have been intensively studied; some indications to the literature can be found at the end of this chapter. They are not in the scope of our book and we restrict ourselves to one example which is both very simple and important.

Consider the energy operator

$$E = \frac{1}{2m}(p^2 + m^2\omega^2 q^2), \tag{3.9.79}$$

which corresponds to the classical expression for the energy of a harmonic oscillator with the mass m and the angular frequency ω. Putting $m = 1$ to simplify formulas we get the formal Hamiltonian

$$H = \frac{1}{2\hbar}\left(-\hbar^2 \frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \omega^2\xi^2\right) \tag{3.9.80}$$

in the Schrödinger representation. Since H is the sum of two noncommuting unbounded operators, the question arises whether (3.9.80) defines an essentially self-adjoint operator.

The expression (3.9.80) is defined at least on $\mathscr{S}(\mathbb{R})$, which is an invariant subspace of p, q and any polynomial in p and q. Introducing on $\mathscr{S}(\mathbb{R})$ the operators

$$a = \frac{1}{\sqrt{2\hbar\omega}}(\omega q + ip), \qquad a^* = \frac{1}{\sqrt{2\hbar\omega}}(\omega q - ip),$$

we have $(\varphi|a\psi) = (a^*\varphi|\psi)$ for $\varphi, \psi \in \mathscr{S}(\mathbb{R})$. Later we shall extend a and a^* so that a^* will be the adjoint of a in the sense of Section 2.4. The relations for a, a^* are the same as those defining the complex amplitude of the classical oscillator.

The Heisenberg CCR (3.3.31) in terms of a, a^* takes the form

$$[a, a^*] = I. \tag{3.9.81}$$

This equation holds at least on $\mathscr{S}(\mathbb{R})$. The energy operator takes the form

$$E = \hbar\omega(a^*a + \tfrac{1}{2}). \tag{3.9.82}$$

Following Dirac we construct the complete orthonormal system of eigenvectors of the operator

$$N = a^*a \tag{3.9.83}$$

and hence obtain the self-adjoint extensions of N and E.

Consider the vector of the *oscillator ground state* $|0) \equiv |0, 0; \hbar/2\omega)$ which according to (3.5.43) is represented by the function

$$(\xi|0) = \left(\frac{\pi\hbar}{\omega}\right)^{-1/4} \exp\left(-\frac{\omega\xi^2}{2\hbar}\right)$$

in the Schrödinger representation. From (3.5.41) it follows that $(\omega q + \mathrm{i}p)|0) = 0$ whence

$$a|0) = 0, \qquad N|0) = 0,$$

so that $|0)$ is eigenvector of N with zero eigenvalue. Define

$$|n) = \frac{1}{\sqrt{n!}}(a^*)^n|0); \qquad n = 0, 1, \ldots \tag{3.9.84}$$

Then by (3.9.81)

$$a|n) = \sqrt{n}|n-1), \qquad a^*|n) = \sqrt{n+1}|n+1). \tag{3.9.85}$$

It follows that

$$N|n) = n|n); \qquad n = 0, 1, \ldots,$$

so that $|n)$ is an eigenvector of N with the eigenvalue n.

In the Schrödinger representation

$$\begin{aligned}(\xi|n) &= \frac{1}{\sqrt{n!}}\left[\frac{1}{\sqrt{2\hbar\omega}}\right]^n\left(\omega\xi - \hbar\frac{\mathrm{d}}{\mathrm{d}\xi}\right)^n\left(\frac{\pi\hbar}{\omega}\right)^{-1/4}\exp\left(-\frac{\omega\xi^2}{2\hbar}\right)\\ &= \sqrt{\frac{\omega}{\pi\hbar}\frac{1}{2^n n!}}H_n\left(\sqrt{\frac{\omega}{\hbar}}\xi\right)\exp\left(-\frac{\omega\xi^2}{2\hbar}\right),\end{aligned} \tag{3.9.86}$$

where $H_n(\cdot)$ are the Hermite polynomials. It is known that the functions (3.9.86) form a complete orthonormal system in $\mathscr{L}^2(\mathbb{R})$ so that

$$\sum_{n=0}^{\infty}|n)(n| = I. \tag{3.9.87}$$

Therefore any vector ψ can be represented as

$$|\psi) = \sum_{n=0}^{\infty}|n)(n|\psi),$$

where $(n|\psi)$; $n = 0, 1, \ldots$, is a square-summable sequence of complex numbers. Therefore states and observables can be represented by matrices acting in l^2. In particular, the CCR (3.3.22) can be written in the matrix form. This is called the Fock representation. The isometric transformations from the Schrödinger representation to the Fock representation and viceversa are give by

$$(n|\psi) = \int(n|\xi)(\xi|\psi)\mathrm{d}\xi, \qquad (\xi|\psi) = \sum_n(\xi|n)(n|\psi),$$

where the kernel $(\xi|n) = \overline{(n|\xi)}$ is defined by (3.9.86).

The self-adjoint extension of N is now given by

$$\mathscr{D}(N) = \left\{\psi : \sum_{n=0}^{\infty}n^2|(n|\psi)|^2 < \infty\right\},$$

$$N|\psi) = \sum_{n=0}^{\infty}n|n)(n|\psi); \qquad \psi \in \mathscr{D}(N),$$

as follows from spectral Theorem 2.4.1. This can be shown to be the unique self-adjoint extension of the operator (3.9.83) defined on $\mathscr{S}(\mathbb{R})$. Defining

$$\mathscr{D}(a) = \mathscr{D}(a^*) = \left\{ \psi : \sum_{n=0}^{\infty} n|(\psi|n)|^2 < \infty \right\},$$

we have $(a)^* = a^*$, $(a^*)^* = a$ and $N = a^*a$.

According to (3.9.82) we obtain the self-adjoint extension of the energy operator E. Obviously the system $|n)$; $n = 0, 1, \ldots$ forms the basis of eigenvectors of E with

$$E|n) = \hbar\omega \left(n + \frac{1}{2} \right) |n); \qquad n = 0, 1, \ldots$$

This relation shows that the spectrum of the oscillator energy consists of discrete series $(n + \frac{1}{2})\hbar\omega$; $n = 0, 1, \ldots$ Physicists call the variable n number of quanta, so that $|n)$ describes the n-quanta state. The operator N presents the observable *number of quanta*. Since according to (3.9.84) a diminishes the "number of quanta" by one, it is called the *annihilation operator* (of a quantum). For similar reason a^* is called the *creation operator*.

Since the Hamiltonian $H = \hbar^{-1}E$ is diagonal in the Fock representation, dynamics of a quantum oscillator is described very simply in this representation, namely

$$V_t|\psi) = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}\omega(n+1/2)t}|n)(n|\psi). \tag{3.9.88}$$

The n-quanta states are thus the stationary, *i.e.*, time-invariant states. In the Heisenberg picture the quantum dynamics takes the form which is similar to classical equations. Introducing time-varying operators

$$a(t) = V_t^* a V_t, \qquad a(t)^* = V_t^* a^* V_t,$$

we obtain using (3.9.84) and (3.9.87)

$$a(t) = \mathrm{e}^{-\mathrm{i}\omega t} a, \qquad a(t)^* = \mathrm{e}^{\mathrm{i}\omega t} a^*. \tag{3.9.89}$$

Expressing p, q back through a, a^* by the relations

$$p = \sqrt{2\hbar\omega}\frac{a - a^*}{2\mathrm{i}}, \qquad q = \sqrt{\frac{2\hbar}{\omega}}\frac{a + a^*}{2}, \tag{3.9.90}$$

we obtain for $p(t) = V_t^* p V_t$, $q(t) = V_t^* q V_t$

$$p(t) = p\cos\omega t - \omega q \sin\omega t, \quad q(t) = q\cos\omega t + p\omega^{-1}\sin\omega t; \quad (3.9.91)$$

these relations are formally the same as the solutions of the equations of motion for the classical harmonic oscillator.

The oscillator dynamics is periodical in the sense that $V_{t+2\pi/\omega} = V_t$ for all t as follows from (3.9.88). Therefore the transformation (3.8.70) of a state S corresponding to the time shift t is equivalent to the transformation

$$S_\theta = \mathrm{e}^{\mathrm{i}\theta N} S \mathrm{e}^{-\mathrm{i}\theta N}, \quad (3.9.92)$$

where $\theta = (-\omega t)(\mathrm{mod}\, 2\pi)$ varies in $[0, 2\pi)$. Since the eigenvalues of N are integers, then $\exp(\mathrm{i}\theta N) = \exp \mathrm{i}(\theta + 2\pi k)N$ so that $\theta \to V_\theta = \exp(\mathrm{i}\theta N)$ is a unitary representation of the additive group of $[0, 2\pi)$ modulo 2π. This group $\mathbb{T}$ is the same as the group of rotations of the unit circle. We shall call θ in (3.9.92) the *phase parameter*.

To elucidate what measurements should be associated with the phase parameter consider the covariance condition

$$V_\theta^* M(B) V_\theta = M(B_{-\theta}); \qquad \theta \in [0, 2\pi),$$

where $V_\theta = \exp(\mathrm{i}\theta N)$, and $B_{-\theta}$ is the shift $(\mathrm{mod}\, 2\pi)$ of the set $B \subset [0, 2\pi)$. A resolution of identity satisfying this condition is given by the symbolic matrix elements in the Fock representation

$$(n|M(\mathrm{d}\theta)|n') = \mathrm{e}^{\mathrm{i}(n-n')\theta} \frac{\mathrm{d}\theta}{2\pi}. \quad (3.9.93)$$

As we shall show in Chapter 4 this measurement can be singled out from all covariant measurements of the phase parameter by an optimality requirement. We call (3.9.93) the *canonical phase measurement* in the Fock representation.

We shall also show that there is no simple covariant measurement of θ. Thus there is no phase observable in the conventional sense. To show that the resolution of identity (3.9.93) is not orthogonal consider the operators

$$P = \int_0^{2\pi} \mathrm{e}^{\mathrm{i}\theta} M(\mathrm{d}\theta), \qquad P^* = \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\theta} M(\mathrm{d}\theta). \quad (3.9.94)$$

From (3.9.93), taking into account that

$$(2\pi)^{-1} \int_0^{2\pi} \exp \mathrm{i}(n - n')\theta \, \mathrm{d}\theta = \delta_{nn'}$$

we get the matrix representation

$$P = \sum_{n=1}^{\infty} |n-1)(n|, \qquad P^* = \sum_{n=1}^{\infty} |n)(n-1|. \tag{3.9.95}$$

It follows that

$$PP^* = I, \qquad P^*P = I - |0)(0|,$$

so that P^* is isometric but not unitary as it would be if (3.9.93) were an orthogonal resolution of identity.

From (3.9.85) and (3.9.95) one gets the relations

$$a = P|a|, \qquad a^* = |a|P^*,$$

where $|a| = \sqrt{a^*a} = \sqrt{N}$, which are the noncommutative analog of the definition of phase for the classical oscillator: $a = |a| \exp \mathrm{i}\theta$. Also from (3.9.95) follow the relations between the unitary group $\{V_\theta; 0 \le \theta < 2\pi\}$ and the discrete semigroup of isometric operators $\{(P^*)^n; n = 0, 1, \ldots\}$

$$\begin{aligned} &V_\theta^*(P^{*n})^n V_\theta = \mathrm{e}^{-\mathrm{i}n\theta}(P^*)^n; \qquad P^n V_\theta (P^*)^n = \mathrm{e}^{\mathrm{i}n\theta} V_\theta; \\ &0 \le \theta < 2\pi, \qquad n = 0, 1, \ldots, \end{aligned} \tag{3.9.96}$$

which are analogous to the time-energy relations (3.8.78). Thus, though there is no strict canonical conjugateness between phase and number of quanta, a weakened form of the conjugateness expressed by the last equation holds.

Similarly to the time representation one can introduce the *phase representation* in the space of the functions of the form

$$\psi(\theta) = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}n\theta}(n|\psi); \qquad 0 \le \theta < 2\pi.$$

The Hilbert space obtained in this way from l^2 is called the Hardy class $\mathscr{H}^2$ for the unit circle. The operator P^* acts as multiplication by $\exp(-\mathrm{i}\theta)$ in this space. The displacement operators $W_{x,v}$ and thus the representation of the CCR can be constructed in this space but we shall not pursue this matter.

3.10. The coherent-state representation

An important role in the theory of quantum oscillator and related topics is played by the minimum-uncertainty states which are reparametrized

by introducing complex variable $\zeta = (2\hbar\omega)^{-12}(\omega\overline{Q} + i\hbar\overline{P})$. Putting $W_\zeta = W_{\overline{Q},\hbar\overline{P}}$ and taking into account that $\hbar = \mu^{-1}$ since $m = 1$ we rewrite the CCR (3.3.22) in the form

$$W_{\zeta_1} W_{\zeta_2} = \mathrm{e}^{\mathrm{i}\,\mathrm{Im}\,\zeta_1\bar{\zeta}_2} W_{\zeta_1+\zeta_2}. \tag{3.10.97}$$

Then denoting $|\zeta) = |\overline{P}, \overline{Q}; \hbar/2\omega)$ we can rewrite (3.5.44) as

$$|\zeta) = W_\zeta |0), \tag{3.10.98}$$

where $|0)$ is the vector of the oscillator ground state. The states $|\zeta)(\zeta|$ are called the oscillator *coherent states*, the explanation of the name to be found in quantum optics where they are especially useful.

Rewriting (3.5.41) in the complex form we get

$$a|\zeta) = \zeta|\zeta), \qquad \zeta \in \mathbb{C}, \tag{3.10.99}$$

so that $\{|\zeta)\}$ forms a continual system of eigenvectors of the annihilation operator a. Distinct from eigenvectors of a self-adjoint operator they are not orthogonal; namely, (3.10.98) and (3.10.97) imply

$$(\zeta_1|\zeta_2) = \exp[-\tfrac{1}{2}(|\zeta_1|^2 + |\zeta_2|^2 - 2\bar{\zeta}_1\zeta_2)]. \tag{3.10.100}$$

The completeness relation (3.5.45) takes the form

$$\int_{\mathbb{C}} |\zeta)(\zeta| \frac{\mathrm{d}^2\zeta}{2\pi} = I, \tag{3.10.101}$$

where $\mathrm{d}^2\zeta = \frac{1}{2}\mathrm{d}\overline{P}\,\mathrm{d}\overline{Q}$. Applying this to a vector $\psi \in \mathscr{H}$ we obtain

$$|\psi) = \int_{\mathbb{C}} |\zeta)(\zeta|\psi) \frac{\mathrm{d}^2\zeta}{\pi}.$$

This formula establishes the one-to-one correspondence between the vectors ψ of the Hilbert space of an irreducible representation of CCR and the functions $\psi(\zeta) = (\zeta|\psi)$ of the complex variable. Since by (3.10.101)

$$(\psi|\varphi) = \int_{\mathbb{C}} (\psi|\zeta)(\zeta|\varphi) \frac{\mathrm{d}^2\zeta}{2\pi},$$

the space $\mathscr{H}$ is isometrically embedded into the space $\mathscr{L}^2(\mathbb{C})$ of complex-valued square-integrable functions of ζ. The image of $\mathscr{H}$ is a proper subspace of $\mathscr{L}^2(\mathbb{C})$ which we are going to describe.

Putting $\zeta_2 = 0$ in (3.10.100) we obtain that the oscillator ground state vector gives rise to the function $(\zeta|0) = \exp(-|\zeta|^2/2)$. Therefore by (3.10.99) the functions representing n-quanta state vectors (3.9.83) are

$$(\zeta|n) = \frac{1}{\sqrt{n!}}(\zeta|(a^*)^n|0) = \frac{\bar{\zeta}^n}{\sqrt{n!}}\mathrm{e}^{-|\zeta|^2/2}; \qquad n = 0, 1, \dots \quad (3.10.102)$$

Since $\psi \in \mathscr{H}$ is uniquely presented as

$$|\psi) = \sum_{n=0}^{\infty} c_n|n), \qquad \sum_{n=0}^{\infty} |c_n|^2 = (\psi|\psi) < \infty,$$

then the function $\psi(\zeta) = (\zeta|\psi)$ has the form

$$\psi(\zeta) = \overline{f(\zeta)}\exp(-|\zeta|^2/2), \quad (3.10.103)$$

where $f(\zeta) = \sum \bar{c}_n\zeta^n/\sqrt{n!}$ is a holomorphic function of $\zeta \in \mathbb{C}$. Moreover $\int |f(\zeta)|^2 \exp(-|\zeta|^2)\mathrm{d}^2\zeta < \infty$. Thus the image of $\mathscr{H}$ is the Hilbert space $\mathscr{E}^2(\mathbb{C})$ of the complex-valued functions (3.10.103) with the inner product

$$(\psi_1|\psi_2) = \int_C \overline{\psi_1(\zeta)}\psi_2(\zeta)\frac{\mathrm{d}^2\zeta}{\pi} = \int_{\mathbb{C}} \overline{f_1(\zeta)}f_2(\zeta)\mathrm{e}^{-|\zeta|^2}\frac{\mathrm{d}^2\zeta}{\pi}. \quad (3.10.104)$$

Any bounded operator X in $\mathscr{E}^2(\mathbb{C})$ is an integral operator with the kernel $(\zeta_1|X|\zeta_2)$ since by (3.10.101)

$$(\zeta_1|X\psi) = \int_{\mathbb{C}} (\zeta_1|X|\zeta_2)(\zeta_2|\psi)\frac{\mathrm{d}^2\zeta_2}{\pi}.$$

In particular the displacement operator W_ζ has the kernel

$$(\zeta_1|W_\zeta|\zeta_2) = \exp[-\tfrac{1}{2}(|\zeta_1|^2 + |\zeta|^2 + |\zeta_2|^2) + \bar{\zeta}_1\zeta_2 + \bar{\zeta}_1\zeta - \bar{\zeta}\zeta_2],$$

as follows from (3.10.97) and (3.10.100). This defines the representation of the CCR in $\mathscr{E}^2(\mathbb{C})$, which is sometimes called the *coherent-state representation*. The transformations from this representation to the Fock representation and viceversa are given by the kernel (3.10.102); those to the Schrödinger representation by the kernel

$$(\xi|\zeta) = (\pi\hbar/\omega)^{-1/4}\exp\left[-\left(\sqrt{\frac{\omega}{2\hbar}}\xi - \zeta\right)^2 \Big/ 2 - |\zeta|^2/2\right],$$

which can be derived from (3.5.42).

The coherent-state representation diagonalizes the creation operator a^* since in $\mathscr{E}^2(\mathbb{C})$ it is just the operator of multiplication by $\bar{\zeta}$: $(\zeta|a^*\psi) = \bar{\zeta}(\zeta|\psi)$. The annihilation operator a is $\partial/\partial\bar{\zeta} + \frac{1}{2}\zeta$. Let us show that the nonorthogonal resolution of identity

$$M(\mathrm{d}^2\zeta) = |\zeta)(\zeta|\frac{\mathrm{d}^2\zeta}{\pi}$$

is in a sense a spectral measure of a^*. Applying formally to (3.10.101) operators a^* and a we obtain

$$a^* = \int_{\mathbb{C}} \bar{\zeta}\, M(\mathrm{d}^2\zeta), \qquad a = \int_{\mathbb{C}} \zeta\, M(\mathrm{d}^2\zeta).$$

To give these relations precise meaning we remark that for $\psi \in \mathscr{D}(a^*)$

$$\|a^*\psi\|^2 = \int |\zeta|^2 |(\psi|\zeta)|^2 \frac{\mathrm{d}^2\zeta}{\pi}.$$

Indeed, by (3.10.104) and (3.10.99)

$$\|a^*\psi\|^2 = \int |(a^*\psi|\zeta)|^2 \frac{\mathrm{d}^2\zeta}{\pi} = \int |\zeta|^2 |(\psi|\zeta)^2 \frac{\mathrm{d}^2}{\pi}\zeta.$$

Thus

$$\mathscr{D}(a^*) = \left\{\psi : \int |\zeta|^2 |(\psi|\zeta)|^2 \frac{\mathrm{d}^2\xi}{\pi} < \infty\right\}.$$

For $\psi \in \mathscr{D}(a^*)$ the integral $\int |\zeta|^2 (\psi|M(\mathrm{d}^2\zeta)\psi)$ is convergent and so are the integrals $(\varphi|a^*\psi) = \int \bar{\zeta}(\varphi|M(\mathrm{d}^2\zeta)\psi)$, $(\varphi|a\psi) = \int \zeta(\varphi|M(\mathrm{d}^2\zeta)\psi)$. To sum up, the operators a^* and $M(\mathrm{d}^2\zeta)$ are connected by the relations

$$\begin{aligned}
&\mathscr{D}(a^*) = \left\{\psi : \int |\zeta|^2 (\psi|M(\mathrm{d}^2\zeta)\psi) < \infty\right\};\\
&\|a^*\psi\|^2 = \int |\zeta|^2 (\psi|M(\mathrm{d}^2\zeta)\psi), \qquad \psi \in \mathscr{D}(a^*); \qquad (3.10.105)\\
&(\varphi|a^*\psi) = \int \bar{\zeta}(\varphi|M(\mathrm{d}^2\zeta)\psi), \qquad \varphi, \psi \in \mathscr{D}(a^*),
\end{aligned}$$

which are similar to those connecting a maximal symmetric operator with its spectral measure (Section 2.4). Applying to (3.10.101) the operator a^m from the left and the operator $(a^*)^n$ from the right we obtain

$$a^m(a^*)^n = \int_{\mathbb{C}} \zeta^m \bar{\zeta}^n M(\mathrm{d}^2\zeta).$$

The order of the operators a and a^* in this formula is important since they do not commute. An operator expression in which all a^* follow all a is called *normally ordered*. The above relation can be generalized to

$$F(a, a^*) = \int F(\zeta, \bar{\zeta}) M(\mathrm{d}^2\zeta),$$

where $F(\zeta, \bar{\zeta})$ is a function representable by a power series in $\zeta, \bar{\zeta}$ which converges rapidly enough and $F(a, a^*)$ means the normally ordered expression.

One may wonder what is the property of the creation operator a^* which enables the spectral representation (3.10.105). Recall that a bounded operator X in $\mathscr{H}$ is called *normal* if $[X, X^*] = 0$; this is equivalent to $\|X\psi\| = \|X^*\psi\|$, $\psi \in \mathscr{H}$. A densely defined (unbounded) operator X is normal if $\mathscr{D}(X) = \mathscr{D}(X^*)$ and $\|X\psi\| = \|X^*\psi\|$, $\psi \in \mathscr{D}(X)$. For a normal operator X there is a unique orthogonal resolution of identity $E(\mathrm{d}^2\zeta)$ on $\mathbb{C}$ such that

$$\mathscr{D}(X) = \left\{\psi : \int |\zeta|^2 (\psi | E(\mathrm{d}^2\zeta)\psi) < \infty\right\};$$

$$\|X\psi\|^2 = \int |\zeta|^2 (\psi | E(\mathrm{d}^2\zeta)\psi); \qquad \psi \in \mathscr{D}(X);$$

$$(\varphi | X\psi) = \int \zeta (\varphi | E(\mathrm{d}^2\zeta)\psi); \qquad \varphi, \psi \in \mathscr{D}(X).$$

Loosely speaking, normal operators are diagonalizable but may have a complex spectrum.

A densely defined operator Y in $\mathscr{H}$ is called *subnormal* if it can be extended to a normal operator X in a Hilbert space $\widetilde{\mathscr{H}} \supset \mathscr{H}$, *i.e.*, $Y\psi = X\psi$, $\psi \in \mathscr{D}(Y) \subset \mathscr{D}(X)$. The operator Y is subnormal if and only if there is a resolution of identity $M(\mathrm{d}^2\zeta)$ such that

$$\mathscr{D}(Y) \subseteq \left\{\psi : \int |\zeta|^2 (\psi | M(\mathrm{d}^2\zeta)\psi) < \infty\right\};$$

$$\|Y\psi\|^2 = \int |\zeta|^2 (\psi | M(\mathrm{d}^2\zeta)\psi); \qquad \psi \in \mathscr{D}(Y); \tag{3.10.106}$$

$$(\varphi | Y\psi) = \int \zeta (\varphi | M(\mathrm{d}^2\zeta)\psi); \qquad \varphi, \psi \in \mathscr{D}(Y).$$

Indeed if X is a normal extension of Y with the spectral measure $E(\mathrm{d}^2\xi)$, then the resolution of identity

$$M(\mathrm{d}^2\zeta) = \widetilde{E} E(\mathrm{d}^2\zeta) \widetilde{E},$$

where $\widetilde{E}$ is the projection from $\tilde{\mathscr{H}}$ onto $\mathscr{H}$, satisfies (3.10.106). Conversely let $M(\mathrm{d}^2\zeta)$ be a resolution of identity satisfying (3.10.106), and let $E(\mathrm{d}^2\zeta)$ be its Naimark extension (see Section 2.5) in a Hilbert space $\tilde{\mathscr{H}}$. Then $X = \int \zeta E(\mathrm{d}^2\zeta)$ (with the corresponding domain) is a normal operator satisfying

$$\|X\psi\|^2 = \int |\zeta|^2(\psi|E(\mathrm{d}^2\zeta)\psi) = \int |\zeta|^2(\psi|M(\mathrm{d}^2\zeta)\psi) = \|Y\psi\|^2$$

for $\psi \in \mathscr{D}(Y)$. From the last relation in (3.10.106) it follows that $\widetilde{E}X\psi = Y\psi$, $\psi \in \mathscr{D}(Y)$. Therefore by the second relation $\|X\psi\|^2 = \|Y\psi\|^2 = \|\widetilde{E}X\psi\|^2$ whence $X\psi = \widetilde{E}X\psi = Y\psi$, $\psi \in \mathscr{D}(Y)$, and X is an extension of Y.

Any densely defined symmetric operator is subnormal since by Theorem 2.4.3 it admits a spectral representation (3.10.106) with $M(\mathrm{d}^2\zeta)$ concentrated on $\mathbb{R}$. In particular the operator (3.8.75) in $\mathscr{L}^2_{\mathbb{K}}(0,\infty)$ representing the time observable is subnormal; its normal extension is the self-adjoint operator $\hbar\mathrm{i}\,\mathrm{d}/\mathrm{d}\epsilon$ in $\mathscr{L}^2_{\mathbb{K}}(\mathbb{R})$. An example of a bounded subnormal operator is the operator P^* given by (3.9.95). Its normal extension is the unitary operator $\sum_{n=-\infty}^{\infty} |n)(n-1|$ in the Hilbert space $\tilde{\mathscr{H}}$ spanned by the orthonormal basis $\{|n);\, n = 0, \pm 1, \ldots\}$ which is composed of the basis in the Fock space $\mathscr{H}$, corresponding to nonnegative n and by additional orthonormal vectors, corresponding to negative values of n. Since (3.10.105) is (3.10.106) for the creation operator a^* it follows that a^* is subnormal. To construct its normal extension consider the commuting extension (3.6.54) of the canonical pair Q, P. In terms of a, a^* it is

$$\tilde{a} = a \otimes I_0 + I \otimes a_0^*, \qquad \tilde{a}^* = a^* \otimes I_0 + I \otimes a_0,$$

where $\tilde{a}$, $\tilde{a}^*$ are normal. Let $|0)_{00}(0|$ be the ground state in $\mathscr{H}_0$. Identifying $\mathscr{H}$ with the subspace of $\tilde{\mathscr{H}} = \mathscr{H} \otimes \tilde{\mathscr{H}}_0$ consisting of vectors of the form $|\psi) \otimes |0)_0$, $\psi \in \mathscr{H}$, we have $\mathscr{H} \subseteq \tilde{\mathscr{H}}$. The operator $\tilde{a}^*$ is then an extension of a^* since

$$\tilde{a}^*[|\psi) \otimes |0)_0] = a^*|\psi) \otimes |0)_0 + |\psi) \otimes a_0|0)_0 = a^*|\psi).$$

We have seen that to different physical quantities including those which are not represented by self-adjoint operators correspond different representations in which the quantity has the especially simple "diagonal" form. All these representations are unitary equivalent; moreover transformations from one representation to another are given by explicit kernels. The choice of the representation plays a secondary role simplifying calculations; in fact any calculation can be translated from the language of one representation to another as well as performed in abstract form involving only the CCR (3.3.22).

3.11. Representations of the rotation group and angular momenta

In the case of the three-dimensional Cartesian space one looks for irreducible representations of the Galilean group of transformations (3.1.1). We shall restrict ourselves to kinematical aspects of a quantum particle in three dimensions and consider the transformations

$$\boldsymbol{\xi}' = \boldsymbol{R}\boldsymbol{\xi} + \boldsymbol{x} + \boldsymbol{v}\tau. \tag{3.11.107}$$

The parameters of the kinematical group are the spatial shift $\boldsymbol{x}$, the relative velocity $\boldsymbol{v}$ and the rotation matrix $\boldsymbol{R}$ and we look for irreducible representations $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{R}) \to W_{\boldsymbol{x},\boldsymbol{v},\boldsymbol{R}}$ of the group. An essentially new element as compared to the one-dimensional case is rotation and we first put $\boldsymbol{x} = 0$, $\boldsymbol{v} = 0$ and consider the representations $\boldsymbol{R} \to W_{\boldsymbol{R}}$ of the rotation group.

By Euler's theorem any rotation $\boldsymbol{R}$ is a rotation $\boldsymbol{R}_{n,\varphi}$ around an axis $\boldsymbol{n}$ through an angle φ. Here $\boldsymbol{n}$ is the unit vector in the direction of the axis. For a fixed $\boldsymbol{n}$ the family $\{\boldsymbol{R}_{\boldsymbol{n},\varphi}\}$ constitutes the one-parameter group and the operators $V_\varphi = W_{\boldsymbol{R}_{\boldsymbol{n},\varphi}}$ form a unitary representation of this group. By Stone's theorem

$$V_\varphi = \exp(-\mathrm{i}\varphi L_n),$$

where L_n is a self-adjoint operator. Let $\{\boldsymbol{e}_j\}$ be a Cartesian frame so that $\boldsymbol{n} = n_1\boldsymbol{e}_1 + n_2\boldsymbol{e}_2 + n_2\boldsymbol{e}_3$ with $\sum n_j^2 = 1$. For small values of φ the matrix of the rotation $\boldsymbol{R}_{\boldsymbol{n},\varphi}$ is approximately

$$\boldsymbol{R}_{\boldsymbol{n},\varphi} \approx \boldsymbol{I} - \varphi \sum_j n_j \boldsymbol{D}_j, \tag{3.11.108}$$

where

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{D}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix},$$

$$\boldsymbol{D}_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{D}_3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

in the frame $\{\boldsymbol{e}_j\}$. The matrices $\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3$ satisfying the commutation relations

$$[\boldsymbol{D}_1, \boldsymbol{D}_2] = -\boldsymbol{D}_3, \quad [\boldsymbol{D}_2, \boldsymbol{D}_3] = -\boldsymbol{D}_1, \quad [\boldsymbol{D}_3, \boldsymbol{D}_1] = -\boldsymbol{D}_2 \tag{3.11.109}$$

constitute what is called the *Lie algebra of the rotation group*. Let $L_j \equiv L_{\boldsymbol{e}_j}$ be the infinitesimal generator of the one-parameter group of rotations around the coordinate axis $\boldsymbol{e}_j$; $j = 1, 2, 3$. From the general representation theory it follows that the operators $\mathrm{i}L_j$; $j = 1, 2, 3$ constitute a representation of the Lie algebra, *i.e.*,

$$[L_1, L_2] = \mathrm{i}L_3, \qquad [L_2, L_3] = \mathrm{i}L_1, \qquad [L_3, L_1] = \mathrm{i}L_2. \quad (3.11.110)$$

Moreover

$$L_n = \sum_{j=1}^{3} n_j L_j, \quad (3.11.111)$$

so that

$$W_{\boldsymbol{R}_{\boldsymbol{n},\varphi}} = \exp\left[-\mathrm{i}\sum_{j=1}^{3} \varphi_j L_j\right], \quad (3.11.112)$$

where $\varphi_j = \varphi \boldsymbol{n}_j$; $j = 1, 2, 3$.

A usual procedure for finding representations of the group is first to find representations of the Lie algebra, *i.e.*, concrete operators L_j; $j = 1, 2, 3$, satisfying (3.11.110), which is a simpler problem, and then to construct the representations of the group by formula (3.11.112). In this way it is found that for any finite dimension $d \geq 2$ there exists exactly one irreducible representation J_1, J_2, J_3 of the Lie algebra and the corresponding irreducible projective unitary representation of the rotation group

$$\boldsymbol{R} \to U(\boldsymbol{R}) = \exp\left[-\mathrm{i}\sum_{j=1}^{3} \varphi_j J_j\right] \quad (3.11.113)$$

in d-dimensional Hilbert space. We shall describe this representation in some detail in the simplest case $d = 2$.

Introducing $\sigma_j = 2J_j$ we can rewrite the commutation relations (3.11.110) in the form

$$[\sigma_1, \sigma_2] = 2\mathrm{i}\sigma_3, \qquad [\sigma_2, \sigma_3] = 2\mathrm{i}\sigma_1, \qquad [\sigma_3, \sigma_1] = 2\mathrm{i}\sigma_2.$$

The solution of these equation is unique to a factor and is given by matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \qquad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

which are called the *Pauli matrices*. They satisfy

$$\sigma_1\sigma_2 = \mathrm{i}\sigma_3, \qquad \sigma_2\sigma_3 = \mathrm{i}\sigma_1, \qquad \sigma_3\sigma_1 = \mathrm{i}\sigma_2. \quad (3.11.114)$$

Any Hermitean (2×2)-matrix can be represented as a unique linear combination of the unit matrix and the Pauli matrices. In particular, for a density matrix

$$S = \tfrac{1}{2}(I + \theta_1\sigma_1 + \theta_2\sigma_2 + \theta_3\sigma_3),$$

where θ_j are the Stokes parameters introduced in Section 1.2. The relations (3.11.114) make the calculations with matrices presented in this form very convenient. In particular the exponents (3.11.113) are given by

$$\begin{aligned} U(\boldsymbol{R}) &= \exp\left(-\frac{\mathrm{i}}{2}\sum_{j=1}^{3}\varphi_j\sigma_j\right) \\ &= I\cdot\cos\frac{\varphi}{2} - \mathrm{i}\sum_{j=1}^{3}\varphi_j\sigma_j\varphi^{-1}\sin\frac{\varphi}{2}, \\ \varphi &= \sqrt{\varphi_1^2+\varphi_2^2+\varphi_3^2}. \end{aligned} \tag{3.11.115}$$

The representation is essentially projective since, *e.g.*, for $\varphi_1 = 2\pi$, $\varphi_2 = \varphi_3 = 0$ relation (3.11.115) gives $-I$ and not I.

For any other value of d there is an orthonormal basis in the representation space in which, say, J_3 has the diagonal form. Namely, the eigenvalues of J_3 are $-j, -j+1, \dots j-1, j$, where $2j+1 = d$, so that

$$J_3 = \sum_{n=-j}^{j} m|m)(m|, \tag{3.11.116}$$

where $|m)$ are the eigenvectors. That the differences of the eigenvalues are integral numbers should be expected; indeed the rotation through the angle 2π results in the initial position and therefore $\exp(-2\pi\,\mathrm{i}J_3) = \gamma I$, where $\gamma = \exp(\mathrm{i}\alpha)$ with α real. It follows that the eigenvalues of J_3 have the form $\alpha' + n$ where n are integers. The expression for $U(\boldsymbol{R})$ in this basis will appear later in Section 4.10. The irreducible representations of the rotation group are conventionally indexed by the number j, taking values $0, \frac{1}{2}, 1, \frac{3}{2}, \dots$ (to $j = 0$ corresponds the trivial one-dimensional representation).

We can now describe irreducible representations of the kinematical group. Let $\mathbb{K} = \mathbb{C}^d$ be d-dimensional complex space and $\mathscr{L}^2_{\mathbb{K}}(\mathbb{R}^3)$ the space of vector-functions $\boldsymbol{\psi}(\boldsymbol{\xi})$ on $\mathbb{R}^3$ with values in $\mathbb{K}$ with the finite squared norm given by

$$\|\psi\|^2 = \iiint \|\boldsymbol{\psi}(\boldsymbol{\xi})\|^2 \mathrm{d}\xi_1\,\mathrm{d}\xi_2\,\mathrm{d}\xi_3.$$

The relation

$$W_{\boldsymbol{x},\boldsymbol{v},\boldsymbol{R}}\boldsymbol{\psi}(\boldsymbol{\xi}) = \exp[\mathrm{i}\mu\boldsymbol{v}\cdot(\boldsymbol{\xi}-\boldsymbol{x}/2)]U(\boldsymbol{R})\psi(\boldsymbol{R}^{-1}(\boldsymbol{\xi}-\boldsymbol{x})) \qquad (3.11.117)$$

where $U(\boldsymbol{R})$ act in the space $\mathbb{K}$, *i.e.*, on components of the vector $\boldsymbol{\psi}(\boldsymbol{\xi})$ at each point $\boldsymbol{\xi}$, defines a projective unitary representation of the kinematical group called the Schrödinger representation. Remark that if we leave aside rotations, *i.e.*, put $\boldsymbol{R} = I$ we obtain an obvious generalization of the Schrödinger representation (3.4.32) for the one-dimensional case. The irreducibility then easily follows from the irreducibility of (3.4.32) and $\{U(\boldsymbol{R})\}$. Any (continuous) projective unitary representation of the kinematical group is unitary equivalent to the Schrödinger representation with some values of μ and d. The type of the representation is therefore completely determined by the pair of "quantum numbers": the mass $m = \hbar\mu$ and the number $j = (d-1)/2$ called *spin* of the object.

To explain the meaning of the number j recall that it is the largest eigenvalue of the operator J_3. We now consider the kinematical meaning of the operators L_k, J_k. Let first $j = 0$. Then putting $\boldsymbol{x} = 0$, $\boldsymbol{v} = 0$ in (3.11.117) we obtain the representation $\boldsymbol{R} \to W_{\boldsymbol{R}}$ of the rotation group in $\mathscr{L}^2(\mathbb{R})$ with

$$W_{\boldsymbol{R}}\psi(\xi) = \psi(\boldsymbol{R}^{-1}\xi).$$

To obtain the expression for L_k in the Schrödinger representation consider the infinitesimal rotations (3.11.108). Taking $n_1 = 0$, $n_2 = 0$, $n_3 = 1$ which corresponds to rotations around e_3 we get

$$\begin{aligned} V_\varphi\psi(\xi_1,\xi_2,\xi_3) &\approx \psi(\xi_1+\varphi\xi_2, -\xi_2-\varphi\xi_1, \xi_3) \\ &\approx \left[1-\varphi\left(\xi_1\frac{\partial}{\partial\xi_2}-\xi_2\frac{\partial}{\partial\xi_1}\right)\right]\psi(\xi_1,\xi_2,\xi_3), \end{aligned}$$

whence

$$L_3 = \mathrm{i}^{-1}\left(\xi_1\frac{\partial}{\partial\xi_2}-\xi_2\frac{\partial}{\partial\xi_1}\right) \qquad (3.11.118)$$

at least on $\mathscr{S}(\mathbb{R})$, or $l_3 = q_1p_2 - q_2p_1$, where $l_3 = \hbar L_3$ and $q_j = \xi_j$ are the position observables, $p_j = \hbar\mathrm{i}^{-1}\partial/\partial\xi_j$ are the momentum observables in the Schrödinger representation. For $l_1 = \hbar L_1$ and $l_2 = \hbar L_2$ similar relations hold with the circular permutation of the indices. Introducing vector notations $\boldsymbol{l} = [l_1, l_2, l_3]$ etc. we can write them in the form $\boldsymbol{l} = \boldsymbol{q}\times\boldsymbol{p}$, which is the same as the expression for the vector of angular momentum in classical mechanics. The operators l_1, l_2, l_3 are therefore called the *orbital angular momentum observables* about the corresponding axes.

Returning to the case of arbitrary spin we remark that the operators J_k arise from representation of the rotation group in the same way as L_k and satisfy the same commutation relations. One may loosely imagine $s_k = \hbar J_k$ as "internal" angular momenta of the quantum object. They are called the *spin angular momentum observables*. The infinitesimal generators of the representation of rotation group obtained from (3.11.117) are $L_k + J_k$; this is interpreted by saying that the total angular momentum is the vector sum of the orbital and the spin angular momenta.

If the spatial degrees of freedom are irrelevant, then the object is described by the $(2j+1)$-dimensional Hilbert space $\mathbb{K}$ with the representation $\boldsymbol{R} \to U(\boldsymbol{R})$ of the rotation group in $\mathbb{K}$. In particular for $j = \frac{1}{2}$ we arrive at the statistical model of the spin-$\frac{1}{2}$ particle described in Section 1.5. Let us confirm the expression (1.5.23) for $\Pr\{\boldsymbol{\theta}_{\rm out}|\boldsymbol{\theta}_{\rm in}\}$ by calculation based on the representation theory. Let $S = |{\rm in})({\rm in}|$ be the density operator in the two-dimensional space describing the state of the particle after it has passed the first filter and $X = |{\rm out})({\rm out}|$ be the test corresponding to the second filter. If φ is the angle between the directions of the two filters, then up to an irrelevant factor the vector $|{\rm out})$ is $U(\boldsymbol{R}_{\boldsymbol{n},\varphi})|{\rm in})$, where $\boldsymbol{n}$ is an appropriate axis. Taking $({\rm in}| = [1, 0]$ we obtain from (3.11.115)

$$\mathbf{Pr}\{\boldsymbol{\theta}_{\rm out}|\boldsymbol{\theta}_{\rm in}\} = \operatorname{Tr} SX = |({\rm in}|{\rm out})|^2 = \cos^2\frac{\varphi}{2},$$

as required.

Quantum mechanics makes possible the explanation of the Stern-Gerlach phenomenon. In the absence of the magnetic field there are no privileged directions and any state in $\mathbb{K}$ corresponds to one and the same energy value ϵ_0 determined by spatial degrees of freedom. Introduction of the magnetic field breaks the symmetry; the Hamiltonian describing the spin degrees of freedom in the presence of the magnetic field $\boldsymbol{B} = [B_1, B_2, B_3]$ can be shown to be

$$H = -\lambda(\sigma_1 B_1 + \sigma_2 B_2 + \sigma_2 B_3).$$

As expected it is invariant under rotations around the direction of $\boldsymbol{B}$. We can take $\boldsymbol{B} = [0, 0, B]$, then $H = -\lambda B\sigma_3$ and therefore H has two eigenvalues $\pm\lambda B$. Thus instead of states with equal energy ϵ_0 there are two states with the energies $\epsilon_0 \pm \lambda B$. Particles with the different energy values are deflected diversely by the anisotropic magnetic field which results in splitting of the beam in the Stern-Gerlach experiment.

This explanation gives an extremely simplified idea of how the quantum mechanics can explain the structure of energy spectra based on symmetry properties. Consideration of more complicated models requires

more knowledge in the representation theory and approximate methods of quantum mechanics and is out of the scope of our book.

3.12. Measuring the angle of rotation

If the apparatus preparing quantum state S is turned around an axis through an angle φ the new state will be

$$S_\varphi = \mathrm{e}^{-\mathrm{i}\varphi L} S \mathrm{e}^{\mathrm{i}\varphi L}, \tag{3.12.119}$$

where L is the infinitesimal generator of the group $V_\varphi = \exp(-\mathrm{i}\varphi L)$, $\varphi \in \mathbb{R}$, constituting the unitary representation of the group of rotations around the given axis. Since rotations through angles differing by $2\pi k$ result in the same position of the apparatus, we may assume that the parameter φ – the angle of rotation – varies in $[0, 2\pi)$ and the relevant shifts group is the group $\mathbb{T}$ as for the phase parameter. Having in mind to describe measurements of the angle of rotation consider the covariance condition

$$V_\varphi^* M(B) V_\varphi = M(B_{-\varphi}); \qquad \varphi \in [0, 2\pi), \quad B \in \mathscr{A}([0, 2\pi)) \tag{3.12.120}$$

where $\{M(B)\}$ is a measurement with values in the interval $[0, 2\pi)$, $B_{-\varphi}$ is the shift of the set B modulo 2π. We are going to describe a canonical solution of this relation.

Consider first the case of zero spin. Passing to the spherical coordinates

$$\begin{aligned} &\xi_1 = \rho \sin\theta \cos\varphi, && \xi_2 = \rho \sin\theta \sin\varphi, && \xi_3 = \rho \cos\theta; \\ &\rho \geq 0, && 0 \leq \varphi < 2\pi, && 0 \leq \theta < \pi, \end{aligned}$$

and letting $\psi(\boldsymbol{\xi}) = \psi(\varphi, \theta, \rho)$ we get

$$\|\psi\|^2 = \iiint |\psi(\varphi, \theta, \rho)|^2 \mathrm{d}\varphi (\sin\theta \, \mathrm{d}\theta)(\rho^2 \, \mathrm{d}\rho).$$

It follows that $\mathscr{H} = \mathscr{H}_\varphi \otimes \mathscr{H}_\theta \otimes \mathscr{H}_\rho$, where $\mathscr{H}_\varphi = \mathscr{L}^2([0, 2\pi))$ and $\mathscr{H}_\theta$, $\mathscr{H}_\rho$ are Hilbert spaces of functions of θ and ρ correspondingly. The operator $V_{\varphi'}$ act only on the variable φ

$$V_{\varphi'} \psi(\varphi, \theta, \rho) = \psi(\varphi - \varphi', \theta, \rho), \tag{3.12.121}$$

therefore we can restrict our considerations to $\mathscr{H}_\varphi = \mathscr{L}^2([0, 2\pi))$. We may call it the *angular representation*. Consider in $\mathscr{H}_\varphi$ the operator Φ of multiplication by φ. According to Section 2.3 this is a Hermitean operator with the spectral representation of the type (2.3.31)

$$\Phi = \int_0^{2\pi} \varphi E(\mathrm{d}\varphi),$$

with the orthogonal spectral measure

$$E(B) = \mathbf{1}_B(\varphi); \qquad B \in \mathscr{A}([0, 2\pi)), \tag{3.12.122}$$

or, symbolically, $E(\mathrm{d}\varphi) = |\varphi)(\varphi|\mathrm{d}\varphi$. A straightforward verification shows that $\{E(B)\}$ satisfies the covariance requirement (3.12.120). As we shall show in Chapter 4 the measurement (3.12.122) can be singled out from all covariant measurements of the angle of rotation by an optimality property. We call it the *canonical measurement of the angle of rotation* and Φ the *canonical angle observable*. Returning to Cartesian coordinates and taking e_3 for the rotation axis we can write (3.12.122) in the form

$$E(B) = \mathbf{1}_{K(B)}(\xi_1, \xi_2, \xi_3),$$

where $K(B)$ is the wedge in the coordinate space defined by

$$K(B) = \{[\xi_1, \xi_2, \xi_3] : \xi_1 = \rho\cos\varphi, \quad \xi_2 = \rho\sin\varphi; \quad 0 \le \rho; \quad \varphi \in B\}.$$

It is instructive to pass to the angular momentum representation which diagonalizes the operator L. In $\mathscr{H}_\varphi = \mathscr{L}^2([0, 2\pi))$ consider the orthonormal basis $\{|m); m = 0, \pm 1, \ldots\}$ of functions $(2\pi)^{-1/2}\exp(\mathrm{i}m\varphi)$. Then the action of operators $V_{\varphi'}$ according to (3.12.121) is given by

$$V_\varphi|m) = \mathrm{e}^{-\mathrm{i}m\varphi'}|m). \tag{3.12.123}$$

It follows that L is the diagonal operator

$$L|m) = m|m); \qquad m = 0, \pm 1, \ldots$$

with $\mathscr{D}(L) = \{\psi : |\sum_m m^2|(\psi|m)|^2 < \infty\}$. The angle measurement (3.12.122) is given by the matrix elements

$$(m|E(B)|m') = \int_B \mathrm{e}^{\mathrm{i}(m'-m)\varphi}\frac{\mathrm{d}\varphi}{2\pi}; \qquad B \in \mathscr{A}([0, 2\pi)),$$

or symbolically

$$(m|E(\mathrm{d}\varphi)|m') = \mathrm{e}^{\mathrm{i}(m'-m)\varphi}\frac{\mathrm{d}\varphi}{2\pi}; \qquad m, m' = 0, \pm 1, \ldots \tag{3.12.124}$$

This is analogous to the expression (3.9.93) for the canonical phase measurement, the role of N is now played by $-L$. Mathematically the difference is that the eigenvalues of N are nonnegative integral numbers while for L they extend to all integers. This stipulates the fact that the canonical phase measurement is not simple contrary to the measurement of the angle.

Returning to $\mathscr{H}_\varphi$ we find that L is the operator $\mathrm{i}^{-1}\mathrm{d}/\mathrm{d}\varphi$ (which agrees with (3.11.118)) with $\mathscr{D}(L) = \{\psi : \psi(0) = \psi(2\pi), \int |\psi'(\varphi)|^2 \mathrm{d}\varphi < \infty\}$. Therefore L and Φ satisfy the Heisenberg type commutation relation $[\Phi, L] = \mathrm{i}I$. However in spite of boundedness of Φ this holds not on $\mathscr{D}(L)$ but on the smaller subspace $\mathscr{D}_0(L) = \{\psi : \psi(0) = \psi(2\pi) = 0, \int |\psi'(\varphi)|^2 \mathrm{d}\varphi < \infty\}$ since $\mathscr{D}(L)$ is not an invariant subspace of Φ. Introducing the unitary operator $U = \exp \mathrm{i}\Phi = \int \exp(\mathrm{i}\varphi) E(\mathrm{d}\varphi)$ we obtain from (3.12.124)

$$U = \sum_{m=-\infty}^{\infty} |m)(m-1|, \tag{3.12.125}$$

so that U is the analog of P^* for the phase. Introducing the discrete group of unitary operators U^m; $m = 0, \pm 1, \ldots$ we get from (3.12.123) the relations

$$\begin{aligned} &V_\varphi^* U^m V_\varphi = \mathrm{e}^{\mathrm{i}m\varphi} U^m, \quad U^m V_\varphi U^{-m} = \mathrm{e}^{\mathrm{i}m\varphi} V_\varphi; \\ &0 \le \varphi < 2\pi; \qquad m = 0, \pm 1, \ldots, \end{aligned} \tag{3.12.126}$$

expressing a form of conjugateness between the angle and the angular momentum.

We now turn to the case of non-zero spin j and restrict ourselves to the case where only spin degrees of freedom are involved. This is so if, *e.g.*, all the components of the vector $\boldsymbol{\psi}(\boldsymbol{\xi})$ in (3.11.117) are spherically invariant functions of ξ. Thus we adopt that states are described by density matrices in $(2j+1)$-dimensional space $\mathbb{K}$. The rotation through an angle φ results in the state change

$$S_\varphi = \mathrm{e}^{-\mathrm{i}\varphi J} S \mathrm{e}^{\mathrm{i}\varphi J}, \tag{3.12.127}$$

where J is the spin angular momentum about the axis of rotation. Put $V_{\varphi'} = \exp(-\mathrm{i}\varphi' J)$ and consider the covariance condition (3.12.120).

Let $\{|m); m = -j, -j+1, \ldots, j\}$ be the orthonormal basis in which J has the diagonal form (3.11.116), then $V_{\varphi'}$ acts on $|m)$ as in (3.12.123) with the only difference that m varies now in the finite range. The relation

$$(m|M(\mathrm{d}\varphi)|m') = \mathrm{e}^{\mathrm{i}(m-m')\varphi} \frac{\mathrm{d}\varphi}{2\pi}; \quad m, m' = -j, -j+1, \ldots, j \tag{3.12.128}$$

defines a covariant measurement of the angle of rotation which we call the *canonical angle measurement* for the spin degrees of freedom. Contrary to the case of spatial degrees of freedom considered above this measurement is not simple. To see it introduce the operators

$$E_\pm = \int_0^{2\pi} \mathrm{e}^{\pm \mathrm{i}\varphi} M(\mathrm{d}\varphi). \tag{3.12.129}$$

If $M(\mathrm{d}\varphi)$ were an orthogonal resolution of identity, then $E_\pm$ would be mutually adjoint unitary operators. From (3.12.128) we get

$$E_+ = \sum_{m=-j+1}^{j} |m)(m-1|, \qquad E_- = \sum_{m=-j+1}^{j} |m-1)(m|, \quad (3.12.130)$$

so that E_+ is analogous to the operator $U = \exp(\mathrm{i}\Phi)$ for the spatial degrees of freedom. However

$$E_-E_+ = I - |j)(j|, \qquad E_+E_- = I - |-j)(-j|, \qquad (3.12.131)$$

so that $E_\pm$ are not unitary, nor even isometric. From (3.12.123) and (3.12.130) we get the analog of the relations (3.12.126)

$$\begin{aligned} &V_\varphi^* E_+^m V_\varphi = \mathrm{e}^{\mathrm{i}m\varphi} E_+^m, \quad V_\varphi^* E_-^m V_\varphi = \mathrm{e}^{-\mathrm{i}m\varphi} E_-^m; \\ &0 \le \varphi < 2\pi, \quad m = -j, -j+1, \ldots, j. \end{aligned} \qquad (3.12.132)$$

Taking $m = 1$ and differentiating with respect to φ we obtain the useful commutation relations

$$[E_+, J] = -E_+, \qquad [E_-, J] = E_-. \qquad (3.12.133)$$

3.13. Comments

Section 3.1. The importance of symmetry groups in quantum mechanics is well known (see, *e.g.*, Weyl [141], Wigner [145]). The idea of classifying quantum systems according to representations of a symmetry group belongs to Wigner [144], who applied it first to the relativistic quantum mechanics corresponding to the Poincaré (or inhomogeneous Lorentz) group; in this connection see also Bogoljubov, Logunov and Todorov [16], Gelfand, Minlos and Shapiro [41], Shirokov [124] and Varadarajan [136]. The classification of nonrelativistic quantum "particles" according to the representations of the Galilei group was given by Inönu and Wigner [70] and Bargmann [6]. In the last paper the general theory of projective unitary representations was developed. The decomposition of arbitrary continuous representation into unitary ones takes place for the type I groups [75]. A good introduction to the advanced applications of the group methods in the theory of elementary particles can be found in Bogoljubov's lectures [15]. Of special importance to quantum mechanics are dynamical symmetries, *i.e.*, symmetries of the Hamiltonian related to conservation laws, see, *e.g.*, Malkin and Man'ko [93].

Proofs and discussion of the Wigner theorem and its generalizations can be found in Bargmann [8], Hunziker [68], Kadison [74], Varadarajan [135], Davies [27].

There is an approach to description of unstable quantum particles using Galilei and Poincaré semigroups, which differ from corresponding groups in that they include only positive time shifts (*cf.* in particular Zwanziger [149], Lanz, Lugiato and Ramella [84]). The restriction to positive time shifts has the motivation that a preparation procedure cannot be arbitrarily delayed without interfering with a subsequent measurement.

Section 3.2. For the proof of Proposition 3.2.1 see, *e.g.*, Bargmann [6] and Varadarajan [136]. A survey of the results concerning the Weyl CCR and its relations to (3.2.20) can be found in Putnam [116].

The Mandelstam-Tamm inequality was derived in [96] for the case of the time parameter t. To obtain the time-energy uncertainty relation the authors introduce "standard time" Δt defined by

$$E_{t+\Delta t}(X) - E_t(X) = (\Delta t)^{-1} \int_t^{t+\Delta t} D_\tau(X) \mathrm{d}\tau.$$

Integrating the inequality then gives $\Delta t \cdot \Delta H \geq \frac{1}{2}$ where $\Delta H = \sqrt{D_t(H)}$. However Krylov and Fock [83] pointed out that the resulting inequality cannot be considered as an analog of the Heisenberg coordinate-momentum uncertainty relation; in the definition of Δt the quantities $E_\tau(X)$, $D_\tau(X)$; $t \leq \tau \leq t + \Delta t$ cannot be obtained from a single "statistical ensemble" since measuring X at a time t changes the state and violates the Hamiltonian evolution. This discussion gives an idea of what kind of problems the traditional quantum measurement theory is faced with. The statistical approach from the point of view of quantum estimation theory seems to be free from these difficulties. This approach was first suggested by Helstrom [52, 53] who independently rederived the Mandelstam-Tamm inequality in the context of quantum estimation theory. The more general presentation given here follows Holevo [66].

Section 3.3. Proposition 3.3.1 is a particular case of the theorem describing the general form of a multiplier $\omega(g_1, g_2)$ on the additive group of $\mathbb{R}^d$ (see Bargmann [6], Jauch [71] and Varadarajan [136]). The kinematical meaning of the observables P, Q is discussed in Jauch [71].

Section 3.4. The original proof of the uniqueness theorem was given in von Neumann's paper [138]. This can also be considered as a particular case of Mackey's imprimitivity theorem [91] (see Jauch [71]).

Section 3.5. Minimum-uncertainty states were first introduced by Schrödinger; the completeness relation was obtained by Bargmann [7] and Glauber [42].

Section 3.6. The problem of joint measurement has been considered from different points of view by von Neumann [138], Urbanik [133], Gordon and Louisell [44], She and Heffner [123], Prugovečki [115], Davies [27] and others. The present treatment follows Holevo [58, 61] and Helstrom [53].

Section 3.7. The equation initially discovered by Schrödinger corresponds to the eigenvalue problem $H\psi = \lambda\psi$. Schrödinger came to it attempting to write the equation for the de Broglie stationary "waves of matter". The correspondence principle was introduced as a formal "postulate" prescribing quantum analogs for classical quantities. The fact that it follows from the more fundamental Galilean relativity was realized later (Inönu and Wigner [70], Bargmann [6]). This section follows closely to Jauch [71].

We shall mention here very briefly two interesting connections between quantum mechanics and the probabilistic theory of stochastic processes. Replacing formally it by t transforms the Schrödinger equation into the parabolic equation of stochastic diffusion processes theory, the group $\{V_t\}$ of unitary operators being replaced by a contraction semigroup. There is the important Feynmann-Kac formula relating quantum dynamics with a stochastic diffusion process. Nelson [102, 103], who gave it a simple proof based on the Trotter product formula, considered the stochastic diffusion process as a hidden variables model for quantum dynamics. Use of the Feynmann-Kac formula just as an analytic tool opens interesting possibilities both for quantum mechanics and for stochastic processes (see Simon [128]).

The other connection reveals in the recent work on quantum stochastic processes allowing to describe the irreversible (in particular, Markovian) evolution of "open" quantum systems. For this see in particular Davies [27], Accardi, Frigerio and Lewis [1] and Gorini *et al*. [45].

Section 3.8. Difficulties with the definition of time observable were discussed in great detail by Allcock [3], where a reader can find other references. Attempts to use the expressions like $\mathrm{i}\hbar\,\mathrm{d}/\mathrm{d}\epsilon$ go back to Pauli [107] however there was no room for them in the conventional form of quantum measurement theory. The presentation here follows Holevo [66] where the case of three dimensional free particle was also treated. For other approaches to the time-energy uncertainty relation see Ekstein and Siegert [32], Wigner [146] and Malkin and Man'ko [93].

Section 3.9. Concerning mathematical problems of quantum mechanics see, *e.g*., Reed and Simon [118]. The quantum oscillator is considered in almost all textbooks on quantum mechanics; we follow the original

presentation of Dirac (see also Louisell [88] for interesting formal algebra with creation-annihilation operators).

A correct approach to phase of quantum oscillator via the operator P^* was suggested by Sussking and Glogower [131] (see also Carruthers and Nieto [20], Volkin [137]). The canonical phase measurement appeared in Holevo [67]. For Hardy classes see, *e.g.*, Halmos [48].

The phase observable, similarly to the time observable is an "eternal" topic in foundations of quantum mechanics. One of the most known more recent suggestions – the Pegg-Barnett "selfadjoint phase operator" [193] – is interesting in that it gives a concrete example of approximation of the nonorthogonal resolution of the identity (3.9.93) in the infinite-dimensional Hilbert space by orthogonal ones. The fact that orthogonal resolutions of the identity are weakly dense in the set of all resolutions of the identity in the infinite-dimensional Hilbert space was observed already in Naimark's paper [100] (for more detail see [178], §2.1.3.)

Consider the Hilbert space $\mathcal{H}$ with the basis $\{|n);\ n = 0, 1, \dots\}$, and its finite-dimensional subspaces $\mathcal{H}_N$, generated by the vectors $\{|n);\ n = 0, 1, \dots, N\}$. In $\mathcal{H}_N$ the orthonormal basis of Pegg-Barnett "phase vectors" is introduced

$$|\theta_{N,k}) = \frac{1}{\sqrt{N+1}} \sum_{n=0}^{N} \mathrm{e}^{in\theta_{N,k}} |n), \quad \theta_{N,k} = \frac{2\pi k}{N+1}; \quad k = 0, 1, \dots, N,$$

with the corresponding orthogonal resolution of the identity

$$M_N(B) = \sum_{\theta_{N,k} \in B} |\theta_{N,k})(\theta_{N,k}|.$$

One can then define selfadjoint "phase operator"

$$\Theta_N = \sum_{k=0}^{N} \theta_{N,k} |\theta_{N,k})(\theta_{N,k}|$$

in $\mathcal{H}_N$ or better, the unitary operator $\mathrm{e}^{i\Theta_N} = \sum_{k=0}^{N} \mathrm{e}^{i\theta_{N,k}} |\theta_{N,k})(\theta_{N,k}|$. If $|\psi)$ is arbitrary vector in $\mathcal{H}$ with finitely many nonzero coefficients $(n|\psi)$, then for N large enough

$$\mu_{S_\psi}^{(N)}(B) = (\psi|M_N(B)|\psi) = \sum_{\theta_{N,k} \in B} |(\theta_{N,k}|\psi)|^2$$

$$= \sum_{\theta_{N,k} \in B} \frac{1}{N+1} \left| \sum_{n=0}^{N} \mathrm{e}^{-in\theta_{N,k}} (n|\psi) \right|^2$$

are probability measures which weakly converge to the distribution of the canonical phase measurement (3.9.93):

$$\mu_{S_\psi}(B) = (\psi|M(B)|\psi) = \frac{1}{2\pi}\int_B \left|\sum_{n=0}^{\infty} \mathrm{e}^{-in\theta}(n|\psi)\right|^2 d\theta$$

in the pure state $S_\psi = |\psi)(\psi|$. Recall that weak convergence of probability measures means convergence of integrals of all bounded continuous functions. Similar statement holds also for mixed states.

Section 3.10. The coherent-state representation for many degrees of freedom was considered by Bargmann [7], Glauber [42], Klauder and Sudarshan [76]. For the free field (infinitely many degrees of freedom) an analogous representation was studied by Berezin [12].

Bounded subnormal operators were introduced by Halmos [48]. See also Sz.-Nagy [119].

Section 3.11. On the representations of the rotation group in quantum mechanics see Wigner [145], Jauch [71] and Mackey [92]. For a systematic study of the rotation group see Gelfand, Minlos and Shapiro [41] and Želobenko [148].

Section 3.12. A consideration of the canonical angle observable for the spin-zero case in the angular representation was given by Judge [73], Susskind and Glogower [131] (see also Carruthers and Nieto [20] and the references quoted therein). For the treatment in the angular momentum representation in the case of spin degrees of freedom see Helstrom [52] and Holevo [67].

Chapter 4
Covariant measurements and optimality

4.1. Parametric symmetry groups and covariant measurements

All symmetry groups considered in the previous chapter were *parametric groups* (Lie groups) of transformations. This means that a parametric set $\Theta = \{\theta\}$, *i.e.*, a continuous manifold in a finite-dimensional space is given and elements of the group $G = \{g\}$ act as continuous one-to-one mappings of the set Θ onto itself, $g : \theta \to g\theta$. Moreover the group G is itself parametrized in such a way that the group product $g_1 g_2$ is at least locally continuous in g_1, g_2.

The examples are the additive group $\mathbb{R}$ considered as the shift group for the real line $\Theta = \mathbb{R}$, the group $\mathbb{T}$ of shifts (mod 2π) of the interval $\Theta = [0, 2\pi)$ and the rotation group in the three dimensional Euclidean space $\mathbb{R}^3$. In the last case it is natural to consider the action of the group only on directions, *i.e.*, unit vectors in $\mathbb{R}^3$. Then the rotation group becomes a group of transformations of the unit sphere $\Theta = \mathbb{S}^2$.

Let G be a parametric group of transformations of a set Θ and $g \to V_g$ be a (continuous) projective unitary representation of G in a Hilbert space $\mathscr{H}$. Let $M(\mathrm{d}\theta)$ be a measurement with values in Θ, *i.e.*, a resolution of identity in $\mathscr{H}$ on the σ-field $\mathscr{A}(\Theta)$ of Borel subsets of Θ[1]. The measurement $M(\mathrm{d}\theta)$ is *covariant* with respect to representation $g \to V_g$ if

$$V_g^* M(B) V_g = M(B_{g^{-1}}), \qquad g \in G, \tag{4.1.1}$$

for any $B \in \mathscr{A}(\Theta)$, where

$$B_g = \{\theta : \theta = g\theta', \theta' \in B\}$$

is the image of the set B under the transformation g. The notion of covariant measurement was introduced in the previous chapter for con-

[1] Since Θ is a continuous manifold it is itself a Borel subset in a finite-dimensional space.

crete symmetry groups. Here we wish to study it from a general point of view.

The importance of this notion for quantum theory lies in the fact that it establishes a correspondence between physical parameters and certain classes of quantum measurements. Indeed, assume that θ is a parameter (in general, multidimensional) describing some aspects of the preparation procedure and S the basic state corresponding to the value θ_0. Then the transformation g results in the preparation of the new state $S_\theta = V_g S V_g^*$, where $\theta = g\theta_0$. If the measurement $M(\mathrm{d}\hat{\theta})$ is then made, the probability distribution of the results $\hat{\theta}$ of the measurement will be

$$\mathbf{Pr}\{\hat{\theta} \in B|\theta\} = \operatorname{Tr} S_\theta M(B), \qquad B \in \mathscr{A}(\Theta).$$

If $M(\mathrm{d}\hat{\theta})$ possesses the covariance property (4.1.1), then

$$\operatorname{Tr} S_\theta M(B) = \operatorname{Tr} S V_g^* M(B) V_g = \operatorname{Tr} S M(B_{g^{-1}})$$

whence replacing B by B_g

$$\Pr\{\hat{\theta} \in B_g|g\theta_0\} = \Pr\{\hat{\theta} \in B|\theta_0\}.$$

Thus the change of the value of θ is properly reflected by the change in the resulting probability distribution, and therefore any resolution of identity $M(\mathrm{d}\theta)$ satisfying (4.1.1) corresponds to a theoretically admissible measurement of the parameter θ.

A mathematical problem which naturally arises is to describe all covariant measurements of the given parameter θ. We shall give its general solution in Section 4.2 and 4.8. Then we shall look for "optimal" measurements having the best theoretically possible accuracy among all covariant measurements of the parameter θ. In this way we shall find that the "canonical" measurements for various parameters introduced in the previous chapter are just the typical representatives in the family of the optimal covariant measurements.

We shall need some general knowledge of parametric groups of transformations. The group G *acts transitively* on Θ if any point θ_0 can be transformed into any other point θ by some $g \in G$. In what follows we assume that a parametric group G acts transitively on Θ. Then the continuous mapping $g \to g\theta_0 = \theta(g)$ maps G onto the whole Θ. This mapping is one-to-one if and only if the *stationary subgroup* G_0 of transformations leaving the point θ_0 invariant is trivial, *i.e.*, reduces to the identical transformation. For an example consider the shift group of $\mathbb{R}$ and fix a point $\theta_0 \in \mathbb{R}$. Any point $\theta \in \mathbb{R}$ can be obtained from θ_0 by a shift: $\theta = \theta_0 + x$. The mapping $x \to \theta_0 + x = \theta(x)$ is obviously one-to-one. The same is

true for the group $\mathbb{T}$ acting on $[0, 2\pi)$. Consider now the rotation group acting on $\mathbb{S}^2$. Fix the "pole" $\boldsymbol{n}_0 \in \mathbb{S}^2$. Any direction $\boldsymbol{n}$ can be obtained from $\boldsymbol{n}_0$ by a rotation $R : \boldsymbol{n} = R\boldsymbol{n}_0$; the mapping $R \to R\boldsymbol{n}_0 = \boldsymbol{n}(R)$ is not one-to-one since $n(R) = n(RR_0)$ where R_0 is any rotation about the axis n_0.

A measure $\mu(\mathrm{d}g)$ on the σ-field $\mathscr{A}(G)$ of Borel subsets of G is called *left-*(*right-*)*invariant* if

$$\mu(gA) = \mu(A) \qquad (\text{correspondingly } \mu(Ag) = \mu(A)), \quad g \in G,$$

where $gA = \{g' : g' = g'', g'' \in A\}$ and $Ag = \{g' : g' = g''g, g'' \in A\}$ $A \in \mathscr{A}(G)$.

It is known that any parametric group possesses a left-invariant measure. If there exists an *invariant* (*i.e.*, both left- and right-) invariant measure, then the group is called *unimodular*. The invariant measure is in general not finite, *i.e.*, $\mu(G) \leq +\infty$. If the group is compact (*i.e.*, is a bounded closed manifold in a finite-dimensional space), then any left- or right-invariant measure is invariant. Moreover it is finite and we shall normalize it by $\mu(G) = 1$. We shall deal only with unimodular groups.

Now we are interested in measures ν on $\mathscr{A}(\Theta)$ with are *invariant* in the sense that

$$\nu(B_g) = \nu(B); \qquad g \in G, \quad B \in \mathscr{A}(\Theta).$$

If G_0 is also unimodular, then such ν exists. *If moreover G_0 is compact*, which we assume, *then ν can be explicitly constructed from μ by the relation*

$$\nu(B) = \mu(\theta^{-1}(B)),$$

where $\theta^{-1}(B) = \{g : g\theta_0 \in B\}$ is the pre-image of the Borel set $B \in \mathscr{A}(\Theta)$. In other words, ν is defined by the requirement that

$$\int_G f(g\theta_0)\mu(\mathrm{d}g) = \int_\Theta f(\theta)\nu(\mathrm{d}\theta) \tag{4.1.2}$$

holds for any integrable function f on Θ. The compactness of G_0 ensures the finiteness of the measure ν (otherwise it may be identically infinite as the simple example $G = \mathbb{R}^2$, $\Theta = \mathbb{R}$, $G_0 = \mathbb{R}$ shows). The invariance of ν follows from the left invariance of μ. Right invariance of μ implies that ν *is the same for any choice of* θ_0, since if $\theta_1 = g_1\theta_0$, then

$$\begin{aligned}\nu(B) &= \mu(\{g : g\theta_0 \in B\}) \\ &= \mu(\{gg_1 : gg_1\theta_0 \in B\}) = \mu(\{g : g\theta_1 \in B\}).\end{aligned}$$

An example of invariant measure is the Lebesgue measure on $\mathbb{R}$. For the rotation group acting on $\mathbb{S}^2$ the invariant measure ν on $\mathbb{S}^2$ is up to a factor the Euclidean area on the unit sphere. The expression for the invariant measure μ on the rotation group will appear in Section 4.10.

4.2. Structure of covariant measurements

The following simple result will be useful in the study of covariant measurements.

Proposition 4.2.1. *Let $M(\mathrm{d}\theta)$ be a measurement covariant with respect to a projective unitary representation $g \to V_g$ of the parametric group G of transformations of the set Θ. For any density operator S in the Hilbert space of the representation and for any Borel $B \in \mathscr{A}(\Theta)$*

$$\int_G \operatorname{Tr} V_g S V_g^* M(B)\mu(\mathrm{d}g) = \nu(B). \tag{4.2.3}$$

Proof. By (4.1.1) we get

$$\begin{aligned}\int_G \operatorname{Tr} V_g S V_g^* M(B)\mu(\mathrm{d}g) &= \int_G \operatorname{Tr} S M(B_{g^{-1}})\mu(\mathrm{d}g) \\ &= \int_G \int_\Theta \mathbf{1}_B(g\theta_0)\mu_S(\mathrm{d}\theta_0)\mu(\mathrm{d}g),\end{aligned}$$

where $\mu_S(\mathrm{d}\theta) = \operatorname{Tr} S M(\mathrm{d}\theta)$. Using (4.2.3) and the fact that ν does not depend on the choice of θ_0 we obtain that it is equal to

$$\int_\Theta \mu_S(\mathrm{d}\theta_0) \int_G \mathbf{1}_B(g\theta_0)\mu(\mathrm{d}g) = \nu(B). \tag{4.2.4}$$

□

We shall also need a Radon-Nikodym theorem for operator-valued measures. The following proposition the proof of which is standard will suffice.

Proposition 4.2.2. *Let $\{M(B);\ B \in \mathscr{A}(\Theta)\}$ be an additive operator-valued set function dominated by the scalar measure $\{m(B);\ B \in \mathscr{A}(\Theta)\}$ in the sense that*

$$|(\varphi|M(B)\psi)| \le m(B)\|\varphi\|\,\|\psi\|, \qquad B \in \mathscr{A}(\Theta),$$

for all $\varphi, \psi \in \mathscr{H}$. Then there exists an operator-valued function $P(\cdot)$ defined uniquely for m-almost all $\theta \in \Theta$ (i.e., for all θ with possible exception of a set of zero m measure), satisfying $\|P(\theta)\| \le 1$ and such that

$$(\varphi|M(B)\psi) = \int_B (\varphi|P(\theta)\psi)m(\mathrm{d}\theta), \qquad B \in \mathscr{A}(\Theta),$$

for all $\varphi, \psi \in \mathscr{H}$. *If* $M(B) \geq 0$ *for all* $B \in \mathscr{A}(\Theta)$, *then* $P(\theta) \geq 0$ *for* m*-almost all* θ.

The function $P(\theta)$ is called the *operator density* of $M(\mathrm{d}\theta)$ with respect to $m(\mathrm{d}\theta)$. We shall write simply $M(B) = \int_B P(\theta)m(\mathrm{d}\theta)$ or $M(\mathrm{d}\theta) = P(\theta)m(\mathrm{d}\theta)$ having in mind the weak convergence of the integral.

To avoid certain technical difficulties we now restrict to finite-dimensional representations $g \to V_g$.

Theorem 4.2.3. *Let* P_0 *be a Hermitean positive operator in the representation space, commuting with the operators* $\{V_g; g \in G_0\}$ *and satisfying*

$$\int_G V_g P_0 V_g^* \mu(\mathrm{d}g) = I. \tag{4.2.5}$$

Then putting

$$P(g\theta_0) = V_g P_0 V_g^*, \tag{4.2.6}$$

we get an operator-valued function of θ *such that*

$$M(B) = \int_B P(\theta)\nu(\mathrm{d}\theta), \qquad B \in \mathscr{A}(\Theta), \tag{4.2.7}$$

is a measurement, covariant with respect to the representation $g \to V_g$.

Conversely for any covariant measurement $M(\mathrm{d}\theta)$ *there is a unique operator* P_0 *satisfying the above conditions such that* $M(B)$ *is expressed through* P_0 *by* (4.2.6) *and* (4.2.7).

Proof. By the condition $[P_0, V_g] = 0$, $g \in G_0$, we have $P(g_1\theta_0) = P(g_2\theta_0)$ if $g_1\theta_0 = g_2\theta_0$ and therefore the relation (4.2.6) defines unambiguously an operator-valued function of $\theta \in \Theta$. Positivity of P_0 implies $M(B) \geq 0$ and σ-additivity is a property of the definite integral (4.2.7). Using (4.1.2) we have

$$\int_\Theta P(\theta)\nu(\mathrm{d}\theta) = \int_G V_g P_0 V_g^* \mu(\mathrm{d}g) = I \tag{4.2.8}$$

by (4.2.4) whence $M(\Theta) = I$.

To prove the converse statement let $d = \dim \mathscr{H}$.

Putting $S = d^{-1}I$ in (4.2.3) we get $\operatorname{Tr} M(B) = d^{-1}\nu(B)$, $B \in \mathscr{A}(\Theta)$. It follows that $(\varphi|M(B)\varphi) \leq d^{-1}\nu(B)$ for any unit vector $\varphi \in \mathscr{H}$. By positivity of $M(B)$ and the Cauchy inequality

$$|(\varphi|M(B)\psi)| \leq \sqrt{(\varphi|M(B)\varphi)(\psi|M(B)\psi)} \leq d^{-1}\nu(B)$$

for unit φ, $\psi \in \mathscr{H}$, so that $\{M(B)\}$ is dominated by $\{d^{-1}\nu(B)\}$. By Proposition 4.2.2 there is density $P(\theta)$ which is uniquely defined and positive for ν-almost all θ, such that

$$M(B) = \int_B P(\theta)\nu(\mathrm{d}\theta). \tag{4.2.9}$$

From the covariance relation (4.1.1)

$$\begin{aligned}\int_B V_g^* P(\theta) V_g \nu(\mathrm{d}\theta) &= \int_{B_{g^{-1}}} P(\theta)\nu(\mathrm{d}\theta) \\ &= \int_B P(g^{-1}\theta)\nu(\mathrm{d}\theta),\end{aligned}$$

whence by uniqueness of the operator density

$$V_g^* P(\theta) V_g = P(g^{-1}\theta), \qquad g \in G,$$

for ν-almost all θ. A measure-theoretic argument shows that the density $P(\theta)$ can be defined in such a way that this equality will hold for all $\theta \in \Theta$. Then putting $P_0 = P(\theta_0)$ we get (4.2.6) and (4.2.7); the relation (4.2.4) follows from the normalization $M(\Theta) = I$. □

The relation (4.2.7) sets up the affine one-to-one correspondence between the two convex sets: the set of all covariant measurements $M(\mathrm{d}\theta)$ and the set $\mathfrak{P}$ of Hermitean operators satisfying the conditions of the theorem.

Note that taking $\nu(B)$ small enough, which is possible in the parametric case, we can make $\|M(B)\| < 1$ and then $M(B)$ cannot be a projection. Thus *in the case of a finite-dimensional representation of a parametric symmetry group there are no simple covariant measurements*.

4.3. The covariant quantum estimation problem

Let θ be a parameter (in general, multidimensional) describing certain aspects of the preparation procedure. Then to each value θ from the admissible set Θ corresponds a quantum state S_θ in $\mathscr{H}$. Assume that there is a symmetry group G acting transitively on Θ, which has a representation $g \to V_g$ in $\mathscr{H}$. The family of states $\{S_\theta\}$ is called *covariant* with respect to the representation if

$$S_{g\theta} = V_g S_\theta V_g^*; \qquad \theta \in \Theta, \quad g \in G.$$

Fix "the initial value" θ_0; then it follows that the basic state $S_{\theta_0} = S$ satisfies: $S = V_g S V_g^*$, $g \in G_0$, so that we can write

$$S_\theta = V_g S V_g^*, \qquad (\theta = g\theta_0). \tag{4.3.10}$$

Assume now that the object is prepared in one of the states S_θ, but the actual value of θ is unknown and the problem is to estimate the value as accurately as possible based on measurements admitted by quantum theory. This is an obvious generalization of the quantum estimation problem for the shift parameter discussed in Section 3.2. We shall consider it following the ideas of statistical estimation adapted to quantum theory.

Let us fix a *deviation function* $W_\theta(\hat{\theta})$ which is interpreted as a measure of deviation of an observed value $\hat{\theta}$ from the actual value θ. We assume here that it is a continuous function of the arguments satisfying $W_\theta(\hat{\theta}) \geq W_\theta(\theta)$. We shall also assume that it is *invariant*

$$W_{g\theta}(g\hat{\theta}) = W_\theta(\hat{\theta}); \qquad g \in G; \quad \theta, \hat{\theta} \in \Theta.$$

For all concrete parametric groups considered here there are some "natural" deviation functions such as the quadratic deviation $(\theta - \hat{\theta})^2$ for $\mathbb{R}$.

The *mean deviation* of the measurement $\boldsymbol{M} = \{M(\mathrm{d}\hat{\theta})\}$ with respect to the actual value θ is

$$\mathscr{R}_\theta\{\boldsymbol{M}\} = \int W_\theta(\hat{\theta})\mu_\theta(\mathrm{d}\hat{\theta}), \tag{4.3.11}$$

where $\mu_\theta(\mathrm{d}\hat{\theta}) = \operatorname{Tr} S_\theta M(\mathrm{d}\hat{\theta})$ is the probability distribution of the measurement $\boldsymbol{M}$ with respect to the state S_θ. Having in mind to find the most accurate measurement of the parameter θ one would like to minimize (4.3.11) for all values of θ. However this is in general impossible, as is well known from the classical mathematical statistics: what is good for one value of θ may be bad for another. To introduce a reasonable concept of optimality one has to compromise by forming a single functional of the quantities $\mathscr{R}_\theta\{\boldsymbol{M}\}$, $\theta \in \Theta$, which would serve as an integral measure of accuracy.

Following the classical estimation theory one can form two different functionals. In Bayes' approach one averages $\mathscr{R}_\theta\{\boldsymbol{M}\}$ with respect to a *prior distribution* $\pi(\mathrm{d}\theta)$. The measurement minimizing the resulting *Bayes mean deviation*

$$\mathscr{R}_\pi\{\boldsymbol{M}\} = \int \mathscr{R}_\theta\{\boldsymbol{M}\}\pi(\mathrm{d}\theta)$$

is called *Bayesian*. The quantity $\mathscr{R}_\pi\{\boldsymbol{M}\}$ represents the total mean deviation in the situation where θ is itself a random parameter with the known

distribution $\pi(\mathrm{d}\theta)$. In particular if Θ is compact and "nothing is known in advance" about θ one may take for $\pi(\mathrm{d}\theta)$ the "uniform distribution", *i.e.*, the normalized invariant measure $\nu(\mathrm{d}\theta)$.

In minimax approach the *maximal mean deviation*

$$\hat{\mathscr{R}}\{\boldsymbol{M}\} = \max_{\theta} \mathscr{R}_\theta\{\boldsymbol{M}\}$$

is subject to minimization. The minimizing measurement is called *minimax*. We shall also call *optimal* measurements minimizing this of the other measure of accuracy.

Restricting ourselves here to the case of compact Θ and G we shall show that for a covariant family of states the minima of both Bayes' and maximal mean deviations coincide and are achieved on a covariant measurement. This is an analogue of the Hunt-Stein theorem in mathematical statistics.

The mean deviation (4.3.11) and therefore the Bayes' mean deviation are affine functionals of measurement $\boldsymbol{M}$. If Θ *is compact and* $W_\theta(\hat{\theta})$ *is continuous, as assumed, then a Bayesian measurement exists*. For the proof of this assertion and for the detailed discussion of the integrability questions in the proof of the following theorem the reader is referred to the comments at the end of this chapter.

Theorem 4.3.1. *In the quantum covariant statistical estimation problem described above the minima of the Bayes' mean deviation* $\mathscr{R}_\nu\{\boldsymbol{M}\}$ *and the maximal mean deviation* $\hat{\mathscr{R}}\{\boldsymbol{M}\}$ *for all* Θ*-measurements are achieved on a covariant measurement. Moreover, for a covariant measurement* $\boldsymbol{M}$

$$\mathscr{R}_\nu\{\boldsymbol{M}\} = \hat{\mathscr{R}}\{\boldsymbol{M}\} = \mathscr{R}_\theta\{\boldsymbol{M}\}, \qquad \theta \in \Theta. \tag{4.3.12}$$

Proof. Introduce the measurement $\boldsymbol{M}_g$ by putting

$$M_g(B) = V_g^* M(B_g) V_g, \qquad B \in \mathscr{A}(\Theta),$$

and note that $\boldsymbol{M}$ is covariant if and only if $\boldsymbol{M}_g = \boldsymbol{M}$, $g \in G$. From the assumed covariance of the family $\{S_\theta\}$ and invariance of $W_\theta(\hat{\theta})$ we get by (4.3.11)

$$\mathscr{R}_\theta\{\boldsymbol{M}_g\} = \mathscr{R}_{g\theta}\{\boldsymbol{M}\}. \tag{4.3.13}$$

In particular for a covariant measurement the mean deviation $\mathscr{R}_\theta\{\boldsymbol{M}\}$ does not depend on θ so that (4.3.12) holds.

Consider the Bayes' mean deviation

$$\mathscr{R}_\nu\{\boldsymbol{M}\} = \int \mathscr{R}_\theta\{\boldsymbol{M}\}\nu(\mathrm{d}\theta).$$

Using (4.3.13) we get $\mathscr{R}_\nu\{\boldsymbol{M}\} = \mathscr{R}_\nu\{\boldsymbol{M}_g\}$. Introduce the "averaged" measurement $\overline{\boldsymbol{M}}$ by

$$\overline{M}(B) = \int_G M_{g^{-1}}(B)\mu(\mathrm{d}g).$$

The integral is easily defined in the sense of the weak convergence and can be verified to determine a covariant measurement. Using (4.3.11) and changing the order of integration we get

$$\mathscr{R}_\nu\{\overline{\boldsymbol{M}}\} = \int \mathscr{R}_\nu\{\boldsymbol{M}_{g^{-1}}\}\mu(\mathrm{d}g) = \mathscr{R}_\nu\{\boldsymbol{M}\}$$

and therefore

$$\max_\theta \mathscr{R}_\theta\{\boldsymbol{M}\} \geq \mathscr{R}_\nu\{\boldsymbol{M}\} = \mathscr{R}_\nu\{\overline{\boldsymbol{M}}\}.$$

Since $\overline{\boldsymbol{M}}$ is covariant, then using (4.3.12)

$$\mathscr{R}_\nu\{\overline{\boldsymbol{M}}\} = \mathscr{R}_\theta\{\overline{\boldsymbol{M}}\} = \max_\theta \mathscr{R}_\theta\{\overline{\boldsymbol{M}}\}.$$

Thus for an arbitrary measurement $\boldsymbol{M}$ we have constructed a covariant measurement $\overline{\boldsymbol{M}}$ with the same value of the Bayes' mean deviation and possibly smaller value of the maximal mean deviation. Therefore the minima of both functionals are achieved on a covariant measurement. □

The relation (4.3.12) shows that the "naive" approach of minimizing the mean deviation for all values of θ simultaneously makes sense if one restricts to covariant measurements. Moreover it is equivalent both to Bayes' and minimax approaches and reduces to the minimization of the mean deviation

$$\mathscr{R}_{\theta_0}\{\boldsymbol{M}\} = \int W_{\theta_0}(\theta)\mu_{\theta_0}(\mathrm{d}\theta)$$

for some particular value θ_0.

Consider this problem assuming $\dim \mathscr{H} < \infty$. Then we are in the conditions of Theorem 4.2.3 and the functional $\mathscr{R}_{\theta_0}\{\boldsymbol{M}\}$ takes the form

$$\begin{aligned}\mathscr{R}_{\theta_0}\{\boldsymbol{M}\} &= \int W_{\theta_0}(\theta)\,\mathrm{Tr}\, V_g^* S V_g P_0 \nu(\mathrm{d}\theta) \\ &= \mathrm{Tr}\, \widehat{W}_0 P_0,\end{aligned}$$

where

$$\widehat{W}_0 = \int_\Theta W_{\theta_0}(\theta) V_g^* S V_g \nu(\mathrm{d}\theta) = \int_G W_{\theta_0}(g\theta_0) V_g^* S V_g \mu(\mathrm{d}g) \qquad (4.3.14)$$

is an operator commuting with all $\{V_g;\, g \in G_0\}$. Based on an analogy with mathematical statistics we can call $\widehat{W}_0$ the *operator of posterior deviation*. Thus one has to find

$$\min \operatorname{Tr} \widehat{W}_0 P_0 \tag{4.3.15}$$

over all Hermitean $P_0 \in \mathfrak{P}$. This problem will be effectively solved for concrete symmetry groups in the following sections.

Finally we shall briefly discuss the noncommutative analog of the classical *maximum-likelihood approach*. Formally it corresponds to Bayes' approach with the uniform prior distribution and the "deviation function"

$$W_{\theta_0}(\theta) = \begin{cases} 0, & \theta \neq \theta_0, \\ -\infty, & \theta = \theta_0. \end{cases}$$

More precisely define the delta-function on Θ by the relation

$$\int \delta_{\theta_0}(\theta) f(\theta) \nu(\mathrm{d}\theta) = f(\theta_0) \tag{4.3.16}$$

for continuous $f(\theta)$. Then minus Bayes' mean deviation corresponding to $W_\theta(\hat{\theta}) = -\delta_\theta(\hat{\theta})$ is equal to

$$\int_\Theta \operatorname{Tr} S_\theta M(\mathrm{d}\theta). \tag{4.3.17}$$

The precise definition of such "trace-integrals" is an interesting mathematical problem which will not be discussed here. Note, however, that if Θ is compact and $\dim \mathscr{H} < \infty$ this can be done rather easily. Then as in the proof of Theorem 4.2.3 one can show that $M(\mathrm{d}\theta)$ is differentiable with respect to $m(\mathrm{d}\theta) = \operatorname{Tr} M(\mathrm{d}\theta)$ so that $M(\mathrm{d}\theta) = P(\theta) m(\mathrm{d}\theta)$ and (4.3.17) can be defined as

$$\int_\Theta (\operatorname{Tr} S_\theta P(\theta)) m(\mathrm{d}\theta). \tag{4.3.18}$$

The measurement maximizing this functional is called the *maximum-likelihood measurement*. As in Theorem 4.3.1 one can show that for a covariant family of states $\{S_\theta\}$ the maximum is achieved on a covariant measurement. For a covariant measurement (4.2.9) the functional (4.3.18) takes the form $\operatorname{Tr} S P_0$. Thus to find the covariant maximum-likelihood measurement one has to solve the problem to find

$$\max \operatorname{Tr} S P_0, \qquad P_0 \in \mathfrak{P},$$

which has the same nature as (4.3.15).

4.4. Measurements of angular parameters

We are going to apply the estimation theory developed in the previous section to one-parameter symmetry groups. The simplest example in estimation is of the angle of rotation by measurements over the spin degrees of freedom. According to Section 3.12 the family of states obtained from the basic state S by rotations through the angles φ, $0 \leq \varphi < 2\pi$, is represented by

$$S_\varphi = \mathrm{e}^{-\mathrm{i}\varphi J} S \mathrm{e}^{\mathrm{i}\varphi J}; \qquad 0 \leq \varphi < 2\pi, \tag{4.4.19}$$

where J is the operator of spin angular momentum, corresponding to the axis of rotation, which has the diagonal form (3.11.116) in the basis of eigenvectors $\{|m)\}$. The index m varies from $-j$ to j where j is the value of spin.

The relevant symmetry group G is the group of rotations about the axis or the group $\mathbb{T}$ of shifts (mod 2π) of the interval $\Theta = [0, 2\pi)$. Operators $V_\varphi = \exp(-\mathrm{i}\varphi J)$ constitute a finite dimensional unitary representation of the group. The stationary subgroup G_0 is trivial in this case and according to Theorem 4.2.3 any covariant measurement of the angle of rotation has the form

$$M(\mathrm{d}\varphi) = \mathrm{e}^{-\mathrm{i}\varphi J} P_0 \mathrm{e}^{\mathrm{i}\varphi J} \frac{\mathrm{d}\varphi}{2\pi},$$

where P_0 is a positive operator satisfying (4.2.5), *i.e.*,

$$(2\pi)^{-1} \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\varphi J} P_0 \mathrm{e}^{\mathrm{i}\varphi J} \mathrm{d}\varphi = I.$$

In the basis $\{|m)\}$

$$(m|M(\mathrm{d}\varphi)|m') = \mathrm{e}^{\mathrm{i}(m'-m)\varphi} p_{mm'} \frac{\mathrm{d}\varphi}{2\pi}, \tag{4.4.20}$$

where $p_{mm'} = (m|P_0|m')$ and (4.2.5) reduced to $p_{mm} = 1$. Thus any covariant measurement of the angle of rotation for spin degrees of freedom is given by the matrix elements (4.4.20) where the matrix $[p_{mm'}]$, with m, m' varying from $-j$ to j, belongs to the convex set

$$\mathfrak{P} = \{[p_{mm'}] : [p_{mm'}] \geq 0,\ p_{mm} = 1\}.$$

We now turn to the covariant estimation problem, restricting to the case of pure states

$$S_\varphi = \mathrm{e}^{-\mathrm{i}\varphi J} |\psi)(\psi| \mathrm{e}^{\mathrm{i}\varphi J}; \qquad 0 \leq \varphi < 2\pi. \tag{4.4.21}$$

We shall look for optimal covariant measurement which minimizes the mean deviation for, say, $\varphi_0 = 0$. According to Theorem 4.3.1 it is also optimal in both Bayes' and minimax senses.

Put $\gamma_m = (m|\psi)/|(m|\psi)|$ if $(m|\psi) \neq 0$; otherwise let γ_m be an arbitrary complex number of unit modulus.

Theorem 4.4.1. *The covariant measurement*

$$(m|M_*(\mathrm{d}\varphi)|m') = \mathrm{e}^{\mathrm{i}(m'-m)\varphi}\gamma_m\bar{\gamma}_{m'}\frac{\mathrm{d}\varphi}{2\pi} \tag{4.4.22}$$

is the optimal measurement of angle of rotation φ in the family of states (4.4.21) *for any deviation function of the form $W_\varphi(\hat{\varphi}) = W(\varphi-\hat{\varphi})$, where $W(\cdot)$ is an even 2π-periodic function on $\mathbb{R}$, satisfying*

$$\int_0^{2\pi} W(\varphi)\cos k\varphi\,\mathrm{d}\varphi \leq 0; \qquad k = 1, 2, \ldots \tag{4.4.23}$$

Proof. Any function $W(\cdot)$ satisfying the conditions of the theorem can be represented by Fourier series

$$W(\varphi) = w_0 - \sum_{k=1}^{\infty} w_k \cos k\varphi,$$

with $w_k \geq 0$; $k = 1, 2, \ldots$ Using (4.4.20) we obtain the probability distribution of the covariant measurement $\boldsymbol{M}$ with respect to the state $S = |\psi)(\psi|$

$$\mu_0(\mathrm{d}\varphi) = \left[\sum_{m,m'} \mathrm{e}^{\mathrm{i}(m'-m)\varphi} p_{mm'}(m'|\psi)(\psi|m)\right]\frac{\mathrm{d}\varphi}{2\pi} \tag{4.4.24}$$

where the sum extends over $-j \leq m, m' \leq j$. By (4.3.11) and (4.4.24)

$$\mathscr{R}_0\{\boldsymbol{M}\} = w_0 - \sum_{k=1}^{\infty} w_k \int_0^{2\pi} \cos k\varphi \sum_{m,m'} \mathrm{e}^{\mathrm{i}(m'-m)\varphi} p_{mm'}(m'|\psi)(\psi|m)\frac{\mathrm{d}\varphi}{2\pi}.$$

The factor multiplying w_k is equal to

$$\frac{1}{2} \sum_{\substack{m,m': \\ |m-m'|=k}} (\psi|m)\, p_{mm'}(m'|\psi).$$

By the positive definiteness

$$|p_{mm'}| \leq \sqrt{p_{mm} \cdot p_{m'm'}} = 1, \tag{4.4.25}$$

so that

$$\frac{1}{2}\sum_{\substack{m,m':\\ |m-m'|=k}}(\psi|m)p_{mm}(m'|\psi) \le \frac{1}{2}\sum_{\substack{m,m':\\ |m-m'|=k}}|(\psi|m)|\,|(m'|\psi)|,$$

with the equality achieved if and only if $p_{mm'} = \gamma_m\bar{\gamma}_{m'}$. It follows that $\mathscr{R}_0\{\boldsymbol{M}\} \ge \mathscr{R}_0\{\boldsymbol{M}_*\}$, where

$$\mathscr{R}_0\{\boldsymbol{M}_*\} = w_0 - \frac{1}{2}\sum_{k=1}^{\infty} w_k \sum_{\substack{m,m'\\ |m-m'|=k}}|(\psi|m)|\cdot|(m'|\psi)|. \tag{4.4.26}$$

□

Several comments are in order. The examples of the deviation functions admitted by the conditions of the theorem are (see Figure 4.1)

$$4\sin^2\frac{\varphi}{2} = 2 - 2\cos\varphi,$$

$$\min\{\varphi, 2\pi-\varphi\} = \frac{\pi}{2} - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\cos(2k+1)\varphi}{(2k+1)^2},$$

$$\left|\sin\frac{\varphi}{2}\right| = \frac{2}{\pi} - \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{\cos k\varphi}{4k^2-1}.$$

The quadratic $(\mathrm{mod}\,2\pi)$ deviation $\min\{(\varphi-\hat{\varphi})^2, (2\pi-\varphi+\hat{\varphi})^2\}$ does not satisfy (4.4.23). Formally the condition (4.4.23) is satisfied by the periodic delta-function

$$-\delta(\varphi(\mathrm{mod}\,2\pi)) = -\frac{1}{2\pi} - \frac{1}{\pi}\sum_{k=1}^{\infty}\cos k\varphi,$$

which corresponds to the maximum-likelihood approach. In fact the proof extends to arbitrary periodic generalized functions with $w_k \ge 0$ since the value of k in the series defining $\mathscr{R}_0\{\boldsymbol{M}\}$ cannot exceed $2j$, so that the series contains only a finite number of nonzero terms. Thus (4.4.22) *is also the covariant maximum-likelihood measurement for the family* (4.4.21).

It is not difficult to give an alternative abstract characterization of the functions satisfying (4.4.23). Note that $w_0 - W(\varphi)$, $\varphi \in \mathbb{R}$, is the Fourier transform of an even measure (concentrated at the integral points) and therefore is positive definite. It follows that a 2π-periodic function $W(\cdot)$ satisfies (4.4.23) if and only if

$$\sum_{j,k} W(\varphi_j - \varphi_k)\bar{c}_j c_k \le 0$$

for any complex c_j, such that $\sum_j c_j = 0$ and any $\varphi_k \in \mathbb{R}$.

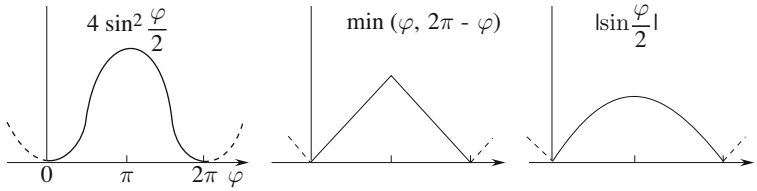

Figure 4.1.

The theorem shows that the optimal measurement is rather insensitive to the choice of the deviation functions – a property which is desirable from the physical point of view.

If the basic state is such that $\arg(m|\psi)$ is the same for all values of m, in particular, if $(m|\psi) \geq 0$ for all m, then $\gamma_m = \text{const.}$ and the optimal covariant measurement is just the canonical measurement of the angle of rotation in the angular momentum representation defined by (3.12.128). However if $\arg(m|\psi)$ varies with m, then the canonical measurement is no longer optimal and its accuracy is worse then the best possible (4.4.26). The optimal measurement takes into account the prior information contained in the variations of phase of the complex number $(m|\psi)$ determined by the basic state $S = |\psi)(\psi|$.

Note that transforming the basis $\{|m)\}$ to $|m)' = \gamma_m|m)$ we reduce the optimal measurement (4.4.22) to the canonical form. The canonical measurement is thus a tantamount representative in the family of optimal covariant measurements which are obtained one from another by the aforementioned "gauge" transformations.

The conclusion of this section apply to the phase measurements (see Section 3.9) and the angle measurements for spatial degrees of freedom (see Section 3.12) with the only essential modification that the index m should vary not from $-j$ to j but from 0 to ∞ in the case of phase and from $-\infty$ to ∞ in the second case. The representation of covariant measurements has the same form (4.4.20) with m, m' varying in the corresponding ranges and Theorem 4.4.1 extends to these cases taking care of convergence of the series. However in obtaining (4.4.20) we cannot refer to Theorem 4.2.3 since the representation space is not finite-dimensional in these cases. The results will follow from the more general theory for arbitrary representation of the group $\mathbb{T}$ to be discussed in Section 4.6.

4.5. Uncertainty relations for angular quantities

To establish the uncertainty relations for such quantities as angle and phase we must first introduce a proper measure of uncertainty. The fact that the angular quantities vary periodically from 0 to 2π makes us reject

the variance which serves as the measure of uncertainty for quantities running freely on the whole real line like coordinate and time. In fact the angular values $\varphi = 0$ and $\varphi = 2\pi$ must be considered as "identical" and the values $\varphi \approx 0$, $\varphi \approx 2\pi$ as "close" to each other. If we consider an angular distribution $\mu(\mathrm{d}\varphi)$ which is symmetric with respect to $\varphi = \pi$ and concentrated at the ends of the interval $[0, 2\pi)$, then its variance $D(\varphi) = \int_0^{2\pi} (\varphi - E(\varphi))^2 \mu(\mathrm{d}\varphi)$ will be $\approx \pi^2$ while any reasonable measure of uncertainty should be almost zero.

Consider the complex random variable $\exp(\mathrm{i}\varphi)$ whose range is the unit circle. Its variance is equal to

$$D(\mathrm{e}^{\mathrm{i}\varphi}) = \int_0^{2\pi} |\mathrm{e}^{\mathrm{i}\varphi} - E(\mathrm{e}^{\mathrm{i}\varphi})|^2 \mu(\mathrm{d}\varphi) = 1 - |E(\mathrm{e}^{\mathrm{i}\varphi})|^2,$$

where $E(\mathrm{e}^{\mathrm{i}\varphi}) = \int_0^{2\pi} \mathrm{e}^{\mathrm{i}\varphi} \mu(\mathrm{d}\varphi)$, μ being the angular probability distribution. We adopt the following measure of uncertainty

$$\Delta(\varphi) = \frac{D(\mathrm{e}^{\mathrm{i}\varphi})}{|E(\mathrm{e}^{\mathrm{i}\varphi})|^2} \equiv |E(\mathrm{e}^{\mathrm{i}\varphi})|^{-2} - 1. \tag{4.5.27}$$

Note that $|E(\mathrm{e}^{\mathrm{i}\varphi})|$ is the distance of the center of mass of the distribution of $\exp(\mathrm{i}\varphi)$ concentrated on the unit circle from the center of the circle. Therefore $\Delta(\varphi) = 0$ for the δ-distributions (complete certainty) and $\Delta(\varphi) = \infty$ for the uniform distribution (complete uncertainty) as is desirable.

Let $\mu(\mathrm{d}\varphi) = (\psi|M(\mathrm{d}\varphi)\psi)$ be the probability distribution of the measurement $\boldsymbol{M} = \{M(\mathrm{d}\varphi)\}$ with respect to the pure state $S = |\psi)(\psi|$. We denote the corresponding uncertainty by $\Delta_S\{\boldsymbol{M}\}$. Since the state S will be mostly kept fixed we shall omit it from the notations. Then for any covariant measurement of the angle φ

$$\Delta\{\boldsymbol{M}\} \geq \Delta_*(\varphi) = \left[\sum_m |(\psi|m)\|(m-1|\psi)| \right]^{-2} - 1, \tag{4.5.28}$$

with the equality achieved for the optimal measurement (4.4.22). Indeed calculating the expectation of $\exp(\mathrm{i}\varphi)$ with respect to the distribution (4.4.24) we find

$$E(\mathrm{e}^{\mathrm{i}\varphi}) = \sum_m (\psi|m)\, p_{m,m-1}(m-1|\psi),$$

whence by (4.4.25)

$$|E(\mathrm{e}^{\mathrm{i}\varphi})| \leq \sum_m |(\psi|m)\|(m-1|\psi)|,$$

and (4.5.28) follows by using (4.5.27). Note that the range of m is arbitrary here so that the result applies to the angular measurements for both spin and spatial degrees of freedom and to phase measurements.

Note that for a covariant measurement $\boldsymbol{M}$ the uncertainty $\Delta\{\boldsymbol{M}\}$ is the same for all the states S_φ obtained from S using (4.4.21). Thus the optimal covariant measurement for the family (4.4.21) is the minimum-uncertainty measurement for the states of the family. For the canonical measurement $p_{mm'} = 1$ and $E(\mathrm{e}^{\mathrm{i}\varphi}) = \sum(\psi|m)(m-1|\psi)$. Thus it is optimal if and only if

$$\left|\sum(\psi|m)(m-1|\psi)\right| = \sum|(\psi|m)(m-1|\psi)|$$

which holds if and only if $\arg(\psi|m)(m-1|\psi) \equiv \varphi_0$ does not depend on m. This means that $\mathrm{Arg}(m|\psi) = -m\varphi_0 + \text{const.}$, *i.e.*, the state belongs to the family (4.4.21) generated from the basic state satisfying $\arg(m|\psi) = \text{const}$. Otherwise there is a covariant measurement having smaller uncertainty than the canonical one. The quantity $\Delta_*(\varphi)$ in the right-hand side of (4.5.28) thus represents the intrinsic minimal uncertainty of the angle φ in the pure state $S = |\psi)(\psi|$. We now are going to establish the uncertainty relation "angle-angular momentum",

$$\Delta_*(\varphi) \cdot D(J) > \tfrac{1}{4}, \tag{4.5.29}$$

for an arbitrary pure state $S = |\psi)(\psi|$.

Performing the "gauge" transformation $|m)' = \gamma_m|m)$ with $\gamma_m = (m|\psi)/|(m|\psi)|$ we have $(m|\varphi) \geq 0$. Thus we can take $(m|\psi) \geq 0$. Then

$$\sum_m |(\psi|m)(m-1|\psi)| = \sum_m (\psi|m)(m-1|\psi) = (\psi|E_+\psi) \equiv \overline{E}_+,$$

where E_+ is given by (3.12.130) and $\overline{X}$ denotes the mean value of X with respect to the state $|\psi)(\psi|$. Therefore by (4.5.28)

$$\Delta_*(\varphi) = \frac{1-|\overline{E}_+|^2}{|\overline{E}_+|^2} = \frac{1-(\overline{C}^2+\overline{S}^2)}{\overline{C}^2+\overline{S}^2},$$

where C and S are the "cosine" and "sine" operators $C = \frac{1}{2}(E_+ + E_-)$, $S = \frac{1}{2}\mathrm{i}(E_+ - E_-)$. From (3.12.131)

$$C^2 + S^2 = I - \tfrac{1}{2}[|-j)(-j| + |j)(j|].$$

Using the fact that $\overline{E}_+ = \overline{C} - \mathrm{i}\overline{S}$ and $D(C) = \overline{C^2} - \overline{C}^2$, $D(S) = \overline{S^2} - \overline{S}^2$ we therefore get

$$\Delta_*(\varphi) = [D(C) + D(S) + \tfrac{1}{2}(\overline{|-j)(-j|} + \overline{|j)(j|})] \cdot (\overline{C}^2 + \overline{S}^2)^{-1}.$$

From (3.12.133) we obtain

$$[C, J] = \mathrm{i}S, \qquad [S, J] = -\mathrm{i}C.$$

By the uncertainty relation (2.6.63) with the right-hand side written in the commutator form (2.6.65) we obtain

$$D(C) \cdot D(J) \geq \tfrac{1}{4}\overline{S}^2, \qquad D(S) \cdot D(J) \geq \tfrac{1}{4}\overline{C}^2. \tag{4.5.30}$$

Therefore

$$\Delta_*(\varphi) \geq \tfrac{1}{4}D(J)^{-1} + \tfrac{1}{2}[|(\psi| - j)|^2 + |(\psi|j)|^2] \cdot (\overline{C}^2 + \overline{S}^2)^{-1}.$$

Since $(\overline{C}^2 + \overline{S}^2)^{-1} = |\overline{E}_+|^{-2} = \Delta_*(\varphi) + 1$, we get

$$\begin{aligned} \Delta_*(\varphi) \geq\ & [1 - \tfrac{1}{2}[|(\psi| - j)|^2 + |(\psi|j)|^2]]^{-1} \\ & \cdot [\tfrac{1}{4}D(J)^{-1} + \tfrac{1}{2}[|(\psi| - j)|^2 + |(\psi|j)|^2]]. \end{aligned} \tag{4.5.31}$$

Omitting the nonnegative term $\frac{1}{2}[|(\psi| - j)|^2 + |(\psi|j)|^2]$ we get the inequality $\Delta_*(\varphi) \geq [4D(J)]^{-1}$.

Let us show that the equality is never achieved here. If this would be the case, then $(\psi| - j) = (\psi|j) = 0$ since we have omitted these terms in (4.5.31), and (4.5.30) turn into equalities. According to (2.6.64) this is the case if and only if

$$[(C - \overline{C}) + \mathrm{i}\alpha(J - \overline{J})]\psi = 0, \qquad [(S - \overline{S}) + \mathrm{i}\beta(J - \overline{J})]\psi = 0$$

for some real α and β (we exclude the case $D(J) = 0$). Denoting $z = \beta - \mathrm{i}\alpha$ we have $(E_- - \overline{E}_-)\psi = z(J - \overline{J})\psi$ whence $(m+1|\psi) = c_m(m|\psi)$ where $c_m = zm + \overline{E}_- - z\overline{J}$. Since $(-j|\psi) = 0$ we obtain by induction $(m|\psi) = 0$ for all m which is impossible for a unit vector ψ. Thus the strict inequality in (4.5.29) is proved.

In the same way we can prove the uncertainty relation for the phase θ of a harmonic oscillator (see Section 3.9). It follows from (4.5.28) for any covariant measurement M of θ and for any pure state $S = |\psi)(\psi|$

$$\Delta\{M\} \geq \Delta_*(\theta) = \left[\sum_{n=0}^{\infty} |(\psi|n)(n-1|\psi)|\right]^{-2} - 1.$$

Using the operators P^*, P given by (3.9.95) in place of E_+, E_- and the corresponding cosine and sine operators we get

$$\Delta_*(\theta) \geq [1 - \tfrac{1}{2}|(\psi|0)|^2]^{-1} \cdot [\tfrac{1}{4}D(N)^{-1} + \tfrac{1}{2}|(\psi|0)|^2],$$

where N is the number of quanta operator, and the "phase-number of quanta" uncertainty relation

$$\Delta_*(\theta)D(N) \geq \tfrac{1}{4}.$$

Finally, for covariant angle measurements in the case of spatial degrees of freedom (see Section 3.12) we have the inequality (4.5.28) with m running from $-\infty$ to ∞ and the uncertainty relation (4.5.29) with J replaced by the operator of orbital angular momentum L. This is obtained by using the operators U, U^* given by (3.12.125) in place of E_+, E_-.

In Section 6.3 we shall derive a general inequality for the uncertainty of angular parameter valid for arbitrary states.

4.6. Covariant measurements of angular parameter in the case of arbitrary representation of the group $\mathbb{T}$

In this section we shall considerably generalize the results of Section 4.4. This will also help us to study the covariant measurements of a shift parameter in $\mathbb{R}$. Let $\varphi \to V_\varphi$, $0 \leq \varphi \leq 2\pi$, be a projective unitary representation of the group $\mathbb{T}$. Since the value $\varphi = 2\pi$ corresponds to the identical state automorphism, we must have $V_{2\pi} = \exp(\mathrm{i}a_0)$ with a_0 real. Putting $V_{2\pi k+\varphi} = V_{2\pi}^k \cdot V_\varphi$ we get a projective unitary representation of the additive group of $\mathbb{R}$. According to Proposition 3.2.1 it reduces to the unitary representation $\varphi \to \exp(-\mathrm{i}\varphi A)$, where A is self-adjoint. The condition $V_{2\pi} = \exp(\mathrm{i}a_0)$ implies that the spectrum of A consists of eigenvalues of the form $m = k + a$, where k are integral numbers, *i.e.*, A has the spectral representation

$$A = \sum_m m E_m, \tag{4.6.32}$$

where E_m is the projection onto the invariant subspace $\mathscr{H}_m$ corresponding to the eigenvalue m. The space $\mathscr{H}$ is the direct orthogonal sum of these subspace $\mathscr{H}_m$

$$\mathscr{H} = \bigoplus_m \mathscr{H}_m, \tag{4.6.33}$$

which means that for any $\psi \in \mathscr{H}$

$$\psi = \sum_m \psi_m, \qquad \|\psi\|^2 = \sum_m \|\psi_m\|^2, \tag{4.6.34}$$

where $\psi_m = E_m\psi$ is the *component* of the vector ψ in the subspace $\mathscr{H}_m$. We shall write briefly $\psi = [\psi_m]$. Then

$$\exp(\mathrm{i}\varphi A)\psi = [\exp(\mathrm{i}m\varphi)\psi_m].$$

We shall study measurements $M(\mathrm{d}\varphi)$ with values in $[0, 2\pi)$ which are covariant with respect to the representation of the group $\mathbb{T}$. Assuming that $\dim \mathscr{H}_m \equiv 1$ we obtain in particular the case of spin angular measurements considered in Section 4.4 if $m = -j, -j+1, \ldots, j$, the case of phase measurements if $m = 0, 1, \ldots$ and the case of spatial angular measurement if $m = 0, \pm 1, \ldots$

By a *kernel* we shall mean here a matrix $[K_{mm'}]$ where $K_{mm'}$ is a bounded operator from $\mathscr{H}_{m'}$ to $\mathscr{H}_m$. To a bounded operator K in $\mathscr{H}$ we ascribe the kernel $[E_m K E_{m'}]$. In particular, the unit operator has the kernel $[\delta_{mm'} E_m]$. The kernel is *positive definite*, $[K_{mm'}] \geq 0$ if

$$\sum_{m,m'} (\psi_m | K_{mm'} \psi_{m'}) \geq 0$$

for any $\psi = [\psi_m]$. Obviously a positive operator has a positive definite kernel.

Theorem 4.6.1. *A measurement covariant with respect to the representation of the group $\mathbb{T}$ is determined by the kernels*

$$M(B) = \left[K_{mm'} \int_B \mathrm{e}^{\mathrm{i}(m'-m)\varphi} \frac{\mathrm{d}\varphi}{2\pi} \right]; \qquad B \in \mathscr{A}([0, 2\pi)),$$

where $[K_{mm'}]$ is a positive definite kernel satisfying $K_{mm} = E_m$.

Proof. Applying (4.2.3) to the representation $\varphi \to \exp(-\mathrm{i}\varphi A)$ we have

$$\int_0^{2\pi} \mathrm{Tr}\, \mathrm{e}^{-\mathrm{i}\varphi A} S \mathrm{e}^{\mathrm{i}\varphi A} M(B) \mathrm{d}\varphi = \mathrm{mes}\, B,$$

where mes denotes the Lebesgue measure on $[0, 2\pi)$. Take for S a density operator S_m acting in $\mathscr{H}_m$, *i.e.*, $S_m E_m = E_m S_m = S_m$. Then

$$\exp(-\mathrm{i}\varphi A) S_m \exp(\mathrm{i}\varphi A) = S_m$$

and denoting by Tr_m the trace in the space $\mathscr{H}_m$ we have

$$\mathrm{Tr}_m\, S_m M_{mm}(B) = (2\pi)^{-1}\, \mathrm{mes}\, B,$$

where $M_{mm'}(B) = E_m M(B) E_{m'}$. Since it holds for arbitrary S_m in $\mathscr{H}_m$ we have

$$M_{mm}(B) = E_m \frac{\mathrm{mes}\, B}{2\pi}. \tag{4.6.35}$$

From positivity of $M(B)$ and the Cauchy inequality

$$|(\psi_m|M_{mm'}(B)\psi_{m'})| \le \frac{\text{mes}\, B}{2\pi}\|\psi_m\| \cdot \|\psi_{m'}\|,$$

for all ψ_m, $\psi_{m'}$. By Proposition 4.2.2

$$M_{mm'}(B) = \int_B P_{mm'}(\varphi)\frac{\mathrm{d}\varphi}{2\pi},$$

where $P_{mm'}(\cdot)$ is the operator density satisfying $\|P_{mm'}(\varphi)\| \le 1$. Since $M_{mm'}(B)$ maps $\mathscr{H}_{m'}$ into $\mathscr{H}_m$ so does $P_{mm'}(\varphi)$. From (4.6.35) $P_{mm}(\varphi) \equiv E_m$; positivity of $M(B)$ implies that the kernel $[P_{mm'}(\varphi)]$ is positive definite for almost all φ. As in the proof of Theorem 4.2.3 the covariance of the measurement $\{M(B)\}$ implies

$$P_{mm'}(\varphi) = \mathrm{e}^{\mathrm{i}(m'-m)\varphi}P_{mm'}(0)$$

for the proper definition of the densities $P_{mm'}(\cdot)$. Putting $K_{mm'} = P_{mm'}(0)$ we get the assertion of the theorem. □

Conversely, any kernel $[K_{mm'}]$ satisfying the conditions of the theorem defines the corresponding covariant measurement, but we shall not prove it here.

Consider now the *covariant estimation problem* for the family of states

$$S_\varphi = \mathrm{e}^{-\mathrm{i}\varphi A}S\mathrm{e}^{\mathrm{i}\varphi A}, \qquad 0 \le \varphi < 2\pi, \tag{4.6.36}$$

where $S = |\psi)(\psi|$ is a pure state. Assume that the deviation function $W(\varphi - \hat{\varphi})$ is continuous and satisfies (4.4.23), then

$$W(\varphi) = w_0 - \sum_{k=1}^{\infty} w_k \cos k\varphi \equiv \sum_{k=-\infty}^{\infty} v_k \mathrm{e}^{\mathrm{i}k\varphi},$$

where $v_k \le 0$ for $k \ne 0$ and $\sum |v_k| < \infty$. Let $\boldsymbol{M}$ be a covariant measurement defined according to Theorem 4.6.1 by the kernel $[K_{mm'}]$. Then arguing as in the proof of Theorem 4.4.1 we get

$$\begin{aligned}\mathscr{R}_0\{\boldsymbol{M}\} &= \sum_{m,m'} v_{m-m'}(\psi_m|K_{mm'}\psi_{m'}) \\ &\ge \sum_{m,m'} v_{m-m'}\|\psi_m\| \cdot \|\psi_{m'}\|,\end{aligned}$$

where ψ_m are the components of the vector ψ. The equality here is achieved for the positive definite kernel

$$K_{mm'} = \frac{|\psi_m)(\psi_{m'}|}{\|\psi_m\| \cdot \|\psi_{m'}\|}. \tag{4.6.37}$$

Similar argument can be given for the deviation function $W(\hat{\varphi} - \varphi) = -\delta((\hat{\varphi} - \varphi) \pmod{2\pi})$ for which $v_m \equiv -(2\pi)^{-1}$, if we suppose additionally that $\sum \|\psi_m\| < \infty$.

The operator-valued measure corresponding to the kernel (4.6.37)

$$M_0(\mathrm{d}\varphi) = \left[\frac{|\psi_m)(\psi_{m'}|}{\|\psi_m\| \cdot \|\psi_{m'}\|} \mathrm{e}^{\mathrm{i}(m'-m)\varphi} \right] \frac{\mathrm{d}\varphi}{2\pi}$$

is not a measurement unless $\dim \mathscr{H}_m \equiv 1$, for otherwise

$$E_0 \equiv \int M_0(\mathrm{d}\varphi) = \sum_m \frac{|\psi_m)(\psi_m|}{(\psi_m|\psi_m)} \neq I.$$

The operator E_0 is the projection onto the subspace $\mathscr{H}_0$ generated by the components ψ_m of the basic vector ψ. Therefore the measure $M_0(\mathrm{d}\varphi)$ can be extended to a resolution of identity in $\mathscr{H}$ by the formula

$$M_*(\mathrm{d}\varphi) = M_0(\mathrm{d}\varphi) \oplus M_1(\mathrm{d}\varphi), \tag{4.6.38}$$

where $M_1(\mathrm{d}\varphi)$ is arbitrary resolution of identity in the orthogonal complement to the subspace $\mathscr{H}_0$. Since the states S_φ are concentrated on $\mathscr{H}_0$, the measurement probability distribution is given by

$$\begin{aligned} \mu_\varphi(\mathrm{d}\hat{\varphi}) &\equiv \operatorname{Tr} S_\varphi M_*(\mathrm{d}\hat{\varphi}) = (\mathrm{e}^{-\mathrm{i}\varphi A}\psi | M_0(\mathrm{d}\hat{\varphi}) \mathrm{e}^{-\mathrm{i}\varphi A}\psi) \\ &= \left| \sum_n \|\psi_m\| \mathrm{e}^{\mathrm{i}m(\hat{\varphi}-\varphi)} \right|^2 \frac{\mathrm{d}\hat{\varphi}}{2\pi}, \end{aligned}$$

so that the additional term $M_1(\mathrm{d}\varphi)$ has no influence on the statistics of measurement and on the value of mean deviation. In particular, we can take

$$M_1(\mathrm{d}\varphi) = (I - E_0)\mu(\mathrm{d}\varphi),$$

where μ is a probability distribution. Of course $\boldsymbol{M}_*$ will be covariant only if so in $\boldsymbol{M}_1$, which holds for the uniform distribution μ.

If $\dim \mathscr{H}_m$ is the same for all m, then the subspaces $\mathscr{H}_m$ are mutually isomorphic. Let for each pair m, m' $U_{mm'}$ be an isometric map of $\mathscr{H}_{m'}$ onto $\mathscr{H}_m$ and the system $\{U_{mm'}\}$ be consistent in the sense that $U_{mm} = E_m$, $U_{mm'}U_{m'm''} = U_{mm''}$. Then the kernel $[U_{mm'}]$ satisfies the conditions of Theorem 4.6.1. Consider the corresponding covariant measurement

$$M(\mathrm{d}\varphi) = [U_{mm'}\mathrm{e}^{\mathrm{i}(m'-m)\varphi}]\frac{\mathrm{d}\varphi}{2\pi}.$$

If we agree that the maps $U_{mm'}$ "identify" $\mathscr{H}_{m'}$ and $\mathscr{H}_m$ we obtain the "canonical" covariant measurement

$$M(\mathrm{d}\varphi) = [E_m \mathrm{e}^{\mathrm{i}(m'-m)\varphi}]\frac{\mathrm{d}\varphi}{2\pi}. \tag{4.6.39}$$

Of course it depends on the way of identifying of the subspaces $\{\mathscr{H}_m\}$.

The measurement (4.6.39) will be optimal for the family (4.6.36) if and only if

$$(\psi_m|U_{mm'}|\psi_{m'}) \equiv (\psi_m|\psi_{m'}) = \|\psi_m\|\,\|\psi_{m'}\|.$$

This holds if and only if $\psi_m = \alpha_m e$ under the given identification of $\{\mathscr{H}_m\}$, where $\alpha_m \geq 0$ and e is a fixed vector in $\mathscr{H}_m$. Otherwise there is a covariant measurement which has smaller mean deviation.

4.7. Covariant measurements of a shift parameter

The most important examples of shift parameters on $\mathbb{R}$ are the coordinate x and the time t. We shall first consider covariant measurements of the shift parameters from a general point of view.

Consider a projective unitary representation of the shift group of $\mathbb{R}$. According to Proposition 3.2.1 it reduces to the unitary representation

$$\theta \to \mathrm{e}^{-\mathrm{i}\theta A}, \tag{4.7.40}$$

where A is a self-adjoint operator in $\mathscr{H}$. There is a form of the spectral theorem which says that A can be represented as the operator of multiplication by an independent variable in a direct orthogonal integral of Hilbert spaces. We shall describe this construction in the particular case when A has *purely Lebesgue spectrum*.

In this section we shall proceed rather heuristically considering the situation as a continuous analog of the discrete one discussed in the previous section, since a rigorous treatment would take too much place here.

Let Λ be an interval of the real line with the Lebesgue measure on it and let for almost all $\lambda \in \Lambda$ a Hilbert space $\mathscr{H}_\lambda$ be given. The inner product and the norm in $\mathscr{H}_\lambda$ will be denoted $(\cdot|\cdot)_\lambda$ and $\|\cdot\|_\lambda$. The *direct orthogonal integral of the Hilbert spaces $\mathscr{H}_\lambda$ with respect to the Lebesgue measure*

$$\mathscr{H} = \int_\Lambda \bigoplus \mathscr{H}_\lambda \,\mathrm{d}\lambda \tag{4.7.41}$$

is the space of functions $\psi = [\psi_\lambda]$, where $\psi_\lambda \in \mathscr{H}_\lambda$, satisfying

$$\|\psi\|^2 \equiv \int_\Lambda \|\psi_\lambda\|_\lambda^2 \mathrm{d}\lambda < \infty. \tag{4.7.42}$$

ψ_λ is called *component* of vector ψ in $\mathscr{H}_\lambda$. The inner product in $\mathscr{H}$ is defined by

$$(\varphi|\psi) = \int_\Lambda (\varphi_\lambda|\psi_\lambda)_\lambda \, \mathrm{d}\lambda. \tag{4.7.43}$$

We assume that *the representation space* $\mathscr{H}$ *decomposes into a direct orthogonal integral* (4.7.41) *such that A acts in* (4.7.41) *as multiplication by* λ:

$$A\psi = [\lambda\psi_\lambda]; \qquad \psi = [\psi_\lambda].$$

If follows that $\exp(\mathrm{i}\theta A)\psi = [\exp(\mathrm{i}\theta\lambda)\psi_\lambda]$. The relations (4.7.41) and (4.7.42) are continuous analogs of (4.6.33) and (4.6.34).

A *kernel* $[K(\lambda, \lambda')]$ is a function defined for almost all $\lambda, \lambda' \in \Lambda$ such that $K(\lambda, \lambda')$ is a bounded operator from $\mathscr{H}_{\lambda'}$ to $\mathscr{H}_\lambda$. An operator K in $\mathscr{H}$ *is determined by the kernel* $[K(\lambda, \lambda')]$ if for any $\psi \in \mathscr{H}$

$$K\psi = \left[\int_\Lambda K(\lambda, \lambda')\psi_{\lambda'} \, \mathrm{d}\lambda'\right]. \tag{4.7.44}$$

This correspondence can be given a strict mathematical sense if K is, say, Hilbert-Schmidt (*cf.* Section 2.7); for a Hermitean Hilbert-Schmidt operator $K = \sum \kappa_j |\psi^j)(\psi^j|$ the kernel is $K(\lambda, \lambda') = \sum \kappa_j |\psi^j_\lambda)_{\lambda\lambda'}(\psi^j_{\lambda'}|$ where ψ^j_λ is the component of ψ^j in $\mathscr{H}_\lambda$. Thus $K^j(\lambda, \lambda') = |\psi^j_\lambda)_{\lambda\lambda'} \cdot (\psi^j_{\lambda'}|$ is an operator from $\mathscr{H}_{\lambda'}$ to $\mathscr{H}_\lambda$ acting according to the formula

$$K^j(\lambda, \lambda')\psi_{\lambda'} = \psi^j_\lambda (\psi^j_{\lambda'}|\psi_{\lambda'})_{\lambda'}.$$

However we shall use the correspondence between the kernels and the operators formally in a broader context; in particular the unit operator in $\mathscr{H}$ will be described by the kernel $[\delta(\lambda - \lambda')I_\lambda]$ where I_λ is the unit operator in $\mathscr{H}_\lambda$.

The kernel is called *positive definite* if

$$\iint (\psi_\lambda|K(\lambda, \lambda')\psi_{\lambda'})_\lambda \mathrm{d}\lambda \, \mathrm{d}\lambda' \geq 0; \qquad \psi = [\psi_\lambda]. \tag{4.7.45}$$

This corresponds to positivity of the operator determined by (4.7.44). Substituting $f(\lambda)\psi_\lambda$ in place of ψ_λ, where f is an arbitrary scalar function we get that (4.7.45) holds if and only if for any $\psi = [\psi_\lambda]$ the scalar kernel $[(\psi_\lambda|K(\lambda, \lambda')\psi_{\lambda'})_\lambda]$ is positive definite. Then by the Cauchy inequality

$$|(\psi_\lambda|K(\lambda, \lambda')\psi_{\lambda'})_\lambda|^2 \leq (\psi_\lambda|K(\lambda, \lambda)\psi_\lambda)_\lambda (\psi_{\lambda'}|K(\lambda', \lambda')\psi_{\lambda'})_{\lambda'} \tag{4.7.46}$$

for almost all $\lambda, \lambda' \in \Lambda$, if $K(\lambda, \lambda)$ is properly defined (*cf.* the remark before (2.7.83)).

The general form of a measurement covariant with respect to the representation (4.7.40) is given by the continuous analog of Theorem 4.6.1:

$$M(B) = \left[K(\lambda, \lambda') \int_B \mathrm{e}^{\mathrm{i}(\lambda'-\lambda)\theta} \frac{\mathrm{d}\theta}{2\pi} \right]; \qquad B \in \mathscr{A}(\mathbb{R}), \tag{4.7.47}$$

where $[K(\lambda, \lambda')]$ *is a positive definite kernel satisfying* $K(\lambda, \lambda) = I_\lambda$. It is easy to check at the formal level that (4.7.47) is a resolution of identity: $M(\mathrm{d}\theta) \geq 0$ due to the positive definiteness of the kernel

$$[K(\lambda, \lambda') \exp \mathrm{i}(\lambda' - \lambda)\theta]$$

and

$$\int M(\mathrm{d}\theta) = [\delta(\lambda - \lambda') I_\lambda] = I.$$

Assume now that $\dim \mathscr{H}_\lambda$ is the same for almost all $\lambda \in \Lambda$. Let $\{U(\lambda, \lambda')\}$ be a consistent family of isometric maps of $\mathscr{H}_{\lambda'}$ onto $\mathscr{H}_\lambda$. Then $[U(\lambda, \lambda')]$ is a kernel determining by (4.7.47) a covariant measurement. If we agree that $\{U(\lambda, \lambda')\}$ identifies the spaces $\mathscr{H}_\lambda$ we obtain the "canonical" covariant measurement

$$M(\mathrm{d}\theta) = [I_\lambda \mathrm{e}^{\mathrm{i}(\lambda'-\lambda)\theta}] \frac{\mathrm{d}\theta}{2\pi}.$$

Let $S = |\psi)(\psi|$ be a pure state with $\psi = [\psi_\lambda]$. Then denoting by $\Phi_\psi(\lambda) = \int_{-\infty}^{\infty} \exp(\mathrm{i}\lambda\theta)(\psi|M(\mathrm{d}\theta)\psi)$ the characteristic function of the measurement probability distribution with respect to the state S we get using (4.7.47)

$$\Phi_\psi(\lambda) = \int (\psi_\mu | K(\mu, \mu - \lambda) \psi_{\mu-\lambda}) \mathrm{d}\mu. \tag{4.7.48}$$

The integral converges, since by (4.7.46) and the condition $K(\lambda, \lambda) = I_\lambda$

$$|(\psi_\mu | K(\mu, \mu - \lambda) \psi_{\mu-\lambda})| \leq \|\psi_\mu\|_\mu \|\psi_{\mu-\lambda}\|_{\mu-\lambda}, \tag{4.7.49}$$

and both functions in the right-hand side are square-integrable. If we suppose additionally that

$$\int_\Lambda \|\psi_\lambda\|_\lambda \mathrm{d}\lambda < \infty, \tag{4.7.50}$$

then the characteristic function is integrable by (4.7.49) and the measurement probability distribution has the continuous density

$$\begin{aligned} p_\psi(\theta) &\equiv \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}\theta\lambda} \Phi_\psi(\lambda) \mathrm{d}\lambda \\ &= \frac{1}{2\pi} \iint (\psi_\lambda | K(\lambda, \lambda') \psi_{\lambda'})_\lambda \mathrm{e}^{\mathrm{i}(\lambda'-\lambda)\theta} \mathrm{d}\lambda \, \mathrm{d}\lambda'. \end{aligned}$$

This agrees with the expression for $M(\mathrm{d}\theta)/\mathrm{d}\theta$ which follows formally from (4.7.47).

We now turn to the *covariant estimation problem* for the shift parameter θ in the family

$$S_\theta = \mathrm{e}^{-\mathrm{i}\theta A} S \mathrm{e}^{\mathrm{i}\theta A}; \qquad \theta \in \mathbb{R},$$

where $S = |\psi)(\psi|$ is a pure state. Since the symmetry group is noncompact we cannot apply Theorem 4.3.1. Moreover since the "uniform distribution" $\mathrm{d}\theta$ is infinite there is no covariant Bayes formulation. One can prove the noncompact analog of Theorem 4.3.1 for the minimax approach, but we shall instead proceed with minimizing the mean deviation

$$\mathscr{R}_\theta\{\boldsymbol{M}\} = \int_{-\infty}^{\infty} W(\hat{\theta} - \theta)\mu_\theta(\mathrm{d}\hat{\theta})$$

restricting essentially to covariant measurements. Due to covariance this quantity does not depend on θ and we have to minimize

$$\mathscr{R}_0\{\boldsymbol{M}\} = \int_{-\infty}^{\infty} W(\theta)\mu_0(\mathrm{d}\theta).$$

Similarly to Theorem 4.4.1 we assume that the deviation function is a real continuous negatively definite function, *i.e.*,

$$W(\theta) = -\int \mathrm{e}^{\mathrm{i}\theta\lambda}\widetilde{W}(\mathrm{d}\lambda),$$

where $\widetilde{W}(\mathrm{d}\lambda)$ is an even finite measure on $\mathbb{R}$. An example is given by $W(\theta - \hat{\theta}) = -\exp[-(\theta - \hat{\theta})^2]$. Then for the covariant measurement (4.7.47)

$$\mathscr{R}_0\{\boldsymbol{M}\} = -\int \Phi_\psi(\lambda)\widetilde{W}(\mathrm{d}\lambda).$$

By the inequality (4.7.49)

$$\mathrm{Re}\,\Phi_\psi(\lambda) \leq \Phi_*(\lambda) \equiv \int \|\psi_\mu\|_\mu \|\psi_{\mu-\lambda}\|_{\mu-\lambda}\mathrm{d}\mu,$$

so that

$$\mathscr{R}_0\{\boldsymbol{M}\} \geq -\iint \|\psi_\mu\|_\mu \|\psi_{\mu-\lambda}\|_{\mu-\lambda}\mathrm{d}\mu\,\widetilde{W}(\mathrm{d}\lambda) = \mathscr{R}_0\{\boldsymbol{M}_0\},$$

where

$$M_0(\mathrm{d}\theta) = \left[\frac{|\psi_\lambda)_{\lambda\lambda'}(\psi_{\lambda'}|}{\|\psi_\lambda\|_\lambda \cdot \|\psi_{\lambda'}\|_{\lambda'}}\mathrm{e}^{\mathrm{i}(\lambda'-\lambda)\theta}\right]\frac{\mathrm{d}\theta}{2\pi}. \tag{4.7.51}$$

As in Section 4.6 the optimal operator-valued measure $M_0(\mathrm{d}\theta)$ is not, in general, a resolution of identity since $\int M_0(\mathrm{d}\theta) \equiv E_0$ is equal to the unit operator if and only if $\dim \mathscr{H}_\lambda \equiv 1$ for almost all $\lambda \in \Lambda$. However it can be extended to a resolution of identity by adding an arbitrary resolution of identity in the orthogonal complement to the subspace $\mathscr{H}_0 = E_0(\mathscr{H})$:

$$M_*(\mathrm{d}\theta) = M_0(\mathrm{d}\theta) \oplus M_1(\mathrm{d}\theta). \tag{4.7.52}$$

The term $M_1(\mathrm{d}\theta)$ has no influence on the measurement statistics with respect to the states S_θ, which is determined by the probability distributions

$$\mu_\theta(\mathrm{d}\hat{\theta}) = \left| \int_\Lambda \|\psi_\lambda\|_\lambda \mathrm{e}^{\mathrm{i}\lambda(\hat{\theta}-\theta)} \mathrm{d}\lambda \right|^2 \frac{\mathrm{d}\hat{\theta}}{2\pi},$$

following from (4.7.51). However, it may affect the covariance property of the measurement (4.7.52). In any case $M_*(\mathrm{d}\theta)$ is "essentially covariant", *i.e.*, satisfies the covariance condition on the subspace $\mathscr{H}_0$, on which the family $\{S_\theta\}$ is concentrated. Thus, essentially, *the measurement* (4.7.52) *is optimal for any deviation function of the required form.*

The same is true for the maximum-likelihood approach. Assume that the basis state $S = |\psi)(\psi|$ satisfies (4.7.50) so that the density $p_\psi(\mathrm{d}\hat{\theta})$ is continuous. Then the formal Bayes functional with the deviation function $W(\theta) = -\delta(\theta)$ and the prior distribution $\pi(\mathrm{d}\theta) = \mathrm{d}\theta$ reduces to

$$\begin{aligned} -p_\psi(0) &= -\int \Phi_\psi(\lambda)\mathrm{d}\lambda \\ &= -\iint (\psi_\lambda | K(\lambda, \lambda')\psi_{\lambda'})_\lambda \mathrm{d}\lambda \, \mathrm{d}\lambda'. \end{aligned}$$

It has to be minimized over positive definite kernels $[K(\lambda,\lambda')]$ satisfying $K(\lambda,\lambda) = I_\lambda$. The solution of this problem is apparently given by the same formulas (4.7.51) and (4.7.52)).

Consider now the case of quadratic deviation

$$W(\hat{\theta} - \theta) = (\hat{\theta} - \theta)^2,$$

which corresponds to the estimation problem formulated in Section 3.2. It is natural to restrict to measurements for which

$$\mathscr{R}_0\{\boldsymbol{M}\} = \int \theta^2 \mu_S(\mathrm{d}\theta)$$

is finite. As is known from a course of probability the finiteness of the second moment is equivalent to the existence of the second derivative of the characteristic function at zero point, and

$$\int \theta^2 \mu_S(\mathrm{d}\theta) = -\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} \Phi_\psi(\lambda)\Big|_{\lambda=0} = -\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} \operatorname{Re} \Phi_\psi(\lambda)\Big|_{\lambda=0}.$$

Assume that the basic state satisfies $\int |\mathrm{d}/\mathrm{d}\lambda \|\psi_\lambda\|_\lambda|^2 \mathrm{d}\lambda < \infty$ and moreover $\psi_\lambda = 0$ at the ends of the interval Λ (if the interval extends to infinity this is automatically satisfied). Then the characteristic function $\Phi^*_\psi(\lambda)$ for the measurement (4.7.52) is twice differentiable at zero point with

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\Phi^*_\psi(\lambda)\bigg|_{\lambda=0} = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\int_\Lambda \|\psi_\mu\|_\mu \|\psi_{\mu-\lambda}\|_{\mu-\lambda}\mathrm{d}\mu$$
$$= -\int_\Lambda \left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\|\psi_\lambda\|_\lambda\right|^2 \mathrm{d}\lambda,$$

where we have used the assumed properties of $[\psi_\lambda]$. Consider the functions $\mathrm{Re}\,\Phi_\psi(\lambda)$ and $\Phi^*_\psi(\lambda)$. They satisfy $\mathrm{Re}\,\Phi_\psi(\lambda) \leq \Phi^*_\psi(\lambda)$, $\mathrm{Re}\,\Phi_\psi(0) = \Phi^*_\psi(0)$ and are twice differentiable at $\lambda = 0$. It follows that

$$(\mathrm{d}^2/\mathrm{d}\lambda^2)\Phi^*_\psi(0) \geq (\mathrm{d}^2/\mathrm{d}\lambda^2)\,\mathrm{Re}\,\Phi_\psi(0)$$

so that $\mathscr{R}_0\{\boldsymbol{M}\} \geq \mathscr{R}_0\{\boldsymbol{M}_*\}$. Thus *the formulas* (4.7.51) *and* (4.7.52) *determine the measurement which is optimal in the sense of mean-square deviation.*

A measurement $\boldsymbol{M} = \{M(\mathrm{d}\theta)\}$ with finite second moment is an *unbiased measurement* of the parameter θ if $E_\theta\{\boldsymbol{M}\} = \theta$, $\theta \in \mathbb{R}$, where the subscript θ refers to the state S_θ. As in Section 3.2 any covariant measurement satisfies $E_\theta\{\boldsymbol{M}\} = E_0\{\boldsymbol{M}\} + \theta$ and so is unbiased up to a constant. Assuming that $E_0\{\boldsymbol{M}\} = 0$ we have $D_\theta\{\boldsymbol{M}\} = \mathscr{R}_0\{\boldsymbol{M}\}$, $\theta \in \mathbb{R}$. We therefore have obtained *a bound for the variance of a covariant measurement*

$$D\{\boldsymbol{M}\} \geq D_*(\theta) \equiv \int_\Lambda \left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\|\psi_\lambda\|_\lambda\right|^2 \mathrm{d}\lambda \tag{4.7.53}$$

which is achieved on the optimal measurement determined by (4.7.51) and (4.7.52). The quantity $D_*(\theta)$ is the intrinsic minimal uncertainty of the shift parameter $\theta \in \mathbb{R}$ in the pure state $S = |\psi)(\psi|$. This is analogical to the quantity $\Delta_*(\varphi)$ of (4.5.28) giving the uncertainty of an angular parameter. Observing that

$$\int_\Lambda \left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\|\psi_\lambda\|_\lambda\right|^2 \mathrm{d}\lambda \geq \frac{1}{4}\left[\int_\Lambda (\lambda - c)^2 \|\psi_\lambda\|^2_\lambda \mathrm{d}\lambda\right]^{-1},$$

(this is just the Heisenberg uncertainty relation (3.3.29) in the momentum representation for the state function $\psi(\lambda) = \|\psi_\lambda\|_\lambda$, $\lambda \in \Lambda$ and $\psi(\lambda) = 0$, $\lambda \bar{\in} \Lambda$), we obtain the uncertainty relation

$$D_*(\theta) \cdot D(A) \geq \tfrac{1}{4},$$

which agrees with (3.2.12). A more general result will be established rigorously in Section 6.3.

As a first application we consider the covariant measurement of the coordinate parameter x in the family

$$S_x = \mathrm{e}^{-\mathrm{i}xP}|\psi)(\psi|\mathrm{e}^{\mathrm{i}xP}; \qquad x \in \mathbb{R}.$$

The role of A is played by momentum operator P (we take $\hbar = 1$), which is diagonal in the momentum representation

$$\psi = [\tilde{\psi}(\eta)], \qquad \|\psi\|^2 = \int_{-\infty}^{\infty} |\tilde{\psi}(\eta)|^2 \mathrm{d}\eta;$$

$$P\psi = [\eta\tilde{\psi}(\eta)].$$

The representation space $\mathscr{H} = \mathscr{L}^2(\mathbb{R})$ is the direct orthogonal integral of the one-dimensional spaces $\mathscr{H}_\eta$, $\eta \in \Lambda \equiv \mathbb{R}$. The optimal covariant measurement is given by the kernel

$$M_*(\mathrm{d}x) = [\gamma_\eta \bar{\gamma}_{\eta'} \mathrm{e}^{\mathrm{i}(\eta'-\eta)x}] \frac{\mathrm{d}x}{2\pi}, \tag{4.7.54}$$

which is the continuous analog of (4.4.22). Here $\gamma_\eta = \tilde{\psi}(\eta)/|\tilde{\psi}(\eta)|$. If $\arg \tilde{\psi}(\eta) = \text{const.}$ for almost all η, we get

$$M_*(\mathrm{d}x) = [\mathrm{e}^{\mathrm{i}(\eta'-\eta)x}] \frac{\mathrm{d}x}{2\pi}, \tag{4.7.55}$$

which is just a symbolic form of (2.4.45) to within the notations for the arguments. Thus the canonical measurement (4.7.55) is the spectral measure of the self-adjoint operator $\mathrm{i}\,\mathrm{d}/\mathrm{d}\eta$ representing the canonical coordinate observable Q in the momentum representation. Otherwise the optimal covariant measurement is given by (4.7.54) which is the spectral measure of the self-adjoint operator $\gamma_\eta \mathrm{i}(\mathrm{d}/\mathrm{d}\eta)\bar{\gamma}_\eta$. This gives the covariant estimate of coordinate x in the family $\{S_x\}$ with the minimal variance

$$D_*(x) = \int_{-\infty}^{\infty} \left(\frac{\mathrm{d}}{\mathrm{d}\eta} |\tilde{\psi}(\eta)| \right)^2 \mathrm{d}\eta.$$

The measurements (4.7.54) corresponding to different ψ are obtained one from another by "gauge" transformations $\tilde{\psi}(\eta) \to \gamma_\eta \tilde{\psi}(\eta)$ with $|\gamma_\eta| \equiv 1$.

The second example is the covariant measurements of time parameter t in the family

$$S_t = \mathrm{e}^{-\mathrm{i}tE}|\psi)(\psi|\mathrm{e}^{\mathrm{i}tE}, \qquad t \in \mathbb{R}.$$

We assume that the energy operator E has purely Lebesgue spectrum occupying $\Lambda = (0, \infty)$. Namely the Hilbert space is the direct orthogonal integral

$$\mathscr{H} = \int_0^\infty \bigoplus \mathscr{H}_\epsilon \, \mathrm{d}\epsilon,$$

and E acts as multiplication by $\epsilon : E\psi = [\epsilon\psi_\epsilon]$ where ψ_ϵ is the component of ψ in $\mathscr{H}_\epsilon$. The optimal covariant time measurement is given by the formula

$$M_*(\mathrm{d}t) = \left[\frac{|\psi_\epsilon)_\epsilon \cdot {}_{\epsilon'}(\psi_{\epsilon'}|}{\|\psi_\epsilon\|_\epsilon \cdot \|\psi_{\epsilon'}\|_{\epsilon'}} \mathrm{e}^{\mathrm{i}(\epsilon'-\epsilon)t} \right] \frac{\mathrm{d}t}{2\pi} \oplus M_1(\mathrm{d}t). \tag{4.7.56}$$

If $\mathscr{H} = \mathscr{L}^2_{\mathbb{K}}(0, \infty)$ as in Section 3.8, then we can take $\mathscr{H}_\epsilon = \mathbb{K}$, $\epsilon \in (0, \infty)$. This gives a special way of identification of $\mathscr{H}_\epsilon$, $\epsilon \in (0, \infty)$ to which corresponds the canonical measurement

$$M(\mathrm{d}t) = [\mathrm{e}^{\mathrm{i}(\epsilon'-\epsilon)t}] \frac{\mathrm{d}t}{2\pi},$$

where we have omitted the unit operator in $\mathbb{K}$ from the squared brackets. This is just a symbolic form of the general relation analogous to (3.8.74), representing the (nonorthogonal) spectral measure of the canonical time observable $T = \mathrm{i}\,\mathrm{d}/\mathrm{d}\epsilon$. On the other hand, the first, essential, term in the optimal time measurement (4.7.56) is the spectral measure of the operator

$$\frac{|\psi_\epsilon)_\epsilon}{\|\psi_\epsilon\|_\epsilon} \mathrm{i} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \frac{{}_\epsilon(\psi_\epsilon|}{\|\psi_\epsilon\|_\epsilon}$$

differing from T by the factors involving the basic state.

4.8. The case of irreducible representation

We now pass to quantum estimation problems with multidimensional parameter θ. Representations of symmetry groups in these problems will be irreducible due to either noncommutativity of the group or essential projectivity of the representation. On the other hand, the irreducibility considerably simplifies the description of the covariant measurements.

Let $g \to V_g$ be an irreducible representation of the group G in the Hilbert space $\mathscr{H}$. Assume first that G is compact; then by a general theorem of the representation theory $d = \dim \mathscr{H} < \infty$. For an irreducible representation the following *orthogonality relations* hold

$$\int (\psi_1|V_g\varphi_1)(\varphi_2|V_g^*\psi_2)\mu(\mathrm{d}g) = c(\varphi_2|\varphi_1)(\psi_1|\psi_2), \tag{4.8.57}$$

where the factor c, depending on the normalization of the invariant measure μ, is equal to d^{-1} if $\mu(G) = 1$.

Since any operator T in a finite-dimensional space is the finite-rank operator $T = \sum_i |\varphi_1^i)(\varphi_2^i|$, (4.8.57) and (2.1.17) imply

$$\int (\psi_1|V_g T V_g^* \psi_2)\mu(\mathrm{d}g) = d^{-1}(\psi_1|\psi_2)\operatorname{Tr} T, \qquad \psi_j \in \mathscr{H},$$

or

$$\int V_g T V_g^* \mu(\mathrm{d}g) = d^{-1}\operatorname{Tr} T \cdot I. \tag{4.8.58}$$

In particular, if S is a density operator, then

$$\int V_g S V_g^* \mu(\mathrm{d}g) = d^{-1} I. \tag{4.8.59}$$

Taking trace here we get the identity $\operatorname{Tr} V_g S V_g^* \equiv \operatorname{Tr} S = 1$, which confirms the correctness of the choice $c = d^{-1}$. If we take $c = 1$ this would correspond to normalization

$$\mu(G) = \dim \mathscr{H} = d. \tag{4.8.60}$$

Turning to Theorem 4.2.3 which gives a description of covariant measurements we see that for an irreducible representation the condition (4.2.4) due to (4.8.59) reduces to $\operatorname{Tr} P_0 = d$; together with positivity of P_0 this means that $S_0 = d^{-1} P_0$ is a density operator. We thus obtain:

Proposition 4.8.1. *Let $g \to V_g$ be an irreducible representation of a compact group G of transformations of the set Θ. The relation*

$$M(\mathrm{d}\theta) = d \cdot V_g S_0 V_g^* \nu(\mathrm{d}\theta) \qquad (\theta = g\theta_0), \tag{4.8.61}$$

establishes the one-to-one affine correspondence between the covariant measurements $M(\mathrm{d}\theta)$ and the density operators S_0, commuting with $\{V_g;\, g \in G_0\}$.

In particular, if G_0 is trivial, then S_0 may be an arbitrary density operator and the extreme points of the set of covariant measurements correspond precisely to pure states $S_0 = |\psi_0)(\psi_0|$:

$$M(\mathrm{d}\theta) = d \cdot V_g|\psi_0)(\psi_0|V_g^* \nu(\mathrm{d}\theta).$$

Consider now the *covariant estimation problem* for the family of states (4.3.10). According to Section 4.3 it reduces to solving the problem (4.3.15) where $\widehat{W}_0$ is the operator of posterior deviation expressed

through the data of the estimation problem by the formula (4.3.14). This has a simple solution in the irreducible case. The set $\mathfrak{P}$ consists now of the operators P_0 commuting with $\{V_g;\, g \in G_0\}$ and satisfying

$$P_0 \geq 0, \qquad \operatorname{Tr} P_0 = d.$$

Let $\hat{w}_{\min}$ be the least eigenvalue of $\widehat{W}_0$, and $\widehat{E}_{\min}$ be the projection onto the corresponding invariant subspace. Then $\widehat{W}_0 \geq \hat{w}_{\min} \cdot I$ whence

$$\operatorname{Tr} \widehat{W}_0 P_0 \geq \hat{w}_{\min} \operatorname{Tr} P_0 = \hat{w}_{\min} d.$$

The equality is achieved if

$$P_0 = \widehat{E}_{\min} \frac{d}{d_{\min}}, \tag{4.8.62}$$

where $d_{\min}$ is the dimensionality of the invariant subspace. Since $\widehat{W}_0$ commutes with $\{V_g;\, g \in G_0\}$, so does $\widehat{E}_{\min}$ and the operator (4.8.62) satisfies all the required conditions. We thus get

Proposition 4.8.2. *Let $g \to V_g$ be an irreducible representation of a compact group G acting on the set Θ. The optimal covariant measurement of the parameter θ is given by the relation*

$$M_*(\mathrm{d}\theta) = \frac{d}{d_{\min}} V_g \widehat{E}_{\min} V_g^* \nu(\mathrm{d}\theta) \qquad (\theta = g\theta_0),$$

where $\widehat{E}_{\min}$ is the projection onto the invariant subspace of the operator of posterior deviation (4.3.14)*, corresponding to the least eigenvalue $\hat{w}_{\min}$, and $d_{\min}$ is the dimensionality of the subspace. The minimal mean deviation is equal to $\hat{w}_{\min} d$.*

An analogous result holds for the maximum-likelihood approach: the covariant maximum-likelihood measurement of the parameter θ in the family $\{S_\theta\}$ is

$$M(\mathrm{d}\theta) = \frac{d}{d_{\max}} V_g E_{\max} V_g^* \nu(\mathrm{d}\theta) \qquad (\theta = g\theta_0),$$

where $E_{\max}$ is the projection onto the invariant subspace corresponding to the greatest eigenvalue of the density operator $S = S_{\theta_0}$, $d_{\max}$ is its dimensionality.

These results are sufficient for the concrete estimation problems to be considered in the following two paragraphs. We conclude this section

with a discussion of non-compact groups which has an application to joint measurements of coordinate and velocity considered in Section 3.6.

The orthogonality relations (4.8.57) hold for irreducible representations of an arbitrary parametric group under the condition of square-integrability of the matrix elements $(\psi_1|V_g\psi_2)$, $g \in G$, for all $\psi_j \in \mathscr{H}$. Normalizing μ in such a way that $c = 1$ we obtain

$$\int V_g S V_g^* \mu(\mathrm{d}g) = I \tag{4.8.63}$$

for an arbitrary density operator S, where the integral converges weakly. Using this fact we shall generalize Proposition 4.8.1. We now assume that the invariant measures μ and ν are normalized so that (4.8.63) holds (for a compact group this corresponds to (4.8.60)).

Theorem 4.8.3. *The relation*

$$M(B) = \int_B V_g S_0 V_g^* \nu(\mathrm{d}\theta), \qquad B \in \mathscr{A}(\Theta) \quad (\theta = g\theta_0), \tag{4.8.64}$$

establishes the one-to-one affine correspondence between the measurements $\{M(B)\}$ covariant with respect to an irreducible square-integrable representation $g \to V_g$ of the group G acting on the set Θ and the density operators S_0 commuting with $\{V_g;\, g \in G_0\}$. The integral is understood in the sense of weak convergence.[2]

We shall first prove two lemmas involving properties of trace-class operators from Section 2.7.

Lemma 4.8.4. *Let M be a Hermitean positive operator and $\{T_n\}$ be a monotonely non-decreasing sequence of Hermitean operators weakly converging to the unit operator. Then $\mathrm{Tr}\, T_n M \uparrow \mathrm{Tr}\, M$; in particular $\sup_n \mathrm{Tr}\, T_n M < \infty$ implies that M is a trace-class operator.*

Proof. Let $\{e_j\}$ be a basis in $\mathscr{H}$, then

$$\mathrm{Tr}\, T_n M = \mathrm{Tr}\, \sqrt{M} T_n \sqrt{M} = \sum_j (\sqrt{M} e_j | T_n \sqrt{M} e_j).$$

By the properties of $\{T_n\}$ we have

$$(\sqrt{M} e_j | T_n \sqrt{M} e_j) \uparrow (\sqrt{M} e_j | \sqrt{M} e_j) = (e_j | M e_j).$$

[2] Moreover if $\nu(B) < \infty$, then this converges as the Bochner's integral of a function with values in the Banach space of trace-class operators, see Dunford and Schwartz [31].

By the monotone convergence theorem

$$\sum_j (\sqrt{M}e_j|T_n\sqrt{M}e_j) \uparrow \sum_j (e_j|Me_j) = \operatorname{Tr} M.$$

□

Lemma 4.8.5. *Let* $\{M(B)\}$ *be a covariant measurement. Then* $\operatorname{Tr} M(B) = \nu(B)$. *In particular if* $\nu(B) < \infty$, *then* $M(B)$ *is trace-class.*

Proof. Let $\{B_n\}$ be a non-decreasing sequence of compact subsets, covering Θ. Consider

$$T_n = \int_{B_n} V_g S V_g^* \mu(\mathrm{d}g), \tag{4.8.65}$$

where S is a density operator. By positivity of the integrand $T_n \geq 0$ and

$$\operatorname{Tr} T_n \leq \int_{B_n} \operatorname{Tr} V_g S V_g^* \mu(\mathrm{d}g) = \mu(B_n) < \infty$$

so that T_n is trace-class. Since $B_n \subseteq B_{n+1}$ we have $T_n \leq T_{n+1}$. Moreover $T_n \to I$ weakly by (4.8.63). The sequence $\{T_n\}$ satisfies Lemma 4.8.4, so that $\operatorname{Tr} M(B) = \lim_n \operatorname{Tr} T_n M(B)$. But according to (4.2.3)

$$\lim_n \operatorname{Tr} T_n M(B) = \lim_n \int_{B_n} \operatorname{Tr} V_g S V_g^* M(B) \mu(\mathrm{d}g) = \nu(B),$$

which proves the lemma. □

Proof of Theorem 4.8.3. Let S_0 be a density operator satisfying the conditions of the theorem. Defining $M(B)$ by the weakly converging integral (4.8.64) we get a resolution of identity. Indeed $M(B) \geq 0$ since the integrand is positive. The weak σ-additivity is a property of the integral. The normalization $M(\Theta) = I$ follows from (4.8.63) and the covariance is a straightforward calculation.

Conversely, let $\{M(B)\}$ be a covariant measurement. Then by Lemma 4.8.5 it satisfies the condition of Proposition 4.2.2, which is established as in the proof of Theorem 4.2.3. Therefore

$$M(B) = \int_B P(\theta)\nu(\mathrm{d}\theta),$$

where $P(\theta)$ is an operator-valued density, positive and satisfying $\|P(\theta)\| \leq 1$ for ν-almost all θ. As in Theorem 4.2.3 covariance implies $P(\theta) = V_g P(\theta_0) V_g^*$, $\theta = g\theta_0$, for the proper definition of the density. By (4.8.63) $P(\theta_0)$ must have the unit trace, *i.e.*, it must be a density operator. The theorem is proved. □

As an example consider the group G of shifts $(\xi, \eta) \to (\xi + x, \eta + v)$ of the plane $\Theta = \mathbb{R}^2$. The stationary subgroup G_0 is trivial in this case. It follows from the uniqueness theorem of Section 3.4 that any (continuous) irreducible projective unitary representation $(x, v) \to W_{x,v}$ of the group is unitary equivalent to the Schrödinger representation with $\mu \neq 0$. Then by Proposition 3.5.1 it is square-integrable; the orthogonality relations hold with $c = 1$ corresponding to the invariant measure $\mu\, \mathrm{d}x\, \mathrm{d}v/2\pi$. Theorem 4.8.3 implies the *any measurement* $M(\mathrm{d}x\, \mathrm{d}v)$, *covariant with respect to the representation* $(x, v) \to W_{x,v}$ has the form

$$M(\mathrm{d}x\, \mathrm{d}v) = W_{x,v} S_0 W_{x,v}^* \frac{\mu\, \mathrm{d}x\, \mathrm{d}v}{2\pi}.$$

Thus (3.6.52) describes precisely the extreme points of the convex set of covariant measurements.

Turning to the optimality question we note that $\mathscr{R}\{\boldsymbol{M}\} = g_x D_x\{\boldsymbol{M}\} + g_v D_v\{\boldsymbol{M}\}$ is an affine functional of a measurement $\boldsymbol{M}$, which is finite if $\boldsymbol{M}$ has finite second moments; by a general theorem of convex analysis (see comments to this chapter) it achieves its minimum at an extreme point of the convex set of covariant measurements $M(\mathrm{d}x\, \mathrm{d}v)$ having finite second moments. Together with (3.6.58) this leads to

Proposition 4.8.6. *The canonical measurement* (3.6.59) *is the optimal covariant joint measurement of coordinate and velocity, minimizing the measure of accuracy* $\mathscr{R}\{\boldsymbol{M}\}$.

4.9. Estimation of pure state

Let $\mathscr{H}$ be a Hilbert space of finite dimensionality n. Denote by Θ the unit sphere in $\mathscr{H}$; the elements of Θ are the vectors $|\theta) \in \mathscr{H}$ satisfying $(\theta|\theta) = 1$. The set Θ can be parametrized as follows: let $\{e_j\}$ be a basis in $\mathscr{H}$, then

$$|\theta) = \sum_{j=1}^{n} \theta_j |e_j),$$

where $\theta_j = (e_j|\theta)$; $j = 1, \ldots, n$, are complex numbers satisfying

$$\sum_{j=1}^{n} |\theta_j|^2 = \sum_{j=1}^{n} [(\mathrm{Re}\,\theta_j)^2 + (\mathrm{Im}\,\theta_j)^2] = 1. \tag{4.9.66}$$

To any $\theta \in \Theta$ corresponds a pure state

$$S_\theta = |\theta)(\theta|. \tag{4.9.67}$$

The parametrization of pure states is not one-to-one, since vectors differing by a complex factor of unit modulus describe the same pure state, but this is not important here.

Assume that a quantum object described by the Hilbert space $\mathscr{H}$ is prepared in a pure state of which "nothing is known" and the problem is to estimate the state as accurately as possible using measurements admitted by quantum theory. In other words we need to estimate the parameter θ in the family (4.9.67).

Consider the group G of all unitary operators U in $\mathscr{H}$. Since the unit vector is transformed by U into a unit vector, this is a compact group of transformations of the set Θ. The normalized invariant measure $\nu(\mathrm{d}\theta)$ on Θ is up to a factor the Euclidean area on the real $2n$-dimensional unit sphere (4.9.66). Considering the action of the unitary operators in $\mathscr{H}$ as the representation of G we see that the family (4.9.67) is covariant

$$US_\theta U^* = |U\theta)(U\theta| = S_{U\theta}.$$

The representation $U \to U$ is apparently irreducible.

Let us describe measurements, covariant with respect to this representation. Fix $\theta_0 = e_1$, and consider the stationary subgroup G_0 of θ_0. Clearly it consists of unitary operators of the form

$$U_0 = \begin{bmatrix} \lambda & 0 \\ 0 & U_0' \end{bmatrix},$$

where $|\lambda| = 1$ and U_0' is a unitary operator in the orthogonal complement to $|\theta_0)$. A Hermitean operator S_0 commutes with all $U_0 \in G_0$ if and only if

$$S_0 = \alpha|\theta_0)(\theta_0| + \beta I,$$

where α, β are real, *i.e.*,

$$S_0 = \begin{bmatrix} \alpha+\beta & & & 0 \\ & \beta & & \\ & & \beta & \\ 0 & & & \ddots \end{bmatrix}.$$

If S_0 is a density operator: $S_0 \geq 0$ and $\mathrm{Tr}\, S_0 = 1$, then $\alpha + \beta \geq 0$, $\beta \geq 0$ and $n\beta + \alpha = 1$, *i.e.*,

$$\beta = (1-\alpha)/n; \qquad -(n-1)^{-1} \leq \alpha \leq 1.$$

By Proposition 4.8.1 *any covariant measurement of parameter θ has the form*

$$M(\mathrm{d}\theta) = U[\alpha n|\theta_0)(\theta_0| + (1-\alpha)I]U^*\nu(\mathrm{d}\theta), \qquad (\theta = U\theta_0),$$

where α is a real number running over the segment $-(n-1)^{-1} \leq \alpha \leq 1$. It follows that the set of covariant measurements, as a convex set, is a segment. An extreme point of this set corresponding to $\alpha = 1$ is

$$M_*(\mathrm{d}\theta) = nU|\boldsymbol{\theta}_0)(\boldsymbol{\theta}_0|U^*\nu(\mathrm{d}\theta) = n|\boldsymbol{\theta})(\boldsymbol{\theta}|\nu(\mathrm{d}\theta). \tag{4.9.68}$$

This is apparently the maximum-likelihood measurement for the family (4.9.67). Another important measurement corresponding to $\alpha = 0$ is

$$M^*(\mathrm{d}\theta) = I \cdot \nu(\mathrm{d}\theta).$$

This resolution of identity describes simple guessing, when the result of measurement is chosen at random with respect to the uniform distribution $\nu(\mathrm{d}\theta)$.

The simplest invariant deviation function on Θ is

$$\delta = 1 - |(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})|^2;$$

we shall consider more general deviation functions

$$W_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = W(\delta). \tag{4.9.69}$$

Proposition 4.9.1. *The covariant measurement* (4.9.68) *is optimal for a deviation function of the form* (4.9.69) *where $W(\cdot)$ is an arbitrary monotonely nondecreasing function, which is not identically constant.*

Proof. Since the mean deviation is an affine functional of measurement it is sufficient to show that

$$\mathscr{R}\{\boldsymbol{M}_*\} < \mathscr{R}\{\boldsymbol{M}^*\},$$

i.e., that

$$\int W_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) n|(\boldsymbol{\theta}_0|\boldsymbol{\theta})|^2 \nu(\mathrm{d}\theta) < \int W_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\nu(\mathrm{d}\theta).$$

Denoting $r = |(\boldsymbol{\theta}_0|\boldsymbol{\theta})|$ rewrite it in the form

$$n \int W(1-r^2) r^2 \nu(\mathrm{d}\theta) < \int W(1-r^2)\nu(\mathrm{d}\theta).$$

By Lemma 4.9.2 proven below this is equivalent to

$$n \int_0^1 W(1-r^2) r^2 \mathrm{d}(1-r^2)^{n-1} > \int_0^1 W(1-r^2)\mathrm{d}(1-r^2)^{n-1},$$

or

$$\int_0^1 W(1-r^2)[(n-1)(1-r^2)^{n-2} - n(1-r^2)^{n-1}]\mathrm{d}r^2 < 0.$$

Passing to the new variable $\delta = 1 - r^2$ we have

$$\int_0^1 W(\delta)\mathrm{d}(\delta^{n-1} - \delta^n) < 0.$$

Integrating by parts we obtain the relation

$$\int_0^1 (\delta^{n-1} - \delta^n)\mathrm{d}W(\delta) > 0,$$

which apparently holds for a nonconstant nondecreasing $W(\cdot)$, since $\delta^{n-1} - \delta^n > 0$ for $0 < \delta < 1$. □

To demonstrate the gain due to the optimal measurement we give the values of mean deviation in case of the simplest deviation function $W(\delta) = \delta$;

$$\mathscr{R}\{\boldsymbol{M}_*\} = n\int_0^1 \delta(1-\delta)\mathrm{d}\delta^{n-1} = \frac{n-1}{n+1},$$
$$\mathscr{R}\{\boldsymbol{M}^*\} = \int_0^1 \delta\,\mathrm{d}\delta^{n-1} = \frac{n-1}{n}.$$

The ratio $\mathscr{R}\{\boldsymbol{M}_*\}/\mathscr{R}\{\boldsymbol{M}^*\} = n/(n+1)$ is minimal and equal to $\frac{2}{3}$ for the two-dimensional Hilbert space and increases to 1 as $n \to \infty$. Thus in an infinite-dimensional Hilbert space there is no better may to estimate the unknown pure state than simple guessing.

Lemma 4.9.2. *For any function* $F(r)$

$$\int_\Theta F(|\theta_0|\theta)|)\nu(\mathrm{d}\theta) = -\int_0^1 F(r)\mathrm{d}(1-r^2)^{n-1}.$$

Proof. Choose a basis $\{e_j\}$ such that $e_1 = \theta_0$ and denote $(e_j|\theta) = \alpha_j + \mathrm{i}\beta_j$. Then

$$\Theta = \left\{\alpha_j, \beta_j : \sum_{j=1}^n (\alpha_j^2 + \beta_j^2) = 1\right\}$$

and $r = \sqrt{\alpha_1^2 + \beta_1^2}$. We shall prove the lemma if we prove that

$$\int_{\alpha_1^2+\beta_1^2 \geq \rho^2} \nu(\mathrm{d}\theta) = (1-\rho^2)^{n-1}. \tag{4.9.70}$$

Consider the auxiliary integral

$$F_m(\rho, R) = \int\cdots\int_{\substack{x_1^2+x_2^2\geq\rho^2\\ x_1^2+\cdots+x_m^2\leq R^2}} \mathrm{d}x_1\cdots\mathrm{d}x_m.$$

The area of the unit sphere is then $(\partial F_n(0, R)/\partial R)|_{R=1}$; therefore the normalized area of the figure cut of the unit sphere by the inequality $x_1^2 + x_2^2 \geq \rho^2$ is given by the expression

$$\left.\frac{\partial F_m(\rho, R)}{\partial R}\right|_{R=1} : \left.\frac{\partial F_m(0, R)}{\partial R}\right|_{R=1}.$$

But this normalized area is the left-hand side of (4.9.70) if $m = 2n$. We have

$$\begin{aligned} F_m(\rho, R) &= \int\cdots\int\left\{\iint_{\rho^2\leq x_1^2+x_2^2\leq R^2-x_3^2-\cdots-x_m^2}\mathrm{d}x_1\,\mathrm{d}x_2\right\}\mathrm{d}x_3\cdots\mathrm{d}x_m \\ &= \int\cdots\int\pi(R^2-\rho^2-x_3^2-\cdots-x_m^2)\mathrm{d}x_3\cdots\mathrm{d}x_m, \end{aligned}$$

the integration being performed over the domain where the integrand is nonnegative, *i.e.*, $x_3^2+\cdots+x_m^2 \leq R^2-\rho^2$. Denoting by $S_{n-2}(r) = cr^{m-3}$ the area of the sphere $x_3^2 + \cdots + x_m^2 = r^2$, we have

$$F_m(\rho, R) = c\pi\int_0^{\sqrt{R^2-\rho^2}}(R^2-\rho^2-r^2)S_{m-2}(r)\mathrm{d}r = c_1(R^2-\rho^2)^{n/2},$$

whence $(\partial/\partial R)F_m(\rho, R)|_{R=1} = c_2(1-\rho^2)^{(m-2)/2}$, so that

$$\left.\frac{\partial F_{2n}(\rho, R)}{\partial R}\right|_{R=1} : \left.\frac{\partial F_{2n}(0, R)}{\partial R}\right|_{R=1} = (1-\rho^2)^{n-1},$$

and (4.9.70) is proved. □

4.10. Measuring parameters of orientation

Here we shall consider the estimation of orientation of a quantum object by measurements involving only spin degrees of freedom. We first assume the prepared basic state S to be invariant under rotations about the given symmetry axis $\boldsymbol{n}_0$:

$$S = \sum_{m=-j}^{j} s_m|m)(m|,$$

where j is the spin of the object, $\{|m)\}$ are the eigenvectors of the spin angular momentum J_0 corresponding to the axis $\boldsymbol{n}_0$ (see Section 3.11). If the preparing apparatus is then rotated so that the symmetry axis is transformed into $\boldsymbol{n} = g\boldsymbol{n}_0$, where g is the element of the rotation group, the new state will be $S_{\boldsymbol{n}} = V_g S V_g^*$, where $g \to V_g$ is the irreducible projective unitary representation of the rotation group in the $(2j+1)$-dimensional Hilbert space spanned by $\{|m)\}$.

We thus have the covariant family of quantum states

$$S_{\boldsymbol{n}} = V_g S V_g^* \qquad (\boldsymbol{n} = g\boldsymbol{n}_0), \quad n \in \mathbb{S}^2,$$

where $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$.

Orientation of the object is now described by the unit vector $\boldsymbol{n}$ pointing in the direction of the symmetry axis. Assume that the actual direction $\boldsymbol{n}$ is unknown and the problem is to estimate it basing on quantum measurement $M(\mathrm{d}n)$.

We shall adopt the deviation function

$$W_{\boldsymbol{n}}(\hat{\boldsymbol{n}}) = |\boldsymbol{n} - \hat{\boldsymbol{n}}|^2 = 2(1 - \boldsymbol{n} \cdot \hat{\boldsymbol{n}}), \tag{4.10.71}$$

which is apparently invariant. As the prior distribution in the Bayes' approach we shall take the normalized invariant measure $\nu(\mathrm{d}n)$ on $\mathbb{S}^2$.

According to Proposition 4.8.1 any covariant measurement of the direction $\boldsymbol{n}$ is given by the formula

$$M(\mathrm{d}n) = (2j+1) V_g S_0 V_g^* \nu(\mathrm{d}n) \qquad (\boldsymbol{n} = g\boldsymbol{n}_0),$$

where S_0 is a density operator commuting with $\{V_g; g \in G_0\}$. Since G_0 consists of rotations about the axis $\boldsymbol{n}_0$, this is equivalent for S_0 to have the diagonal form in the basis $\{|m)\}$. To find the optimal covariant measurement we need to calculate the operator of posterior deviation

$$\widehat{W}_0 = 2 \int_G (1 - \boldsymbol{n}_0 \cdot g\boldsymbol{n}_0) V_g^* S V_g \mu(\mathrm{d}g),$$

where $\mu(\mathrm{d}g)$ is the normalized invariant measure on the rotation group. It is shown at the end of this section that

$$\widehat{W}_0 = \frac{2}{2j+1} \left[I - \frac{\operatorname{Tr} S J_0}{j(j+1)} J_0 \right]. \tag{4.10.72}$$

According to Proposition 4.8.2 we need to know the least eigenvalue and the corresponding eigenvector of (4.10.72). Since J_0 has the known spectral representation (3.11.116) we find the eigenvalue $\hat{w}_{\min} = 2(1 -$

$|\bar{J}_0|(j+1)^{-1})$, where $\bar{J}_0 = \operatorname{Tr} S J_0$ and the corresponding eigenvector $|j)$ if $\bar{J}_0 > 0$ or $|-j)$ if $\bar{J}_0 < 0$.

Denoting $|m, \boldsymbol{n}) \equiv V_g|m)$, where $\boldsymbol{n} = g\boldsymbol{n}_0$, we see from Proposition 4.8.2 that the *optimal covariant measurement of direction of the symmetry axis* $\boldsymbol{n}$ *is*

$$\begin{aligned} M_*(\mathrm{d}n) &= (2j+1)V_g|\pm j)(\pm j|V_g^*\nu(\mathrm{d}n) \\ &= (2j+1)|\pm j; \boldsymbol{n})(\pm j; \boldsymbol{n}|\nu(\mathrm{d}n) \qquad (\boldsymbol{n} = g\boldsymbol{n}_0), \end{aligned}$$

where the sign $\pm$ *corresponds to the sign of* $\bar{J}_0$. *The minimal mean deviation is equal to* $\hat{w}_{\min} = 2(1 - |\bar{J}_0|(j+1)^{-1})$. Note that simple guessing which is described by the resolution of identity $M^*(\mathrm{d}n) = I \cdot \nu(\mathrm{d}n)$ gives the mean deviation $\mathscr{R}\{\boldsymbol{M}^*\} = \operatorname{Tr} \widehat{W}_0 = 2$ so that

$$\frac{\mathscr{R}\{\boldsymbol{M}_*\}}{\mathscr{R}\{\boldsymbol{M}^*\}} = 1 - \frac{|\bar{J}_0|}{j+1}.$$

This shows that the gain due to the optimal measurement increases with the value of ratio $|\bar{J}_0|(j+1)^{-1}$ and vanishes if $\bar{J}_0 = 0$.

According to Section 4.3 the measurement $M_*(\mathrm{d}n)$ is optimal in both Bayes and minimax senses. The maximum-likelihood approach would give, in general, a different result. As follows from the remark after Proposition 4.8.2 the covariant maximum-likelihood measurement coincides with $M_*(\mathrm{d}n)$ only if the vector $|\pm j)$ corresponds to the greatest eigenvalue of the basic density operator S.

Consider now the problem of estimation of orientation without assuming rotational symmetry of the basic state S. The orientation is then described by fixing a Cartesian frame $\theta = \{\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3\}$ in $\mathbb{R}^3$ bound to the preparing apparatus. Let $\theta_0 = \{\boldsymbol{n}_1^0, \boldsymbol{n}_2^0, \boldsymbol{n}_3^0\}$ be a fixed basic frame. Rotation g is uniquely defined by the transformed frame $\theta = g\theta_0 \equiv \{g\boldsymbol{n}_1^0, g\boldsymbol{n}_2^0, g\boldsymbol{n}_3^0\}$. Thus the set Θ of all frames can be identified with the rotation group G, and the relevant covariant family of states in $S_\theta = V_g S V_g^*$, $\theta = g\theta_0$, where g runs over the rotation group.

The deviation of the frame $\hat{\theta} = \{\hat{\boldsymbol{n}}_1, \hat{\boldsymbol{n}}_2, \hat{\boldsymbol{n}}_3\}$ from the frame θ will be taken in the form

$$W_\theta(\hat{\theta}) = \sum_{i=1}^{3} |\boldsymbol{n}_i - \hat{\boldsymbol{n}}_i|^2 = 2\sum_{i=1}^{3}(1 - \boldsymbol{n}_i \cdot \hat{\boldsymbol{n}}_i). \tag{4.10.73}$$

The corresponding operator of posterior deviation

$$\widehat{W}_0 = 2\sum_{i=1}^{3} \int [1 - \boldsymbol{n}_i^0 \cdot g\boldsymbol{n}_i^0] V_g^* S V_g \mu(\mathrm{d}g)$$

using (4.10.72) is equal to

$$\widehat{W}_0 = \frac{2}{2j+1}\sum_{i=1}^{3}\left[I - \frac{\bar{J}_i}{j(j+1)}J_i\right] = \frac{2}{2j+1}\left[3 - \frac{\sqrt{\sum_{i=1}^{3}\bar{J}_i^2}}{j(j+1)}\sum_{i=1}^{3}\alpha_i J_i\right],$$

where J_i is the spin angular momentum corresponding to the axis $\boldsymbol{n}_i^0$, $\bar{J}_i = \operatorname{Tr} S J_i$ and $\alpha_i = \bar{J}_i(\sum_i \bar{J}_i^2)^{-1/2}$. The operator $\sum_i \alpha_i J_i$ represents spin angular momentum corresponding to the axis $\bar{\boldsymbol{n}} = \alpha_1 \boldsymbol{n}_1^0 + \alpha_2 \boldsymbol{n}_2^0 + \alpha_3 \boldsymbol{n}_3^0$. It follows from Section 3.11 that its maximal eigenvalue is equal to j; the corresponding eigenvector is apparently $|j;\bar{\boldsymbol{n}})$. Then from Proposition 4.8.2 *the optimal covariant measurement of orientation* $\theta = \{\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3\}$ *is given by*

$$M_*(\mathrm{d}\theta) = (2j+1)V_g|j;\bar{\boldsymbol{n}})(\bar{\boldsymbol{n}};\, j|V_g^*\nu(\mathrm{d}\theta) \qquad (\theta = g\theta_0).$$

The minimal mean deviation is equal to $2[3 - \sqrt{\sum_i \bar{J}_i^2}(j+1)^{-1}]$, so that the gain due to the optimal measurement is

$$\frac{\mathscr{R}\{\boldsymbol{M}_*\}}{\mathscr{R}\{\boldsymbol{M}^*\}} = 1 - \frac{1}{3}\frac{\sqrt{\sum_{i=1}^{3}\bar{J}_i^2}}{j+1},$$

and has the same nature as in the previous problem.

To prove (4.10.72) we calculate the matrix elements of $\widehat{W}_0$ in the basis $\{|m)\}$. For this we need the matrix elements of the representation operators $\{V_g\}$. Let $0 \le \theta \le \pi, 0 \le \psi < 2\pi, 0 \le \varphi < 2\pi$ be the Euler angles describing the rotation g. Then

$$\mu(\mathrm{d}g) = (8\pi^2)^{-1}\mathrm{d}(\cos\theta)\mathrm{d}\psi\,\mathrm{d}\varphi.$$

Moreover

$$(n|V_g|m) = \mathrm{e}^{-\mathrm{i}m\psi - \mathrm{i}n\varphi}P_{mn}^{j}(\cos\theta),$$
$$P_{mn}^{j}(t) = K(1-t)^{\alpha/2}(1+t)^{\beta/2}P_s^{\alpha\beta}(t),$$

where $P_s^{\alpha\beta}$ are the Jacobi polynomials, $\alpha = |n-m|$, $\beta = |n+m|$, $s = j - \frac{1}{2}(\alpha+\beta)$ (see Comments). Integrating by φ and ψ and using the

relation $1 - \boldsymbol{n}_0 \cdot g\boldsymbol{n}_0 = 1 - \cos\theta$ we get

$$(m|\widehat{W}_0|m') = \delta_{mm'} \sum_{n=-j}^{j} (n|S|n) \int_{-1}^{1} (1-t)|P_{mn}^{j}(t)|^2 \mathrm{d}t.$$

It remains to calculate

$$\int_{-1}^{1} (1-t)|P_{mn}^{j}(t)|^2 \mathrm{d}t = K^2 \int_{-1}^{1} (1-t)^{\alpha+1}(1+t)^{\beta}|P_{s}^{\alpha\beta}(t)|^2 \mathrm{d}t, \quad (4.10.74)$$

where the constant K is defined by the normalization

$$K^2 \cdot \int_{-1}^{1} (1-t)^{\alpha}(1+t)^{\beta}|P_{s}^{\alpha\beta}(t)|^2 \mathrm{d}t = 2(2j+1)^{-1}.$$

Using recurrent formulas for the Jacobi polynomials, we obtain

$$(1-t)P_{s}^{\alpha\beta}(t) = \left[1 + \frac{\alpha^2 - \beta^2}{4j(j+1)}\right] P_{s}^{\alpha\beta}(t) + AP_{s+1}^{\alpha\beta}(t) + BP_{s-1}^{\alpha\beta}(t),$$

where A and B are constants. Since $\{P_s^{\alpha\beta}; s = 0, 1, \ldots\}$ is an orthogonal system on the interval $(-1, 1)$ with the weight $(1-t)^{\alpha}(1+t)^{\beta}$ then by the normalization, the integral is equal to the factor of $P_s^{\alpha\beta}$ in the right-hand side of (4.10.74) multiplied by $2(2j+1)^{-1}$, *i.e.*,

$$\frac{2}{2j+1}\left[1 + \frac{\alpha^2 - \beta^2}{4j(j+1)}\right] = \frac{2}{2j+1}\left[1 - \frac{nm}{j(j+1)}\right].$$

It follows that

$$\begin{aligned}\widehat{W}_0 &= \sum_{m,m'} |m)(m|\widehat{W}_0|m')(m'| \\ &= \frac{2}{2j+1}\left[I - \frac{1}{j(j+1)} \sum_{n=-j}^{j} n(n|S|n) \sum_{m=-j}^{j} m|m)(m|\right]\end{aligned}$$

which is the required formula (4.10.72).

4.11. Comments

Section 4.1. The classical treatise on continuous groups is Pontrjagin's book [113]. A detailed exposition of the compact parametric groups and their representations addressed both to mathematicians and physicists is presented in the book of Želobenko [148]. Kirillov [75] gives a concentrated account of the general representation theory and related topics of functional analysis.

The notion of covariant measurement extends onto general resolutions of identity the important notion of *imprimitivity system* (which is in the present terminology just simple covariant measurement) introduced by Murray and von Neumann and studied by Mackey [91]. Presentations of Mackey's results with applications of quantum theory can be found in Varadarajan [136] and Jauch [71]. The covariant measurements arise also from the "covariant instruments" of Davies [26].

Section 4.2. The results of this section are due to Davies [26] and Holevo [59, 64, 67]. Concerning operator-valued integration and the corresponding Radon-Nikodym theorem see, *e.g.*, Dunford and Schwarz [31].

Section 4.3. Mathematical statistics (see, *e.g.*, Cramer [23], Ferguson [34]) presents a natural framework for statistical measurement theory of classical systems. If the unknown parameter θ is finitely-valued, then one speaks of "hypotheses testing"; if it is continuous one speaks of "estimation".

A possibility of fruitful application of the ideas of mathematical statistics to quantum measurement problems was first demonstrated by Helstrom in the paper [50] devoted to optimal discrimination between two quantum states. In Holevo [57, 58] the general resolutions of identity were introduced as the noncommutative analog of classical randomized procedures to develop quantum statistical decision theory. A detailed physical account of quantum estimation and hypotheses testing theory is presented in Helstrom's book [53]. Mathematical consideration of the general Bayes' problem including existence questions and the relevant integration theory can be found in the author's work [59, 64]. The noncommutative Hunt-Stein theorem appeared in [67]. Quantum maximum-likelihood approach was developed in [59]. For relations with the classical maximum-likelihood see [64].

Section 4.4. Optimal angular measurements (4.4.22) were derived by Helstrom [52] from the maximum-likelihood principle. He also showed that they are Bayesian if the deviation function is $4\sin^2(\varphi - \hat{\varphi})/2$. Theorem 4.4.1 was proved by Holevo [67].

Section 4.5. There were several attempts to extend uncertainty relations to angular quantities. Some of them use a variational measure of uncertainty as in Judge [73], others are content with the uncertainty relations (4.5.30) for C and S operators, see Louisell [87], Carruthers and Nieto [20]. The inequality based on the covariant uncertainty (4.5.27) which is one of the type used in classical statistics of angular observations (see *e.g.*, Mardia [98]) seems to be rather new.

Section 4.6. The results of this and the next section are generalized naturally to the case of parameter with values in arbitrary Abelian locally compact group, see §2.3 of [178].

Section 4.7. Concerning the direct orthogonal integral of Hilbert spaces see, *e.g.*, Gelfand and Vilenkin [40]. The material of this section is taken from Holevo [66], where the coordinate measurements for the three-dimensional non-relativistic particle and for the photon are also discussed. The study of Newton and Wigner [105] (see also Wightman [142], Varadarajan [136]) shows that a relativistic zero-mass object such as photon is "nonlocalizable" in the sense that it has no coordinate observable, *i.e.*, spectral measure satisfying the appropriate covariance condition. Jauch and Piron [72] pointed out that this disagrees with experimental evidence for photon localizability and discussed two possibilities to describe it theoretically. The first one, which was ultimately adopted by the authors, uses a nonadditive projection-valued set function, see also Amrein [5]. The alternative approach using a covariant non-orthogonal resolution of identity was developed by Holevo [66], where a rigorous uncertainty relation for photon coordinates was also obtained.

Section 4.8. The covariant measurements in the case of irreducible representation were studied by Holevo [59, 64, 67].

The well-known theorem of Bauer [9] says that an affine upper semicontinuous functional on a compact convex subset of a locally convex Hausdorff topological space attains its minimum at an extreme point of the subset. The applicability of this theorem to the functional $\mathscr{R}\{\boldsymbol{M}\} = g_x D_{(x)}\{\boldsymbol{M}\} + g_v D_{(v)}\{\boldsymbol{M}\}$ follows from Theorem 7.1 of [64].

Section 4.9. The pure state estimation problem was discussed by Helstrom [52, 53] who observed that (4.9.68) is the maximum-likelihood measurement. The results of this section were obtained by the author. The "full model" in which the multidimensional parameter is the quantum state itself deserves special attention both for its importance for applications and for special mathematical properties. One considers the analog of the sample of repeated independent observations $S_\theta^{\otimes n} = S_\theta \otimes \cdots \otimes S_\theta$, where S_θ is an unknown state in d-dimensional Hilbert space $\mathcal{H}$. The case of unknown pure state $S_\theta = |\theta)(\theta|$ is the most studied one, with special attention paid to qubit state (i. e. $d = 2$). For the deviation functions of the type (4.9.69) the optimal resolution of the identity is found in the symmetrized tensor product of n copies of the space $\mathcal{H}$ and has the form $M_*(d\theta) = S_\theta^{\otimes n} \nu_n(d\theta)$, where $\nu_n(d\theta)$ is unitarily invariant measure on the variety of pure states. These estimates are consistent as $n \to \infty$. The asymptotic theory for the full model in the case of pure

states based on Rao-Cramér type inequalities was developed in [165], while [172], [182] used the large deviations approach.

A fundamental distinction of the quantum estimation theory appears in consideration of the asymptotic properties of estimates in a sample of independent identical systems. In the paper [177], devoted to asymptotic estimation of the shift parameter, it was pointed out that statistical information in quantum models with independent observations can be strictly superadditive. This property, similar to superadditivity of the Shannon information in quantum information theory (see Section 5.1 in [179]), means that statistical information for a composite system with independent components can be strictly greater than sum of informations from the subsystems. This property has a deep physical meaning: it is due to the existence of entangled measurements in the composite system and as such is dual to Einstein-Podolsky-Rosen correlations for entangled quantum states (see Section 2.4 of the Supplement). This superadditivity was established for the full model [165, 172, 188].

The full model clearly displays another feature of quantum estimation problem: the complexity sharply increases with passage from pure to mixed states. (Note that in the classical statistics estimation of pure states is trivial since pure classical states are just distributions degenerated at different points.) The estimation problem for an arbitrary state can be split into two: first, estimation of the spectrum i. e. eigenvalues of the density matrix and second, estimation of the eigenvectors. Estimation of the spectrum requires new ideas: in [182] a solution based on representation theory for the permutation group of n elements was proposed. A consistent and asymptotically efficient estimate is provided by lengths of the rows of the Young diagram related to an irreducible representation of the permutation group.

The noncommutative analog of the famous Le Cam's Local Asymptotic Normality for the problem of estimation of an arbitrary mixed state of a finite dimensional quantum system was established by Guta and Kahn following the work of Hayashi and Matsumoto (see [170] and references therein). The proofs involve diverse mathematical tools, notably representation theory for SU(d) and its tensor degrees as well as for the symmetric group. It is remarkable that while the permutation symmetry is present already in the classical estimation problem with i.i.d. observation, it is only the quantum case which requires the full power of advanced representation theory.

Section 4.10. The material of this section is taken from Holevo [67]. The vectors $|j;\boldsymbol{n})$; are the "coherent state vectors" for the representation of the rotation group introduced by Radcliffe [117] and Perelo-

mov [108,109]. The formula for the matrix elements of the representation is derived, *e.g.*, in Wigner [145] or in Gelfand, Minlos and Shapiro [41].

Chapter 5
Gaussian states

5.1. Quasiclassical states of the quantum oscillator

Consider a quantum degree of freedom, *e.g.*, an oscillator, described by the canonical observables $q = Q$, $p = \hbar P$. The ground state $|0)(0|$ is the minimum-uncertainty state in which Q and P have zero mean values. The state

$$|\overline{P}, \overline{Q})(\overline{Q}, \overline{P}| = W_{\overline{Q},\hbar\overline{P}}|0)(0|W^*_{\overline{Q},\hbar\overline{P}} \tag{5.1.1}$$

(where to simplify notations we put $|\overline{P}, \overline{Q}) \equiv |\overline{P}, \overline{Q}; \hbar/2\omega)$, ω being the oscillator frequency) can be regarded as the result of an external influence onto the object in ground state, shifting the mean values of the canonical observables but leaving their uncertainties unchanged.

Let us now assume that the influence has random nature, *i.e.*, the parameters $\overline{P}$ and $\overline{Q}$ are random variables with probability distribution $\mu(\mathrm{d}\overline{P}\,\mathrm{d}\overline{Q})$. From the point of view of an experimenter who observes the given quantum object but not the source of influence, determining the values $\overline{P}$, $\overline{Q}$ in an individual experiment, the object's state is statistically described by the density operator

$$S = \int |\alpha, \beta)(\beta, \alpha|\mu(\mathrm{d}\alpha\,\mathrm{d}\beta), \tag{5.1.2}$$

where we changed $\overline{P}$, $\overline{Q}$ to α, β. This represents averaging of (5.1.1) with respect to the distribution of $\overline{P}$ and $\overline{Q}$. The mean value of any real-valued measurement $\boldsymbol{M}$ is then the average of the means $E_{\overline{P},\overline{Q}}\{\boldsymbol{M}\}$ corresponding to the states (5.1.1):

$$E_S\{\boldsymbol{M}\} = \int E_{\alpha,\beta}\{\boldsymbol{M}\}\mu(\mathrm{d}\alpha\,\mathrm{d}\beta).$$

In particular the mean values of the canonical observables are equal to those of the classical probability distribution μ

$$E_S(P) = \int \alpha\mu(\mathrm{d}\alpha\,\mathrm{d}\beta), \qquad E_S(Q) = \int \beta\mu(\mathrm{d}\alpha\,\mathrm{d}\beta).$$

The states represented by (5.1.2) are called *quasiclassical*; the relation (5.1.2) defines a map from the simplex of probability distributions μ into the convex set of quantum states. This map can be shown to be one-to-one; it follows that there are quantum states which cannot be represented in the quasiclassical form (5.1.2) with a probability distribution μ. Otherwise the set of quantum states would be simplex.

Assume that be object is influenced by a large number k of independent identical sources, so that the resulting influence is characterized by parameters

$$\overline{P} = \sum_{j=1}^{k} \overline{P}_j, \qquad \overline{Q} = \sum_{j=1}^{k} \overline{Q}_j,$$

where $(\overline{P}_j, \overline{Q}_j)$ are independent identically distributed pairs of random variables. Then by the classical central limit theorem the distribution $\mu(\mathrm{d}\alpha\,\mathrm{d}\beta)$ of $(\overline{P}, \overline{Q})$ will be approximately Gaussian as $k \to \infty$. The resulting state is a particular case of quantum Gaussian states to be discussed in this chapter.

Introducing the complex variable $\zeta = (2\hbar\omega)^{-1/2}(\omega\beta + \mathrm{i}\hbar\alpha)$ so that $|\alpha, \beta) = |\zeta)$ is the vector of coherent state (see Section 3.10), consider the special quasiclassical Gaussian state

$$S = \frac{1}{\pi\overline{N}} \int |\zeta)(\zeta| \mathrm{e}^{-|\zeta|^2/\overline{N}} \mathrm{d}^2\zeta. \tag{5.1.3}$$

The real parameter $\overline{N}$ is equal to the mean value of the number of quanta N. To show this we calculate the matrix elements of S in the basis $\{|n)\}$ of the eigenvectors of N. We have

$$(n|S|m) = \frac{1}{\pi\overline{N}} \int (n|\zeta)(\zeta|m) \mathrm{e}^{-|\zeta|^2/\overline{N}} \mathrm{d}^2\zeta.$$

Using (3.10.102) and denoting $\zeta = r\mathrm{e}^{\mathrm{i}\varphi}$ we get

$$\begin{aligned}(n|S|m) &= \frac{1}{\pi\overline{N}} \int_0^{2\pi} \mathrm{e}^{\mathrm{i}(n-m)\varphi} \mathrm{d}\varphi \int_0^{\infty} r\,\mathrm{d}r \cdot r^{n+m} \mathrm{e}^{-r^2(\overline{N}+1)/\overline{N}} \\ &= \delta_{nm} \frac{1}{\overline{N}+1} \left(\frac{\overline{N}}{\overline{N}+1}\right)^n.\end{aligned}$$

It follows that the density operator S is diagonal in the Fock representation, namely

$$S = \frac{1}{\overline{N}+1} \sum_{n=0}^{\infty} \left(\frac{\overline{N}}{\overline{N}+1}\right)^n |n)(n|. \tag{5.1.4}$$

Calculating the mean value of number of quanta we get

$$\frac{1}{\overline{N}+1}\sum_{n=0}^{\infty} n\left(\frac{\overline{N}}{\overline{N}+1}\right)^n = \overline{N},$$

as required.

The basic role in statistical physics is played by the so called Gibbs states. If the energy of the object can take discrete series of values $\{E_n\}$, then the Gibbs state is defined as the mixture of the states S_n corresponding to definite values of the energy, the weight of the state S_n in the mixture being proportional to $\exp(-E_n/kT)$. Here T is the absolute temperature, k is Boltzmann's constant. A common belief strongly supported by numerous model considerations is that the Gibbs state is the equilibrium state to which the object is driven by infinitely long interaction with surrounding media at the temperature T. Since for the harmonic oscillator $E_n = \hbar\omega(n+\frac{1}{2})$, we see that (5.1.4) is the Gibbs state at the temperature T, if we put

$$\overline{N} = \frac{1}{\mathrm{e}^{\hbar\omega/kT}-1}.$$

For this reason one says that the state (5.1.4) or (5.1.3) describes "thermal noise" of the quantum harmonic oscillator in thermal equilibrium with the media at the temperature T. The mean values of canonical observables are zero for this state since the Gaussian distribution in (5.1.3) has zero mean.

If the oscillator which was initially in the equilibrium state S suffers an external influence described by the displacement operator $W_{\overline{Q},\hbar\overline{P}} \equiv W_{\bar{a}}$, where $\overline{a} = (2\hbar\omega)^{-1/2}(\omega\overline{Q} + \mathrm{i}\hbar\overline{P})$, then the new state is

$$S_{\bar{a}} = W_{\bar{a}} S W_{\bar{a}}^*. \tag{5.1.5}$$

Taking into account (3.10.97) we get

$$S_{\bar{a}} = \frac{1}{\pi\overline{N}}\int |\zeta)(\zeta|\mathrm{e}^{-|\zeta-\bar{a}|^2/\overline{N}}\mathrm{d}^2\zeta, \tag{5.1.6}$$

so that the mean values of the canonical observables are

$$E_S(Q) \equiv \overline{Q} = \sqrt{2\hbar/\omega}\,\mathrm{Re}\,\overline{a}, \quad E_S(P) \equiv \overline{P} = \sqrt{2\omega/\hbar}\,\mathrm{Im}\,\bar{a}.$$

The states of the form (5.1.6) and their analogs for many degrees of freedom are widely used in quantum optics for the description of radiation fields, both chaotic as natural light and coherent as those generated by lasers. The free electromagnetic field is known to be mathematically

equivalent to an infinite collection of harmonic oscillators. For our purposes it is sufficient to consider a finite collection of oscillators with frequencies ω_j; $j = 1, \ldots, s$. Let P_j, Q_j; $j = 1, \ldots s$, be the canonical observables of this "cut-off radiation field". Then the thermal background radiation in the absence of external sources is described by the density operator

$$S = \bigotimes_j S^{(j)}, \tag{5.1.7}$$

where $S^{(j)}$ is the equilibrium state (5.1.3) for the j^{th} oscillator with the mean number of quanta equal to $\overline{N}_j = (\exp(\hbar\omega_j/kT) - 1)^{-1}$. Influence of a source results in a change of the state of the radiation field. If the simplest model of the influence (5.1.5) is adopted, then the resulting state will be described by the density operator

$$S_{\bar{a}} = \bigotimes_j S^{(j)}_{\bar{a}_j}, \tag{5.1.8}$$

where $S^{(j)}_{\bar{a}_j}$ is the state of the form (5.1.6) for the j^{th} oscillator. The multidimensional complex parameter $\bar{\boldsymbol{a}} = [\bar{a}_j]$ characterizes the source of influence. Thus the state (5.1.8) describes "signal plus noise".

The states of the type (5.1.7), (5.1.8) possess certain attractive analytical properties which are interesting both for studying physical models and also from a mathematical point of view. Since these properties are essentially due to the "Gaussian" character of the states, it is convenient to look upon them from a more general point of view and abstract from the concrete representation (5.1.7), (5.1.8) characteristic for the states of the radiation field. In this chapter we introduce and study a general class of quantum Gaussian states displaying remarkable analogies with the classical Gaussian probability distributions.

5.2. The CCR for many degrees of freedom

Let us rearrange the CCR (3.3.22) for one degree of freedom by introducing two-component vectors $z = [x, y]$ and the skew-symmetric form

$$\Delta(z, z') = xy' - x'y.$$

By putting $V(z) = W_{-x,y/\mu}$, the CCR takes the form

$$V(z)V(z') = \mathrm{e}^{\mathrm{i}\Delta(z,z')/2}V(z + z'). \tag{5.2.9}$$

In the case of s degrees of freedom we act in the same way. Let x_k, y_k be a pair of real numbers; we put $z_k = [x_k, y_k]$ and $z = [z_1, \ldots, z_s]$.

Thus z will denote real $2s$-dimensional vectors. Introduce the bilinear skew-symmetric form

$$\Delta(z, z') = \sum_{k=1}^{s} (x_k y'_k - x'_k y_k).$$

As in the case $s = 1$ we call the *representation of the CCR* (with s degrees of freedom) any continuous family of unitary operators $z \to V(z)$ in a Hilbert space $\mathscr{H}$ satisfying (5.2.9).

The canonical observables P_k, Q_k; $k = 1, \ldots, s$, arise from $\{V(z)\}$ in the following way. Since Δ is skew-symmetric,

$$\Delta(z, z) \equiv 0, \tag{5.2.10}$$

and therefore by (5.2.9) the family $\{V(tz), t \in \mathbb{R}\}$ for a fixed z is a group of unitary operators. By Stone's theorem

$$V(tz) = \mathrm{e}^{\mathrm{i}tR(z)}, \tag{5.2.11}$$

where $R(z)$ is a self-adjoint operator. From (5.2.9), (5.2.11)

$$\begin{aligned} \mathrm{e}^{\mathrm{i}tR(z)}\mathrm{e}^{\mathrm{i}t'R(z')} &= \mathrm{e}^{\mathrm{i}tt'\Delta(z,z')/2}\mathrm{e}^{\mathrm{i}R(tz+t'z')} \\ &= \mathrm{e}^{\mathrm{i}tt'\Delta(z,z')}\mathrm{e}^{\mathrm{i}t'R(z')}\mathrm{e}^{\mathrm{i}tR(z)}. \end{aligned} \tag{5.2.12}$$

Differentiation with respect to t and t' at the point $t = t' = 0$ gives formally

$$[R(z), R(z')] = -\mathrm{i}\Delta(z, z')I. \tag{5.2.13}$$

A rigorous version of this commutation relation will be obtained in Section 5.4. Let e_k be the vector $z = [z_1, \ldots, z_s]$ such that $z_j = 0$ for $j \neq k$ and $z_k = [1, 0]$, and h_k be the similar vector with $z_k = [0, 1]$. Putting $R(e_k) = P_k$, $R(h_k) = Q_k$ and observing that

$$\Delta(e_k, h_l) = \delta_{kl}, \qquad \Delta(e_k, e_l) = \Delta(h_k, h_l) = 0, \tag{5.2.14}$$

we get

$$[P_k, Q_l] = -\mathrm{i}\delta_{kl}, \qquad [P_k, P_l] = [Q_k, Q_l] = 0,$$

i.e., the Heisenberg commutation relations (3.6.50) for s degrees of freedom.

From (5.2.12) formally $R(z + z') = R(z) + R(z')$ so that

$$R(z) = \sum_k (x_k P_k + y_k Q_k),$$

since $z = \sum_k (x_k e_k + y_k h_k)$. This relation can be given precise meaning if the right-hand side is understood as the (unique) self-adjoint extension of the sum defined on a common dense domain (we shall not go into detail here). Therefore we can write

$$V(z) = \exp\left[\mathrm{i}\sum_{k=1}^{s}(x_k P_k + y_k Q_k)\right].$$

Operators $R(z)$ will also be called the *canonical observables*.

It is useful to look at this construction from the "coordinate-free" point of view. Let Z be an arbitrary real linear space and $\Delta(z, z')$ a bilinear skew-symmetric form on Z. We assume that it is *nondegenerate*, *i.e.*, if $\Delta(z, z') = 0$ for all $z \in Z$, then $z' = 0$. The pair (Z, Δ) is called a *symplectic space*. For any symplectic space we can define a *representation of the CCR* as the family of unitary operators $\{V(z), z \in Z\}$ satisfying (5.2.9) for all $z, z' \in Z$ (and an appropriate continuity condition). Then the *canonical observables* $\{R(z), z \in Z\}$ can be defined as in (5.2.11).

This approach allows Z to have infinite dimensionality. However since we agreed to confine ourselves to a finite number of degrees of freedom, we assume from now on that Z is finite-dimensional. It follows that the dimensionality is necessarily even: $\dim Z = 2s$. To see it introduce an inner product α on Z and denote the resulting Euclidean space by (Z, α). Let $\mathscr{D}$ be the *associated operator of the form* Δ in (Z, α), *i.e.*,

$$\Delta(z, z') = \alpha(z, \mathscr{D} z'); \qquad z, z' \in Z. \tag{5.2.15}$$

By the properties of the form Δ the operator $\mathscr{D}$ is a nondegenerate skew-symmetric ($\mathscr{D}^* = -\mathscr{D}$) operator in (Z, α). By a theorem in linear algebra there is an orthonormal basis $\tilde{e}_1, \tilde{h}_1; \tilde{e}_2, \tilde{h}_2; \ldots$ in (Z, α) in which $\mathscr{D}$ has the matrix of the form

$$\left[\begin{array}{cc|cc|c} 0 & d_1 & 0 & & \\ -d_1 & 0 & & & \\ \hline & & 0 & d_2 & \\ 0 & & -d_2 & 0 & \\ \hline & & & & \ddots \end{array}\right], \quad d_j > 0. \tag{5.2.16}$$

In particular, since $\mathscr{D}$ is nondegenerate, Z needs to have even dimensionality.

A basis $\{e_j, h_j;\ j = 1, \ldots, s\}$ in (Z, Δ) is called *symplectic* if it satisfies (5.2.14). In a symplectic basis the form Δ has the canonical coordinate representation $\Delta(z, z') = \sum(x_j y'_j - x'_j y_j)$, where $z = \sum(x_j e_j +$

$y_j h_j)$, $z' = \sum(x'_j e_j + y'_j h_j)$. Thus symplectic bases in a symplectic space play the same role as orthonormal bases in an Euclidean space. For any symplectic basis the observables $P_j = R(e_j)$, $Q_j = R(h_j)$ satisfy the Heisenberg commutation relations.

A particular symplectic basis $\{e_j, h_j\}$ in the space of real $2s$-dimensional vectors was described before (5.2.14). In any symplectic space there are plenty of symplectic bases. Indeed, let α be an arbitrary inner product on Z; then the basis $e_j = d_j^{-1/2}\tilde{e}_j$, $h_j = d^{-1/2}\tilde{h}_j$; $j = 1, \ldots, s$ is symplectic by (5.2.15) and (5.2.16). Summing up and denoting $a_j = d_j^{-1}$ we have

Proposition 5.2.1. *A finite-dimensional symplectic space (Z, Δ) has necessarily even dimensionality $2s$. For any inner product α there is a symplectic basis $\{e_j, h_j;\ j = 1, \ldots, s\}$ in (Z, Δ) in which α has the diagonal matrix of the form*

$$\left[\begin{array}{cc|cc|c} a_1 & 0 & 0 & & \\ 0 & a_1 & & & \\ \hline 0 & & a_2 & 0 & \\ & & 0 & a_2 & \\ \hline & & & & \ddots \end{array}\right], \qquad a_j > 0.$$

Transition from one symplectic basis to another is described by a *symplectic operator* T, satisfying

$$\Delta(Tz, Tz') = \Delta(z, z'); \qquad z, z' \in Z. \tag{5.2.17}$$

For any symplectic operator $|\det T| = 1$. Indeed, let α be an inner product in Z; then (5.2.17) reads

$$\alpha(Tz, \mathscr{D}\, Tz') = \alpha(z, \mathscr{D}\, z'); \qquad z, z' \in Z,$$

i.e., $T^* \mathscr{D}\, T = \mathscr{D}$, where T^* is the operator adjoint to T in the Euclidean space (Z, α). Since $\det T^* = \det T$ and $\det \mathscr{D} \neq 0$, $(\det T)^2 = 1$.

Let $z = \sum(x_j e_j + y_j h_j)$ be the coordinate representation of vector z in a symplectic basis. Introduce the Lebesgue measure in Z putting

$$d^{2s}z = \mathrm{d}x_1\, \mathrm{d}y_1 \cdots \mathrm{d}x_s\, \mathrm{d}y_s.$$

It follows that the measure is invariant under the symplectic transformations and the definition does not depend on the particular choice of a symplectic basis.

As an illustration consider the simplest case of $\dim Z = 2$ (one degree of freedom). Let $\{e, h\}$ be a symplectic basis in (Z, Δ), *i.e.*, $\Delta(e, h) = 1$

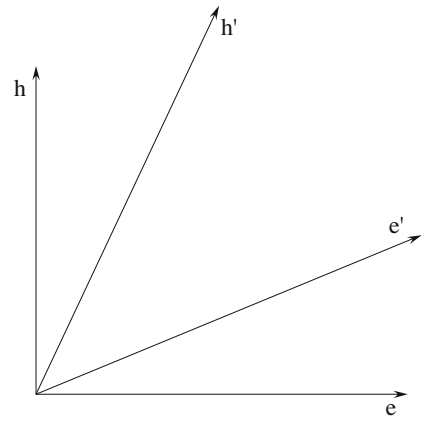

Figure 5.1.

and x, y be the components of vector z. Let us choose a Cartesian frame on the plane and take the unit vectors of the frame for e and h (Figure 5.1). The fact that e and h are orthogonal and have unit length at this picture does not reflect any geometrical property of symplectic basis, and is just a matter of arbitrariness. However, this can be interpreted by introducing the inner product $\alpha(z, z') = xx' + yy'$ in Z which is related to the basis as in Proposition 5.2.1. From what was said above it follows that a different pair of vectors $\{e', h'\}$ constitutes a symplectic basis if and only if the area of the oriented parallelogram with the sides e', h' is equal to $+1$.

5.3. Proof of the Stone-von Neumann uniqueness theorem. The Weyl transform

Let $z \to V(z)$ be a representation of the CCR, $f(z)$ a complex Lebesgue integrable function on a symplectic space (Z, Δ). The integral

$$V(f) = (2\pi)^{-s} \int f(z) V(-z) \mathrm{d}^{2s} z \tag{5.3.18}$$

is well defined in the sense of weak convergence[1]. The correspondence $f \to V(f)$ is often called the *Weyl transform*. The following properties follow easily from the definition and the CCR (5.2.9):

(1) $$V(f(z))^* = V(\overline{f(-z)});$$

(2) $$V(f_1)V(f_2) = V(f_1 \times f_2),$$

[1] Moreover it converges as the Bochner's integral of a function with values in the Banach space of all bounded operators.

where

$$f_1 \times f_2(z) = (2\pi)^{-s} \int f_1(w) f_2(z-w) \mathrm{e}^{\mathrm{i}\Delta(w,z)/2} \mathrm{d}^{2s} w$$

$$V(f(z))V(w) = V(f(z+w)\mathrm{e}^{\mathrm{i}\Delta(w,z)/2}), \tag{3}$$

$$V(w)^* V(f(z))V(w) = V(f(z)\mathrm{e}^{\mathrm{i}\Delta(w,z)}).$$

Moreover the correspondence $f \to V(f)$ is one-to-one: $V(f) = 0$ implies $f(z) = 0$ for almost all $z \in Z$. Indeed, from (3)

$$\int \mathrm{e}^{\mathrm{i}\Delta(w,z)} f(z)(\varphi | V(-z)\psi) \mathrm{d}^{2s} z = 0; \quad w \in Z;\ \varphi, \psi \in \mathscr{H},$$

whence by the uniqueness property of the usual Fourier transform $f(z)(\varphi | V(-z)\psi) = 0$ so that $f(z) = 0$ for almost all $z \in Z$.

It follows that if $z \to V_j(z)$, $j = 1, 2$ are two representations of the CCR, then the correspondence $V_1(f) \leftrightarrow V_2(f)$ is one-to-one. According to the properties (1) and (2), this correspondence preserves algebraic operations and the Hermitean conjugation. In fact there is the stronger Stone-von Neumann uniqueness theorem (for finite number degrees of freedom);

Theorem 5.3.1. *Any two* (*continuous*) *irreducible representation of the CCR are unitary equivalent. Any representation is the direct orthogonal sum of irreducible representations.*

Proof. Introducing the inner product

$$j(z, z') = \sum_{k=1}^{s} (x_k x'_k + y_k y'_k)$$

in Z where $[x_k, y_k]$, $[x'_k, y'_k]$ are the components of the vectors z, z' in a fixed symplectic basis, consider the function

$$f_0(z) = \mathrm{e}^{-j(z,z)/4}.$$

Put $P = V(f_0)$. Since $f_0(z) > 0$, $P \neq 0$. Using the properties (2) and (3) we obtain after some calculation the important identity

$$PV(w)P = f_0(w)P; \qquad w \in Z. \tag{5.3.19}$$

It follows that $P^2 = P$. Moreover since f_0 is real, $P^* = P$ by (1). Thus P is a projection onto a subspace $\mathscr{M}$ of the representation space $\mathscr{H}$.

If $\varphi, \psi \in \mathscr{M}$, then using (5.2.9) and (5.3.19)

$$\begin{aligned}(V(z)\varphi|V(w)\psi) &= (V(z)P\varphi|V(w)P\psi)\\ &= \mathrm{e}^{\mathrm{i}\Delta(w,z)/2}(\varphi|PV(w-z)P\psi)\\ &= \mathrm{e}^{\mathrm{i}\Delta(w,z)/2}f_0(w-z)(\varphi|\psi).\end{aligned} \tag{5.3.20}$$

Let $\{e_\alpha\}$ be an orthonormal basis in $\mathscr{M}$. Then (5.3.20) implies that the subspaces $\mathscr{M}_\alpha = [V(z)e_\alpha]$ generated by the vectors of the form $V(z)E_\alpha$, $z \in Z$, are orthogonal for different α. By construction $\mathscr{M}_\alpha$ are invariant subspaces of the representation operators $V(z)$, $z \in Z$. Therefore if the representation is irreducible then $\dim \mathscr{M} = 1$. The converse is also true: if the representation $z \to V(z)$ is reducible, then $\mathscr{H} = \mathscr{H}_1 \oplus \mathscr{H}_2$ where $\mathscr{H}_j$ are invariant subspaces; then applying the above construction to $\mathscr{H}_j$ instead of $\mathscr{H}$ we would get the projections P_j in $\mathscr{H}_j$ with $P = P_1 \oplus P_2$ so that $\dim \mathscr{M} > 1$.

Now from (5.3.19)

$$PV(z)e_\alpha = PV(z)Pe_\alpha = f_0(z)e_\alpha;$$

so that P is an operator of rank 1 in $\mathscr{M}_\alpha$

$$P\psi = c(\psi)e_\alpha, \qquad \psi \in \mathscr{M}_\alpha,$$

therefore $z \to V(z)$ acts irreducibly in $\mathscr{M}_\alpha$. Let us prove that $\mathscr{H} = \bigoplus_\alpha \mathscr{M}_\alpha$. Denote by $\mathscr{H}_0$ the orthogonal complement of $\bigoplus_\alpha \mathscr{M}_\alpha$ in $\mathscr{H}$. Then $\mathscr{H}_0$ is an invariant subspace of $\{V(z)\}$; moreover $P\mathscr{H}_0 = 0$. It follows that $\mathscr{H}_0 = 0$. Otherwise applying the whole construction to $\mathscr{H}_0$ instead of $\mathscr{H}$ we would obtain $P = 0$ which contradicts $f_0 \not\equiv 0$. This proves that any representation is the direct orthogonal sum of irreducible representations acting in the subspace $\mathscr{M}_\alpha$.

Let $z \to V_j(z)$ be irreducible representations of the CCR in Hilbert spaces $\mathscr{H}_j$, $j = 1, 2$. Then $P_j = V_j(f_0)$ is the one-dimensional projection onto the unit vector $e_j \in \mathscr{H}_j$ and $\mathscr{H}_j = [V(z)e_j]$. Define the operator U from $\mathscr{H}_2$ to $\mathscr{H}_1$ putting $UV_2(z)e_2 = V_1(z)e_1$. Then U maps a dense set in $\mathscr{H}_2$ onto a dense set in $\mathscr{H}_1$, preserving the values of inner products, since by (5.3.20)

$$\begin{aligned}(V_1(z)e_1|V_1(w)e_1) &= \exp\left[\frac{1}{2}\mathrm{i}\Delta(w,z)\right]f_0(z-w)\\ &= (V_2(z)e_2|V_2(w)e_2).\end{aligned}$$

Therefore it can be extended by continuity to an isometric map from $\mathscr{H}_2$ onto $\mathscr{H}_1$. By construction

$$U^* V_1(z) U = V_2(z); \qquad z \in Z,$$

and the theorem is proved. □

We now restrict to irreducible representations. The following multidimensional analog of Proposition 3.5.1 is proved in the same way as for $s = 1$.

Proposition 5.3.2. *Let $z \to V(z)$ be irreducible representation of the CCR in a Hilbert space $\mathscr{H}$. Then the matrix elements $(\varphi|V(z)\psi)$ are square-integrable functions of z. If $\{e_j\}$ is an orthonormal basis in $\mathscr{H}$, then the functions $\{(2\pi)^{-1/2}(e_j|V(z)e_k)\}$ form an orthonormal basis in the space $\mathscr{L}^2(Z)$ of complex square-integrable functions on Z.*

Let T be a trace-class operator in the representation space. Define

$$\mathscr{F}_z[T] = \operatorname{Tr} T V(z); \qquad z \in Z. \tag{5.3.21}$$

As we shall see, the "noncommutative Fourier transform" $T \to \mathscr{F}_z[T]$ is inverse to the Weyl transform (5.3.18). The following properties result from the CCR and general properties of trace (see Section 2.7):

(1) $\mathscr{F}_0[T] = \operatorname{Tr} T; \qquad |\mathscr{F}_z[T]| \le \|T\|_1;$

(2) $\mathscr{F}_z[T^*] = \overline{\mathscr{F}_{-z}[T]};$

(3) $\mathscr{F}_z[T V(w)] = \mathscr{F}_{z+w}[T] \cdot \mathrm{e}^{\mathrm{i}\Delta(z,w)/2},$

$\mathscr{F}_z[V(w)^* T V(w)] = \mathscr{F}_z[T] \cdot \mathrm{e}^{\mathrm{i}\Delta(z,w)}.$

The transform (5.3.21) satisfies the following "noncommutative Parceval relation".

Theorem 5.3.3. *The map $T \to \mathscr{F}_z[T]$ extends uniquely to an isometric map from the Hilbert space $\mathfrak{T}^2(\mathscr{H})$ of Hilbert-Schmidt operators in $\mathscr{H}$ onto $\mathscr{L}^2(Z)$, so that*

$$\operatorname{Tr} T_1^* T_2 = (2\pi)^{-s} \int \overline{\mathscr{F}_z[T_1]} \mathscr{F}_z[T_2] \mathrm{d}^{2s} z; \qquad T_j \in \mathfrak{T}^2(\mathscr{H}). \tag{5.3.22}$$

Proof. If we prove that for any Hermitean trace-class operator

$$\operatorname{Tr} T^2 = (2\pi)^{-s} \int |\mathscr{F}_z[T]|^2 \mathrm{d}^{2s} z,$$

then (5.3.22) for trace-class operators will follow by polarization and linearity. Let $\{t_j\}$ be the eigenvalues of T, $\{e_j\}$ the corresponding eigenvectors. Then by Theorem 2.7.1 the trace of $TV(z)$ is equal to

$$\mathscr{F}_z[T] = \sum_j t_j(e_j|V(z)e_j), \tag{5.3.23}$$

where $\sum |t_j| < \infty$. Using Proposition 5.3.2 we see that the series converges in $\mathscr{L}^2(Z)$ and

$$(2\pi)^{-s}\int |\mathscr{F}_z[T]|^2 \mathrm{d}^{2s}z = \sum_j t_j^2 |(e_j|e_j)|^2 = \mathrm{Tr}\, T^2.$$

Thus $T \to \mathscr{F}_z[T]$ maps isometrically the set of trace-class operators, which is a dense subspace of the Hilbert space $\mathfrak{T}^2(\mathscr{H})$ into the Hilbert space $\mathscr{L}^2(Z)$. Since the functions $\mathscr{F}_z[|e_j)(e_k|] = (e_k|V(z)e_j)$ form a basis in $\mathscr{L}^2(Z)$, the range of this map is dense in $\mathscr{L}^2(Z)$. Therefore the map $T \to \mathscr{F}_z[T]$ uniquely extends by continuity to the isometric map of $\mathfrak{T}^2(\mathscr{H})$ onto $\mathscr{L}^2(Z)$ so that (5.3.22) holds for Hilbert-Schmidt operators. □

Corollary 5.3.4. *The state S in pure if and only if*

$$(2\pi)^{-s}\int |\mathscr{F}_z[S]|^2 \mathrm{d}^{2s}z = 1.$$

Proof. By Theorem 5.3.3

$$(2\pi)^{-s}\int |\mathscr{F}_z[S]|^2 \mathrm{d}^{2s}z = \mathrm{Tr}\, S^2 = \sum_j s_j^2,$$

where s_j are the eigenvalues of the density operator S. But $s_j \geq 0$, $\sum s_j = 1$ whence $\sum s_j^2 \leq 1$ with the sign of equality achieved if one and only if one of s_j is equal to 1, and the others are zero. This means that S is a one-dimensional projection, *i.e.*, a density operator of a pure state. □

Corollary 5.3.5. *For any Hilbert-Schmidt operator T*

$$T = (2\pi)^{-s}\int \mathscr{F}_z[T]V(-z)\mathrm{d}^{2s}z, \tag{5.3.24}$$

where the integral converges weakly.

Proof. Putting $T_1 = |\varphi)(\psi|$ in (5.3.22) and taking into account that

$$(\psi|V(z)\varphi) = (\varphi|V(z)^*\psi) = (\varphi|V(-z)\psi)$$

we get

$$(2\pi)^{-s}\int \mathscr{F}_z[T](\varphi|V(-z)\psi)\mathrm{d}^{2s}z = \mathrm{Tr}\,|\psi)(\varphi|T = (\varphi|T\psi), \quad (5.3.25)$$

which is just the detailed form of (5.3.24). The "inversion formula" (5.3.24) relates the transforms $T \to f(z) = \mathscr{F}_z[T]$ and $f \to T = V(f)$ showing that they are mutually inverse:

$$T = V(\mathscr{F}_z[T]).$$

In particular the map $T \to \mathscr{F}_z[T]$ in one-to-one. □

The inversion formula enables us to obtain the expression for the kernel of an operator T in any irreducible representation through the function $\mathscr{F}_z[T]$. As an example we consider the Schrödinger representation for one degree of freedom, and prove the following relation for the kernel of a Hilbert-Schmidt operator T in $\mathscr{L}^2(\mathbb{R})$:

$$(\xi|T|\xi') = (2\pi)^{-1}\int \mathscr{F}_{\xi-\xi',y}[T]\mathrm{e}^{-\mathrm{i}(\xi+\xi')y/2}\mathrm{d}y, \quad (5.3.26)$$

where

$$\mathscr{F}_{x,y}[T] = \mathrm{Tr}\,TV(x,y) \equiv \mathrm{Tr}\,TW_{-x,y/\mu}.$$

Indeed from (3.4.32)

$$(\varphi|V(-x,-y)\psi) = \int (\varphi|\xi)\mathrm{e}^{-\mathrm{i}y(\xi-x/2)}(\xi - x|\psi)\mathrm{d}\xi \quad (5.3.27)$$

for any $\varphi, \psi \in \mathscr{H}$. Therefore by (5.3.25)

$$\iint (\varphi|\xi)(\xi|T|\xi')(\xi'|\psi)\mathrm{d}\xi\,\mathrm{d}\xi'$$
$$= (2\pi)^{-1}\iiint \mathscr{F}_{x,y}[T](\varphi|\xi)\mathrm{e}^{-\mathrm{i}y(\xi-x/2)}(\xi - x|\psi)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\xi$$

with all the functions being square-integrable so that it is possible to perform integration in arbitrary order. Putting $\xi' = \xi - x$ and using arbitrariness of φ, ψ we get (5.3.26).

In the same way using (5.3.21) and (2.7.83) we obtain

$$\mathscr{F}_{x,y}[T] = \int (\xi + x|T|\xi)\mathrm{e}^{\mathrm{i}y(\xi+x/2)}\mathrm{d}\xi.$$

Using this relation and (3.5.42) we get the Weyl transform of the density operator (5.1.1):

$$\begin{aligned}&\mathscr{F}_{x,y}[|\overline{P},\overline{Q})(\overline{P},\overline{Q}|]\\&= \exp\left[\mathrm{i}(x\overline{P} + y\overline{Q}) - \frac{1}{4}\left(\frac{\hbar}{\omega}x^2 + \frac{\omega}{\hbar}y^2\right)\right].\end{aligned} \tag{5.3.28}$$

We now also have tools for

Proof of Proposition 3.6.1. Consider the classical characteristic function of the probability distribution of the measurement $E(\mathrm{d}x\,\mathrm{d}v)$ with respect to the state $S \otimes S_0$. According to (2.7.78) it is

$$\iint \mathrm{e}^{\mathrm{i}(\xi x+\eta\mu v)}\mu^{\boldsymbol{E}}_{S\otimes S_0}(\mathrm{d}x\,\mathrm{d}v) = \mathrm{Tr}\, S\otimes S_0\mathrm{e}^{\mathrm{i}(\xi\overline{Q}+\eta\overline{P})}.$$

Using (3.6.54) we get

$$\mathrm{e}^{\mathrm{i}(\xi\overline{Q}+\eta\overline{P})} = \mathrm{e}^{\mathrm{i}(\xi Q+\eta P)} \otimes \mathrm{e}^{\mathrm{i}(-\xi Q_0+\eta P_0)},$$

so that the characteristic function is equal to

$$\mathrm{Tr}\, S\mathrm{e}^{\mathrm{i}(\xi Q+\eta P)} \cdot (\overline{\psi}|\mathrm{e}^{\mathrm{i}(-\xi Q_0+\eta P_0)}\overline{\psi}).$$

Using (3.4.32), (3.4.37) we can reduce the second factor to the form

$$\begin{aligned}(\overline{\psi}|W_{-\eta,-\xi/\mu}\overline{\psi}) &= \int \psi(\lambda)\mathrm{e}^{-\mathrm{i}\xi(\lambda+\eta/2)}\overline{\psi(\lambda+\eta)}\mathrm{d}\lambda\\&= \overline{(\psi|W_{-\eta,\xi/\mu}\psi)} = \overline{\mathrm{Tr}\, S_\psi \mathrm{e}^{\mathrm{i}(\xi Q_0+\eta P_0)}}.\end{aligned}$$

Since

$$\mathrm{Tr}\, S\mathrm{e}^{\mathrm{i}(\eta P+\xi Q)} = \mathscr{F}_{\eta,\xi}[S],$$

the characteristic function of the probability distribution $\mu^{\boldsymbol{E}}_{S\otimes S_0}(\mathrm{d}x\,\mathrm{d}v)$ is equal to

$$\mathscr{F}_{\eta,\xi}[S] \cdot \overline{\mathscr{F}_{\eta,\xi}[S_\psi]}.$$

Since by Theorem 5.3.3 both factors are Lebesgue square-integrable their product is an integrable function. Therefore the inverse Fourier transform

$$\frac{\mu}{(2\pi)^2}\iint \mathrm{e}^{-\mathrm{i}(\xi x+\eta\mu v)}\mathscr{F}_{\eta,\xi}[S] \cdot \overline{\mathscr{F}_{\eta,\xi}[S_\psi]}\mathrm{d}\eta\,\mathrm{d}\xi$$

is defined, giving the probability density of $\mu^{\boldsymbol{E}}_{S\otimes S_0}$. Using property (3) of the "noncommutative Fourier transform" we get

$$\mathrm{e}^{\mathrm{i}(\xi x+\eta\mu v)}\mathscr{F}_{\eta,\xi}[S_\psi] = \mathscr{F}_{\eta,\xi}[W_{x,v}S_\psi W^*_{x,v}].$$

By the "Parceval relation" (5.3.22) the probability density of $\mu^{\boldsymbol{E}}_{S\otimes S_0}$ is

$$\begin{aligned}&\frac{\mu}{(2\pi)^2}\iint \mathscr{F}_{\eta,\xi}[S]\cdot\overline{\mathscr{F}_{\eta,\xi}[W_{x,v}S_\psi W^*_{x,v}]}\mathrm{d}\eta\,\mathrm{d}\xi\\ &= \frac{\mu}{2\pi}\operatorname{Tr} S W_{x,v}S_\psi W^*_{x,v} = (\psi|W^*_{x,v}SW_{x,v}\psi)\frac{\mu}{2\pi},\end{aligned}$$

and thus it is the same as the probability density (5.6.58) of the measurement $\boldsymbol{M}$. This proves the proposition. □

5.4. Characteristic function and moments of state

Consider the transform $\mathscr{F}_z[T]$ of a Hermitean trace-class operator T. If $T \geq 0$, then $\mathscr{F}_z[T]$ possesses the following property of Δ-*positive definiteness*: for any n; $z_1, \ldots z_n \in Z$ and $c_1, \ldots c_n \in \mathbb{C}$

$$\sum_{j,k=1}^{n} c_j\bar{c}_k\mathscr{F}_{z_j-z_k}[T]\exp\left[\frac{1}{2}\mathrm{i}\Delta(z_j,z_k)\right] \geq 0. \tag{5.4.29}$$

Indeed by (5.2.9) and (5.3.22) this is nothing but

$$\operatorname{Tr} T\left[\sum_k c_k V(z_k)\right]^*\left[\sum_j c_j V(z_j)\right]$$

which is always nonnegative, if $T \geq 0$.

We call the transform $\mathscr{F}_z[S]$ of a density operator S the *characteristic function* of S, having in mind the analogy with characteristic function of classical probability distribution. The following is the noncommutative analog of the Bochner-Khinchin theorem.

Theorem 5.4.1. *For $\mathscr{F}(z)$ to be characteristic function of a quantum state the following conditions are necessary and sufficient*:

(1) $\mathscr{F}(0) = 1$, $\mathscr{F}(z)$ *is continuous at* $z = 0$;
(2) $\mathscr{F}(z)$ *is* Δ*-positive definite.*

Proof. Let S be a density operator; then $\mathscr{F}_0[S] = \operatorname{Tr} S = 1$. The condition (2) follows from positivity of S. To prove continuity note that

$$\mathscr{F}_z[S] = \sum_j s_j(e_j|V(x)e_j),$$

where s_j are the eigenvalues and e_j the eigenvectors of S. By the assumed continuity of the representation $z \to V(z)$ each term in the series is a continuous function of z; moreover the series converges uniformly since $|(e_j|V(z)e_j)| \leq 1$ and $\sum s_j < \infty$. Thus $\mathscr{F}(z)$ is continuous for all $z \in Z$, and the necessity is proved.

Turning to sufficiency we first show that continuity at $z = 0$ and Δ-positive definiteness imply uniform continuity of $\mathscr{F}(z)$ for all z (in analogy with ordinary characteristic functions). For this take $n = 3$ in (5.4.29) and the values of z equal to 0, z_1, z_2. Then (5.4.29) signifies positive definiteness of the Hermitean form in c_1, c_2, c_3 with the matrix

$$\begin{bmatrix} 1 & \mathscr{F}(-z_1) & \mathscr{F}(-z_2) \\ \mathscr{F}(z_1) & 1 & \mathscr{F}(z_1 - z_2)\mathrm{e}^{\mathrm{i}\Delta(z_1,z_2)/2} \\ \mathscr{F}(z_2) & \mathscr{F}(z_2 - z_1)\mathrm{e}^{\mathrm{i}\Delta(z_2,z_1)/2} & 1 \end{bmatrix}.$$

By Sylvester's criterion

$$\begin{aligned} &1 - \mathscr{F}(z_1)\mathscr{F}(-z_1) \geq 0, \\ &1 + 2\,\mathrm{Re}\,\overline{\mathscr{F}(z_1)}\mathscr{F}(z_2)\overline{\mathscr{F}(z_2 - z_1)}\mathrm{e}^{\mathrm{i}\Delta(z_1,z_2)/2} \\ &- |\mathscr{F}(z_2)|^2 - |\mathscr{F}(z_1)|^2 - |\mathscr{F}(z_1 - z_2)|^2 \geq 0. \end{aligned}$$

The first inequality implies

$$\mathscr{F}(-z) = \overline{\mathscr{F}(z)} \tag{5.4.30}$$

and

$$|\mathscr{F}(z)|^2 \leq 1. \tag{5.4.31}$$

Rearranging the second inequality we get

$$\begin{aligned} &|\mathscr{F}(z_2) - \mathscr{F}(z_1)|^2 \\ &\leq 1 - |\mathscr{F}(z_2 - z_1)|^2 - 2\,\mathrm{Re}\,\overline{\mathscr{F}(z_1)}\mathscr{F}(z_2)[1 - \overline{\mathscr{F}(z_2 - z_1)}\mathrm{e}^{\mathrm{i}\Delta(z_2,z_1)/2}]. \end{aligned}$$

Using (5.4.30), (5.4.31) we obtain the final inequality

$$|\mathscr{F}(z_2) - \mathscr{F}(z_1)|^2 \leq 4|1 - \overline{\mathscr{F}(z_2 - z_1)}\mathrm{e}^{\mathrm{i}\Delta(z_2,z_1)/2}|,$$

which proves the assertion.

We now construct a Hilbert space and a density operator in it for which $\mathscr{F}(z)$ is the characteristic function. Consider the operator $\widehat{V}_0(z)$ acting on a function $\psi(w)$, $w \in Z$, by the formula

$$\widehat{V}_0(z)\psi(w) = \exp\left[-\frac{1}{2}\mathrm{i}\Delta(z, w)\right]\psi(z + w).$$

It is easy to check that the operators $\{\widehat{V}_0(z); z \in Z\}$ satisfy the CCR (5.2.9). We are now going to define the inner product with respect to which these operators will be unitary. Consider the linear space $\hat{\mathscr{H}}_0$ of functions on Z of the form

$$\psi(w) = \left[\sum_k c_k V(z_k)\right] 1(w) \equiv \sum_k c_k \exp\left[-\frac{1}{2}\mathrm{i}\Delta(z_k, w)\right], \qquad w \in Z,$$

where $1(w)$ is the function which is identically equal to one. Introduce in $\hat{\mathscr{H}}_0$ the sesquilinear form

$$(\psi^{(1)}|\psi^{(2)}) = \sum_{j,k} c_j^{(2)} \bar{c}_k^{(1)} \mathscr{F}(z_j^{(2)} - z_k^{(1)}) \exp\left[\frac{1}{2}\mathrm{i}\Delta(z_j^{(2)}, z_k^{(1)})\right],$$

where $\psi^{(\alpha)} = [\sum_j c_j^{(\alpha)} V(z_j^{(\alpha)})]1; \alpha = 1, 2$. By Δ-positive definiteness

$$(\psi|\psi) \geq 0, \qquad \psi \in \hat{\mathscr{H}}_0,$$

so that $(\cdot|\cdot)$ is pre-inner product on $\hat{\mathscr{H}}_0$. Moreover

$$(\widehat{V}_0(z)\psi_1|\widehat{V}_0(z)\psi_2) = (\psi_1|\psi_2); \qquad \psi_1, \psi_2 \in \hat{\mathscr{H}}_0 \tag{5.4.32}$$

for all $z \in Z$.

Denote by $\hat{\mathscr{H}}$ the completion of $\hat{\mathscr{H}}_0$ with respect to this pre-inner product. By (5.4.32) the operators $\widehat{V}_0(z)$ extend uniquely by continuity to unitary operators $\widehat{V}(z)$ in $\hat{\mathscr{H}}$. Thus $z \to V(z)$ is a representation of the CCR in $\hat{\mathscr{H}}$, if we prove its continuity. For this it is sufficient to check continuity of the functions $(\psi^{(1)}|\widehat{V}(\cdot)\psi^{(2)}) = (\psi^{(1)}|\widehat{V}_0(\cdot)\psi^{(2)})$ for $\psi^{(\alpha)}$ lying in $\hat{\mathscr{H}}_0$. But for such ψ's

$$\begin{aligned}&(\psi^{(1)}|\widehat{V}_0(z)\psi^{(2)})\\ &= \sum_{j,k} c_j^{(z)} \bar{c}_k^{(1)} \exp\left[\frac{1}{2}\mathrm{i}\Delta(z_j^{(2)} + z, z^{(1)} z_k - z)\right] \mathscr{F}(z_j^{(2)} + z - z_k^{(1)}),\end{aligned}$$

and the required continuity follows from continuity of $\mathscr{F}(z)$.

From the definitions of $\widehat{V}(z)$ and the inner product

$$(1|\widehat{V}(z)1) = \mathscr{F}(z). \tag{5.4.33}$$

By Theorem (5.3.18) the constructed representation $z \to \widehat{V}(z)$ is unitarily equivalent to the direct orthogonal sum of copies of an irreducible representation $z \to V(z)$ in a Hilbert space $\mathscr{H}$:

$$\widehat{V}(z) = U^{-1} \begin{bmatrix} V(z) & & 0 \\ & V(z) & \\ 0 & & \ddots \end{bmatrix} U.$$

The operator U maps isometrically $\hat{\mathscr{H}}$ onto $\mathscr{H} \oplus \mathscr{H} \oplus \cdots$. Put

$$U1 = \psi_1 \oplus \psi_2 \oplus \cdots \tag{5.4.34}$$

and define $S = \sum_j |\psi_j)(\psi_j|$ in $\mathscr{H}$. Apparently $S \geq 0$ and

$$\operatorname{Tr} S = \sum_j (\psi_j|\psi_j) = (U1|U1) = (1|1) = 1.$$

Thus S is a density operator. From (5.4.33) and (5.4.34) we get

$$\mathscr{F}(z) = (1|\widehat{V}(z)1) = \sum_j (\psi_j|V(z)\psi_j) = \operatorname{Tr} SV(z),$$

so that $\mathscr{F}(z)$ is the characteristic function of S. The theorem is proved. □

There is an interesting peculiarity which has no analogy in probability theory. Since any trace-class operator is Hilbert-Schmidt, $\mathscr{F}_z[S]$ is square-integrable by Theorem 5.3.3. This means that continuity and Δ-positive definiteness imply square integrability of $\mathscr{F}(z)$. Notice that if the square integrability is postulated the proof of sufficiency becomes much simpler. Indeed, basing on Theorem 5.3.3 introduce the Hilbert-Schmidt operator $S = V(\mathscr{F})$. The conditions (1) and (2) imply $S \geq 0$ and $\operatorname{Tr} S = 1$ so that S is a density operator and $\mathscr{F}(z) = \mathscr{F}_z[S]$ by the inversion formula.

Moments of a probability distribution are easily expressed through the derivatives of its characteristic function. An analogous relation exists in the noncommutative theory. Let

$$R(z) = \int \lambda E_z(\mathrm{d}\lambda)$$

be the spectral representation of the canonical observable $R(z)$. Consider its probability distribution

$$\mu_S^z(B) = \operatorname{Tr} SE_z(B); \qquad B \in \mathscr{A}(\mathbb{R}),$$

with respect to a state S. The function $\mathscr{F}_{tz}[S]$, $t \in \mathbb{R}$, is a classical characteristic function of $\mu_S^z(\mathrm{d}\lambda)$ since by (2.7.78)

$$\mathscr{F}_{tz}[S] = \operatorname{Tr} S\mathrm{e}^{\mathrm{i}tR(z)} = \int \mathrm{e}^{\mathrm{i}t\lambda}\mu_S^z(\mathrm{d}\lambda).$$

Assume that n^{th} absolute moment of the distribution μ_S^z is finite; then it is well known from probability theory that the function $\mathscr{F}_{tz}[S]$ is n times differentiable and the n^{th} moment of μ_s^z is equal to

$$m_n(z) = \mathrm{i}^{-n}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\mathscr{F}_{tz}[S]\bigg|_{t=0}. \tag{5.4.35}$$

If n is even, them, conversely, the existence of the n^{th} derivative at $t = 0$ implies the finiteness of the n^{th} moment.

From (5.4.35) it is clear that $m_n(z)$ is a homogeneous polynomial in z of degree n. The main interest for us is the *mean value* of the state

$$m_1(z) \equiv m(z) = E_S(R(z)) = \int \lambda \mu_S^z(\mathrm{d}\lambda),$$

which is linear in z and the *second moment* $m_2(z) = \int \lambda^2 \mu_S^z(\mathrm{d}\lambda)$, which is a quadratic form in z. Introducing the corresponding real symmetric bilinear form

$$m_2(z, z') = -\frac{\partial^2}{\partial t \partial s} \mathscr{F}_{tz+sz'}[S]\Big|_{t=s=0}$$

we can define the *correlation function* of the state as

$$\alpha(z, z') = m_2(z, z') - m(z)m(z'), \tag{5.4.36}$$

so that

$$\alpha(z, z) = D_S(R(z)) = \int (\lambda - m(z))^2 \mu_S^z(\mathrm{d}\lambda).$$

We call S the *state with finite second moments* if $m_2(z) < \infty$ for all $z \in Z$. Recall that $\mathscr{L}_h^2(S)$ is the real Hilbert space of the operators which are square-summable with respect to S. Then according to Section 2.9 $m_2(z) < \infty$ implies $R(z) \in \mathscr{L}_h^2(S)$ and the relation (2.9.100) and (2.9.101) show that

$$m(z) = \langle I, R(z) \rangle_S, \qquad \alpha(z, z) = \langle R(z) - m(z), R(z) - m(z) \rangle_S.$$

By polarization the correlation function is equal to

$$\alpha(z, z') = \langle R(z) - m(z), R(z') - m(z') \rangle_S. \tag{5.4.37}$$

Now we turn to a rigorous version of the Heisenberg commutation relation (5.2.13) which reads

$$[R(z), R(z')]_S = \Delta(z, z'); \qquad z, z' \in Z. \tag{5.4.38}$$

To prove it we need

Lemma 5.4.2. *Let $M(\mathrm{d}\lambda)$ be a measurement with finite second moments and $X_M = \int \lambda M(\mathrm{d}\lambda)$ is defined as in Section* 2.9. *Consider the family of bounded operators*

$$V_t = \int \mathrm{e}^{\mathrm{i}t\lambda} M(\mathrm{d}\lambda)$$

as a function of t with values in $\mathscr{L}^2_\pm(S)$. Then

$$X_{\boldsymbol{M}} = \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}V_t\Big|_{t=0},$$

the derivative being taken in $\mathscr{L}^2_\pm(S)$.

Proof. We need to show that

$$\frac{V_t - I}{\mathrm{i}t} \to X_M \qquad \text{in } \mathscr{L}^2_\pm(S) \qquad \text{as } t \to 0.$$

By the inequality (2.9.109) it is sufficient to prove that

$$\frac{\mathrm{e}^{\mathrm{i}\lambda t} - 1}{\mathrm{i}t} \to \lambda \qquad \text{in } \mathscr{L}^2(\mu_S).$$

The pointwise convergence is evident; moreover

$$\left|\lambda - \frac{\mathrm{e}^{\mathrm{i}\lambda t} - 1}{\mathrm{i}t}\right|^2 \le 4\lambda^2,$$

since

$$\left|\frac{\mathrm{e}^{\mathrm{i}\lambda t} - 1}{\mathrm{i}t}\right| = \left|\frac{\sin \lambda t/2}{t/2}\right| \le \lambda.$$

By the finiteness of second moment $\int \lambda^2 \mu_S(\mathrm{d}\lambda) < \infty$, and by the Lebesgue dominated convergence theorem

$$\int \left|\lambda - \frac{\mathrm{e}^{\mathrm{i}\lambda t} - 1}{\mathrm{i}t}\right|^2 \mu_S(\mathrm{d}\lambda) \to 0,$$

what is required. □

The lemma implies that if $m_2(z) < \infty$, then

$$R(z) = \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}V(tz)\Big|_{t=0} \qquad \text{in } \mathscr{L}^2_\pm(S). \tag{5.4.39}$$

Multiplying (5.2.12) by S, taking trace and using $V(z)^* = v(-z)$, we get

$$\langle V(-tz),\, V(sz')\rangle^-_S = \mathrm{e}^{\mathrm{i}ts\Delta(z,z')}\langle V(-tz),\, V(sz')\rangle^+_S.$$

Differentiating with the help of (5.4.39), we obtain

$$-\langle R(z),\, R(z')\rangle^-_S = \mathrm{i}\Delta(z,z') - \langle R(z),\, R(z')\rangle^+_S.$$

Taking into account the second relation in (2.8.97) we obtain (5.4.38). Taking into account (2.8.92) we can also write

$$[R(z) - m(z), R(z') - m(z')]_S = \Delta(z, z'). \tag{5.4.40}$$

Proposition 2.8.3 together with (5.4.37) and (5.4.40) implies that the correlation function of a state with finite second moments satisfies the equivalent inequalities

$$\begin{aligned} &\alpha(z, z)\alpha(z', z') \geq \tfrac{1}{4}\Delta(z, z')^2, \\ &\alpha(z, z) + \alpha(z', z') \geq \Delta(z, z'); \qquad z, z' \in Z. \end{aligned} \tag{5.4.41}$$

The first of these is just the rigorous Heisenberg uncertainty relation for the canonical observables $R(z)$, $R(z')$. Moreover it follows from (2.8.98) that

$$[\alpha(z_j, z_k) \pm \tfrac{1}{2}\mathrm{i}\Delta(z_j, z_k)] \geq 0 \tag{5.4.42}$$

for any n and $z_1, \ldots z_n \in Z$.

Lemma 5.4.3. *Let S be density operator of a state with finite second moments, then*

$$\begin{aligned} &\mathscr{F}_z[SR(z_1)] = [-\tfrac{1}{2}\Delta(z, z_1) - \mathrm{i}\nabla_{z_1}]\mathscr{F}_z[S], \\ &\mathscr{F}_z[R(z_1)S] = [\tfrac{1}{2}\Delta(z, z_1) - \mathrm{i}\nabla_{z_1}]\mathscr{F}_z[S], \\ &\mathscr{F}_z[S \circ R(z_1)] = -\mathrm{i}\nabla_{z_1}\mathscr{F}_z[S], \qquad (5.4.43) \\ &\mathscr{F}_z[[R(z_1), S]] = \Delta(z, z_1)\mathscr{F}_z[S], \qquad (5.4.44) \end{aligned}$$

where ∇_{z_1} is the derivative along the direction z_1:

$$\nabla_{z_1}\mathscr{F}(z) = \frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(z + tz_1)\bigg|_{t=0}.$$

Proof. Since $R(z_1) \in \mathscr{L}^2(S)$, the operators $SR(z_1), \ldots$ in the left-hand sides are trace-class by Proposition 2.8.2 and have square-integrable transforms $\mathscr{F}_z[SR(z_1)], \ldots$ It is sufficient to check the first relation. We have

$$\begin{aligned} \mathscr{F}_z[SR(z_1)] &= \mathrm{Tr}(SR(z_1))V(z) \\ &= \langle V(-z), R(z_1)\rangle_S^0. \end{aligned}$$

According to (5.4.39)

$$\begin{aligned} \mathscr{F}_z[SR(z_1)] &= \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\langle V(-z), V(tz_1)\rangle_S^-\bigg|_{t=0} \\ &= \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}\,SV(tz_1)V(z)\bigg|_{t=0}. \end{aligned}$$

Taking into account the CCR we get

$$\mathscr{F}_z[SR(z_1)] = \mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}_{z+tz_1}[S]\mathrm{e}^{\mathrm{i}t\Delta(z_1,z)/2}\Big|_{t=0},$$

as required. □

It is not difficult to extend these formulas to a wider class of operators S but we shall not do it.

5.5. Structure of general Gaussian states

Let us calculate the characteristic function of the quasiclassical state (5.1.2). Using (5.3.28) we get

$$\begin{aligned}\mathscr{F}_{x,y}[S] &= \exp\left[-\frac{1}{4}\left(\frac{\hbar}{\omega}x^2+\frac{\omega}{\hbar}y^2\right)\right]\int \exp[\mathrm{i}(\alpha x+\beta y)]\mu(\mathrm{d}\alpha\,\mathrm{d}\beta)\\ &= \mathscr{F}_0(x,y)\cdot\bar{\mu}(x,y),\end{aligned} \tag{5.5.45}$$

where $\mathscr{F}_0(x,y)$ is the characteristic function of the ground state $|0)(0|$, and $\tilde{\mu}(x,y)$ is the classical characteristic function of the probability distribution μ. Consider the quasiclassical Gaussian state (5.1.7). Returning to real variables by letting

$$\zeta = (2\hbar\omega)^{-1/2}(\omega\beta+\mathrm{i}\hbar\alpha), \qquad \bar{a} = (2\hbar\omega)^{-1/2}(\omega\overline{Q}+\mathrm{i}\hbar\overline{P}),$$

we can write its density operator in the form

$$\begin{aligned}S_{\overline{P},\overline{Q}} = \frac{1}{2\pi\overline{N}}\iint\ &|\alpha,\beta)(\beta,\alpha|\\ &\times\exp\left\{-\frac{1}{2\overline{N}}\left[\frac{\hbar}{\omega}(\alpha-\overline{P})^2+\frac{\omega}{\hbar}(\beta-\overline{Q})^2\right]\right\}\mathrm{d}\alpha\,\mathrm{d}\beta.\end{aligned} \tag{5.5.46}$$

The characteristic function of the Gaussian distribution is

$$\bar{\mu}(x,y) = \exp\left[\mathrm{i}(\overline{P}x+\overline{Q}y)-\frac{\overline{N}}{2}\left(\frac{\omega}{\hbar}x^2+\frac{\hbar}{\omega}y^2\right)\right]$$

so that by (5.5.45)

$$\mathscr{F}_{x,y}[S_{\overline{P},\overline{Q}}] = \exp[\mathrm{i}(\overline{P}x+\overline{Q}y)-\tfrac{1}{2}(\sigma_P^2x^2+\sigma_Q^2y^2)], \tag{5.5.47}$$

where

$$\sigma_P^2 = \frac{\omega}{\hbar}\left(\overline{N} + \frac{1}{2}\right), \qquad \sigma_Q^2 = \frac{\hbar}{\omega}\left(\overline{N} + \frac{1}{2}\right)$$

with $\overline{N} = \sigma_P \sigma_Q - \frac{1}{2}$.

For many degrees of freedom the characteristic function of the state (5.1.8) is apparently the product of the factors of the type (5.5.46):

$$\prod_k \exp\left[\mathrm{i}(\overline{P}_k x_k + \overline{Q}_k y_k) - \frac{1}{2}(\sigma_{Pk}^2 x_k^2 + \sigma_{Qk}^2 y_k^2)\right]. \tag{5.5.48}$$

We thus see that the characteristic function of the quantum state (5.5.46) has the same analytic form as the characteristic function of a Gaussian distribution though the variances are not arbitrary positive numbers but are subject to the uncertainty relation $\sigma_P^2 \sigma_Q^2 \geq \frac{1}{4}$.

Basing on this analogy we introduce the following general definition. Let $z \to V(z)$ be an irreducible representation of the CCR on a symplectic space (Z, Δ). The state S in the representation space $\mathscr{H}$ is called *Gaussian* if its characteristic function has the form

$$\mathscr{F}_z[S] = \exp[\mathrm{i}m(z) - \tfrac{1}{2}\alpha(z, z)], \tag{5.5.49}$$

where $m(z)$ is a linear functional, $\alpha(z, z')$ is a bilinear symmetric form on Z. The function (5.5.49) is infinitely differentiable and therefore all moments of the Gaussian state are finite. The relations (5.4.35) and (5.4.36) imply that $m(z)$ is the mean value and $\alpha(z, z')$ is the correlation function of S as anticipated in the notation.

Theorem 5.5.1. *For* (5.5.49) *to be the characteristic function of a quantum state it is necessary and sufficient that* $\alpha(z, z')$ *satisfy one of the equivalent relations* (5.4.41), (5.4.42).

Proof. The necessity follows from the fact that α is the correlation function of a quantum state. To prove sufficiency it is enough to check Δ-positive definiteness of the function (5.5.49), *i.e.*,

$$\sum_{j,k} c_j \bar{c}_k \exp\left[\mathrm{i}m(z_j) - \mathrm{i}m(z_k) - \frac{1}{2}\alpha(z_j - z_k, z_j - z_k) + \frac{1}{2}\mathrm{i}\Delta(z_j, z_k)\right] \geq 0$$

for any $z_j \in Z$, $c_j \in \mathbb{C}$. Putting $b_j = c_j \exp[\mathrm{i}m(z_j) - \frac{1}{2}\alpha(z_j, z_j)]$ rewrite it in the form

$$\sum_{j,k} b_j \bar{b}_k \exp\left[\alpha(z_j, z_k) + \frac{1}{2}\mathrm{i}\Delta(z_j, z_k)\right] \geq 0.$$

Note that the matrix

$$[\alpha(z_j, z_k) + \tfrac{1}{2}\mathrm{i}\Delta(z_j, z_k)]$$

is positive definite by (5.4.42); and referring to a result of Schur which says that positive definiteness of a matrix $[a_{jk}]$ implies positive definiteness of $[\exp a_{jk}]$ the theorem is proved. □

Denote by S_m the Gaussian state with the mean $m(z)$ and fixed correlation function $\alpha(z, z')$. We shall show that S_m can be regarded as a result of an "influence" on the state $S = S_0$ described by a unitary displacement operator. Since the form Δ is nondegenerate, there exists the unique vector $m_\Delta \in Z$ such that

$$m(z) = \Delta(m_\Delta, z), \qquad z \in Z$$

(this is easily seen in the coordinate representation). Then

$$S_m = V(m_\Delta)^* S V(m_\Delta).$$

For this it is sufficient to check that characteristic functions of both sides are equal. Using property (3) of the transform $T \to \mathscr{F}_z[T]$ (see Section 5.3), we have

$$\mathscr{F}_z[V(m_\Delta)^* S V(m_\Delta)] = \mathscr{F}_z[S]\mathrm{e}^{\mathrm{i}\Delta(m_\Delta, z)} = \mathscr{F}_z[S_m],$$

as required.

Consider now the Euclidean space (Z, α). By the Riesz-Frechet lemma there is a unique $m_\alpha \in Z$ satisfying

$$m(z) = \alpha(m_\alpha, z), \qquad z \in Z$$

(in the finite-dimensional case this is easily seen in the coordinate representation). Introduce $\mathscr{D}$ as the associated operator of the form Δ, *i.e.*, the operator satisfying (5.2.15) where α is the correlation function. Then $m_\alpha = -\mathscr{D} m_\Delta$ so that

$$S_m = V(\mathscr{D}^{-1} m_\alpha) S V(\mathscr{D}^{-1} m_\alpha)^*. \tag{5.5.50}$$

The condition (5.4.41) for the correlation function sets certain restriction onto the operator $\mathscr{D}$. Putting $z' = -\frac{1}{2}\mathscr{D} z$ in (5.4.41) we get

$$I + \tfrac{1}{4}\mathscr{D}^2 \geq 0 \qquad \text{in } (Z, \alpha). \tag{5.5.51}$$

Let $\{e_j, h_j\}$ be the symplectic basis in which $\mathscr{D}$ has the matrix (5.2.15). From (5.5.51)

$$a_j \equiv d_j^{-1} \geq \tfrac{1}{2}.$$

Since

$$\alpha(z, z) = \sum_j a_j (x_j^2 + y_j^2),$$

where $[x_j, y_j]$ are the components of the vector z in this basis, the characteristic function of the state S_m takes the form

$$\prod_j \exp\left[\mathrm{i}(\overline{P}'_j x_j + \overline{Q}'_j y_j) - \frac{a_j}{2}(x_j^2 + y_j^2)\right]. \tag{5.5.52}$$

Here we put

$$\overline{P}'_j = E_{S_m}(P'_j), \qquad \overline{Q}'_j = E_{S_m}(Q'_j), \qquad a_j = D_{S_m}(Q'_j) = D_{S_m}(P'_j),$$

where $P'_j = R(e_j)$, $Q'_j = R(h_j)$.

Note that if we start with a coordinate form of the CCR related to some fixed set of canonical observables P_k, Q_k as in Section 5.2, then the symplectic basis in which the characteristic function of a Gaussian state has the simplest form (5.5.52) may be arbitrary. The new canonical observables P'_j, Q'_j need not coincide with P_k, Q_k and are in general related to them through a linear transformation preserving the Heisenberg commutation relations (such transformations are called canonical). For example, the characteristic function (5.5.48) is reduced to the form (5.5.52) by the symplectic transformation

$$x_j \to (\sigma_{Pj}/\sigma_{Qj})^{1/2} x_j = (\omega_j/\hbar)^{1/2} x_j,$$
$$y_j \to (\sigma_{Qj}/\sigma_{Pj})^{1/2} y_j = (\hbar/\omega_j)^{1/2} y_j,$$

to which corresponds the canonical transformation

$$P'_j = (\hbar/\omega_j)^{1/2} P_j, \qquad Q'_j = (\omega_j/\hbar)^{1/2} Q_j.$$

Then the characteristic function $\operatorname{Tr} S \exp \mathrm{i}[\sum_j (P'_j x_j + Q'_j y_j)]$ will have the form (5.5.52) with $a_j = \sigma_{Pj}\sigma_{Qj} = \overline{N}_j + \frac{1}{2}$. In general the transformation can be much more complicated.

The fact that the characteristic function (5.5.49) can be always decomposed into the factors corresponding to mutually commuting pairs of the canonical observables $\{P'_j, Q'_j\}$ means that the space $\mathscr{H}$ of the irreducible representation $z \to V(z)$ can be represented as the tensor product

$$\mathscr{H} = \bigotimes_j \mathscr{H}'_j, \tag{5.5.53}$$

so that $V(z) = \bigotimes_j V_j(z_j, y_j)$ with $V_j(x_j, y_j) = \exp \mathrm{i}(P'_j x_j + Q'_j y_j)$ acting irreducibly in $\mathscr{H}'_j$; moreover

$$S_m \bigotimes_j S'_j,$$

where S'_j are the Gaussian states in $\mathscr{H}'_j$ with characteristic functions of the simplest form

$$\exp\left[\mathrm{i}(\overline{P}'_j x_j + \overline{Q}'_j y_j) - \frac{a_j}{2}(x_j^2 + y_j^2)\right].$$

We emphasize again that the decomposition (5.5.53) is defined by the Gaussian state itself (more definitely, by its correlation function) and need not be the same as the "natural" decomposition related to initial canonical observables.

Finally let us calculate the eigenvalues of S_m. Since the transformation (5.5.50) does not alter the eigenvalues of S we can take $m(z) \equiv 0$. By (5.1.4) the zero-mean density operator S'_j for the j^{th} degree of freedom has the eigenvalues

$$\frac{1}{\overline{N}_j + 1}\left(\frac{\overline{N}_j}{\overline{N}_j + 1}\right)^n; \qquad n = 0, 1, \ldots, \tag{5.5.54}$$

where $\overline{N}_j = \sigma_{Pj}\sigma_{Qj} - \frac{1}{2} = a_j - \frac{1}{2}$. The tensor product of such states will have for eigenvalues the numbers

$$\prod_{j=1}^{s} \frac{1}{\overline{N}_j + 1}\left(\frac{\overline{N}_j}{\overline{N}_j + 1}\right)^{n_j};$$

corresponding to all combinations of $n_j = 0, 1, \ldots$ In particular the maximal eigenvalue is equal to

$$\prod_{j=1}^{s} \frac{1}{\overline{N}_j + 1} = \prod_{j=1}^{s} \frac{1}{a_j + \frac{1}{2}}.$$

The state is pure if and only if this value is equal to 1. Since $a_j \geq \frac{1}{2}$, this can hold only if all $a_j = \frac{1}{2}$ or

$$\det \tfrac{1}{2}\mathscr{D} = 1.$$

5.6. A characteristic property of Gaussian states

Let $z \to V(z)$ be an irreducible representation of the CCR in $\mathscr{H}$ and S be a state in $\mathscr{H}$. Consider the Hilbert space $\mathscr{L}^2(S)$ associated with the state S.

Lemma 5.6.1. *The linear span of the operators* $\{V(z), z \in Z\}$ *is dense in* $\mathscr{L}^2(S)$.

Proof. Let $X \in \mathscr{L}^2(S)$ satisfy

$$\langle V(z), X\rangle_S = 0, \qquad z \in Z. \tag{5.6.55}$$

Since $\mathscr{L}^2(S)$ is the complexification of $\mathscr{L}^2_h(S)$, then $X = X_1 + \mathrm{i}X_2$ with $X_j \in \mathscr{L}^2_h(S)$. Therefore we can use (2.8.88) to obtain

$$\mathrm{Tr}(X \circ S)V(z) = 0, \qquad z \in Z.$$

Thus $\mathscr{F}_z[X \circ S] = 0$; and by the inversion formula $X \circ S = 0$. Therefore using again (2.8.88) we obtain for any bounded Y

$$\langle Y, X\rangle_S = \mathrm{Tr}\, Y^*(S \circ X) = 0$$

so that $X = 0$ in $\mathscr{L}^2(S)$ and the lemma is proved. □

Assume now that the state S has finite second moments so that $R(z) \in \mathscr{L}^2_h(S)$ for all $z \in Z$. Denote by $\mathfrak{R}$ the subspace of $\mathscr{L}^2_h(S)$ generated by the operators

$$c + R(z); \qquad c \in \mathbb{R}, \quad z \in Z.$$

If $\mathfrak{R}_0$ is the one-dimensional subspace of $\mathfrak{R}$ consisting of multiples of the unit operator, then $\mathfrak{R} = \mathfrak{R}_0 \oplus \mathfrak{R}_1$ where $\mathfrak{R}_1$ is the subspace of operators

$$R(z) - m(z), \qquad z \in Z,$$

where $m(z)$ is the mean of the state S. This follows from the fact that

$$\langle R(z) - m(z), I\rangle_S = m(z) - m(z) \equiv 0.$$

By (5.4.37)

$$z \to R(z) - m(z) \tag{5.6.56}$$

is an isometric map of the Euclidean space (Z, α) onto the subspace $\mathfrak{R}_1 \subset \mathscr{L}^2_h(S)$. Consider the commutation operator $\mathscr{D}$ of S defined by (2.10.113) and denote by $\mathfrak{D}_1$ its restriction onto the subspace $\mathfrak{R}_1$, so that

$$[Y, X]_S = \langle Y, \mathfrak{D}_1 X\rangle_S; \qquad X, Y \in \mathfrak{R}_1.$$

Now let $\mathscr{D}$ be the operator defined through the correlation function α of the state according to (5.2.15). Then the relation (5.4.40) can be written in the form

$$\alpha(z, \mathscr{D} z') = \langle R(z) - m(z), \mathfrak{D}(R(z') - m(z'))\rangle_S; \qquad z, z' \in Z.$$

This means that the operator $\mathscr{D}$ corresponds to $\mathfrak{D}_1$ under the isometric map (5.6.56), *i.e.*,

$$R(\mathscr{D} z) - m(\mathscr{D} z) = \mathfrak{D}_1(R(z) - m(z)). \tag{5.6.57}$$

Theorem 5.6.2. *A state S with finite second moments is Gaussian if and only if $\mathfrak{R}$ (or $\mathfrak{R}_1$) is an invariant subspace of the commutation operator $\mathfrak{D}$.*

Proof. Observing that $\mathfrak{D}(\mathfrak{R}_0) = [0]$ by (2.10.115) we see that invariance of $\mathfrak{R}_1$ is equivalent to invariance of $\mathfrak{R}$.

Let S be a Gaussian state. We shall show that $\mathfrak{D} = \mathfrak{D}_1$ on $\mathfrak{R}_1$ and thus $\mathfrak{R}_1$ is an invariant subspace of $\mathfrak{D}$. Taking into account (5.6.57) we have to show that

$$\mathfrak{D}(R(z) - m(z)) = R(\mathscr{D} z) - m(\mathscr{D} z), \qquad z \in Z. \tag{5.6.58}$$

By the definition of commutation operator and the fact that $\mathfrak{D} I = 0$ this is equivalent to the identity

$$[X, R(z)]_S = \langle X, R(\mathscr{D} z) - m(\mathscr{D} z)\rangle_S, \qquad X \in \mathscr{L}^2(S). \tag{5.6.59}$$

By Lemma 5.5.1 we need to check this only for $X = V(-w)$, $w \in Z$. But (5.4.44) with (2.8.90) implies

$$[V(-w), R(z)]_S = \mathrm{i}\Delta(w, z)\mathscr{F}_w[S],$$

and (5.4.43) with (2.8.88) implies

$$\langle V(-w), R(z)\rangle_S = -\mathrm{i}\nabla_z \mathscr{F}_w[S].$$

We thus need to check that the characteristic function of a Gaussian state $\mathscr{F}_w[S]$ given by (5.5.49) satisfies

$$\mathrm{i}\Delta(w, z)\mathscr{F}_w[S] = -[\mathrm{i}\nabla_{\mathscr{D} z} + m(\mathscr{D} z)]\mathscr{F}_w[S], \tag{5.6.60}$$

and this is confirmed by a straightforward computation.

To prove the converse assume that $\mathfrak{D}(\mathfrak{R}_1) \subset \mathfrak{R}_1$. Then $\mathfrak{D}(R(z) - m(z)) = R(\mathscr{D}_1 z) - m(\mathscr{D}_1 z)$ where $\mathscr{D}_1$ is a linear operator in Z. Denoting by α the correlation function of the state S we have

$$\begin{aligned}\alpha(z, \mathscr{D}_1 w) &= \langle R(z) - m(z), \mathscr{D}(R(w) - m(w))\rangle_S \\ &= [R(z) - m(z), R(w) - m(w)]_S = \Delta(z, w),\end{aligned}$$

so that $\mathscr{D}_1 = \mathscr{D}$ where $\mathscr{D}$ is defined by (5.2.15). Thus (5.6.58) holds and therefore the characteristic function of the state S satisfies the differential equation (5.6.60). Since $\mathscr{D}$ is nondegenerate we can substitute $\mathscr{D}^{-1} z$ for z in (5.6.60) to obtain

$$-\alpha(w, z)\mathscr{F}_w[S] = [\nabla_z - \mathrm{i}m(z)]\mathscr{F}_w[S]. \tag{5.6.61}$$

Let $\{z_j\}$ be an orthonormal basis in the Euclidean space (Z, α) and w_j be the components of the vector w in this basis. Then the coordinate form of (5.6.61) reads

$$\left[\frac{\partial}{\partial w_j} - \mathrm{i}m(z_j)\right]\mathscr{F}_w[S] = -w_j \mathscr{F}_w[S], \qquad j = 1, \ldots, 2s.$$

The unique solution of this equation satisfying $\mathscr{F}_0[S] = 1$ is

$$\exp\left[\mathrm{i}\sum_j w_j m(z_j) - \frac{1}{2}\sum_j w_j^2\right] = \exp\left[\mathrm{i}m(w) - \frac{1}{2}\alpha(w, w)\right].$$

The theorem is proved. □

Consider the relation (5.6.59) for bounded X. Then using (2.8.88) and (2.8.90) we obtain a useful identity

$$\mathrm{i}[R(z), S_m] = (R(\mathscr{D} z) - m(\mathscr{D} z)) \circ S_m \tag{5.6.62}$$

for the density operator of Gaussian state with mean $m(z)$ and correlation function $\alpha(z, z') = \Delta(z, \mathscr{D}^{-1} z')$.

Let $\{S_{\boldsymbol{\theta}}\}$ be the family of Gaussian states with the fixed correlation function α and the mean value of the form

$$m(z) = \sum_{j=1}^{n} \theta_j m_j(z),$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n] \in \mathbb{R}^n$, and $m_j(z)$ are fixed linear functionals on Z. Introduce $m_j \in Z$ defined by the Riesz-Frechet lemma according to

$$m_j(z) = \alpha(m_j, z), \qquad z \in Z.$$

Proposition 5.6.3. *The family* $\{S_{\boldsymbol{\theta}}\}$ *is differentiable*[2] *as a function of* $\boldsymbol{\theta}$ *with values in the Banach space* $\mathfrak{T}^1(\mathscr{H})$ *of trace-class operators and*

$$\frac{\partial S_{\boldsymbol{\theta}}}{\partial \theta_j} = \mathrm{i}[R(\mathscr{D}^{-1}m_j), S_{\boldsymbol{\theta}}] = (R(m_j) - m(m_j)) \circ S_\theta. \tag{5.6.63}$$

Proof. Using (5.5.50) we can write

$$S_{\boldsymbol{\theta}} = V\left(\sum_{j=1}^n \theta_j \mathscr{D}^{-1}m_j\right) S_0 V\left(\sum_{j=1}^n \theta_j \mathscr{D}^{-1}m_j\right)^*.$$

Giving an increment t to the parameter θ_j we get

$$S_{\boldsymbol{\theta}+t\boldsymbol{\delta}_j} = \mathrm{e}^{\mathrm{i}tR(\mathscr{D}^{-1}m_j)} S_{\boldsymbol{\theta}} \mathrm{e}^{-\mathrm{i}tR(\mathscr{D}^{-1}m_j)},$$

where $\boldsymbol{\delta}_j$ is the vector with the j^{th} component being the unit, the others being zero. Thus putting $S_t = S_{\boldsymbol{\theta}+t\boldsymbol{\delta}_j}$ we obtain the one-parameter family $\{S_t;\, t \in \mathbb{R}\}$ of the form (3.2.7), the infinitesimal generator $R(\mathscr{D}^{-1}m_j)$ of the corresponding unitary group $\{\exp \mathrm{i}tR(\mathscr{D}^{-1}m_j);\, t \in \mathbb{R}\}$ belonging to $\mathscr{L}_h^2(S)$. Therefore the family $\{S_t\}$ satisfies the conditions of Proposition 6.3.1 which will be proved in the next chapter. This implies the differentiability of $\{S_t\}$ and the expression (6.3.19) for the derivative which gives the first equality in (5.6.63). The second follows from (5.6.62). □

Let us illustrate these results by applying them to the quasi-classical Gaussian state (5.5.46) with the characteristic function (5.5.47), the role of the parameters $[\theta_j]$ being played by $\overline{P}$, $\overline{Q}$. Then z is two-dimensional vector with components x, y so that

$$\Delta(z, z') = xy' - x'y.$$

The mean value and the correlation function are

$$m(z) = \overline{P}x + \overline{Q}y, \qquad \alpha(z, z) = \sigma_P^2 x^2 + \sigma_Q^2 y^2,$$

whence $m(z) = \overline{P}\alpha(m_P, z) + \overline{Q}\alpha(m_Q, z)$ with

$$m_P = \sigma_P^{-2}[1, 0], \qquad m_Q = \sigma_Q^{-1}[0, 1].$$

[2] By differentiability of a function with values in a Banach space we always mean strong (*i.e.*, norm) differentiability.

Therefore

$$R(m_P) - m(m_P) = \sigma_P^{-2}(P - \overline{P}),$$
$$R(m_Q) - m(m_Q) = \sigma_Q^{-2}(Q - \overline{Q}).$$

Moreover

$$\mathscr{D} = \begin{bmatrix} 0 & -\sigma_Q^{-2} \\ \sigma_P^{-2} & 0 \end{bmatrix},$$

and the relations (5.6.63) become

$$\begin{aligned} \frac{\partial S_{\overline{P},\overline{Q}}}{\partial \overline{P}} &= \mathrm{i}[Q, S_{\overline{P},\overline{Q}}] = \sigma_P^{-2}(P - \overline{P}) \circ S_{\overline{P},\overline{Q}}, \\ \frac{\partial S_{\overline{P},\overline{Q}}}{\partial \overline{Q}} &= -\mathrm{i}[P, S_{\overline{P},\overline{Q}}] = \sigma_Q^{-2}(Q - \overline{Q}) \circ S_{\overline{P},\overline{Q}}. \end{aligned} \tag{5.6.64}$$

Since $R(z) = Px + Qy$, the relation (5.6.58) is equivalent to

$$\mathfrak{D}(P) = -\sigma_Q^{-2}(Q - \overline{Q}), \qquad \mathfrak{D}(Q) = \sigma_P^{-2}(P - \overline{P}). \tag{5.6.65}$$

5.7. Comments

Section 5.1. The relation (5.1.2) gives the so called Glauber's P-representation [42] (see also Klauder and Sudarshan [76]). Quantization of the electromagnetic field, *i.e.*, representation of the field by an infinite collection of quantum oscillators is due to Dirac [30]; it is presented in Louisell [88], Klauder and Sudarshan [76] and Helstrom [53] in the form convenient for applications in quantum optics. There one can find also a discussion of states of the radiation field.

Section 5.2. The coordinate-free approach to the CCR was developed by Segal [127] (for a finite number of degrees of freedom see Castler [21]). For symplectic spaces and skew-symmetric forms see, *e.g.*, Maltzev [94].

Section 5.3. The Stone-von Neumann uniqueness theorem is essentially finite-dimensional: it does not hold for an infinite number of degrees of freedom (see, *e.g.*, Segal [127]). To this fact are related some of the "divergences" of the quantum field theory. A thorough consideration of the problems of this theory can be found in Wightman [143] and Bogoljubov, Logunov and Todorov [16]. The Weyl transform introduced in [141] was studied by Loupias and Miracle-Sole [89], Pool [114] and Holevo [55] among others. The Weyl transform reveals most distinctly both similarity and disparateness between quantum and classical mechanics (see, *e.g.*, Shirokov [125]).

Section 5.4. The characteristic function of a state was introduced for quantum fields by Segal [126] who generalized a particular construction of Moyal [99].

The fact that the characteristic function determines the density operator uniquely underlies the method of homodyne tomography in quantum optics [184]. In the case of one mode

$$\mathrm{Tr}S\mathrm{e}^{i(Px+Qy)} = \mathrm{Tr}S\mathrm{e}^{irR(\phi)} = \int \mathrm{e}^{irt} p_\phi(t)dt,$$

where $R(\phi) = P\sin\phi + Q\cos\phi$ is the so called field quadrature corresponding to the phase ϕ. For each fixed value of ϕ measurement of the quadrature $R(\phi)$ can be accomplished via optical homodyning: the field of the mode with given frequency is combined with intense reference field (so called local oscillator) from the auxiliary laser of the same frequency by means of a balanced 50:50 beam splitter and then the difference of the intensities of the outgoing fields is measured. In the limit of infinite intensity this difference turns out to be proportional to the quadrature $R(\phi)$ [200]. The value of ϕ is determined by the parameters of the local oscillator. Knowledge of the probability densities $p_\phi(t)$ for all values $\phi \in [0, \pi]$ is equivalent to knowledge of the characteristic function of the state and hence, of the density operator ($p_\phi(t)$ is just the Radon transform of the Wigner function of the state which explains the use of the term "tomography"). Corresponding transformations can be obtained from the inversion formulas, see [161]. By making measurements of the distribution of quadratures for sufficiently many values of ϕ one can obtain consistent estimates of the density operator. Let us stress that in this case one deals with measurements of the incompatible observables $R(\phi)$ performed on different subensembles of the whole statistical ensemble.

For consideration of the mathematical aspects of this problem as a nonparametric inverse statistical problem see [166].

Section 5.5. The general definition of the Gaussian state (for a field) was given by Manuceau and Verbeure [97] under the name of quasi-free states, though particular Gaussian states appeared earlier in various sources (see, *e.g.*, Glauber [42] and Louisell [88]). There is a noncommutative central limit theorem (Cushen and Hudson [24]) in which the Gaussian states appear as limit law.

For a proof of the Schur's result see, *e.g.*, Polya and Szegö [112, Section VII, Problem 36].

Section 5.6. The characteristic property of Gaussian states was established by Holevo [65].

Chapter 6
Unbiased measurements

6.1. Quantum communication channel

Consider the idealized scheme of information transmission presented at Figure 6.1. In the absence of signal the physical carrier of information $\mathscr{C}$ (say, electromagnetic field) is in a state S. Usually it is adopted that S is the equilibrium (Gibbs) state at the given temperature. Transmission of a signal is accomplished through an influence of the communication source $\mathscr{S}$ onto the system $\mathscr{C}$, which forces a definite change in its state. If there are parameters of the source $\mathscr{S}$ which can be varied, then the resulting state S_θ will depend on the values of these parameters θ.

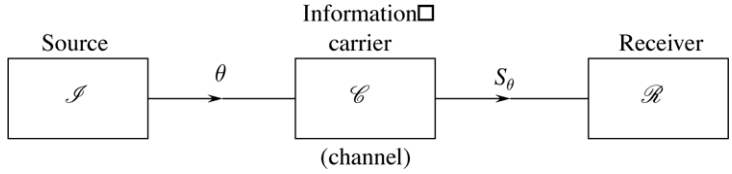

Figure 6.1.

If the information carrier is described classically, then its states are probability distributions $P_\theta(\mathrm{d}\omega)$ on the phase space Ω of the system $\mathscr{C}$. Such information channels are considered in the classical information theory. On the other hand if the information carrier is a quantum-mechanical system, then its states are described by density operators S_θ in the corresponding Hilbert space and one speaks of a quantum communication channel. The advent of coherent light sources such as laser opened a possibility for the creation of communication systems working in the optical range of frequencies. While for the radio frequencies the "energy quantum" $\hbar\omega$ is quite small as compared to the mean energy of the thermal fluctuations kT and the radiation field can be described classically, at the optical frequencies quantum fluctuations become significant and a consistent description of the information carrier – the radiation field

– requires quantum theory. Assuming the simplified picture of the field as a finite collection of quantum oscillators (which is usually justified here) we see that in the absence of signal the field is described by the quasiclassical Gaussian state (5.1.7) with zero mean, and the influence of the source is reflected by the appearance of non-zero mean $\bar{\boldsymbol{a}}$ which plays the role of the transmitted signal. This is the quantum analog of the conventional information-theoretic model of the signal distorted by additive Gaussian noise.

The final element of the communication system is receiver $\mathscr{R}$ destined for obtaining an estimate $\hat{\theta}$ of the actual value θ of the transmitted signal, basing on observations over the system $\mathscr{C}$. Abstracting from details of realization of the estimation procedure, we can say that the receiver performs a measurement of the parameter θ in the family of states $\{S_\theta\}$. Of principal importance are the problems of optimal reception of the signal and the theoretical bounds to accuracy of measurements.

In Chapter 4 we have considered this kind of problems for measurements of kinematical parameters of quantum states, using Bayes' and minimax approaches. Here we develop a different approach which is based on the notion of unbiasedness. This approach does not assume existence and knowledge of the prior distribution and is applicable to an arbitrary family of states without any symmetry. The most complete results are obtained in this way for Gaussian states.

Let $\{S_{\boldsymbol{\theta}}\}$ be a family of quantum states where the parameter

$$\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]$$

runs over a domain $\Theta \subset \mathbb{R}^n$, and $M(\mathrm{d}^n\theta)$, with $\mathrm{d}^n\theta = \mathrm{d}\theta_1 \dots \mathrm{d}\theta_n$ be a measurement with values in Θ. Throughout this section we shall assume that the second moments are finite:

$$\int \hat{\theta}_j^2 \mu_{\boldsymbol{\theta}}(\mathrm{d}^n\hat{\theta}) < \infty; \qquad j = 1, \dots, n, \tag{6.1.1}$$

where $\mu_{\boldsymbol{\theta}}$ is the measurement probability distribution with respect to the state $S_{\boldsymbol{\theta}}$:

$$\mu_{\boldsymbol{\theta}}(B) = \operatorname{Tr} S_{\boldsymbol{\theta}} M(B), \qquad B \in \mathscr{A}(\Theta).$$

The measurement is called *unbiased* if

$$\int \hat{\theta}_j \mu_{\boldsymbol{\theta}}(\mathrm{d}^n\hat{\theta}) = \theta_j; \qquad j = 1, \dots, n, \tag{6.1.2}$$

for all $\theta \in \Theta$. This means that there is no systematic deviation in the results of the measurement. Differentiating this equation formally we

obtain

$$\int \hat{\theta}_j \frac{\partial \mu_{\boldsymbol{\theta}}}{\partial \theta_k}(\mathrm{d}^n\hat{\theta}) = \delta_{jk}; \qquad j,k = 1,\ldots,n. \tag{6.1.3}$$

The precise meaning of this relation will be established later. What we shall need in the most part of this section is the set of conditions (6.1.1)-(6.1.3) satisfied at a fixed point $\boldsymbol{\theta} \in \Theta$. We shall refer to it as *local unbiasedness* at the point $\boldsymbol{\theta}$.

For the deviation function we shall use a quadratic form

$$W_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \sum_{j,k} g_{jk}(\theta_j - \hat{\theta}_j)(\theta_k - \hat{\theta}_k),$$

where $\boldsymbol{G} = [g_{jk}]$ is a real nondegenerate positive definite *weight matrix*. The accuracy of an unbiased measurement $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ is then described by the *total mean-square deviation*

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \int W_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\mu_{\boldsymbol{\theta}}(\mathrm{d}^n\hat{\theta}).$$

For one-dimensional parameter, G is a positive number and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = G\boldsymbol{D}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ where $\boldsymbol{D}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ is the variance of $\boldsymbol{M}$ with respect to the state S_θ. In the general case, introducing the covariance matrix of the unbiased measurement

$$\boldsymbol{B}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \left[\int (\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)\mu_\theta(\mathrm{d}^n\hat{\theta})\right] \equiv [b_{jk}\{\boldsymbol{M}\}], \tag{6.1.4}$$

we have

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \mathbf{Tr}\,\boldsymbol{G}\boldsymbol{B}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \sum_{j,k} g_{jk}b_{jk}\{\boldsymbol{M}\},$$

where **Tr** denotes the trace of an $(n \times n)$-matrix. Thus for $n > 1$ the weight matrix enters essentially in the mean deviation.

If the measurement minimizes $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ it is called the *best* (*unbiased or locally unbiased*) *measurement* of the parameter $\boldsymbol{\theta}$. If there exists a measurement minimizing $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ simultaneously for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ it is called the *uniformly best*. In this chapter we shall obtain some general lower bounds for the mean-square deviation and apply them to estimation of the mean value of Gaussian states.

6.2. A lower bound for the variance in one-dimensional case

Let $\{S_\theta\}$ be a family of states parametrized by a one-dimensional parameter θ. The bound which will be established here is a noncommutative analog of the well-known Cramér-Rao inequality in mathematical statistics.

Concerning the family $\{S_\theta\}$ we assume:

(1) *the family* $\{S_\theta\}$ *is differentiable as a function of* θ *with values in the Banach space of trace-class operators.*

Using this condition we can write for any bounded X

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta(X) = \mathrm{Tr}\,\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta \cdot X, \tag{6.2.5}$$

where $(\mathrm{d}/\mathrm{d}\theta)S_\theta$ is the derivative of the family. The possibility of interchanging the differentiation and the trace follows from the inequality (2.7.72) as generalized in Theorem 2.7.2. We further assume:

(2) *the linear functional of* X *defined by* (6.2.5) *can be extended to a linear continuous functional on the Hilbert space* $\mathscr{L}_h^2(S_\theta)$; *that is*

$$\left|\mathrm{Tr}\,\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta \cdot X\right|^2 \le c\,\mathrm{Tr}\,S_\theta X^2, \qquad X \in \mathfrak{B}_h(\mathscr{H}),$$

where c *is a constant.*

By the Riesz-Frechet lemma there is $L_\theta \in \mathscr{L}_h^2(S_\theta)$ such that $\mathrm{Tr}(\mathrm{d}/\mathrm{d}\theta)S_\theta \cdot X = \langle L_\theta, X\rangle_\theta$ for all bounded X, where $\langle\cdot,\cdot\rangle_\theta \equiv \langle\cdot,\cdot\rangle_{S_\theta}$. By (2.8.88) this is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta = S_\theta \circ L_\theta \equiv \frac{1}{2}(S_\theta L_\theta + L_\theta S_\theta). \tag{6.2.6}$$

An operator $L_\theta \in \mathscr{L}_h^2(S_\theta)$ satisfying (6.2.6) is called the *symmetric logarithmic derivative* of the family $\{S_\theta\}$ at the point θ. For future use we note that

$$\langle I, L_\theta\rangle_\theta = E_\theta(L_\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta(I) = 0. \tag{6.2.7}$$

Consider a measurement $M(\mathrm{d}\theta)$ with the probability distribution $\mu_\theta(\mathrm{d}\hat{\theta}) = \mathrm{Tr}\,S_\theta M(\mathrm{d}\hat{\theta})$. As we have agreed in Section 6.1 we assume that it has finite second moment; moreover we assume that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int \hat{\theta}\mu_\theta(\mathrm{d}\hat{\theta}) = \int \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}), \tag{6.2.8}$$

where $\mathrm{d}\mu_\theta/\mathrm{d}\theta$ is the real σ-additive set function (of finite total variation) defined by the relation

$$\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(B) = \mathrm{Tr}\,\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta M(B), \qquad B \in \mathscr{A}(\Theta).$$

Proposition 6.2.1. *Let the family* $\{S_\theta\}$ *satisfy the condition* (1) *and* (2) *and a measurement* $\boldsymbol{M} = \{M(\mathrm{d}\hat{\theta})\}$ *satisfy* (6.2.8) *at a point* θ. *Then*

$$D_\theta\{\boldsymbol{M}\} D_\theta(L_\theta) \geq \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{E_\theta}\{\boldsymbol{M}\}\right]^2. \tag{6.2.9}$$

Proof. From (6.2.5) and the definition of L_θ

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta(X) = \langle L_\theta, X\rangle_\theta$$

for all bounded X. We first extend this relation to measurements $\boldsymbol{M} = \{M(\mathrm{d}\theta)\}$ with finite second moments. Putting $X = M(B)$ we get

$$\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(B) = \langle L_\theta, M(B)\rangle_\theta. \tag{6.2.10}$$

Consider the operator

$$X_{\boldsymbol{M}} = \int \hat{\theta} M(\mathrm{d}\hat{\theta}) \in \mathscr{L}_h^2(S_\theta),$$

defined as in Section 2.8. Since the integral is the limit in $\mathscr{L}_h^2(S_\theta)$ of finite integral sums, then from (6.2.10)

$$\begin{aligned}\langle X_{\boldsymbol{M}}, L_\theta\rangle_\theta &= \left\langle \int \hat{\theta} M(\mathrm{d}\hat{\theta}), L_\theta \right\rangle_\theta \\ &= \int \hat{\theta}\langle M(\mathrm{d}\hat{\theta}), L_\theta\rangle_\theta = \int \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}).\end{aligned} \tag{6.2.11}$$

Using (6.2.8) we get

$$\langle L_\theta, X_{\boldsymbol{M}}\rangle_\theta = \frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta\{\boldsymbol{M}\}. \tag{6.2.12}$$

Now by (2.9.106)

$$D_\theta\{\boldsymbol{M}\} \geq \langle X_{\boldsymbol{M}} - E_\theta\{\boldsymbol{M}\}, X_{\boldsymbol{M}} - E_\theta\{\boldsymbol{M}\}\rangle_\theta.$$

Using the Cauchy inequality and (6.2.7) we get

$$\begin{aligned}&\langle X_{\boldsymbol{M}} - E_\theta\{\boldsymbol{M}\}, X_{\boldsymbol{M}} - E_\theta\{\boldsymbol{M}\}\rangle_\theta \cdot \langle L_\theta, L_\theta\rangle_\theta \\ &\geq \langle X_{\boldsymbol{M}} - E_\theta\{\boldsymbol{M}\}, L_\theta\rangle_\theta^2 = \langle X_{\boldsymbol{M}}, L_\theta\rangle_\theta^2.\end{aligned}$$

Substituting (6.2.12) we obtain the inequality (6.2.9). □

For an unbiased measurement $\boldsymbol{M}$ (6.2.9) gives

$$D_\theta\{\boldsymbol{M}\} \geq D_\theta(L_\theta)^{-1}. \tag{6.2.13}$$

The above proof shows that in fact all that is needed for (6.2.13) is the relation (6.1.3) which takes the form

$$\int \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}) = 1$$

or, due to (6.2.11)

$$\langle L_\theta, X_{\boldsymbol{M}}\rangle_\theta = 1. \tag{6.2.14}$$

We thus have

Corollary 6.2.2. *The inequality* (6.2.13) *holds for any family* $\{S_\theta\}$ *satisfying the conditions* (1) *and* (2) *and any measurement* $\boldsymbol{M} = \{M(\mathrm{d}\hat{\theta})\}$ *satisfying* (6.2.14)*, in particular for measurements, locally unbiased at the point* θ*.*

The quantity $J_\theta = D_\theta(L_\theta)$ appearing in the right-hand side of (6.2.13) is the noncommutative analog of the Fisher information in mathematical statistics. It is a measure of information about the parameter θ contained in the family $\{S_\theta\}$.

In what follows we shall directly use the conditions of the type (6.2.14) which are both easy to check and relevant for the inequalities we need. However for completeness we shall give a sufficient condition for (6.2.8) which allows to deduce local unbiasedness from unbiasedness by differentiation.

Proposition 6.2.3. *Let the family* $\{S_\theta\}$ *satisfy the condition* (1) *in an interval of values of* θ *and* $-T \leq \mathrm{d}S_\theta/\mathrm{d}\theta \leq T$*, where* T *is a positive trace-class operator. Let a measurement* $\boldsymbol{M} = \{M(\mathrm{d}\hat{\theta})\}$ *be such that* $\int |\hat{\theta}|\mu(\mathrm{d}\hat{\theta}) < \infty$ *where* $\mu(\mathrm{d}\hat{\theta}) = \mathrm{Tr}\, T M(\mathrm{d}\hat{\theta})$*. Then* (6.2.8) *holds, for any* θ *in the interval.*

Proof. By the Lagrange formula for finite increment

$$\begin{aligned}\frac{\mu_{\theta+\Delta\theta}(B) - \mu_\theta(B)}{\Delta\theta} &= \frac{\mathrm{d}}{\mathrm{d}\theta}\mu_{\theta+h\Delta\theta}(B)\\ &= \mathrm{Tr}\,\frac{\mathrm{d}}{\mathrm{d}\theta}S_{\theta+h\Delta\theta}M(B), \qquad 0 < h < 1,\end{aligned}$$

whence

$$\left|\frac{\mu_{\theta+\Delta\theta}(B) - \mu_\theta(B)}{\Delta\theta}\right| \leq \mathrm{Tr}\, T M(B) = \mu(B), \qquad B \in \mathscr{A}(\Theta),$$

and thus $|\mathrm{d}/\mathrm{d}\theta\mu_\theta(B)| \le \mu(B)$. Therefore the integral $\int \hat{\theta}(\mathrm{d}\mu_\theta/\mathrm{d}\theta)(\mathrm{d}\hat{\theta})$ is well defined and converges. For a finite c,

$$\int_{|\hat{\theta}|\le c} \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}) = \mathrm{Tr}\,\frac{\mathrm{d}}{\mathrm{d}\theta}\int_{|\hat{\theta}|\le c}\hat{\theta}M(\mathrm{d}\hat{\theta}) = \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{|\hat{\theta}|\le c}\hat{\theta}\mu_\theta(\mathrm{d}\hat{\theta}).$$

On the other hand

$$\left|\frac{1}{\Delta\theta}\left[\int_{|\hat{\theta}|>c}\hat{\theta}\mu_{\theta+\Delta\theta}(\mathrm{d}\hat{\theta}) - \int_{|\hat{\theta}|>c}\hat{\theta}\mu_\theta(\mathrm{d}\hat{\theta})\right] - \int_{|\hat{\theta}|>c}\hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta})\right| \le 2\int_{|\hat{\theta}|>c}|\hat{\theta}|\mu(\mathrm{d}\theta).$$

Thus the left-hand side converges uniformly to zero as $c \to \infty$ which implies (6.2.8). □

Finally we shall apply the results of this section to estimation of parameter $\overline{Q}$ in the family of the quasiclassical Gaussian states (5.5.46) (the parameter $\overline{P}$ is assumed to be fixed). By Proposition 5.6.3 this family satisfies the conditions (1) and (2) with the symmetric logarithmic derivative

$$L_{\overline{Q}} = \sigma_Q^{-1}(Q - \overline{Q})$$

found from the second of the relations (5.6.64). Therefore $D_{\overline{Q}}(L_{\overline{Q}}) = \sigma_Q^{-2}$ and the inequality (6.2.13) gives

$$D_{\overline{Q}}\{\boldsymbol{M}\} \ge \sigma_Q^2 = \frac{\hbar}{\omega}\left(\overline{N} + \frac{1}{2}\right) \tag{6.2.15}$$

for any locally unbiased measurement of the parameter $\overline{Q}$.

Apparently this bound is attained (for any value of $\overline{Q}$) by the simple measurement $E(\mathrm{d}\overline{Q})$ corresponding to the coordinate observable Q. Using Dirac's notation we can write

$$E(\mathrm{d}\overline{Q}) = |\overline{Q})(\overline{Q}|\mathrm{d}\overline{Q}, \tag{6.2.16}$$

where $|\overline{Q})$ are formal eigenvectors of the operator Q. Since $E(\mathrm{d}\overline{Q})$ is the spectral measure of Q, $X_{\boldsymbol{E}} = Q$ and the condition (6.2.14) is satisfied since

$$\langle L_{\overline{Q}}, Q\rangle_{\overline{Q}} = \sigma_Q^{-2}\langle Q - \overline{Q}, Q\rangle_{\overline{Q}} = \sigma_Q^{-2}\langle Q - \overline{Q}, Q - \overline{Q}\rangle_{\overline{Q}} = 1.$$

Thus the *canonical coordinate observable Q gives the uniformly best locally unbiased measurement of the coordinate parameter $\overline{Q}$ in the family $\{S_{\overline{P},\overline{Q}}\}$ of quasiclassical Gaussian states*. Similar result holds for the momentum parameter $\overline{P}$.

6.3. The case of shift parameter

Here we shall prove the inequality

$$D_\theta\{\boldsymbol{M}\}D_\theta(A) \geq \frac{1}{4}\left|\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta\{\boldsymbol{M}\}\right|^2 \tag{6.3.17}$$

for the variance of measurements of the shift parameters θ in the family

$$S_\theta = \mathrm{e}^{\mathrm{i}A\theta}S\mathrm{e}^{-\mathrm{i}A\theta}. \tag{6.3.18}$$

This is generalization of the Mandelstam-Tamm inequality (3.2.9) which is connected with uncertainty relations for shift parameters, and we wish to compare it with the noncommutative Cramér-Rao inequality (6.2.9) originating from analogies with mathematical statistics.

First we establish the condition ensuring the properties (1) and (2) of Section 6.2 for the family (6.3.18).

Proposition 6.3.1. *Let the infinitesimal generator A of the unitary group $\{V_\theta = \exp \mathrm{i}\theta A; \theta \in \mathbb{R}\}$ be square-summable with respect to the basic state S. Then it is square-summable with respect to S_θ, $\theta \in \mathbb{R}$. The family $\{S_\theta\}$ satisfies the condition* (1) *and its derivative is given by*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta = \mathrm{i}[A, S_\theta], \tag{6.3.19}$$

where the commutator is defined as in Section 2.8. *Moreover $\{S_\theta\}$ satisfies condition* (2) *and its symmetric logarithmic derivative is equal to*

$$L_\theta = \mathrm{e}^{\mathrm{i}\theta A}\mathfrak{D}(A)\mathrm{e}^{-\mathrm{i}\theta A},$$

where $\mathfrak{D}$ is the commutation operator of the basic state S.

Proof. We first prove that $A \in \mathscr{L}_h^2(S)$ implies $A \in \mathscr{L}_h^2(S_\theta)$. As it was observed before the formulation of Stone's theorem in Section 2.4, $\psi \in \mathscr{D}(A)$ implies $V_\theta\psi \in \mathscr{D}(A)$. Moreover since V_θ is a function of A, $AV_\theta\psi = V_\theta A\psi$. Therefore $\mathscr{D}(A) \supset \mathscr{R}(\sqrt{S})$ implies

$$\mathscr{D}(A) \supset \mathscr{R}(V_\theta\sqrt{S}V_\theta^*) = \mathscr{R}(\sqrt{S_\theta})$$

and the operator $A\sqrt{S_\theta} = AV_\theta\sqrt{S}V_\theta^* = V_\theta A\sqrt{S}V_\theta^*$ is Hilbert-Schmidt. It remains to refer to Proposition 2.8.2.

Put now $S_\theta = T_\theta R_\theta$, where $T_\theta = V_\theta\sqrt{S_\theta}$, $R_\theta = \sqrt{S_\theta}V_\theta^*$. We shall show that the families $\{T_\theta\}$, $\{R_\theta\}$ are differentiable as functions of θ with values in the Hilbert space $\mathfrak{T}^2(\mathscr{H})$, and

$$\frac{\mathrm{d}}{\mathrm{d}\theta}T_\theta = \mathrm{i}AT_\theta, \qquad \frac{\mathrm{d}R_\theta}{\mathrm{d}\theta} = -\mathrm{i}R_\theta A. \tag{6.3.20}$$

We restrict to $\{T_\theta\}$ since consideration of $\{R_\theta\}$ is similar. We have

$$\left\| \frac{T_{\theta+\Delta\theta} - T_\theta}{\Delta\theta} - \mathrm{i}AT_\theta \right\|_2 = \|F_{\Delta\theta}(A)AT_\theta\|_2,$$

where the function

$$F_{\Delta\theta}(x) = (\Delta\theta \cdot x)^{-1}(\mathrm{e}^{\mathrm{i}\Delta\theta\cdot x} - 1 - \mathrm{i}\Delta\theta \cdot x)$$

has the properties

$$|F_{\Delta\theta}(x)| \leq \text{const.};$$
$$\lim_{\Delta\theta\to 0} F_{\Delta\theta}(x) = 0, \qquad x \in \mathbb{R}.$$

By using the spectral representation of $F_{\Delta\theta}(A)$ one can show that

$$\|F_{\Delta\theta}(A)\| \leq \text{const.}; \tag{6.3.21}$$
$$\lim_{\Delta\theta\to 0} F_{\Delta\theta}(A)\psi = 0, \qquad \psi \in \mathscr{H}. \tag{6.3.22}$$

Let us show now that $\|F_{\Delta\theta}(A)AT_\theta\|_2 \to 0$ as $\Delta\theta \to 0$. Since $Q = AT_\theta \in \mathfrak{T}^2(\mathscr{H})$, Q can be approximated by finite-rank operators $\widetilde{Q}$. From (6.3.22) it follows that $\|F_{\Delta\theta}(A)\widetilde{Q}\|_2 \to 0$ as $\Delta\theta \to 0$. On the other hand by the inequality (2.7.81)

$$\|F_{\Delta\theta}(A) \cdot (Q - \widetilde{Q})\|_2 \leq \|F_{\Delta\theta}(A)\| \cdot \|Q - \widetilde{Q}\|_2,$$

which can be made arbitrarily small using (6.3.21). This proves the first relation in (6.3.20).

The relation (6.3.19) now follows from the simple fact that if $S_\theta = T_\theta \cdot R_\theta$ where $\{T_\theta\}$, $\{R_\theta\}$ are differentiable as functions with values in $\mathfrak{T}^2(\mathscr{H})$, then $\{S_\theta\}$ is differentiable as function with values in $\mathfrak{T}^1(\mathscr{H})$ and

$$\frac{\mathrm{d}}{\mathrm{d}\theta}S_\theta = \frac{\mathrm{d}T_\theta}{\mathrm{d}\theta}R_\theta + T_\theta\frac{\mathrm{d}R_\theta}{\mathrm{d}\theta}.$$

In fact

$$\frac{S_{\theta+\Delta\theta} - S_\theta}{\Delta\theta} = \frac{T_{\theta+\Delta\theta} - T_\theta}{\Delta\theta}R_{\theta+\Delta\theta} + T_\theta\frac{R_{\theta+\Delta\theta} - R_\theta}{\Delta\theta}.$$

We have

$$(\Delta\theta)^{-1}(T_{\theta+\Delta\theta} - T_\theta) \to \frac{\mathrm{d}T_\theta}{\mathrm{d}\theta}, \qquad (\Delta\theta)^{-1}(R_{\theta+\Delta\theta} - R_\theta) \to \frac{\mathrm{d}R_\theta}{\mathrm{d}\theta}$$

in $\mathfrak{T}^2(\mathscr{H})$: the family $\{R_\theta\}$, being differentiable, is continuous so that $R_{\theta+\Delta\theta} \to R_\theta$ in $\mathfrak{T}^2(\mathscr{H})$. Passing to the limit as $\Delta\theta \to 0$ and using (2.7.80) we see that $\{S_\theta\}$ satisfies condition (1) and the expression (6.3.19) for $\mathrm{d}S_\theta/\mathrm{d}\theta$ follows.

Consider the expectation $E_\theta(X)$ of bounded observable X. Using (6.2.5), (6.3.19) and (2.8.90) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta(X) = [X, A]_\theta, \tag{6.3.23}$$

where $[\cdot,\cdot]_\theta \equiv [\cdot,\cdot]_{S_\theta}$. It follows that the family $\{S_\theta\}$ satisfies condition (2) of Section 6.2 since

$$\left|\mathrm{Tr}\,\frac{\mathrm{d}S_\theta}{\mathrm{d}\theta}X\right|^2 = |[A, X]_\theta|^2 \le 4\langle A, A\rangle_\theta\langle X, X\rangle_\theta$$

by the inequality (4) of Proposition 2.8.3.

Comparing (6.2.6) and (6.3.19) we get

$$S_\theta \circ L_\theta = \mathrm{i}[A, S_\theta], \tag{6.3.24}$$

or, which is the same

$$\langle X, L_\theta\rangle_\theta = [X, A]_\theta$$

for all $X \in \mathscr{L}_h^2(S_\theta)$. By the definition (2.10.113) of the commutation operator this means that

$$L_\theta = \mathfrak{D}_\theta(A), \tag{6.3.25}$$

where $\mathfrak{D}_\theta$ is the commutation operator of S_θ. From (6.3.24) and (6.3.18)

$$S \circ (\mathrm{e}^{-\mathrm{i}\theta A}L_\theta\mathrm{e}^{\mathrm{i}\theta A}) = \mathrm{i}[A, S],$$

whence $\mathfrak{D}_\theta(A) = \mathrm{e}^{\mathrm{i}\theta A}\mathfrak{D}(A)\mathrm{e}^{-\mathrm{i}\theta A}$ and we get the desired relation for L_θ. □

Proposition 6.3.2. *Under the condition of Proposition* 6.3.1 *the inequality* (6.3.17) *holds for any measurement* $\boldsymbol{M} = \{M(\mathrm{d}\theta)\}$ *with finite second moment satisfying* (6.2.8).

Proof. Putting $X = M(B)$ in (6.3.23) we get

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mu_\theta(B) = [M(B), A]_\theta, \qquad B \in \mathscr{A}(\Theta).$$

Then we can extend (6.3.23) onto measurements $\boldsymbol{M}$ satisfying the conditions of the proposition:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta\{\boldsymbol{M}\} = [X_{\boldsymbol{M}}, A]_\theta. \tag{6.3.26}$$

Indeed by (6.2.8)

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E_\theta\{\boldsymbol{M}\} = \int \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}).$$

Arguing as in the proof of Proposition 6.2.1 we get

$$\int \hat{\theta}\frac{\mathrm{d}\mu_\theta}{\mathrm{d}\theta}(\mathrm{d}\hat{\theta}) = \int \hat{\theta}[M(\mathrm{d}\hat{\theta}),\, A]_\theta = [X_{\boldsymbol{M}},\, A]_\theta,$$

as required. Combining (2.9.104) with (2.9.106) we get

$$D_\theta\{\boldsymbol{M}\}D_\theta(A) \geq \frac{1}{4}[X_{\boldsymbol{M}},\, A]_\theta^2.$$

Substituting (6.3.26) into the right-hand side of this inequality we get (6.3.17). □

If $\boldsymbol{M}$ is locally unbiased, then (6.3.17) takes the form

$$D_\theta\{\boldsymbol{M}\} \geq [4D_\theta(A)]^{-1}. \tag{6.3.27}$$

This generalizes the uncertainty relation for covariant measurements of the shift parameter displayed in Section 4.7. Thus under the condition of Proposition 6.3.1 there are two different bounds for the variance of a locally unbiased measurement $\boldsymbol{M}$ of the shift parameter θ: the "uncertainty relation" (6.3.27) and the "Cramér-Rao inequality" (6.2.13).

We shall show that *the last inequality is in general more informative* (*i.e., provides a greater lower bound*) *than* (6.3.27) *and that the two bounds coincide in the case of pure state family* $\{S_\theta\}$. First of all we remark that both bounds in fact do not depend on θ. While this is clear for (6.3.27), for (6.2.13) this follows from the expression for L_θ obtained in Proposition 6.3.1 and (6.3.18). Therefore we can assume $\theta = 0$. Using (6.3.25), (6.2.7) and (2.10.115) we get

$$\begin{aligned}\frac{1}{4}D_s(L_0) &= \frac{1}{4}\langle\mathfrak{D}(A), \mathfrak{D}(A)\rangle_S \\ &= \frac{1}{4}\langle\mathfrak{D}(A-\overline{A}), \mathfrak{D}(A-\overline{A})\rangle_S = -\frac{1}{4}\langle(A-\overline{A}), \mathfrak{D}^2(A-\overline{A})\rangle_S,\end{aligned}$$

where $\overline{A} = E_S(A)$. Since $1 + \frac{1}{4}\mathfrak{D}^2 \geq 0$,

$$\frac{1}{4}D_S(L_0) \leq \langle A-\overline{A},\, A-A\rangle_S = D_S(A),$$

which proves that (6.2.13) is at least as good as (6.3.27). The equality holds if an only if

$$\left(I + \frac{1}{4}\mathfrak{D}^2\right)(A-\overline{A}) = 0. \tag{6.3.28}$$

To explain the meaning of this condition consider the matrix representation of the operator $A \in \mathscr{L}^2(S)$, as described by (2.10.110). According to the last relation in (2.10.118) the operator $I + \frac{1}{4}\mathfrak{D}^2$ acts multiplying the matrix element in the j^{th} row and the k^{th} column by $4s_j s_k (s_j + s_k)^{-1}$, where $\{s_j\}$ are the eigenvalues of S. Therefore $(I + \frac{1}{4}\mathfrak{D}^2)X = 0$ if and only if $SXS = 0$, so that (6.3.28) is equivalent to

$$SAS = \overline{A} \cdot S^2.$$

This is apparently satisfied if S is a pure state.

Let us return to the example considered at the end of Section 6.2. According to the results of Section 5.6 the Gaussian states $\{S_{\overline{P},\overline{Q}}\}$ satisfy

$$S_{\overline{P},\overline{Q}} = \mathrm{e}^{-\mathrm{i}\overline{Q}P} S \mathrm{e}^{\mathrm{i}\overline{Q}P},$$

where $S = S_{\overline{P},0}$. Applying (6.3.27) with $A = -P$ we get only

$$D_{\overline{Q}}\{\boldsymbol{M}\} \geq (2\sigma_P)^{-2},$$

as compared to the bound σ_Q^2 of (6.2.15). By the Heisenberg uncertainty relation $\sigma_Q^2 \geq (2\sigma_P)^{-2}$ and the equality holds only for the pure minimum-uncertainty state.

Being less informative, the bound (6.3.27) may still be useful since it is expressed strictly in terms of the infinitesimal operator A and does not require knowledge of the symmetric logarithmic derivative. But if the latter is available, as in the Gaussian case, then the more informative bound (6.2.13) should be used.

Although (6.3.17) applies as well to the angular parameter φ in the family $\{S_\varphi\}$ defined by (4.6.36), it is not the inequality which should be used in this case. As we have already seen in Section 4.5 the variance $D_\varphi\{\boldsymbol{M}\}$ need to be replaced by the uncertainty defined by (6.5.45). Moreover the unbiasedness condition must be also modified. One easily sees that a covariant measurement $\boldsymbol{M} = \{M(\mathrm{d}\varphi)\}$ of an angular quantity is necessarily biased. To find the substitute for the unbiasedness condition we introduce the bounded operator

$$U_M = \int_0^{2\pi} \mathrm{e}^{\mathrm{i}\varphi} M(\mathrm{d}\varphi).$$

Then denoting $E_\varphi(U) = \operatorname{Tr} S_\varphi U$ we have

$$E_\varphi(U_M) = \mathrm{e}^{\mathrm{i}\varphi} E_0(U_{\boldsymbol{M}}),$$

as a consequence of the covariance property of $\boldsymbol{M}$. It follows that

$$\mathrm{i}^{-1}\frac{\mathrm{d}}{\mathrm{d}\varphi}\ln E_\varphi(U_{\boldsymbol{M}}) = 1. \tag{6.3.29}$$

This replaces the local unbiasedness property in the angular case.

We now prove the inequality

$$\Delta_\varphi\{\boldsymbol{M}\}\cdot D_\varphi(A) \geq \frac{1}{4}\left|\frac{\mathrm{d}}{\mathrm{d}\varphi}\ln E_\varphi(U_{\boldsymbol{M}})\right|^2 \tag{6.3.30}$$

under the same assumption of square-summability of the operator A in the family (4.6.36), for any measurement $\boldsymbol{M} = \{M(\mathrm{d}\varphi)\}$.

From (6.3.23), $(\mathrm{d}/\mathrm{d}\varphi)E_\varphi(U_{\boldsymbol{M}}) = [A, U_{\boldsymbol{M}}]_\varphi$, so that

$$\frac{\mathrm{d}}{\mathrm{d}\varphi}\ln E_\varphi(U_{\boldsymbol{M}}) = [A, U_{\boldsymbol{M}}]_\varphi E_\varphi(U_{\boldsymbol{M}})^{-1}.$$

Using (2.8.99) we get

$$\frac{1}{4}|[A, U_{\boldsymbol{M}}]_\varphi|^2 \leq \langle A-\overline{A}, A-\overline{A}\rangle_\varphi \langle U_{\boldsymbol{M}}-\overline{U}_{\boldsymbol{M}}, U_{\boldsymbol{M}}-\overline{U}_{\boldsymbol{M}}\rangle_\varphi,$$

where $\overline{A} = E_\varphi(A)$ etc. Applying (2.9.109):

$$\begin{aligned}\langle U_{\boldsymbol{M}}-\overline{U}_{\boldsymbol{M}}, U_{\boldsymbol{M}}-\overline{U}_{\boldsymbol{M}}\rangle_\varphi &\leq \int |\mathrm{e}^{\mathrm{i}\hat{\varphi}}-\overline{U}_{\boldsymbol{M}}|^2\mu_\varphi(\mathrm{d}\hat{\varphi}) \\ &\equiv \Delta_\varphi\{\boldsymbol{M}\}\cdot|E_\varphi(U_{\boldsymbol{M}})|^2.\end{aligned}$$

Summing up we arrive to (6.3.30).

If now the measurement $\boldsymbol{M}$ satisfies (6.3.29) (in particular if $\boldsymbol{M}$ is covariant), then we obtain the modification of (6.3.27) to the angular parameter

$$\Delta_\varphi\{\boldsymbol{M}\} \geq [4D_\varphi(A)]^{-1}.$$

Arguing as in Section 6.2 we can improve this bound to

$$\Delta_\varphi\{\boldsymbol{M}\} \geq D_\varphi(L_\varphi)^{-1}$$

where L_φ is the symmetric logarithmic derivative of the family (4.6.36).

6.4. Estimation of force by measurements over a trial object

Consider a quantum object of mass m under the action of constant force F. Since the potential energy is equal to $V(x) = -Fx$ the Hamiltonian of the object is

$$H = \frac{1}{\hbar}\left(\frac{p^2}{2m} - Fq\right).$$

Writing the dynamical equation (3.7.69) in the momentum representation (3.4.38) we get

$$\mathrm{i}\frac{\partial\tilde{\psi}_t(\eta)}{\partial t} = -\frac{\eta^2\tilde{\psi}_t(\eta)}{2m\hbar} - \mathrm{i}F\frac{\partial\tilde{\psi}_t(\eta)}{\partial\eta}.$$

This is a linear first order partial differential equation and its solution is easily found to be

$$\tilde{\psi}_t(\eta) \equiv V_t\tilde{\psi}_0(\eta) = \tilde{\psi}_0(\eta - Ft) \cdot \exp\left(-\frac{\mathrm{i}\eta^2 t}{2m\hbar} + \frac{\mathrm{i}\eta Ft^2}{2m\hbar} - \frac{\mathrm{i}F^2t^3}{6m\hbar}\right).$$

Using the relation (3.4.39) we can write the operator of time evolution $V_t = \exp[-(\mathrm{i}t/\hbar)(p^2/2m - Fq)]$ in the form

$$V_t = W_{Ft^2/2m,Ft/m}V_t^0 \cdot \exp\left(\frac{\mathrm{i}F^2t^3}{12m\hbar}\right). \tag{6.4.31}$$

Here

$$W_{Ft^2/2m,Ft/m} = \exp\frac{\mathrm{i}}{\hbar}\left(Ftq - \frac{Ft^2}{2m}p\right) \tag{6.4.32}$$

is the displacement operator corresponding to the kinematical transformation

$$(x, v) \to \left(x + \frac{Ft^2}{2m}, v + \frac{Ft}{m}\right);$$

$V_t^0 = \exp(-\mathrm{i}p^2/2m\hbar)$ is the operator of free evolution and the last exponent is an inessential scalar factor of unit modulus.

The relation (6.4.31) has simple physical meaning. In classical mechanics the motion in the field of constant force F is described by the relations

$$p(t) = p + Ft, \qquad q(t) = q + \frac{p}{m}t + \frac{Ft^2}{2m}. \tag{6.4.33}$$

The transformation $(p, q) \to (p(t), q(t))$ can be composed of the kinematical displacement $(p, q) \to (p+Ft, q+Ft^2/2m)$ and the free evolution $(p, q) \to (p, q+(p/m)t)$. The relation (6.4.31) is the quantum analog of this fact expressed in the Schrödinger picture. Introducing time-dependent observables $p(t) = V_t^* p V_t$, $q(t) = V_t^* q V_t$ one can derive from (6.4.31) the quantum relations (6.4.33) in the Heisenberg picture.

Summing up, we can say that if the object prepared initially in the state S, was exposed to the action of force F during the time t, then the resulting state will be

$$V_t S_t V_t^* = W_{Ft^2/2m, Ft/m} S_t^0 W_{Ft^2/2m, Ft/m}^*,$$

where $S_t^0 = V_t^0 S (V_t^0)^*$ is the basic state which would result if the force were absent. Denoting the resulting state by S_F and using (6.4.32) we can write

$$S_F = \mathrm{e}^{\mathrm{i}FA} S_t^0 \mathrm{e}^{-\mathrm{i}FA}, \tag{6.4.34}$$

where

$$A = \frac{t}{\hbar}\left(q - \frac{p}{2m}t\right). \tag{6.4.35}$$

We now turn to the question – what is the extremal accuracy with which the value of F can be estimated by quantum measurements after the time t. The foregoing discussion reduces this question to estimation of the shift parameter F in the family (6.4.34), so that we can apply the results of Section 6.3.

If the operator (6.4.35) has finite second moment with respect to S_t^0, then the variance of any locally unbiased measurement $\boldsymbol{M}$ of the force F satisfies (6.3.27). Let us express the quantity $D_F(A)$ which enters into (6.3.27) through the variances corresponding to the initial state S. Passage from Schrödinger to Heisenberg picture gives

$$D_F(A) \equiv D_{S_t^0}(A) = \frac{t^2}{\hbar^2} D_S\left(q^0(t) - \frac{p^0(t)}{2m}t\right), \tag{6.4.36}$$

where $q^0(t) = (V_t^0)^* q V_t^0$, $p^0(t) = (V_t^0)^* p V_t^0$ are the time-dependent canonical observables for the free motion. From (3.7.66) and (3.7.65)

$$p^0(t) = p, \qquad q^0(t) = q + \frac{p}{m}t. \tag{6.4.37}$$

Substituting it into (6.4.36) we obtain *the inequality for the variance of a locally unbiased measurement* $\boldsymbol{M}$ *of the force* F

$$D_F\{\boldsymbol{M}\} \geq \frac{\hbar^2}{4t^2 D_S\left(q + \frac{p}{2m}t\right)}, \tag{6.4.38}$$

which holds under the assumption that p and q have finite second moments with respect to the initial state S.

Assuming further that q and p are *uncorrelated*, *i.e.*,

$$\langle q - E_S(q),\, p - E_S(p)\rangle_S = 0 \tag{6.4.39}$$

for the initial state S we get from (6.4.38)

$$D_F\{\boldsymbol{M}\} \geq \frac{\hbar^2}{4t^2}\left[D_S(q) + \frac{t^2}{4m^2}D_S(p)\right]^{-1}. \tag{6.4.40}$$

By the Heisenberg uncertainty relation, the right-hand side can not exceed the value $\hbar m/2t^3$ which is achieved for $D_S(p) = \hbar m/t$, $D_s(q) = \hbar t/4m$. Thus *under the most unfavorable* conditions the bound becomes

$$D_F\{\boldsymbol{M}\} \geq \frac{\hbar m}{2t^3} \tag{6.4.41}$$

However if the initial state is properly prepared, the extremal quantum accuracy of the force measurements can be made arbitrarily high for a fixed t. To show this we introduce the observables of the form

$$B = \alpha\frac{2mq}{t^2} + \beta\frac{p}{t}, \qquad \alpha + \beta = 1,$$

which are canonically conjugate in the sense of Section 3.2 to A as follows easily from the expressions for A and B, which are just linear combinations of p and q, and from the CCR. Then as it was shown in Section 3.2, B corresponds to a covariant measurement of parameter F and therefore it is up to a constant an unbiased estimate of F. A strictly unbiased estimate is obtained by subtracting from p and q their basic mean values which according to (6.4.37) are

$$E_{S_t^0}(p) = E_S(p), \qquad E_{S_t^0}(q) = E_S(q) + \frac{t}{m}E_S(p).$$

Using (6.4.37) and (6.4.39) we obtain the variance of B, which is the same for all values of F

$$\begin{aligned} D_{S_t^0}(B) &= D_S\left(\alpha\cdot 2m\left(q + \frac{t}{m}p\right)t^{-1} + \beta p t^{-1}\right) \\ &= \frac{4m^2}{t^4}D_S(q)\alpha^2 + \frac{1}{t^2}D_S(p)(2\alpha+\beta)^2. \end{aligned}$$

The minimum of this quantity under the restriction $\alpha+\beta = 1$ is achieved for

$$\alpha = -D_S(p)\left[\frac{4m^2}{t^2}D_S(q) + D_S(p)\right]^{-1}$$

which corresponds to the unbiased estimate[1]

$$B_* = t^{-1}\left[D_S(p)^{-1} + \frac{t^2}{4m^2}D_S(q)^{-1}\right]^{-1} \cdot \left[\frac{p - E_S(p)}{D_S(p)} - \frac{t}{2m}\frac{q - \frac{p}{m}t - E_S(q)}{D_S(q)}\right], \tag{6.4.42}$$

having the variance

$$D_{S_i^0}(B_*) = t^{-2}\left[D_S(p)^{-1} + \frac{t^2}{4m^2}D_S(q)^{-1}\right]^{-1} = \frac{D_S(q)\cdot D_S(p)}{t^2\left[D_S(q) + \frac{t^2}{4m^2}D_S(q)\right]}. \tag{6.4.43}$$

As it should be expected this quantity satisfies (6.4.40); moreover the equality is achieved if the initial state S is a minimum-uncertainty state.

The variance of the estimate B_* tends to zero if either $D_S(p)$ or $D_S(q)$ tends to zero. First let $D_S(p) \approx 0$, *i.e.*, the initial state has the sharply determined momentum $E_S(p)$, then (6.4.42) implies that $B_* \approx [p - E_S(p)]t^{-1}$. In this case measuring observable $[p - E_S(p)]t^{-1}$ gives almost definitely the value of the force F. This agrees with semiclassical consideration based on the formula $F \cdot t = \Delta p$ from the Newtonian mechanics.

If on the contrary $D_S(q) \approx 0$, then (6.4.42) gives

$$B_* \approx \frac{2m}{t^2}\left(q - p\frac{t}{m} - E_S(q)\right). \tag{6.4.44}$$

Assume for a moment that p, q are the classical canonical variables. Then $q - pt/m$ is just the value of coordinate starting from which at $t = 0$ the free object will reach the coordinate q at the moment t. To measure the quantity like (6.4.43) one may in principle perform the joint measurement of p and q with arbitrary accuracy and substitute the results in the corresponding expression to obtain a final result. Comparison with (6.4.43)

[1] A reader familiar with mathematical statistics can readily see that the same expression (6.4.42) with p, q replaced by $p(t), q(t)$ describes the best linear unbiased estimate of F in the classical estimation problem using observations $p(t), q(t)$ given by (6.4.33) with p, q regarded as usual random variables.

and (6.4.42) shows that $D_{S_t^0}(L_0)^{-1} \equiv D_0(L_0)^{-1} = D_{S_t^0}(B_*)$ so that the estimate $B_* = L_0/D_0(L_0)$ attains the lower bound in (6.2.13), which now can be written as

$$D_F\{\boldsymbol{M}\} \geq t^{-2}\left[D_S(p)^{-1} + \frac{t^2}{4m^2}D_S(q)^{-1}\right]^{-1}.$$

This bound is more informative than (6.4.40), but it is valid only for Gaussian states.

6.5. A bound for the measurement covariance matrix based on symmetric logarithmic derivatives

Consider a family of states $\{S_{\boldsymbol{\theta}}\}$ where $\boldsymbol{\theta} = [\theta_1, \dots \theta_n]$ is a multidimensional parameter running over a domain Θ. For example, this can be the family of Gaussian states $\{S_{\overline{P},\overline{Q}}\}$ with the two-dimensional parameter $[\overline{P}, \overline{Q}]$. We shall adopt the following multidimensional analog of conditions (1) and (2) of Section 6.2.

(1) *The family $\{S_{\boldsymbol{\theta}}\}$ is differentiable as a function of $\theta_1, \dots \theta_n$ with values in the Banach space of trace-class operators.*
(2) *The linear functionals*

$$\frac{\partial}{\partial\theta_j}E_{\boldsymbol{\theta}}(X) = \operatorname{Tr}\frac{\partial S_{\boldsymbol{\theta}}}{\partial\theta_j}X, \qquad X \in \mathfrak{B}_h(\mathscr{H})$$

can be extended to continuous linear functionals on the Hilbert space $\mathscr{L}_h^2(S_{\boldsymbol{\theta}})$, i.e., there is a constant c such that

$$\left|\operatorname{Tr}\frac{\partial}{\partial\theta_j}S_{\boldsymbol{\theta}}\cdot X\right|^2 \leq c\operatorname{Tr}S_{\boldsymbol{\theta}}X^2; \qquad X \in \mathfrak{B}_h(\mathscr{H}), \qquad j = 1, \dots, n.$$

As in the one-dimensional case these conditions ensure the existence of the *symmetric logarithmic derivatives* $L_{\boldsymbol{\theta}}^j$; $j = 1, \dots, n$, which can be defined as the elements of $\mathscr{L}_h^2(S_{\boldsymbol{\theta}})$ satisfying

$$\frac{\partial S_{\boldsymbol{\theta}}}{\partial\theta_j} = S_{\boldsymbol{\theta}} \circ L_{\boldsymbol{\theta}}^j. \tag{6.5.45}$$

Since all our considerations will refer to a fixed point $\boldsymbol{\theta}$ we shall simplify notations by omitting the subscript $\boldsymbol{\theta}$. We thus write S for $S_{\boldsymbol{\theta}}$, $\langle\cdot,\cdot\rangle_S$ for $\langle\cdot,\cdot\rangle_{S_{\boldsymbol{\theta}}}$ etc. The symmetric logarithmic derivatives will be denoted L^j; $j = 1, \dots, n$.

We are interested in measurements $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ of the multidimensional parameter $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$. From the very beginning we restrict our attention to *measurements which are locally unbiased at the given point* $\boldsymbol{\theta}$, *i.e.*, satisfy (6.1.1)-(6.1.3) with

$$\frac{\partial \mu_{\boldsymbol{\theta}}}{\partial \theta_k}(B) = \mathrm{Tr}\,\frac{\partial S_{\boldsymbol{\theta}}}{\partial \theta_k} M(B); \qquad B \in \mathscr{A}(\Theta).$$

By finiteness of the second moments the integrals

$$X^j_{\boldsymbol{M}} = \int \hat{\theta}_j M(\mathrm{d}^n\hat{\theta})$$

are defined as elements of $\mathscr{L}^2_h(S)$. Arguing as in (6.2.11) we get an alternative formulation of the condition (6.1.3)

$$\langle X^j_M, L^k\rangle_S = \delta_{jk}.$$

Since as in (6.2.7)

$$\langle I, L^j\rangle_S = \frac{\partial}{\partial \theta_j} E_\theta(I) = 0,$$

this is equivalent to

$$\langle L^j, X_k\rangle_S = \delta_{jk}, \tag{6.5.46}$$

where we have introduced the new variables

$$X_j = X^j_{\boldsymbol{M}} - \theta_j = \int (\hat{\theta}_j - \theta_j) M(\mathrm{d}^n\hat{\theta}). \tag{6.5.47}$$

Introducing the Gram matrix of the system $\{L^j\}$ in $\mathscr{L}^2_h(S)$

$$\boldsymbol{J} = [\langle L^j, L^k\rangle_S],$$

we call it the *information matrix*. *If there exists a locally unbiased measurement, then* $\boldsymbol{J}$ *is nondegenerate and the covariance matrix* (6.1.4) *of any locally unbiased measurement satisfies the multidimensional analog of the inequality* (6.2.13):

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq \boldsymbol{J}^{-1}. \tag{6.5.48}$$

This means that for any row-vector $\boldsymbol{v} = [v_1, \ldots, v_n]$

$$\boldsymbol{v}\boldsymbol{B}\{\boldsymbol{M}\}\boldsymbol{v}^* \geq \boldsymbol{v}\boldsymbol{J}^{-1}\boldsymbol{v}^*,$$

where $\boldsymbol{v}^*$ is the Hermitean conjugate column-vector. Since both matrices in (6.5.48) are real, we can confine ourselves to real vectors $\boldsymbol{v}$.

To prove (6.5.48) we first note that the covariance matrix $\boldsymbol{B}\{\boldsymbol{M}\}$ satisfies

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq [\langle X_j, X_k\rangle_S], \tag{6.5.49}$$

where $\{X_j\}$ are defined by (6.5.47). This follows from (2.9.108) by taking $f(\hat{\theta}) = \sum v_j(\hat{\theta}_j - \theta_j)$ with $\{v_j\}$ being arbitrary real numbers. The remaining follows from

Lemma 6.5.1. *Let X_j, L^j; $j = 1, \ldots, n$ be a pair of systems of vectors in a linear space $\mathscr{L}$, satisfying*

$$\langle L^j, X_k\rangle = \delta_{jk}; \qquad j, k = 1, \ldots, n,$$

where $\langle\cdot,\cdot\rangle$ is a pre-inner product on $\mathscr{L}$. Then the Gram matrices of both systems $\boldsymbol{\Gamma}_X = [\langle X_j, X_k\rangle]$ and $\boldsymbol{\Gamma}_L = [\langle L^j, L^k\rangle]$ are nondegenerated and satisfy $\boldsymbol{\Gamma}_X \geq \boldsymbol{\Gamma}_L^{-1}$.

Proof. Introducing vectors $X = \sum u_j X_j$, $Y = \sum v_j L^j$ we have $\langle X, Y\rangle = \boldsymbol{v}\boldsymbol{u}^*$. The Cauchy inequality for $\langle\cdot,\cdot\rangle$ gives

$$\boldsymbol{u}\boldsymbol{\Gamma}_X\boldsymbol{u}^* \cdot \boldsymbol{v}\boldsymbol{\Gamma}_L\boldsymbol{v}^* = \langle X, X\rangle\langle Y, Y\rangle \geq (\boldsymbol{v}\boldsymbol{u}^*)^2.$$

Putting $\boldsymbol{v} = \boldsymbol{u}\Gamma_L^{-1}$ we see that $\boldsymbol{\Gamma}_X$, $\boldsymbol{\Gamma}_L$ cannot be degenerate. Putting $\boldsymbol{v} = \boldsymbol{u}\boldsymbol{\Gamma}_L^{-1}$ we get $\boldsymbol{u}\boldsymbol{\Gamma}_X\boldsymbol{u}^* \geq \boldsymbol{u}\boldsymbol{\Gamma}_L^{-1}\boldsymbol{u}^*$ and the lemma is proved. □

Combining (6.5.49) with Lemma 6.5.1 gives the inequality (6.5.48).

This inequality provides at once a simple lower bound for the total mean square deviation $\boldsymbol{\Sigma}\{\boldsymbol{M}\} = \mathbf{Tr}\,\boldsymbol{G}\boldsymbol{B}\{\boldsymbol{M}\}$ with the weight matrix $\boldsymbol{G}$:

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} \geq \mathbf{Tr}\,\boldsymbol{G}\boldsymbol{J}^{-1}. \tag{6.5.50}$$

For the Gaussian family $\{S_{\overline{P},\overline{Q}}\}$ the symmetric logarithmic derivatives obtained through (5.6.64) are

$$L^Q = \sigma_Q^{-1}(Q - \overline{Q}), \qquad L^P = \sigma_P^{-2}(P - \overline{P}), \tag{6.5.51}$$

so that

$$\boldsymbol{J} = \begin{bmatrix} \sigma_P^{-2} & 0 \\ 0 & \sigma_Q^{-2} \end{bmatrix}. \tag{6.5.52}$$

The matrix inequality (6.5.48) then gives the two scalar inequalities

$$D_{(P)}\{\boldsymbol{M}\} \geq \sigma_P^2, \qquad D_{(Q)}\{\boldsymbol{M}\} \geq \sigma_Q^2,$$

where

$$D_{(P)}\{\boldsymbol{M}\} = \int (\alpha - \overline{P})^2 \mu_{\overline{P},\overline{Q}}(\mathrm{d}\alpha\, \mathrm{d}\beta),$$
$$D_{(Q)}\{\boldsymbol{M}\} = \int (\beta - \overline{Q})^2 \mu_{\overline{P},\overline{Q}}(\mathrm{d}\alpha, \mathrm{d}\beta)$$

are the marginal variances of the locally unbiased measurement $\boldsymbol{M} = \{M(\mathrm{d}\alpha \mathrm{d}\beta)\}$ with the probability distribution

$$\mu_{\overline{P},\overline{Q}}(\mathrm{d}\alpha\, \mathrm{d}\beta) = \operatorname{Tr} S_{\overline{P},\overline{Q}} M(\mathrm{d}\alpha\, \mathrm{d}\beta).$$

The inequality (6.5.50) gives

$$g_P D_{(P)}\{\boldsymbol{M}\} + g_Q D_{(Q)}\{\boldsymbol{M}\} \geq g_P \sigma_P^2 + g_Q \sigma_Q^2 \tag{6.5.53}$$

for arbitrary positive weights g_P, g_Q. We shall see in the next section that this bound is never achieved; a more informative bound will be obtained which takes into account the impossibility of an errorless joint measurement of P and Q.

6.6. A bound based on right logarithmic derivatives

As distinct from the classical mathematical statistics, in the noncommutative theory there are several analogs of the Cramér-Rao inequality using different definitions of logarithmic derivative. Let the family $\{S_\theta\}$ satisfy condition (1) of Section 6.5; instead of (2) we assume

(2′) *There is a constant c such that*

$$\left| \operatorname{Tr} \frac{\partial}{\partial \theta_j} S_\theta \cdot X \right|^2 \leq c \operatorname{Tr} S_\theta X X^*; \qquad X \in \mathfrak{B}(\mathscr{H}), \quad j = 1, \ldots, n.$$

This means that the complex linear functionals

$$\frac{\partial}{\partial \theta_j} E_\theta(X) = \operatorname{Tr} \frac{\partial}{\partial \theta_j} S_\theta \cdot X$$

on $\mathfrak{B}(\mathscr{H})$ can be extended to continuous linear functions on the complex Hilbert space $\mathscr{L}_+^2(S_\theta)$. By the Riesz-Frechet lemma there exist unique elements $\widetilde{L}_\theta^j \in \mathscr{L}_+^2(S_\theta)$ such that

$$\operatorname{Tr} \frac{\partial}{\partial \theta_j} S_\theta \cdot X = \langle \widetilde{L}_\theta^j, X \rangle_{S_\theta}^+, \qquad X \in \mathfrak{B}(\mathscr{H}),$$

or equivalently

$$\frac{\partial}{\partial\theta_j}S_{\boldsymbol{\theta}} = (\widetilde{L}^j_{\boldsymbol{\theta}})^* S_{\boldsymbol{\theta}} \quad \text{or} \quad \frac{\partial}{\partial\theta_j}S_{\boldsymbol{\theta}} = S_{\boldsymbol{\theta}}\widetilde{L}^j_{\boldsymbol{\theta}}. \tag{6.6.54}$$

The operators $\widetilde{L}^j \equiv \widetilde{L}^j_{\boldsymbol{\theta}}$ are called the *right logarithmic derivatives* of the family $\{S_{\boldsymbol{\theta}}\}$. They satisfy

$$\langle I, \widetilde{L}^j\rangle^+_S = 0, \qquad j = 1, \dots, n.$$

The *right information matrix* is defined by the relation

$$\tilde{\boldsymbol{J}}_{\boldsymbol{\theta}} = \tilde{\boldsymbol{J}} = [\langle \widetilde{L}^j, \widetilde{L}^k\rangle^+_S].$$

Let $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ be a locally unbiased measurement of the multidimensional parameter $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]$ which we assume to exist. In the same way as in (6.5.46) we get

$$\langle \widetilde{L}^j, X_k\rangle^+_S = \delta_{jk}.$$

Instead of (6.5.49) we get applying the inequality (2.9.109)

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq [\langle X_j, X_k\rangle^{\pm}_S]. \tag{6.6.55}$$

Then applying Lemma 6.5.1 to the systems of vectors $\{\widetilde{L}^j\}$, $\{X_j\}$ in the complex Hilbert space $\mathscr{L}^2_+(S)$ we get a different inequality for the measurement covariance matrix

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq \tilde{\boldsymbol{J}}^{-1}. \tag{6.6.56}$$

It is important here that unlike $\boldsymbol{J}$ in (6.5.48) the matrix $\tilde{\boldsymbol{J}}$ is complex Hermitean, so writing (6.6.56) in the scalar form

$$\boldsymbol{v}\boldsymbol{B}\{\boldsymbol{M}\}\boldsymbol{v}^* \geq \boldsymbol{v}\tilde{\boldsymbol{J}}^{-1}\boldsymbol{v}^*$$

one should consider all complex vectors $\boldsymbol{v}$, since confinement to real vectors would produce a less informative matrix inequality which does not take into account the imaginary part of $\tilde{\boldsymbol{J}}^{-1}$.

We are going to apply (6.6.56) to the Gaussian family $\{S_{\overline{P},\overline{Q}}\}$ but first we shall find an important general relation between the symmetric and the right logarithmic derivatives.

Note that condition (2′) implies condition (2) of Section 6.6.58 since $\langle X, X\rangle^{\pm}_S \leq 2\langle X, X\rangle_S$ as it was mentioned in Section 2.8. Therefore condition (2′) ensures the existence of both logarithmic derivatives. By the definitions of the logarithmic derivatives

$$\langle L^j, X\rangle_S = \langle \widetilde{L}^j, X\rangle^+_S \tag{6.6.57}$$

for all bounded X. Using the first relation in (2.8.97) and the definition of the commutation operator we have

$$\begin{aligned}\langle Y, X\rangle_S^+ &= \langle Y, X\rangle_S + \frac{1}{2}\mathrm{i}[Y, X]_S \\ &= \left\langle Y, \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right) X\right\rangle_S .\end{aligned} \tag{6.6.58}$$

It follows that

$$\left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\widetilde{L}^j = L^j. \tag{6.6.59}$$

If the operator $(I + \frac{1}{2}\mathrm{i}\mathfrak{D})$ is nondegenerated which is the case if the state $S = S_{\boldsymbol{\theta}}$ is exact, then by Proposition 2.10.1 we get

$$\widetilde{L}^j = \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)^{-1} L^j \tag{6.6.60}$$

and

$$\tilde{\boldsymbol{J}} = \left[\left\langle L^j, \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)^{-1} L^k\right\rangle_S\right]. \tag{6.6.61}$$

What is needed for (6.6.56) is the matrix $\tilde{\boldsymbol{J}}^{-1}$. It can be readily calculated in the particular case where *the subspace $\mathscr{L}$ of $\mathscr{L}^2(S)$ generated by the symmetric logarithmic derivatives $\{L^j\}$ is an invariant subspace of the commutation operator $\mathfrak{D}$*:

$$\mathfrak{D}(\mathscr{L}) \subset \mathscr{L}. \tag{6.6.62}$$

Then $I = \frac{1}{2}\mathrm{i}\mathfrak{D}$ is a Hermitean operator for which $\mathscr{L}$ is a finite-dimensional invariant subspace. Therefore we can consider $I + \frac{1}{2}\mathrm{i}\mathfrak{D}$ as an operator acting in $\mathscr{L}$ and $(I + \frac{1}{2}\mathrm{i}\mathfrak{D})^{-1}$ in (6.6.61) as its inverse in $\mathscr{L}$. Then $\tilde{\boldsymbol{J}}$ is the matrix of the quadratic form of the operator $(I + \frac{1}{2}\mathrm{i}\mathfrak{D})^{-1}$ in the basis $\{L^j\}$ and it is known from linear algebra that

$$\tilde{\boldsymbol{J}}^{-1} = \boldsymbol{J}^{-1}\left[\left\langle L^j, \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right) L^k\right\rangle_S\right]\boldsymbol{J}^{-1},$$

since $\boldsymbol{J} = [\langle L^j, L^k\rangle_S]$ is the Gram matrix of the basis $\{L^j\}$ in $\mathscr{L}$. Introducing the real skew-symmetric matrix

$$\boldsymbol{D} = [\langle L^j, \mathfrak{D}L^k\rangle_S] = [[L^j, L^k]_S],$$

we find the desired expression

$$\tilde{\boldsymbol{J}}^{-1} = \boldsymbol{J}^{-1}\left[\boldsymbol{J} + \frac{1}{2}\mathrm{i}\boldsymbol{D}\right]\boldsymbol{J}^{-1} = \boldsymbol{J}^{-1} + \frac{1}{2}\mathrm{i}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1}. \tag{6.6.63}$$

Consider now the example of the two-parameter Gaussian family $\{S_{\overline{P},\overline{Q}}\}$. Using the relations (6.6.59), (6.5.51) and (5.6.65) we find

$$\begin{aligned}\widetilde{L}^Q &= (4\sigma_P^2\sigma_Q^2 - 1)^{-1}[4\sigma_P^2(Q - \overline{Q}) - 2\mathrm{i}(P - \overline{P})],\\ \widetilde{L}^P &= (4\sigma_P^2\sigma_Q^2 - 1)^{-1}[4\sigma_Q^2(P - \overline{P}) - 2\mathrm{i}(Q - \overline{Q})],\end{aligned} \tag{6.6.64}$$

assuming that $4\sigma_P^2\sigma_P^2 \neq 1$. This excludes the case where $S_{\overline{P},\overline{Q}}$ are pure states. Then by (5.5.54) all eigenvalues of the density operator $S_{\overline{P},\overline{Q}}$ are positive, *i.e.*, $S_{\overline{P},\overline{Q}}$ is exact. Thus assuming $4\sigma_P^2\sigma_Q^2 \neq 1$ we just guarantee that the operator $I + \frac{1}{2}\mathrm{i}\mathfrak{D}$ is nondegenerated for all $\overline{P}$, $\overline{Q}$. From (6.6.64) the right information matrix $\tilde{\boldsymbol{J}}$ can be derived but there is a simpler way based on (6.6.63). Indeed the subspace $\mathscr{L}$ generated by the symmetric logarithmic derivatives $L^Q = \sigma_Q^{-2}(Q - \overline{Q})$ and $L^p = \sigma_P^{-2}(P - \overline{P})$ is just $\mathfrak{R}_1$ for the state $S_{\overline{P},\overline{Q}}$ as defined in Section 5.6. By Theorem 6.6.1 it is an invariant subspace of the commutation operator $\mathfrak{D}$ (this is also seen directly from (5.6.65)). Therefore (6.6.63) can be used. The matrix $\boldsymbol{J}$ is given by (6.5.52); since

$$[P, Q]_S = 1, \qquad [Q, Q]_S = [P, P]_S = 0$$

by (5.4.38) with $R(z) = Px + Qy$, $\Delta(z, z') = xy' - x'y$, then

$$\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1} = \begin{bmatrix} 0 & 1\\ -1 & 0\end{bmatrix}.$$

Thus according to (6.6.63)

$$\tilde{\boldsymbol{J}}^{-1} = \begin{bmatrix} \sigma_P^2 & \mathrm{i}/2\\ -\mathrm{i}/2 & \sigma_Q^2\end{bmatrix}.$$

Choosing for $\boldsymbol{v}$ in the scalar inequality following (6.6.56) the complex vector $[\sqrt{g_P}, \mathrm{i}\sqrt{g_Q}]$ we get the bound for the total mean-square deviation of a locally unbiased measurement

$$g_P D_{(P)}\{\boldsymbol{M}\} + g_Q D_{(Q)}\{\boldsymbol{M}\} \geq g_P\sigma_P^2 + g_Q\sigma_Q^2 + \sqrt{g_P g_Q}, \tag{6.6.65}$$

which is apparently better than (6.5.53). This bound is the same as the bound (3.6.58) for the covariant measurement. As we know from Section 3.6 it is achieved by the canonical measurement

$$M(\mathrm{d}\overline{P}, \mathrm{d}\overline{Q}) = |\overline{P}, \overline{Q}; \sigma^2)(\sigma^2; \overline{Q}, \overline{P}|\frac{\mathrm{d}\overline{P}\,\mathrm{d}\overline{Q}}{2\pi}, \tag{6.6.66}$$

with $\sigma^2 = \frac{1}{2}\sqrt{g_P/g_Q}$. According to Proposition 3.6.1 a realization of this measurement is accomplished by the pair of compatible observables

$$\widetilde{P} = P \otimes I_0 + I \otimes P_0, \qquad \widetilde{Q} = Q \otimes I_0 - I \otimes Q_0$$

in the space $\mathscr{H} \otimes \mathscr{H}_0$ where the auxiliary degree of freedom P_0, Q_0 is described by the zero-mean minimum-uncertainty state $|0, 0; \sigma^2)(\sigma^2; 0, 0|$. As we shall show in Section 6.8 this measurement is locally unbiased. Thus it is the uniformly best locally unbiased measurement of the parameters $\overline{P}$, $\overline{Q}$ in the Gaussian family $\{S_{\overline{P},\overline{Q}}\}$.

The bounds given by the inequalities (6.5.48) and (6.6.56) are in general incompatible. We have just seen that for the two-parameter Gaussian family $\{S_{\overline{P},\overline{Q}}\}$ the bound (6.6.56) is more informative than (6.5.48); we now show that *for an arbitrary one-parameter family* $\{S_\theta\}$ *the inequality* (6.5.48) *is always at least as informative as* (6.6.56)*, i.e.*,

$$\langle L, L\rangle_S \leq \langle \widetilde{L}, \widetilde{L}\rangle_S^+.$$

For this use matrix representation of the logarithmic derivatives in the basis $\{\psi_j\}$ of the eigenvectors of $S = S_\theta$. Multiplying (6.5.45) and (6.6.54) by $(\psi_j|$ and $|\psi_k)$ we get

$$(\psi_j|L\psi_k) = 2(s_j + s_k)^{-1}\left(\psi_j\left|\frac{\mathrm{d}S_\theta}{\mathrm{d}\theta}\psi_k\right.\right),$$
$$(\psi_j|\widetilde{L}\psi_k) = s_j^{-1}\left(\psi_j\left|\frac{\mathrm{d}S_\theta}{\mathrm{d}\theta}\psi_k\right.\right).$$

It follows that

$$\langle L, L\rangle_S = 2\sum_{j,k} p_{jk}(s_j + s_k)^{-1},$$
$$\langle \widetilde{L}, \widetilde{L}\rangle_S^+ = \sum_{j,k} p_{jk}s_j^{-1} = \sum_{j,k}\frac{p_{jk}}{2}(s_j^{-1} + s_k^{-1}),$$

where $p_{jk} = |\psi_j|(\mathrm{d}S_\theta/\mathrm{d}\theta)\psi_k)|^2 \geq 0$. Since $2(s_j+s_k)^{-1} \leq \frac{1}{2}(s_j^{-1}+s_k^{-1})$, the required inequality is established.

In the same way the left logarithmic derivative can be introduced. The corresponding left information matrix is easily found to be $\tilde{\boldsymbol{J}}^T$ where ^{T} means transposition of a matrix. Remark that

$$\tilde{\boldsymbol{J}}^{-1} = \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} + \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}, \qquad (\tilde{\boldsymbol{J}}^{-1})^{\mathrm{T}} = \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} - \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1},$$

where $\mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1}$ is real symmetric, $\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}$ is real skew-symmetric. Since $\boldsymbol{B}\{\boldsymbol{M}\}$ is real and symmetric, (6.6.56) implies

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} \pm \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}.$$

This shows in particular that the bounds obtained from the right or left logarithmic derivatives are essentially the same.

In the example of a Gaussian family we were able to obtain the bound (6.6.65) for the mean-square deviation by choosing artificially the vector $\boldsymbol{v}$. We now show how to obtain from (6.6.56) a general lower bound for $\Sigma\{\boldsymbol{M}\}$ with arbitrary weight matrix $\boldsymbol{G}$. According to the last inequality we can write

$$\begin{aligned}\Sigma\{\boldsymbol{M}\} \geq \min\{&\mathbf{Tr}\,\boldsymbol{G}\boldsymbol{B} : \boldsymbol{B} \text{ is real symmetric,}\\ &\boldsymbol{B} \geq \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} \pm \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}\}.\end{aligned} \tag{6.6.67}$$

To obtain an explicit form of the minimum we use the following device. If $\boldsymbol{M}$ is a matrix which is similar to a diagonal matrix, *i.e.*,

$$\boldsymbol{M} = \boldsymbol{T}\begin{bmatrix}\mu_1 & & 0\\ & \ddots & \\ 0 & & \mu_n\end{bmatrix}\boldsymbol{T}^{-1},$$

then one can develop a functional calculus of $\boldsymbol{M}$; in particular we put

$$\mathrm{abs}\,\boldsymbol{M} = \boldsymbol{T}\begin{bmatrix}|\mu_1| & & 0\\ & \ddots & \\ 0 & & |\mu_n|\end{bmatrix}\boldsymbol{T}^{-1}.$$

In general $\mathrm{abs}\,\boldsymbol{M} \neq |\boldsymbol{M}| \equiv \sqrt{\boldsymbol{M}^*\boldsymbol{M}}$; the equality holds if $\boldsymbol{M}$ is Hermitean.

The product of two Hermitean matrices $\boldsymbol{G}\boldsymbol{R}$, one of which, say $\boldsymbol{G}$, is positive definite and nondegenerate, is similar to a diagonal matrix. Indeed $\boldsymbol{G}\boldsymbol{R} = \sqrt{\boldsymbol{G}}(\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}})\sqrt{\boldsymbol{G}}^{-1}$ and the matrix $\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}$, being Hermitean, is similar to a diagonal matrix. From the definitions it follows that

$$\mathrm{abs}\,(\boldsymbol{G}\boldsymbol{R}) = \sqrt{G}|\sqrt{G}\boldsymbol{R}\sqrt{G}|\sqrt{G}^{-1}. \tag{6.6.68}$$

Lemma 6.6.1. *Let $\boldsymbol{R}$ be a complex Hermitean matrix; then*

$$\min\{\mathbf{Tr}\,\boldsymbol{G}\boldsymbol{X} : \boldsymbol{X} \geq \pm\boldsymbol{R}\} = \mathbf{Tr}\ \mathrm{abs}\,(\boldsymbol{G}\boldsymbol{R}),$$

and the minimum is achieved for $\boldsymbol{X} = \boldsymbol{G}^{-1}\mathrm{abs}\,(\boldsymbol{G}\boldsymbol{R})$.

Proof. Since $\boldsymbol{X} \geq \pm\boldsymbol{R}$, then

$$\sqrt{\boldsymbol{G}}\boldsymbol{X}\sqrt{\boldsymbol{G}} \geq \pm\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}} \qquad \text{and}$$
$$\boldsymbol{e}^*\sqrt{\boldsymbol{G}}\boldsymbol{X}\sqrt{\boldsymbol{G}}\boldsymbol{e} \geq |\boldsymbol{e}^*\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}\boldsymbol{e}|$$

for any column-vector $\boldsymbol{e}$. Let $\{\boldsymbol{e}_j\}$ be the basis of eigenvectors of the Hermitean matrix $\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}$ and $\{\mu_j\}$ be its eigenvalues. Then

$$\mathbf{Tr}\,\boldsymbol{G}\boldsymbol{X} = \mathbf{Tr}\,\sqrt{\boldsymbol{G}}\boldsymbol{X}\sqrt{\boldsymbol{G}} = \sum_j \boldsymbol{e}_j^*\sqrt{\boldsymbol{G}}\boldsymbol{X}\sqrt{\boldsymbol{G}}\boldsymbol{e}_j \geq \sum_j |\mu_j|$$
$$= \mathbf{Tr}\,|\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}| = \mathbf{Tr}\ \text{abs}\,(\boldsymbol{G}\boldsymbol{R}).$$

Substituting $\boldsymbol{X} = \boldsymbol{G}^{-1}\text{abs}\,(\boldsymbol{G}\boldsymbol{R})$ gives the lower bound and we need only to check that $\boldsymbol{G}^{-1}$ abs $(\boldsymbol{G}\boldsymbol{R}) \geq \pm\boldsymbol{R}$. Using (6.6.68) we can rewrite the left-hand side as $\sqrt{\boldsymbol{G}}^{-1}|\sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}|\sqrt{\boldsymbol{G}}^{-1}$. Then the required inequality reduces to the inequality $|\boldsymbol{Y}| \geq \pm\boldsymbol{Y}$ for the Hermitean matrix $\boldsymbol{Y} = \sqrt{\boldsymbol{G}}\boldsymbol{R}\sqrt{\boldsymbol{G}}$, which follows from the properties of the functional calculus of Hermitean matrices (*cf.* Section 2.3). □

Assume now that $\boldsymbol{G}$ is real and $\boldsymbol{R}$ is pure imaginary, *i.e.*, $\boldsymbol{R} = \mathrm{i}\boldsymbol{Q}$ where $\boldsymbol{Q}$ is real skew-symmetric. Then the minimizing matrix $\boldsymbol{X}$ is a real matrix. Indeed by (6.6.68) and the definition of $|\boldsymbol{Y}|$ we get $\boldsymbol{X} = \sqrt{\boldsymbol{G}^{-1}}\sqrt{\sqrt{\boldsymbol{G}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Q}\sqrt{\boldsymbol{G}}}\sqrt{\boldsymbol{G}^{-1}}$. This expression involves square roots of real positive matrices which are again real matrices. It follows that the result of Lemma 6.6.1 can be applied to calculation of the minimum in (6.6.67) giving

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} \geq \mathbf{Tr}\,\boldsymbol{G}\,\mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} + \mathbf{Tr}\ \text{abs}\,(\mathrm{i}\boldsymbol{G}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}). \tag{6.6.69}$$

If the condition (6.6.62) is satisfied, then from (6.6.63) we obtain

$$\mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} = \boldsymbol{J}^{-1}, \qquad \mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1} = \frac{1}{2}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1},$$

so that

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} \geq \mathbf{Tr}\,\boldsymbol{G}\boldsymbol{J}^{-1} + \frac{1}{2}\,\mathbf{Tr}\ \text{abs}\,(\mathrm{i}\boldsymbol{G}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1}). \tag{6.6.70}$$

We leave to the reader to show that application of (6.6.70) to the Gaussian family $\{S_{\overline{P},\overline{Q}}\}$ gives the bound (6.6.65).

6.7. A general bound for the total mean-square deviation

We have obtained the two essentially different bounds for the total mean-square deviation. Here we shall derive a general inequality which implies the bounds based on both definitions of logarithmic derivative.

We assume that the family $\{S_{\boldsymbol{\theta}}\}$ satisfies the conditions (1) and (2) of Section 6.5 at the fixed point $\boldsymbol{\theta}$. Let $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ be a measurement of the parameter $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$, locally unbiased at the point $\boldsymbol{\theta}$. Introducing as in Section 6.5

$$X_j = \int (\hat{\theta}_1 - \theta_j) M(\mathrm{d}^n\hat{\theta}) \equiv X^j_{\boldsymbol{M}} - \theta_j \tag{6.7.71}$$

we have the local unbiasedness condition expressed in the form

$$\langle L^j, X_k\rangle_S = \delta_{jk}; \qquad j,k = 1, \ldots, n. \tag{6.7.72}$$

where S as usual denotes $S_{\boldsymbol{\theta}}$.

The covariance matrix of a measurement with finite second moments satisfies the inequality (6.6.55), which using (2.8.97) can be written in the form

$$\boldsymbol{B}\{\boldsymbol{M}\} \geq [\langle X_j, X_k\rangle_S] \pm \frac{1}{2}\mathrm{i}[[X_j, X_k]_S]. \tag{6.7.73}$$

Putting

$$\kappa_{jk} = b_{jk}\{\boldsymbol{M}\} - \langle X_j, X_k\rangle_S \tag{6.7.74}$$

we can write (6.7.73) in the form

$$[\kappa_{jk}] \geq \pm\frac{1}{2}\mathrm{i}[[X_j, X_k]_S]. \tag{6.7.75}$$

The total mean-square deviation is then

$$\begin{aligned}\boldsymbol{\Sigma}\{\boldsymbol{M}\} &\equiv \sum_{j,k} g_{jk} b_{jk}\{\boldsymbol{M}\} \\ &= \sum_{j,k} g_{jk}[\kappa_{jk} + \langle X_j, X_k\rangle_S].\end{aligned} \tag{6.7.76}$$

Measurement $\boldsymbol{M}$ enters into this expression through the variables $X_j \in \mathscr{L}^2_h(S)$ and the real symmetric matrix $[\kappa_{jk}] = \boldsymbol{K}$. We shall obtain the lower bound for the total mean-square deviation by taking minimum of (6.7.76) with respect to all $\{X_j\}$ and $[\kappa_{jk}]$ satisfying the necessary restrictions (6.7.72) and (6.7.75). This bound may not be attainable; this will be the case only if the minimizing arguments $\{X^*_j\}$, $[\kappa^*_{jk}]$ can be obtained from some measurement $\boldsymbol{M}$ by the relations (6.7.71) and (6.7.74).

To find the minimum it is convenient to reparametrize the variables $\{X_j\}$ and $[\kappa_{jk}]$ according to the following lemma. In what follows we shall use the complex Hilbert space $\mathscr{L}^2(S) = \mathscr{L}_h^2(S) \oplus \mathrm{i}\mathscr{L}_h^2(S)$. By *complex extension* of a real linear operator $\mathfrak{A}$ in $\mathscr{L}_h^2(S)$ we mean the operator in $\mathscr{L}^2(S)$ defined by the relation $\mathfrak{A}(X_1 + \mathrm{i}X_2) = \mathfrak{A}X_1 + \mathrm{i}\mathfrak{A}X_2$. If $\mathfrak{A}$ is a bounded real symmetric operator in $\mathscr{L}_h^2(S)$, then its complex extension is Hermitean in $\mathscr{L}^2(S)$ etc.

Lemma 6.7.1. *The elements $X_j \in \mathscr{L}_h^2(S)$; $j = 1, \ldots, n$ and the real symmetric $(n \times n)$-matrix $[\kappa_{jk}]$ satisfy* (6.7.75) *if and only if there are $Y_j \in \mathscr{L}_h^2(S)$; $j = 1, \ldots, n$ and a bounded real symmetric operator $\mathfrak{F}$ in $\mathscr{L}_h^2(S)$ such that*

(1) $X_j = \mathfrak{F}Y_j$; $j = 1, \ldots, n$;
(2) $\kappa_{jk} = \langle Y_j, \mathfrak{F}(I - \mathfrak{F})Y_k\rangle_S$; $j, k = 1, \ldots, n$;
(3) *the complex extension of $\mathfrak{F}$ satisfies*

$$\mathfrak{F} \geq \mathfrak{F}\left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\mathfrak{F}.$$

Proof. Let $\{X_j\}$ and $[\kappa_{jk}]$ satisfy (6.7.75), and let $\mathscr{L}$ be the subspace of $\mathscr{L}_h^2(S)$ generated by X_j; $j = 1, \ldots, n$. Define the real symmetric operator $\mathfrak{K}$ in $\mathscr{L}$ by

$$\langle X_j, \mathfrak{K}X_k\rangle_S = \kappa_{jk}; \qquad j, k = 1, \ldots, n.$$

Then (6.7.75) can be written in the form

$$\langle Y, \mathfrak{K}Y\rangle_S \geq \pm\frac{1}{2}\mathrm{i}\langle Y, \mathfrak{D}Y\rangle, \qquad Y \in \mathscr{L} \oplus \mathrm{i}\mathscr{L}. \tag{6.7.77}$$

Define a bounded real symmetric operator $\mathfrak{F}$ in $\mathscr{L}^2(S)$ by

$$\mathfrak{F} = \begin{cases} (I + \mathfrak{K})^{-1}, & \text{on } \mathscr{L}, \\ 0, & \text{on } \mathscr{L}_n^2(S) \ominus \mathscr{L}. \end{cases}$$

Letting $Y_j = (I + \mathfrak{K})X_j$ we have $X_j = \mathfrak{F}Y_j$ and

$$\begin{aligned} \kappa_{jk} &= \langle X_j, \mathfrak{K}X_k\rangle_S = \langle Y_j, \mathfrak{F}\mathfrak{K}\mathfrak{F}Y_k\rangle_S \\ &= \langle Y, \mathfrak{F}(I - \mathfrak{F})Y_k\rangle_S, \end{aligned}$$

so that the conditions (1) and (2) of Lemma 6.7.1 are satisfied and it remains to check condition (3). Since the range of the operator $\mathfrak{F}$ constructed above is $\mathscr{L}$ we can put $Y = \mathfrak{F}X$, $X \in \mathscr{L}^2(S)$, in (6.7.77). Adding

to both sides of the resulting inequality the term $\langle \mathfrak{F}X, \mathfrak{F}X\rangle_S$ we get

$$\left\langle X, \mathfrak{F}\left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\mathfrak{F}X\right\rangle_S \leq \langle X, \mathscr{F}(I+\mathfrak{K})\mathfrak{F}X\rangle_S = \langle X, \mathfrak{F}X\rangle_S,$$

what was required.

Conversely let $\{Y_j\}$ and $\mathfrak{F}$ satisfy the conditions of the lemma. Defining $\{X_j\}$, $[\kappa_{jk}]$ by (1), (2) we can write the relation (6.7.75) in the form

$$[\langle Y_j, \mathfrak{F}(I-\mathfrak{F})Y_k\rangle_S] \geq \pm\frac{1}{2}\mathrm{i}[[\mathfrak{F}Y_j, \mathfrak{F}Y_k]_S] = \pm\frac{1}{2}\mathrm{i}[\langle Y_j, \mathfrak{F}\mathfrak{D}\mathfrak{F}Y_k\rangle_S],$$

which we need to prove. For this the following inequality will suffice

$$\mathscr{F}(I-\mathscr{F}) \geq \pm\frac{1}{2}\mathrm{i}\mathfrak{F}\mathfrak{D}\mathfrak{F},$$

which follows from (3), and the lemma is proved. □

Expressing (6.7.76) in terms of the new variables we get

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} = \sum_{j,k} g_{jk}\langle Y_j, \mathfrak{F}Y_k\rangle_S, \tag{6.7.78}$$

and the condition (6.7.72) takes the form

$$\langle L^j, \mathfrak{F}Y_k\rangle_S = \delta_{jk}. \tag{6.7.79}$$

The problem now is to minimize (6.7.78) with respect to all $\{Y_j \in \mathscr{L}_h^2(S);\ j=1,\ldots,n\}$ satisfying (6.7.79) and all operators $\mathfrak{F}$ in $\mathscr{L}_h^2(S)$ satisfying condition (3) of Lemma 6.7.1.

We shall proceed first keeping $\mathfrak{F}$ fixed and minimizing with respect to $\{Y_j\}$. For this we remark that $\mathfrak{F}$ is a positive operator, moreover

$$0 \leq \mathfrak{F} \leq I. \tag{6.7.80}$$

Indeed adding the two inequalities corresponding to the different signs in condition (3) we get $\mathfrak{F} \geq \mathfrak{F}^2$ or $\mathfrak{F}(I-\mathfrak{F}) \geq 0$ which is equivalent to (6.7.80). Therefore the bilinear symmetric form $\langle X, \mathfrak{F}Y\rangle_S$ is positive-definite and defines a pre-inner product on $\mathscr{L}_h^2(S)$. By Lemma 6.5.1 and the assumed existence of a locally unbiased measurement the matrix

$$\boldsymbol{F} = [\langle L_j, \mathfrak{F}L_k\rangle_S]$$

is nondegenerate and satisfies the following matrix inequality

$$[\langle Y_j, \mathfrak{F}Y_k\rangle_S] \geq [\langle L_j, \mathfrak{F}L_k\rangle_S]^{-1} = \boldsymbol{F}^{-1}.$$

The equality is achieved here if

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \boldsymbol{F}^{-1} \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix}.$$

Thus we come to the inequality

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} \geq \inf \mathbf{Tr}\, \boldsymbol{G}\boldsymbol{F}^{-1} \tag{6.7.81}$$

where the infimum is takes over all bounded real symmetric operators $\mathfrak{F}$ *in* $\mathscr{L}_h^2(S)$ *satisfying condition* (3) *of Lemma* 6.7.1. Denote by $\mathfrak{F}_*$ the solution of the minimization problem in the right-hand side of (6.7.81) (if it exists) and put $F_* = [\langle L_j, \mathfrak{F}_* L_k\rangle]$. The variables of the original minimization problem are then expressed through $\mathfrak{F}_*$ by the relations

$$\begin{bmatrix} X_1^* \\ \vdots \\ X_n^* \end{bmatrix} = \boldsymbol{F}_*^{-1} \begin{bmatrix} \mathfrak{F}_* L_1 \\ \vdots \\ \mathfrak{F}_* L_n \end{bmatrix}, \qquad \boldsymbol{K}_* = \boldsymbol{F}_*^{-1}[\langle L_j, \mathfrak{F}_*(I - \mathfrak{F}_*)L_k\rangle_S \boldsymbol{F}_*^{-1}. \tag{6.7.82}$$

An explicit solution for $\mathfrak{F}_*$ was found only in particular cases, one of which will be demonstrated at the end of this section. Now we are going to show that (6.7.81) is in general more informative than the two previously obtained inequalities.

Using the second inequality in (6.7.80) we obtain

$$\boldsymbol{F} = [\langle L_j, \mathfrak{F}L_k\rangle_S] \leq [\langle L_j, L_k\rangle_S] = \boldsymbol{J}.$$

Therefore $\boldsymbol{F}^{-1} \geq \boldsymbol{J}^{-1}$ and we get the inequality (6.5.51), corresponding to symmetric logarithmic derivative. To derive (6.6.69) we first note that condition (3) of Lemma 6.7.1 can be written in the form

$$0 \leq \left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}\right) \mathfrak{F} \left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}\right) \leq \left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}\right). \tag{6.7.83}$$

Indeed, multiplying the inequality of condition (3) by $\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}$ from the left and the right we get

$$\left(\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}\mathfrak{F}\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}\right)^2 \leq \sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}\mathfrak{F}\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}},$$

whence $0 \leq \sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}\,\mathfrak{F}\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}} \leq I$ as in (6.7.80). Framing this again with $\sqrt{I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}}$ we get (6.7.83). Assume now that the condition (2′) of Section 6.6 holds providing the existence of the right logarithmic derivatives. Then using (6.6.59)

$$\boldsymbol{F} = \left[\left\langle \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\widetilde{L}^j, \mathfrak{F}\left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\widetilde{L}^k\right\rangle_S\right].$$

By (6.7.83) and (6.6.58)

$$\boldsymbol{F} \leq \left[\left\langle \widetilde{L}^j, \left(I + \frac{1}{2}\mathrm{i}\mathfrak{D}\right)\widetilde{L}^k\right\rangle_S\right] = [\langle \widetilde{L}^j, \widetilde{L}^k\rangle_S^+] = \tilde{\boldsymbol{J}},$$

so that

$$\boldsymbol{F}^{-1} \geq \tilde{\boldsymbol{J}}^{-1} = \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} + \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}.$$

Since $\boldsymbol{F}^{-1}$ is real symmetric, this implies that $\boldsymbol{F}^{-1} \geq \mathrm{Re}\,\tilde{\boldsymbol{J}}^{-1} + \mathrm{i}\,\mathrm{Im}\,\tilde{\boldsymbol{J}}^{-1}$. Therefore the bound of the inequality (6.7.81) is greater or equal to the bound of the inequality (6.6.67), which is the same as (6.6.69).

We now show that *if the subspace $\mathscr{L}$ generated by the symmetric logarithmic derivatives L^j; $j = 1, \ldots, n$, is an invariant subspace of the commutation operator $\mathfrak{D}$ of the state S, then* (6.7.81) *gives the bound* (6.6.70). For this we need the proposition which will be used also in the sequel.

Proposition 6.7.2. *Let $\mathscr{M}$ be a closed invariant subspace of the operator $\mathfrak{D}$, containing the vectors L^j; $j = 1, \ldots, n$. Then the value of infimum in* (6.7.81) *remains unchanged if $\mathfrak{F}$ and $\mathfrak{D}$ are considered as operators acting in $\mathscr{M}$ instead of $\mathscr{L}_h^2(S)$.*

Proof. Denote by $\mathfrak{D}_{\mathscr{M}}$ the restriction of the operator $\mathfrak{D}$ onto $\mathscr{M}$. We need to show that to any bounded real symmetric operator $\mathfrak{F}$ satisfying condition (3) of Lemma 6.7.1 corresponds an operator $\mathfrak{F}_{\mathscr{M}}$ in $\mathscr{M}$ satisfying

$$\mathfrak{F}_{\mathscr{M}}\left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}_{\mathscr{M}}\right)\mathfrak{F}_{\mathscr{M}} \leq \mathfrak{F}_{\mathscr{M}}, \tag{6.7.84}$$

and such that

$$\langle L^j, \mathfrak{F}L^k\rangle_S = \langle L^j, \mathfrak{F}_{\mathscr{M}}L^k\rangle_S, \tag{6.7.85}$$

and conversely.

Let $\mathfrak{F}$ satisfy condition (3) which we take in the form (6.7.83). Let E be the projection onto $\mathscr{M}$, then by the assumption it commutes with

$\mathfrak{D}$. Multiplying (6.7.83) from the left and the right by E and denoting $\mathfrak{F}_{\mathscr{M}} = E\mathfrak{F}E$ we get

$$0 \le \left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}_{\mathscr{M}}\right)\mathfrak{F}_{\mathscr{M}}\left(I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}_{\mathscr{M}}\right) \le I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}_{\mathscr{M}}, \tag{6.7.86}$$

which is equivalent to (6.7.84). The relation (6.7.85) holds since $EL^j = L^j$; $j = 1, \dots, n$.

Conversely if $\mathfrak{F}_{\mathscr{M}}$ is a bounded real symmetric operator in $\mathscr{M}$ satisfying (6.7.84) or (6.7.86) then extending it by zero onto the orthogonal complement of $\mathscr{M}$ we get the operator $\mathfrak{F}$ with the required properties. □

Returning to the case where the subspace $\mathscr{L}$ generated by $\{L^j\}$ is an invariant subspace of $\mathfrak{D}$ we see that we can deal with operators $\mathfrak{F}$ as acting in $\mathscr{L}$. Since the Gram matrix of the basis L^j; $j = 1, \dots, n$ in $\mathscr{L}$ is $\boldsymbol{J}$ then the operators $\mathfrak{F}$ and $I \pm \frac{1}{2}\mathrm{i}\mathfrak{D}$ are represented in this basis by the matrices $\boldsymbol{J}^{-1}\boldsymbol{F}$, $\boldsymbol{J}^{-1}(\boldsymbol{J} \pm \frac{1}{2}\mathrm{i}\boldsymbol{D})$. Therefore the condition (6.7.84) in the matrix form is

$$\boldsymbol{F}\boldsymbol{J}^{-1}\left(\boldsymbol{J} \pm \frac{1}{2}\mathrm{i}\boldsymbol{D}\right)\boldsymbol{J}^{-1}\boldsymbol{F} \le \boldsymbol{F},$$

whence

$$\boldsymbol{F}^{-1} \ge \boldsymbol{J}^{-1} \pm \frac{1}{2}\mathrm{i}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1},$$

since $\boldsymbol{F}$ is nondegenerate. Thus (6.7.81) is equivalent to

$$\boldsymbol{\Sigma}\{\boldsymbol{M}\} \ge \min\left\{ \mathbf{Tr}\,\boldsymbol{G}\boldsymbol{F}^{-1} : \quad \boldsymbol{F}^{-1} \text{ is real symmetric} \right.$$
$$\left. \text{and } \boldsymbol{F}^{-1} \ge \boldsymbol{J}^{-1} \pm \frac{1}{2}\mathrm{i}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1} \right\}.$$

But this is just (6.6.67) with $\tilde{\boldsymbol{J}}$ given by (6.6.63). We thus come again to (6.6.70). According to Lemma 6.6.1 the optimal matrix $\boldsymbol{F}_*$ is given by

$$\boldsymbol{F}_*^{-1} = \boldsymbol{J}^{-1} + \frac{1}{2}\boldsymbol{G}^{-1}\,\mathrm{abs}(\mathrm{i}\boldsymbol{G}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1}).$$

Using (6.7.82) and taking into account the relation

$$\begin{bmatrix} \mathfrak{F}_* L_1 \\ \vdots \\ \mathfrak{F}_* L_n \end{bmatrix} = \boldsymbol{F}_*\boldsymbol{J}^{-1}\begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix},$$

which connects the operator $\mathfrak{F}_*$ with its matrix, we get

$$\begin{bmatrix} X_1^* \\ \vdots \\ X_n^* \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix}, \qquad \boldsymbol{K}_* = \frac{1}{2}\boldsymbol{G}^{-1} \operatorname{abs}(\mathrm{i}\boldsymbol{G}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1}). \tag{6.7.87}$$

If there exists a measurement $\boldsymbol{M}_*$ such that these $\{X_j^*\}$ and $\boldsymbol{K}_* = [\kappa_{jk}]$ are obtained from $\boldsymbol{M}_*$ by the formulas (6.7.71) and (6.7.74), then this is the best locally unbiased measurement at the point θ. As we shall see later such situation takes place in the Gaussian case.

6.8. Linear measurements

In the examples considered in Sections 6.5, 6.6 we have found the best locally unbiased measurements of the mean-value parameters $\overline{P}$ and $\overline{Q}$ in the Gaussian states $\{S_{\overline{P},\overline{Q}}\}$ for one degree of freedom. In the case of one-dimensional parameter the inequality (6.2.13) based on the symmetric logarithmic derivative was used, in the two-parameter case we have to use the inequality (6.6.56) based on the right logarithmic derivative. The common feature for both cases is that the observables describing the optimal measurement are linear functions of the canonical observables P and Q.

Now we are going to generalize these results to arbitrary Gaussian states for the CCR with a finite number of degrees of freedom. We shall show that the best locally unbiased measurement of linear mean-value parameters in a family of Gaussian states can be found in the class of linear measurements to be defined below. Roughly speaking a linear measurement corresponds to a vector observable linearly expressed through the canonical observables but with the account for possible incompatibility of the components. With the last essential supplement the theorem to be proved in Section 6.9 can be regarded as the noncommutative analog of the classical result of mathematical statistics which says that the best unbiased estimates of linear mean-value parameters of Gaussian probability distributions are linear functions of observations.

Let (Z, Δ) be a symplectic space and $z \to V(z)$ be an irreducible representation of the CCR in $\mathscr{H}$. A measurement $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ of a parameter $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]$ is called *linear* if for any state S with finite second moments in $\mathscr{H}$:

(1) $\boldsymbol{M}$ has finite second moments with respect to S so that the elements

of the covariance matrix

$$b_{jk}\{\boldsymbol{M}\} = \int (\theta_j - \overline{\theta}_j)(\theta_k - \overline{\theta}_k)\mu_S(\mathrm{d}^n\theta),$$
$$\left(\overline{\theta}_j = \int \theta_j \mu_S(\mathrm{d}^n\theta)\right)$$

are well defined.

(2) The elements $X^j_{\boldsymbol{M}} = \int \theta_j M(\mathrm{d}^n\theta)$ of $\mathscr{L}^2_h(S)$ are in the subspace of canonical observables, *i.e.*,

$$X^j_{\boldsymbol{M}} = R(z_j); \qquad j = 1, \ldots, n,$$

for some $z_j \in Z$.

(3) The numbers $\kappa_{kj} = b_{jk}\{\boldsymbol{M}\} - \alpha(z_j, z_k)$, where α is the correlation function of S, do not depend on the choice of S.

The variables $\{R(z_j)\}$ and $[\kappa_{jk}]$ are called the *parameters* of the linear measurement. They are restricted by the inequality

$$[\kappa_{jk}] \geq \pm\frac{1}{2}\mathrm{i}[\Delta(z_j, z_k)], \tag{6.8.88}$$

which follows from (6.7.75) by putting $X_j = R(z_j) - \overline{\theta}_j$.

Proposition 6.8.1. *Let z_j; $j = 1, \ldots, n$, be arbitrary elements of Z, $[\kappa_{jk}]$ be a real symmetric $(n \times n)$-matrix satisfying* (6.8.88). *Then a linear measurement with the parameters $\{R(z_j)\}$, $[\kappa_{jk}]$ exists.*

We shall prove this proposition under the simplifying assumption that $\{z_j\}$ form a basis in Z.

Proof. In addition to the representation $z \to V(z)$ in $\mathscr{H}$ consider an irreducible representation $z \to V_0(z)$ in the space $\mathscr{H}_0$ of the different CCR corresponding to the symplectic space $(Z, -\Delta)$. This means that $\{V_0(z)\}$ satisfy

$$V_0(z)V_0(z') = \mathrm{e}^{-\mathrm{i}\Delta(z,z')/2}V_0(z + z'). \tag{6.8.89}$$

For any states S, S_0 with finite second moments

$$[R_0(z), R_0(z')]_{S_0} = -[R(z), R(z')]_S$$

according to (5.4.38).

Consider the family of the operators $\widetilde{V}(z) = V(z) \otimes V_0(z)$, $z \in Z$, in the Hilbert tensor product $\mathscr{H} \otimes \mathscr{H}_0$. By (5.2.9) and (6.8.89), $z \to \widetilde{V}(z)$

is a unitary representation of the additive group of Z, *i.e.*$\widetilde{V}(z)\widetilde{V}(z') = \widetilde{V}(z+z')$; $z, z' \in Z$. Applying Stone's theorem to the one-parameter unitary groups $\{V(tz_j); t \in \mathbb{R}\}$ we get

$$\widetilde{V}(tz_j) = \mathrm{e}^{\mathrm{i}t\widetilde{R}(z_j)}, \tag{6.8.90}$$

where

$$\widetilde{R}(z_j) = R(z_j) \otimes I + I \otimes R_0(z_j); \qquad j = 1, \ldots, n. \tag{6.8.91}$$

Since the operators (6.8.90) commute for different j, then by Section 2.6 the operators (6.8.91) represent compatible observables so that

$$\widetilde{R}(z_j) = \int \theta_j E(\mathrm{d}^n\theta); \qquad j = 1, \ldots, n,$$

where $\boldsymbol{E} = \{E(\mathrm{d}^n\theta)\}$ is the orthogonal resolution of identity representing the joint measurement of $\widetilde{R}(z_j)$; $j = 1, \ldots, n$.

Let S_0 be a state with finite second moments in $\mathscr{H}_0$ satisfying

$$\begin{aligned} &E_{S_0}(R_0(z_j)) = 0; \\ &\langle R_0(z_j), R_0(z_k)\rangle_{S_0} = \kappa_{jk}; \qquad j, k = 1, \ldots, n. \end{aligned} \tag{6.8.92}$$

Under the simplifying assumption such a state always exists. Indeed if $\{z_j\}$ is basis in Z, then we can uniquely define a bilinear symmetric form $\kappa(\cdot,\cdot)$ on Z by requiring $\kappa(z_j, z_k) = k_{kj}$. The condition (6.8.88) ensures that the form $\kappa(\cdot,\cdot)$ satisfies the condition (5.4.42) and by Theorem 5.5.1 $\kappa(\cdot,\cdot)$ is the correlation function of a Gaussian state. Taking for S_0 the Gaussian state with zero mean and the correlation function $\kappa(\cdot,\cdot)$ we get (6.8.92).

Consider the triple $(\mathscr{H}_0, S_0, \boldsymbol{E})$. Using Proposition 2.5.2 there is a measurement $\boldsymbol{M} = \{M(\mathrm{d}^n\theta)\}$ in $\mathscr{H}$ such that

$$\mu_S(B) = \mu^{\boldsymbol{M}}_{S\otimes S_0}(B); \qquad B \in \mathscr{A}(\mathbb{R}^n), \tag{6.8.93}$$

where $\mu_S(\mathrm{d}^n\theta) = \operatorname{Tr} SM(\mathrm{d}^n\theta)$ is the probability distribution of $\boldsymbol{M}$ with respect to the state S. In other words $\boldsymbol{M}$ is the measurement having a realization $(\mathscr{H}_0, S_0, \boldsymbol{E})$. We now show that $\boldsymbol{M}$ is a linear measurement with the parameters $\{R(z_j)\}$ and $[\kappa_{jk}]$.

We need the simple relation

$$\begin{aligned} &\langle X_1 \otimes Y_1, X_2 \otimes Y_2\rangle_{S\otimes S_0} = \langle X_1, X_2\rangle_S \langle Y_1, Y_2\rangle_{S_0} \\ &X_j \in \mathscr{L}^2(S), \qquad Y_j \in \mathscr{L}^2(S_0) \end{aligned} \tag{6.8.94}$$

which follows from the corresponding relation for bounded X_j, Y_j by taking $\mathscr{L}^2$ limits. Since by (2.9.100) $E_{S\otimes S_0}(\widetilde{R}(z_j)) = \langle \widetilde{R}(z_j), I\rangle_{S\otimes S_0}$, then using (6.8.91) and (6.8.94) we get

$$E_{S\otimes S_0}(\widetilde{R}(z_j)) = E_S(R(z_j)) + E_{S_0}(R_0(z_j)) = \overline{\theta}_j$$

since $E_{S_0}(R_0(z_j)) = 0$ by (6.8.92). In the same way we get for the measurement covariance matrix

$$b_{jk}\{\boldsymbol{M}\} = \langle R(z_j) - \overline{\theta}_j, R(z_k) - \overline{\theta}_k\rangle_S + \langle R_0(z_j), R_0(z_k)\rangle_{S_0}.$$

The first term is just the correlation function $\alpha(z_j, z_k)$, thus we see that the difference $b_{jk}\{\boldsymbol{M}\} - \alpha(z_j, z_k)$ according to the second relation in (6.8.92) is κ_{jk} as required.

It remains to show that $X^j_{\boldsymbol{M}} = R(z_j)$. Since the relation (6.8.93) is linear in S, then

$$\operatorname{Tr} T M(B) = \operatorname{Tr}(T \otimes S_0)E(B)$$

for arbitrary trace-class operator T. Putting $T = YS$ where Y is a bounded operator and S is a density operator, we get

$$\langle M(B), Y\rangle_S = \langle E(B), Y \otimes I\rangle_{S\otimes S_0}.$$

It follows that for any simple function $f(\cdot)$

$$\left\langle \int f(\theta)M(\mathrm{d}\theta), Y\right\rangle_S = \left\langle \int f(\theta)E(\mathrm{d}\theta), Y\otimes I\right\rangle_{S\otimes S_0}.$$

Taking the $\mathscr{L}^2$ limits and using (6.8.94) we get

$$\begin{aligned}\langle X^j_{\boldsymbol{M}}, Y\rangle_S &= \left\langle \int \theta_j E(\mathrm{d}^n\theta), Y\otimes I\right\rangle_{S\otimes S_0} = \langle \widetilde{R}(z_j), Y\otimes I\rangle_{S\otimes S_0}\\ &= \langle R(z_j), Y\rangle_S + \langle R_0(z_j), I\rangle_{S_0} = \langle R(z_j), Y\rangle_S,\end{aligned}$$

since S_0 has zero mean. Since this holds for all bounded Y, then $X^j_{\boldsymbol{M}} = R(z_j)$ in $\mathscr{L}^2_h(S)$, and the proposition is proved. □

Returning to the case of one degree of freedom we see that the canonical measurement (6.6.66) is a linear measurement with the parameters

$$\begin{aligned}&X^P_{\boldsymbol{M}} = P, \qquad X^Q_{\boldsymbol{M}} = Q;\\ &\kappa_{QQ} \equiv D_{S_0}(Q_0) = \frac{1}{2}\sqrt{g_P/g_Q}, \qquad \kappa_{PP} \equiv D_{S_0}(P_0) = \frac{1}{2}\sqrt{g_Q/g_P},\\ &\kappa_{PQ} = \kappa_{QP} = 0,\end{aligned}$$

as follows from the realization of this measurement and the relations (6.8.92). The local unbiasedness conditions

$$\langle L^P, X_{\boldsymbol{M}}^P\rangle = \langle L^Q, X_{\boldsymbol{M}}^Q\rangle = 1;$$
$$\langle L^P, X_{\boldsymbol{M}}^Q\rangle = \langle L^Q, X_{\boldsymbol{M}}^P\rangle = 0$$

easily follow by taking into account the expression (6.5.51) for the symmetric logarithmic derivative (the inner product $\langle\cdot,\cdot\rangle$ corresponds here to the state $S_{\overline{P,Q}}$).

Consider now the simple measurement (6.2.16) of the single observable Q. Clearly $X_E^Q = \int \theta E(\mathrm{d}\theta) = Q$ and by (2.9.101) $\kappa_{QQ} = 0$. Therefore $\boldsymbol{E}$ is a linear measurement with parameters

$$X_{\boldsymbol{E}}^Q = Q, \qquad \kappa_{QQ} = 0.$$

In this case Z_Q does not form a basis in Z and the construction of Proposition 6.8.1 does not go without modifications. We can see the reason for this by observing that the measurement of Q can be considered as a vaguely defined limit of the joint measurement of $\widetilde{Q}$ and $\widetilde{P}$ in the realization of (6.6.66) with $\kappa_{QQ} \to 0$, $\kappa_{PP} \to \infty$. This corresponds to $g_P/g_Q \to 0$, *i.e.*, we ultimately neglect the contribution due to measurement of P into the total measure of accuracy. The limit state with $D_{S_0}(Q_0) = 0$ does not exist, and this makes necessary a modification in the proof of Proposition 6.8.1.

A kinematical implementation of the linear measurement (6.6.66) was given in Section 3.6 where Q, P were the position and momentum observables of a particle in one dimension. In the quantum communication theory the canonical observables P, Q arise from the representation of the quantum radiation field as a superposition of harmonic components

$$E(t) \sim q\cos\omega t + p\omega^{-1}\sin\omega t,$$

corresponding to various frequencies ω. We shall give an explanation of the measurement (6.6.66) in this context. Consider the planar monochromatic wave $\boldsymbol{E}(t)$ propagating along the axis z as in Figure 6.2. Since $\boldsymbol{E}(t)$ is orthogonal to the direction z we can decompose it into orthogonal axes x, y to obtain

$$E_x(t) \sim q_x\cos\omega t + p_x\omega^{-1}\sin\omega t,$$
$$E_y(t) \sim q_y\cos\omega t + p_y\omega^{-1}\sin\omega t.$$

Let $E(t) = E_x(t)$ and assume that $E_y(t)$ is described by the ground oscillator state S_0; physically this means that the wave is polarized in the

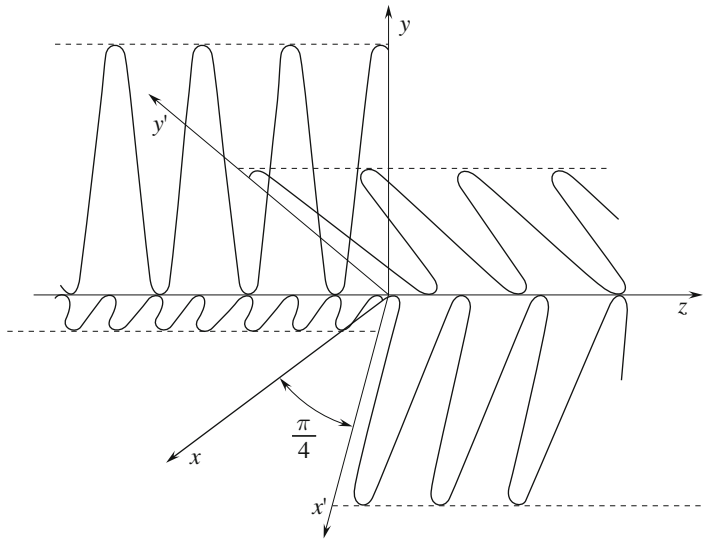

Figure 6.2.

direction x and the y-degree of freedom presents only the inevitable "vacuum" fluctuations. Introducing the new axes x', y' as shown in Figure 6.2, we have

$$E_{x'}(t) \sim (q_x + q_y)\cos\omega t + (p_x + p_y)\omega^{-1}\sin\omega t,$$
$$E_{y'}(t) \sim (q_x - q_y)\cos\omega t + (p_x - p_y)\omega^{-1}\sin\omega t.$$

The components $E_{x'}(t)$ and $E_{y'}(t)$ corresponding to the mutually orthogonal axes can be separated by a double-refracting filter; the result is achieved by the joint measurement of the amplitudes

$$\tilde{p} = p_x + p_y, \qquad \tilde{q} = q_x - q_y.$$

Turning to the general linear measurement (6.8.91) we remark that performing a linear transformation one can reduce the set of observables (6.8.91) to a set of pairs $\widetilde{P}$, $\widetilde{Q}$ in the realization of (6.6.66). Therefore any linear measurement over an optical field can be implemented by using a finite number of certain optical devices (see Comments).

6.9. Mean-value estimation for Gaussian states

Consider the family of Gaussian states $\{S_{\boldsymbol{\theta}}; \boldsymbol{\theta} = [\theta_1, \ldots, \theta_n] \in \mathbb{R}^n\}$ with the fixed correlation function $\alpha(\cdot,\cdot)$ and the mean value of the form

$$m(z) = \sum_{j=1}^{n} \theta_j m_j(z), \tag{6.9.95}$$

where $m_j(\cdot)$ are known linear functionals on the symplectic space (Z, Δ) and θ_j are unknown real parameters subject to statistical estimation through the observation of the quantum system. For example $S_{\boldsymbol{\theta}}$ can be the state of radiation field representing the mixture of the background radiation and the signal, in which the amplitudes θ_j of the components $m_j(\cdot)$ need to be estimated. We assume that $m_j(\cdot)$ are linearly independent.

Theorem 6.9.1. *The uniformly best locally unbiased measurement of the mean-value parameter $\boldsymbol{\theta}$ of Gaussian states $\{S_{\boldsymbol{\theta}}\}$ exists and is in the class of linear measurements.*

Proof. According to Proposition 5.6.3 the family $\{S_{\boldsymbol{\theta}}\}$ satisfies conditions (1) and (2) of Section 6.5 and the symmetric logarithmic derivatives are given by

$$L^j = R(m_j) - m(m_j); \qquad j = 1, \dots, n, \tag{6.9.96}$$

where $m_j \in Z$ are determined by

$$m_j(z) = \alpha(m_j, z), \qquad z \in Z.$$

We now fix the point $\boldsymbol{\theta}$ and consider the bound (6.7.81). Due to the characteristic property of Gaussian states expressed in Theorem 5.6.2, the subspace $\mathfrak{R}_1 \subset \mathscr{L}^2(S_{\boldsymbol{\theta}})$ consisting of observables $R(z) - m(z)$, $z \in Z$, is an invariant subspace of the commutation operator $\mathfrak{D}$ of $S_{\boldsymbol{\theta}}$. Since $\mathfrak{R}_1$ apparently contains the vectors (6.9.96), Proposition 6.7.2 applies, and we can consider $\mathfrak{F}$ in (6.7.81) as an operator in $\mathfrak{R}_1$. Denote by $\mathscr{F}$ the operator in Z corresponding to $\mathfrak{F}$ under the isometric map (5.6.56) so that

$$\mathfrak{F}(R(z) - m(z)) = R(\mathscr{F}z) - m(\mathscr{F}z).$$

Then (6.7.81) can be written as

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} \geq \inf \mathbf{Tr}\, \boldsymbol{G}\boldsymbol{F}^{-1} \equiv \boldsymbol{\Sigma}_*, \tag{6.9.97}$$

where now

$$\boldsymbol{F} = [\alpha(m_j, \mathscr{F}m_k)], \tag{6.9.98}$$

and the infimum in (6.9.97) is taken over $\boldsymbol{F}$ corresponding by (6.9.98) to real symmetric operators $\mathscr{F}$ in Z, such that their complex extensions satisfy

$$0 \leq \left(I + \frac{1}{2}\mathrm{i}\mathscr{D}\right)\mathscr{F}\left(I + \frac{1}{2}\mathrm{i}\mathscr{D}\right) \leq \left(I + \frac{1}{2}\mathrm{i}\mathscr{D}\right) \tag{6.9.99}$$

in the complexification of the Euclidean space (Z, α). Here $\mathscr{D}$ is the complex extension of the operator defined by (5.2.15), where α is the correlation function of $S_{\boldsymbol{\theta}}$.

Let us show that the infimum in (6.9.97) is achieved. The set of the operators $\mathscr{F}$ in the finite-dimensional space Z satisfying (6.9.99) is apparently closed. It is bounded since by (6.7.80) $0 \leq \mathscr{F} \leq I$. Thus this set is compact. The function $\mathscr{F} \to \mathbf{Tr}\, \boldsymbol{G}\boldsymbol{F}^{-1}$ is the composition of the continuous map $\mathscr{F} \to \boldsymbol{F}$ given by (6.9.98) and the function $\boldsymbol{F} \to \mathbf{Tr}\, \boldsymbol{G}\boldsymbol{F}^{-1}$. The last function is the supremum of the family of continuous functions $\boldsymbol{F} \to \mathbf{Tr}\, \boldsymbol{G}(\boldsymbol{F} + \epsilon \boldsymbol{I})^{-1}$, $\epsilon > 0$, and thus is lower semicontinuous. Therefore the function $\mathscr{F} \to \mathbf{Tr}\, \boldsymbol{G}\boldsymbol{F}^{-1}$ is lower semicontinuous and by a generalization of the Weierstrass theorem it attains the minimum on the compact set.

Let $\mathscr{F}_*$ be the operator in Z which furnishes the minimum in (6.9.97) and $\boldsymbol{F}_*$ be the corresponding matrix. Then according to (6.7.82) and (6.9.96) the optimal arguments X_j^* are given by

$$X_j^* = R(z_j^*) - m(z_j^*); \qquad j = 1, \ldots, n \tag{6.9.100}$$

where

$$\begin{bmatrix} z_1^* \\ \vdots \\ z_n^* \end{bmatrix} = \boldsymbol{F}_*^{-1} \begin{bmatrix} \mathscr{F}_* m_1 \\ \vdots \\ \mathscr{F}_* m_n \end{bmatrix}.$$

By definition the system $\{z_j^*\}$ is biorthogonal to the vectors $\{m_j\}$ representing the mean-value components in (Z, α):

$$\alpha(m_k, z_j^*) \equiv m_k(z_j^*) = \delta_{jk}. \tag{6.9.101}$$

The elements $\{X_j^*\}$ and the matrix

$$\boldsymbol{K}_* \equiv [\kappa_{jk}^*] = \boldsymbol{F}_*^{-1}[\alpha(m_j, \mathscr{F}_*(I - \mathscr{F}_*)m_k)]\boldsymbol{F}_*^{-1}$$

satisfy the inequality (6.7.75). Taking into account that by (6.9.100) and (5.4.38)

$$\begin{aligned}[X_j^*, X_k^*]_S &= [R(z_j^*) - m(z_j^*), R(z_k^*) - m(z_k^*)]_S \\ &= \Delta(z_j^*, z_k^*),\end{aligned}$$

we can rewrite (6.7.75) in the form

$$[\kappa_{jk}^*] \geq +\frac{1}{2}\mathrm{i}[\Delta(z_j^*, z_k^*)].$$

By Proposition 6.8.1 there is a linear measurement $\boldsymbol{M}_* = \{M_*(\mathrm{d}^n\theta)\}$ with the parameters $\{R(z_j^*)\}$ and $[\kappa_{jk}^*]$. It is locally unbiased since by (6.9.101)

$$\begin{aligned}\langle L^j, X^k_{\boldsymbol{M}_*}\rangle_S &= \langle R(m_j) - m(m_j), R(z_k^*)\rangle_S \\ &= \alpha(m_j, z_k^*) = \delta_{jk},\end{aligned}$$

and its total mean-square deviation attains the minimal value $\boldsymbol{\Sigma}_*$ allowed by (6.9.97). Therefore $\boldsymbol{M}_*$ is the best locally unbiased measurement at the point $\boldsymbol{\theta}$. Since the parameters of the linear measurement $\boldsymbol{M}_*$ do not depend on $\boldsymbol{\theta}$, it is the uniformly best, and the theorem is proved. □

An explicit solution for $\mathscr{F}_*$ can be obtained in particular cases. Consider the subspace $Z_{\mathscr{L}} \subset Z$ spanned by the vectors m_j; $j = 1, \dots, n$ (it corresponds to the subspace $\mathscr{L} \subset \mathscr{L}_h^2(S)$ under the isometric map $z \leftrightarrow R(z) - m(z)$), and assume that it is an invariant subspace of the operator $\mathscr{D}$. In particular this trivially holds if $n = \dim Z$ so that $Z_{\mathscr{L}} = Z$. Then $\mathscr{L}$ is an invariant subspace of the commutation operator $\mathfrak{D}$. Therefore we can apply the solution (6.7.87) to obtain the linear measurement with the parameters

$$\begin{bmatrix} R(z_1^*) \\ \vdots \\ R(z_n^*) \end{bmatrix} = \boldsymbol{J}^{-1} \begin{bmatrix} R(m_1) \\ \vdots \\ R(m_n) \end{bmatrix}; \qquad \boldsymbol{K}_* = \frac{1}{2}\boldsymbol{G}^{-1}\,\mathrm{abs}(\mathrm{i}\boldsymbol{G}\boldsymbol{J}^{-1}\boldsymbol{D}\boldsymbol{J}^{-1}),$$

where now

$$\begin{aligned}\boldsymbol{J} &\equiv [\langle L^j, L^k\rangle_S] = [\alpha(m_j, m_k)], \\ \boldsymbol{D} &\equiv [[L^j, L^k]_S] = [\Delta(m_j, m_k)].\end{aligned}$$

Recalling the proof of Proposition 6.8.1 we can loosely describe the realization of the best measurement $\boldsymbol{M}_*$ as the joint measurement of the compatible observables

$$\widetilde{R}_j = R(z_j^*) \otimes I_0 + I \otimes R_j^0,$$

where R_j^0 are the auxiliary canonical observables in the space $\mathscr{H}_0$ with the state S_0 satisfying

$$E_{S_0}(R_j^0) = 0, \qquad [\langle R_j^0, R_k^0\rangle_{S_0}] = \boldsymbol{K}_*.$$

If we put formally $\Delta \equiv 0$ we obtain the solution of the corresponding classical problem which is just the measurement of the now compatible observables $R(z_j^*)$; $j = 1, \dots, n$. This solution is independent of the

weight matrix $\boldsymbol{G}$. In the general noncommutative case $Z_{\mathscr{L}}$ need not be an invariant subspace of the optimal operator $\mathscr{F}_*$, and both $\{z_j^*\}$ and $\boldsymbol{K}_*$ may depend on the weight matrix $\boldsymbol{G}$ in the definition of the mean-square deviation.

Theorem 6.9.1 also allows us to establish a useful property of Gaussian states which characterizes them as "the least informative" or "the least favorable" from the point of view of an experimenter. Let $\{\widetilde{S}_{\boldsymbol{\theta}}\}$ be a different family of states having the same first and second moments as the Gaussian family $\{S_{\boldsymbol{\theta}}\}$ and arbitrary in other respects. Denote by $\alpha(\cdot,\cdot)$ their common correlation function. Let $\boldsymbol{M} = \{M(d^n\theta)\}$ be a linear locally unbiased measurement with the parameters $\{R(z_j)\}$ and $[\kappa_{jk}]$. Then its total mean-square deviation with respect to the state $\widetilde{S}_{\boldsymbol{\theta}}$ is equal to

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \sum_{j,k=1}^{n} g_{jk}[\kappa_{jk} + \alpha(z_j, z_k)],$$

by (6.7.76) and by the conditions (1) and (2) in the definition of linear measurement. Since κ_{jk} do not depend on the choice of state by condition (3), and $\alpha(\cdot,\cdot)$ is the same for $\widetilde{S}_\theta$ and S_θ, then $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ refers to the Gaussian state S_θ. Since the minimum of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$ is achieved on a linear measurement and is equal to $\boldsymbol{\Sigma}_*$ from (6.9.97), then the same is true for $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\{\boldsymbol{M}\}$. It follows that

$$\min_{\boldsymbol{M}} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} \le \boldsymbol{\Sigma}_*,$$

where the minimum is taken over all locally unbiased measurements and

$$\max_{\{S_\theta\}} \min_{\boldsymbol{M}} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\{\boldsymbol{M}\} = \boldsymbol{\Sigma}_*,$$

with the maximum attained for the Gaussian family $\{S_\theta\}$.

Thus under the given prior information about the first and second moments the Gaussian states give the greatest mean-square deviation in measurements of the mean-value parameters. Choosing the best linear measurement corresponds to the experimental strategy which minimizes the total mean-square deviation proceeding from the prudent assumption that the choice of Nature may happen to be the least favorable as regards to the measurement accuracy.

6.10. Comments

Section 6.1. A brief account of the mathematical Shannon-Kolmogorov theory of classical communication channels can be found in Kolmogorov's report [79]. The need for consideration of quantum limitations in

optical channels was recognized by Gabor [39]. Mathematical models for quantum communication channels were considered by Holevo [56], Ingarden [69] and Davies [28]. For information-theoretic considerations see Holevo [60] and Lindblad [85]. Concerning the general notion of quantum communication channel and its capacities see [179], also [178], Section 3.2.

Section 6.2, 6.3. For the classical Cramér-Rao inequality see Cramer [23]. A thorough geometrical study of the Cramér-Rao inequality can be found in Chentzov [22]. Proposition 6.2.1 is the rigorous version of the result of Helstrom [52, 53]. The results of Section 6.3 appeared partially in Holevo [66].

In the quantum case estimation problems with multidimensional parameter differ from those with one-dimensional parameter rather in radical way. This happens due to the noncommutativity of the algebra of quantum observables reflecting existence of incompatible observables which in principle cannot be measured in one experiment. This sets the new restrictions for the components of a multidimensional parameter that are absent in the classical case and leads to inherent non-uniqueness of the logarithmic derivatives and the corresponding Rao-Cramér inequality.

As it is well known, in the classical mathematical statistics the Fisher information generates Riemannian metric on the variety of probability distributions which is essentially unique monotone invariant in the category of statistical (Markov) morphisms [158], [22]. This metric has natural quantum analogs, but the uniqueness no longer holds. The expression

$$d(S_1, S_2) = \sqrt{2(1 - \|\sqrt{S_1}\sqrt{S_2}\|_1)}$$

defines metric on the convex set $\mathfrak{S}(\mathcal{H})$ of the density operators. In the more general context of von Neumann algebras such a metric called the *Bures metric* was studied in detail by Araki, Uhlmann and others. If $\{S_\theta\}$ is a family satisfying the conditions 1) and 2) on p. 274, then as $\Delta\theta \to \theta$, one has

$$d(S_\theta, S_{\theta+\Delta\theta})^2 \approx \frac{1}{4} \sum_{i,j=1}^{k} \langle L_\theta^i, L_\theta^j \rangle_\theta \Delta\theta_i \Delta\theta_j. \qquad (**)$$

Thus the Bures metric is locally equivalent to Riemannian metric determined by the quantum analog of the Fisher information matrix. Morozova and Chentsov [158] described metrics in $\mathfrak{S}(\mathcal{H})$, $\dim \mathcal{H} < \infty$, that are monotone invariant in the category of Markov morphisms. In this class the Riemannian metric defined in (**) is the minimal one while the metric related to the right or left logarithmic derivative is maximal. A

different example is provided by Bogoljubov-Kubo-Mori metric which plays an important role in quantum statistical mechanics. Investigations in this direction were continued in [185, 190], and an exhaustive description of monotone invariant Riemannian metrics on the set of density operators was obtained. The significance of these results for quantum estimation theory needs further elucidation. In this connection we mention a result from [173] showing that the Bogoljubov-Kubo-Mori metric appears in the large deviations theory in the bound for superefficient estimates. The noncommutative analog of the Sanov theorem giving an expression for the asymptotic of the error of the optimal discrimination between two quantum states in terms of the relative entropy was obtained in [175,189].

From a general point of view, the studies in the classical geometrostatistics (the term introduced by Kolmogorov) related to the names of Chentsov, Amari, Barndorff-Nielsen and others [151, 158] lead quite naturally to the investigations of differential geometric structure in the more complex state spaces such as quantum state (density operators) space [152, 196].

Section 6.4. The problem is motivated by the work of Braginskij and Vorontzov [19] who discussed the related detection problem and obtained (6.4.41) as a bound for the detection threshold.

Section 6.5, 6.6. The inequality (6.5.48) is the rigorous version of Helstrom's result [51, 53]. The inequality (6.6.56) is that of Yuen and Lax [147]. The optimal property of the canonical measurement was established by Holevo [61] and Yuen and Lax [147].

The inequality of Lemma 6.6.1, due to Belavkin and Grishanin, is adduced in the work of Stratonovich [130].

Section 6.7-6.9. The material is taken from Holevo [59,62–65]. Lemma 6.7.1 appeared in the paper [62] devoted to the Gaussian Bayes problem. The linear measurements were introduced under the name of canonical measurements in [56].

An arbitrary quantum optical system performing a Bogoljubov-type transformation is equivalent to scheme composed of linear devices such as multiport interferometers and few basic nonlinear "squeezing" devices, such as parametric frequency converter [156]. One should add that an arbitrary Gaussian state can be represented as a response of such system to vacuum or thermal equilibrium states, while any canonical measurement as a set of homodyne and heterodyne measurements at the output of such a scheme. In quantum optics an approximate joint measurement of the quadratures q, p is realized by optical heterodyning. Then the field of the measured mode is combined with intense reference field

of a frequency slightly different from the mode frequency. Measuring the intensities of the two fields at the output of the beam splitter of nearly unit transmissivity, in the limit of infinite local oscillator intensity turns out to be equivalent to the joint measurement of $\tilde{q}$, $\tilde{p}$ [200].

In particular, by computing the quantity $\tilde{\phi} = \arctan(\tilde{p}/\tilde{q})$ one obtains an estimate for the phase. This is statistically equivalent to a covariant phase measurement in the sense of Section 6.8 which is of course non-optimal since it carries redundant information about the amplitude. (For more detail on the heterodyne phase measurement see Section 3.5.2 [157].) In the work [199] adaptive homodyne measurement of phase was considered which uses a feedback and electro-optical modulator for the control of the phase of the local oscillator, under which the phase quadrature $R(\phi)$ is measured preferentially. For comparison of the different phase measurements in [199] the criterion $V = \min_\psi \Delta_\psi$ was used, where Δ_ψ is the measure of uncertainty (4.7.2), and the minimum is taken over all pure states with the mean number of quanta less than or equal to N. For $N \gg 1$ the canonical (*i.e.* optimal) distribution of the phase one has $V \approx \pi^2/N^2$, while the heterodyne measurement gives only $V \approx 1/4N$. A special choice of the feedback allows to achieve $V \approx \ln N/4N^2$ which approaches the quality of the canonical measurement, however the problem of its practical realization remains open.

Explicit solutions for $\mathscr{F}_*$ in various particular cases are given in [63, 64]. For applications to concrete spatially-temporal signal models see the book of Helstrom [53].

The Gaussian state is known to have the greatest quantum entropy $-\operatorname{Tr} S \ln S$ among all states with the fixed first and second moments (see, *e.g.*, Louisell [88]). This gives an alternative corroboration of the property of Gaussian states of being "the least informative".

Supplement[1]
Statistical structure of quantum theory and hidden variables

1. Introduction

By the end of XVIII century scientists developed the picture of the material world as a huge mechanism, the evolution of which is subject to a rigid dynamical laws and in principle can be predicted with arbitrary detail and accuracy. This system of conceptions which acquired the name "determinism" was progressive for that time and was based on a triumphant success of the Newtonian mechanics which allowed to give a rational explanation to a number of earlier inexplicable physical facts.

However, the development of natural sciences was leading to increase of the role of statistical concepts. The study of random phenomena started in XVII century with the simplest models of uncertainty emerging in hazard games. The notion of probability introduced in connection with such models was very successfully exploited in the second part of XIX century by creators of statistical mechanics. The laws of heat (thermodynamics) acquired explanation through statistical behavior of models of matter as a huge ensemble of identical interacting particles – molecules. Nevertheless by the beginning of XX century determinism continued to dominate at the foundations of natural sciences. The Nature was considered as fundamentally deterministic, and observed stochasticity was regarded as a secondary phenomenon reflecting incompleteness of our knowledge of the real state of the Nature.

Creation of basics of statistical physics marked a triumph of the ancient idea of atomism. But deeper study of the elementary components of the matter has led to a paradoxical conclusion – these components cannot be considered as particles in the proper, classical sense of the word. Depending on the conditions of observation, they can display

[1] This Supplement is self-contained and can be read independently of the main content of the book.

either corpuscular or wave properties. An electron interacting with vapor in Wilson chamber leaves a trace which is interpreted as a trajectory of a material particle. On the other hand, a beam of electrons passing through a crystal is diffracted making specific interference pattern similar to one produced by the light passing through a system of sufficiently small holes.

Attempts of theoretical explanation of such an unusual, "dualistic" behavior of microobjects has led in the beginning of XX century to the creation of quantum mechanics – the most revolutionary scientific theory of the recent age. Historically Heisenberg's "matrix mechanics" and Schrödinger's "wave mechanics" arose as result of ingenious insights in search of mathematical objects capable to combine discreteness and continuity in microprocesses. Soon after Born suggested statistical interpretation which organically fit the mathematical apparatus of quantum mechanics but excluded its deterministic explanation. A deep physico-philosophical analysis of the content of quantum mechanics developed by Bohr and Heisenberg has led to a conclusion that it represents a radically new type of theoretical model of the reality reflecting in its very structure the wholeness of the system in question and a complementarity between different aspects of its description. According to Bohr, the notion of complementarity is used to characterize the relation between the data obtained under different experimental conditions and may be interpreted only on the basis of mutually exclusive concepts. One might say that the quantum theory gives a "holographic" image of realities of the microworld. However nothing prevents from combining different projections of a three-dimensional object into one picture (remember the portraits of Picasso presenting the same person enface and in profile). But "elementarity" of the microobjects excludes a possibility of combining complementary measuring procedures which presuppose their own specific organization in space and time. For example, it makes no sense to speak about trajectory of electron diffracting on a crystal since any attempt to trace the trajectory changes the conditions of the experiment such that the interference becomes impossible.

From this point of view statistical nature of quantum mechanics becomes closely related to complementarity. The quantities that are measured in complementary conditions "cannot simultaneously be ascribed definite values. In this way, the statistical character of the formalism is displayed as a natural generalization of the description of classical physics" [155]. Thus the statistical nature of microprocesses acquires in quantum mechanics a primary importance. Not only "God does not play dice", but there is a fundamental source of uncertainty in the Nature which cannot be imitated by any classical mechanism of randomness. Of

course, the outcomes of any particular experiment are just a realization of a random variable, but the whole totality of experiments relevant to a given quantum system cannot be given a classical description. The classical method of description reducing in fact to listing the properties of the object turns out to be applicable to the objects of the surrounding macroscopical world to the extent that quantum uncertainties are negligible at the scale of this world.

In the dramatic discussion developed in 1930-s the main opponent of Bohr and other creators of the "orthodox" interpretation of quantum mechanics was Einstein whose views were shared by de Broglie, Schrödinger and some other scientists taking active part in establishing the basic principles of quantum theory. Einstein promoted the idea of incompleteness of quantum mechanics according to which its statistical nature is due to fluctuations of some "hidden variables" yet to be taken into account, and is to be replaced by a deterministic description in a future complete theory.

From that point of view an electron has an individual trajectory whether it is observed or not. The trajectory appears random since we do not know deeper principles governing the electron's motion. Having found these hidden principles, we recover determinism. Such a viewpoint might be appealing from the viewpoint of naive realism. However up to now all efforts to find alternative "deeper" interpretation of quantum mechanics turned out futile; moreover each such an effort led ultimately to strengthening of positions of the statistical interpretation shared by a majority of working physicists.

The hidden variable issue raised in this discussion can be thus formulated as the question: is it possible in principle to reduce the mathematical model of quantum mechanics to this or another form of classical probabilistic description? One should admit that the very analytical apparatus for describing uncertainty in quantum mechanics is so different from the language of probability theory that an idea of a mathematical proof of impossibility of introduction of hidden variables, which would stop once and forever all the controversy, appears naturally.

States and observables are described in quantum mechanics by matrices (operators) $\hat{S}, \hat{X}$; in probability theory the (statistical) states are described by probability distributions $S(d\omega)$ and the observable quantities – by functions $X(\omega)$ on the phase space $\Omega = \{\omega\}$ of the classical system. Thus the question is about possibility or impossibility of establishing a correspondence $S \to \hat{S}$, $X \to \hat{X}$ between classical and quantum states and observables, which would reproduce statistical prediction of quantum mechanics and, of course, satisfy certain important, physically motivated restrictions.

The first attempt of an impossibility proof was made in the famous von Neumann's book "Mathematical Foundations of Quantum Mechanics" which appeared in 1932. For some this was considered as a decisive argument against the hidden variables. However later it was understood that this argument does not solve the problem because it is based on a formal assumption lacking a physical motivation. During the last fifty years the issue was considerably clarified and even was made available for an experimental test. These investigations concerning basic understanding of the nature of physical reality are a substantial argument in favor of the viewpoint according to which introduction of physically meaningful hidden variables in quantum theory is not possible in principle.

In this essay we attempt to give a simple and self-consistent presentation of the hidden variable issue in a historical perspective.

In the first part we shall analyze general properties of description of any statistical experiment and outline the mathematical structures that arise from this description. These are convexity in the state space which is due to possibility of mixing ensembles, and partial order of observables reflecting the degree of informativity in measurement outcomes. Preservation of these structures is a minimal necessary condition for any hidden variables theory, due to their very universality. This general consideration of a statistical experiment will also allow us to distill the mathematical essence of the fundamental notion of complementarity.

The second part starts with the discussion of most important impossibility results for hidden variables. A number of such attempts, starting from von Neumann's theorem, descended from the belief rooted in the orthodox interpretation that it is the complementarity which is the main obstacle for a classical description in quantum mechanics. An important conclusion of our discussion is that complementarity excludes classical description only under additional assumption of one-to-one correspondence between quantum and classical entities. The "technical" condition of injectivity of the correspondence $S \to \hat{S}, X \to \hat{X}$ present in some modifications of von Neumann's theorem was recognized as the special property of "noncontextuality" of classical description. Thus a physically acceptable proof of impossibility can not be based on complementarity alone and requires appeal to other properties of quantum mechanical description. Such a property turns out to be the *quantum nonseparability* which is discussed in the last section. The argument based on the celebrated Bell's inequality shows impossibility of a classical description for a composite quantum system respecting separation into subsystems and hence, also the Einstein locality principle.

2. The structure of statistical theories

2.1. Axiomatic approaches in quantum mechanics

The story of von Neumann's theorem shows well that the hidden variables issue does not reduce to a merely mathematical problem. As far as the matter concerns interrelation between quantum mechanics and probability theory, the decisive role should be played by the choice of basic properties characterizing these structures. Therefore before proceeding to a mathematical consideration of the hidden variable issue we must carefully analyze and select basic postulates for both theories.

Here it is pertinent to remind that the problem of "mathematical formulation of the axioms of physics" was raised by Hilbert in his famous speech at the II-nd Mathematical Congress in 1900. Namely the question was about "axiomatic construction of those physical disciplines, in which mathematics already plays an outstanding role: these are in the first place probability theory and mechanics." Notably, in one place with the logical foundation of probability theory Hilbert put "development of the method of average values in mathematical physics, in particular, kinetic theory of gases," in this way pointing towards the most profound problems of mathematical physics, the investigation of which later led to the mathematical methods in statistical mechanics and theory of dynamical systems.

As it is well known, search for a mathematical basis for probability theory was completed with the publication in 1933 of the Kolmogorov's axiomatic system, giving a set of formally simple and intuitively appealing statements underlying the whole mathematical structure of the theory. Quantum mechanics could not be mentioned by Hilbert for the simple reason that the very physical notion of quantum did not yet exist – the famous Planck's report was made four months later in the same year 1900. The cornerstone for mathematization of quantum theory became the already mentioned monograph of von Neumann, who started investigations in this field in 1926-1927 while being Hilbert's collaborator [176].

Von Neumann's works initiated investigations in axiomatics of quantum theory. Considerable progress and clarification was achieved towards 1960-1970 along the following three mainstreams.

Algebraic approach [33, 127] takes as a basis the "algebra of observables" of the physical system. This approach turned out to be most productive from a mathematical viewpoint: together with group representations it served as a source for the modern highly refined structural theory of operator algebras. Physical applications of this approach concern mainly systems with infinite degrees of freedom – quantum fields and matter.

The quantum logic approach [71, 135, 153, 192] starts with the "lattice of propositions", *i.e.* observables taking only two values (0 and 1). Efforts in this direction culminated with construction of a certain axiom system characterizing the lattice of orthoprojections in a Hilbert space, *i.e.* "propositions" related to a quantum mechanical system. Introduction of certain algebraic structure (essentially the structure of Jordan algebra) in both approaches requires ultimately certain assumptions whose physical motivation is not so clear. In 1950-s American mathematician Mackey made an attempt to present the notions of quantum mechanics starting from some primary properties of statistical description of a physical system [92]. Although unfinished, this attempt had great influence on subsequent investigations.

In 1970-s the third approach to foundations of quantum mechanics appeared in which the notion of state plays a primary or equal role as compared to observables or measurements [27, 64, 171, 186]. The main component of the corresponding mathematical scheme is the convex set of states of the physical system. This approach, sometimes denoted as "convex" or "operational," could be also called "statistical" because it represents a far reaching development of statistical interpretation of quantum mechanics. In particular, it will be shown below that the "statistical ideology" provides also a natural basis for a discussion of hidden variables issue.

Notwithstanding apparent achievements, there is still no final form for a quantum axiomatics. Therefore the main attention in the first part of this essay will be paid to those basic assumptions which in this or the other form underlie any reasonable statistical theory. However first we must discuss the classical picture of statistical experiment which will serve a starting point for further discussion.

2.2. The classical picture of statistical experiment

In any experiment one can conventionally distinguish the two main stages. During the first stage of *preparation* a definite experimental arrangement is set to establish the "input data" of system or object under the experiment. At the subsequent stage of *measurement* the thus prepared system or object interacts with this or another measuring device, resulting in certain "output data" – the *outcome* of the measurement.

One of the most basic features of a scientific experiment is its *reproducibility i.e.* possibility of potentially indefinite independent repetitions of a given measurement in given conditions. Consider a sequence of such repetitions of certain experiment. As a rule, the outcomes of individual experiments will not be strictly the same and will be subject to

some random fluctuations the magnitude of which depend of the nature of the experiment and the system under investigation. Thus, although the outcomes of an experiments are conditioned by both preparation and measurement, this dependence usually is not deterministic but has statistical character. For classical systems described in terms of phase space, this can be expressed in the language of probability theory.

Denote by ω the complete set of variables characterizing the classical system or object. The space of all possible concrete values of ω makes the *phase space* of the object.

Preparation of any physical state of the object is made by certain device which, due to its imperfection or peculiarities of its construction may not be able to ensure exact reproduction of the values of all parameters for different individual representatives of this same object. Moreover, the object may be characterized by such an immense number of variables that there is no way to arrange complete control of all of them. It is assumed, however, that the variations of the values of ω being prepared are characterized by certain stability described by a probability distribution P. This probability distribution ascribing to an elementary phase space volume $d\omega$ its probability $P(d\omega)$ is called the *state* of the object.

Thus, this definition of state is essentially statistical one reflecting the possibility of fluctuations of the object's parameters. Its real content is that if one considers an ensemble, *i.e.* very large (potentially unbounded) collection of independent representatives of a given object, then the fraction (frequency) of those representatives for which the value of ω lies in a subset $B \subset \Omega$ is close to its theoretical value $P(B)$.

Consider two ensembles corresponding to the states P_1 and P_2 consisting of N representatives each and let us form the new ensemble by taking pN representatives of the first ensemble and $(1-p)N$ representatives of the second one, where $0 \leq p \leq 1$. In accordance with the frequency interpretation the new ensemble will be described by the state $pP_1 + (1-p)P_2$, which is called the *mixture* of the states P_1 and P_2 in the proportion $p : (1-p)$. Mixtures $\sum_j p_j P_j$ of arbitrary finite collections of states are introduced similarly. Moreover, one can consider continuous mixtures $\int p(d\alpha) P_\alpha$ where $p(d\alpha)$ is a probability distribution. Such mixtures may describe states prepared by a device with fluctuating parameter α. If the states are represented as elements (points) of a certain set, then arbitrary mixtures of P_1 and P_2 fill out the segment connecting the points P_1 and P_2. Such a set, which contains the segment connecting arbitrary two points of the set, is called *convex*. Thus the set of classical states which we shall denote $\mathfrak{S}(\Omega)$ is convex. A point in a convex set is called *extreme* if it is not an internal point of a segment belonging entirely to $\mathfrak{S}(\Omega)$. Extreme points describe *pure* states which can not

be represented as a mixture of other states. In the classical picture the pure states are just degenerate probability distributions concentrated at the points ω of the phase space.

To make this clear consider the simplest case where Ω consists of finite number of points: $\Omega = \{\omega_1, \ldots, \omega_N\}$. In this case state P is given by the finite probability distribution $[p_1, \ldots, p_N]$, where $p_j \geq 0, \sum_j p_j = 1$. Pure states are the degenerate distributions $[1, 0, \ldots, 0], \ldots, [0, \ldots, 0, 1]$.

An important theorem (proved by Caratheodory in the finite-dimensional case and by Krein and Milman in infinite dimensions) says that in a compact convex set an arbitrary point can be represented as a mixture of extreme points. If such a representation is unique, then this special convex set is called *simplex*. In particular, this is the case in the example above. In the case of arbitrary phase space Ω one should admit continuous mixtures of states. Thus in the classical picture of statistical experiments the state space $\mathfrak{S}(\Omega)$ makes simplex in which every state is a unique mixture of pure states, corresponding to exact values of all system's parameters.

The second, conclusive, stage of the experiment consists of measurement of a certain quantity X. Classically, in the ideal case the measurement does not introduce any changes in the system, *i.e.* amounts to an observation. In such a case *observable* X is described by a function which assigns to every possible $\omega \in \Omega$ its objective value $X(\omega)$. Having thus observed X, one can compute outcomes of observations of quantities $f(X)$, where f is a function, without direct observation of these quantities.

Let for simplicity observable X take finite number of real values $\{x_j\}$. Then

$$X(\omega) = \sum_i x_i E_i(\omega), \tag{2.1}$$

where $E_i(\omega)$ is the indicator function of the subset $\Omega_i \subset \Omega$ on which $X(\omega)$ takes the value x_i, *i.e.* the function equal to 1 on Ω_i and 0 outside Ω_i. The subsets Ω_i are disjoint and form a decomposition of the space Ω, while the family of their indicator functions $E = \{E_i(\omega)\}$ makes an *orthogonal resolution of identity*:

$$\sum_i E_i(\omega) = 1; \quad E_i(\omega)E_j(\omega) = 0 \quad \text{for} \quad i \neq j; \quad E_i(\omega)^2 = E_i(\omega)$$

for all $\omega \in \Omega$. Now consider observable $f(X(\omega))$, where f is a real function. Clearly,

$$f(X(\omega)) = \sum_i f(x_i) E_i(\omega). \tag{2.2}$$

Even if all x_i were different, some of the values $f(x_i)$ may coincide. Therefore, to enable treatment of (2.1) and (2.2) on equal footing, it is convenient to admit from the very beginning that some of x_i may coincide. Then *observation* (measurement without errors) will be described by the resolution of identity E, and to any observable X there correspond a variety of methods of observation, differing by detail of decomposition of the phase space Ω.

From the point of view of statistics, the resolution of the identity $E = \{E_i(\omega)\}$ bears all essential information about the measurement: probability of the $i-$th outcome in the state P is equal to

$$\mu_P^E(i) = P(\Omega_i) = \int_\Omega P(d\omega) E_i(\omega). \tag{2.3}$$

Operationally, E decomposes initial statistical ensemble into subensembles characterized by the properties $\omega \in \Omega_i$.

Hence the *mean value* of the observable (2.1) in the state P (the expectation) is equal to $M_P\{X\} = \int_\Omega P(d\omega) X(\omega)$.

The measurements described by the orthogonal resolutions of the identity are deterministic in the sense that they sharply classify representatives of an ensemble. A nondeterministic (unsharp, fuzzy) measurement provides probabilities $M_i(\omega)$ for i-th outcomes for ensemble representative characterized by the phase space point ω, so that

$$\sum_i M_i(\omega) = 1, \quad M_i(\omega) \geq 0.$$

The collection of operators $M = \{M_i(\omega)\}$ is a *resolution of the identity* in Ω, which is in general, nonorthogonal in the sense that $M_i(\omega) M_j(\omega) \not\equiv 0$ for $i \neq j$. Moreover, $M_i(\omega)^2 \leq M_i(\omega)$. Probability of the i-th outcome in the state P for such a measurement is equal to

$$\mu_P^M(i) = \int_\Omega P(d\omega) M_i(\omega). \tag{2.4}$$

This formula shows how the uncertainty of the outcome in the classical picture arises from two sources: from the uncertainty in the state preparation P and in the measurement procedure M.

The resolution of the identity M gives just the probabilities of the outcomes for an unsharp measurement, but knowing these probabilities, one can model statistical realization of the measurement involving generator of random numbers. Assume that there is such a device allowing to obtain values of random variable λ uniformly distributed in $[0, 1]$ (say, a properly calibrated roulette). Let us describe deterministic measurement

$E = \{E_i(\omega, \lambda)\}$ over the system which consists of the given object and a generator of random numbers, which is statistically equivalent to the measurement $M = \{M_i(\omega)\}$ in the sense that for any state P probabilities of all outcomes for the measurements M and E are the same. For this decompose the phase space of the composite system $\Omega \times \Lambda$ into the subsets $\Omega'_i = \left\{(\omega, \lambda) : \sum_{k=1}^{i-1} M_k(\omega) < \lambda \le \sum_{k=1}^{i} M_k(\omega)\right\}$ and denote by $E_i(\omega, \lambda)$ the indicator function of the subset Ω'_i.

By the construction we have

$$\int d\lambda E_i(\omega, \lambda) = M_i(\omega).$$

Indeed, for a given ω the integral is simply the integral with respect to $d\lambda$ from $\sum_{k=1}^{i-1} M_k(\omega)$ to $\sum_{k=1}^{i} M_k(\omega)$, *i.e.* $M_k(\omega)$. Integrating this identity with respect to $P(d\omega)$ we get

$$\begin{aligned} \mu_P^M(i) &= \int_\Omega P(d\omega) M_i(\omega) \\ &= \int_\Omega \int_\Lambda P(d\omega) d\lambda E_i(\omega, \lambda) = \mu_{P\times d\lambda}^M(i), \end{aligned} \tag{2.5}$$

which means the statistical equivalence of the measurements M and E.

The procedure of obtaining the outcome by using a generator of random numbers, introduced into statistics by Wald, is called randomization [22] and the corresponding measurements can be also called randomized.

From the viewpoint of the statistics, the results of experiment consisting of preparation of the state P and subsequent measurement M are completely described by the probability distribution of measurement outcomes $\mu_P^M = \left\{\mu_P^M(i)\right\}$. Notice that the correspondence $P \to \mu_P^M$ has the characteristic property of *affinity*: if the state P is mixture of states P_1 and P_2 in the proportion $p : (1 - p)$ then the probability distribution μ_P^M is mixture of the probability distributions $\mu_{P_1}^M$ and $\mu_{P_2}^M$ in the same proportion, $\mu_{pP_1+(1-p)P_2}^M = p\mu_{P_1}^M + (1 - p)\mu_{P_2}^M$.

2.3. The main features of statistical description

Having in mind the passage to quantum mechanics, let us try to formulate axiomatically the main features of statistical description of an experiment without using assumption of classicality, *i.e.* without introducing a phase space.

Axiom 2.1. Let there be given a set $\mathfrak{S}$ whose elements are called states and a set $\mathfrak{M}$ whose elements are called measurements. With a measurement $M \in \mathfrak{M}$ is associated the space U of its possible outcomes. For any

pair $S \in \mathfrak{S}$, $M \in \mathfrak{M}$ there is a probability distribution μ_S^M on U called the probability distribution of the measurement M in the state S.

Intuitively S represents a more or less detailed description of the preparation of a "statistical ensemble" and M – a measurement in this ensemble. An attempt to interpret these notions by introducing a phase space leads in fact to "hidden variables", but we shall discuss this later. Here S and M are considered as primary notions. For any (measurable) subset $B \subset U$ the quantity $\mu_S^M(B)$ is interpreted as theoretical value for the fraction of representatives of the ensemble prepared in the state S for which the outcome of the measurement M lies in the subset B.

The first axiom thus formalizes the requirement of reproducibility of the individual experiments and the stability of frequencies. The following axiom says that mixing of ensembles is an admissible way of state preparation.

Axiom 2.2. For arbitrary states S_1, S_2 and an arbitrary number $p, 0 < p < 1$, there exists the state S called mixture of the states S_1 and S_2 in the proportion $p : (1 - p)$ such that $\mu_S^M = p\mu_{S_1}^M + (1 - p)\mu_{S_2}^M$ for all measurements $M \in \mathfrak{M}$.

An outcome of a measurement can be the data obtained from one or several measuring devices as well as any other method of representing the information – for example, a picture on a computer display. Quite often the information obtained as a result of a measurement should be processed in this or another way. The result of such a processing can be regarded as an outcome of a complex measuring procedure including the given transformation. If M_1 is a measurement with the values in U_1 and M_2 – a measurement with the values in U_2, such that there is a (measurable) function $f : U_2 \to U_1$ satisfying the relation

$$\mu_S^{M_1}(B) = \mu_S^{M_2}(f^{-1}(B)); \quad B \subset U_1,$$

for all $S \in \mathfrak{S}$ then this means that the outcomes of the measurement M_1 are obtained from the outcomes of measurement M_2 by a functional transformation f. (Recall that $f^{-1}(B)$ denotes the inverse image of B *i.e.* the set of all such $u_2 \in U_2$, that $f(u_2) \in B$). In this case we say that M_1 is *subordinated* to the measurement M_2. If U_1 and U_2 are finite sets, this means that

$$\mu_S^{M_1}(u_1) = \sum_{u_2 : f(u_2) = u_1} \mu_S^{M_2}(u_2),$$

i.e. subordination means coarse-graining of the measurement outcomes.

Axiom 2.3. For any measurement M, the set $\mathfrak{M}$ contains all the measurements subordinated to M.

A pair of sets $(\mathfrak{S}, \mathfrak{M})$ satisfying the axioms 2.1 - 2.3 will be called *statistical model*. The statistical model is said to be *separated* if

Axiom 2.4. From $\mu_{S_1}^M = \mu_{S_2}^M$ for all $M \in \mathfrak{M}$ it follows that $S_1 = S_2$ and from $\mu_S^{M_1} = \mu_S^{M_2}$ for all $S \in \mathfrak{S}$ it follows that $M_1 = M_2$.

For a separated model both the operation of mixing in $\mathfrak{S}$ and the coarse-graining in $\mathfrak{M}$ are uniquely defined. Then the set of states $\mathfrak{S}$ obtains a convex structure while the set of measurements $\mathfrak{M}$ – a structure of partial order.

To illustrate the general notion of statistical model, let us come back to the classical picture where statistical states of an object are described by probability distributions on a phase space Ω so that $\mathfrak{S} = \mathfrak{S}(\Omega)$. If we consider measurements without errors, which correspond to random variables (see the relation (2.1)), then $\mathfrak{M}$ consists of the orthogonal resolutions of the identity on Ω. Then the probabilities of measurement outcomes in a given state are determined by the formula (2.3). If, on the other hand, we include measurements with random errors, then $\mathfrak{M}$ will consist of all, not only orthogonal, resolutions of the identity on Ω, and the probabilities of measurement outcomes will be determined by the formula (2.4). In this way arise the two basic classical models which differ in the structure of the set $\mathfrak{M}$. The first is the Kolmogorov model while the second may be called the Wald model. Both these models are separated.

2.4. Statistical model of quantum mechanics

In the previous section we have seen that quite general properties of statistical description lead to emergence of the two main mathematical structures: convexity (mixing) in the state space and subordination in the set of measurements. Importance of these structures in the context of quantum mechanics was stressed already in the book of von Neumann. Let us consider this in more detail.

In quantum mechanics to a system is related a complex Hilbert space $\mathcal{H}$. For simplicity we take it finite-dimensional. Then $\mathcal{H}$ consists of column vectors $|\psi\rangle = \begin{bmatrix} \psi_1 \\ \dots \\ \psi_n \end{bmatrix}$, where ψ_j are complex numbers. Denoting $\langle\psi| = \begin{bmatrix} \bar{\psi}_1 & \dots & \bar{\psi}_n \end{bmatrix}$ the Hermitian conjugate row-vector, we can write the inner product in $\mathcal{H}$ in the form $\langle\varphi|\,\psi\rangle$. Any ensemble of quantum systems defining a quantum state is described by a density matrix $\hat{S} = \left[s_{ij}\right]$ satisfying

$$\hat{S} \geq 0, \quad \mathrm{Tr}\hat{S} = 1.$$

The first relation means that the matrix $\hat{S}$ is Hermitian positive semidefinite while $\mathrm{Tr}\hat{S} = \sum_i s_{ii}$ denotes the trace of the matrix. If the ensem-

bles described by the density matrices $\hat{S}_1, \hat{S}_2$ are mixed in the proportion $p : (1-p)$ then the mixed ensemble is described by the matrix $p\hat{S}_1 + (1-p)\hat{S}_2$ which is again a density matrix. Thus the set $\mathfrak{S} = \mathfrak{S}(\mathcal{H})$ of all density matrices is convex, the convex combinations describing mixing of the ensembles. The extreme points of the convex set $\mathfrak{S}(\mathcal{H})$ – pure quantum states – are given by the density matrices of the form $\hat{S}_\psi = |\psi\rangle\langle\psi|$, with ψ a unit vector, $\langle\psi|\psi\rangle = 1$. It is difficult to describe the set $\mathfrak{S}(\mathcal{H})$ in pictorial way, except for the case $n = 2$, when it turns out to be isomorphic to the unit ball in the real 3-dimensional space. Any density matrix can be represented as a mixture of extreme points

$$\hat{S} = \sum_j p_j \hat{S}_{\psi_j}; \quad p_j \geq 0, \sum_j p_j = 1,$$

but contrary to the classical picture such a representation is highly nonunique.

A real-valued quantum observable is described by a Hermitian matrix $\hat{X} = \left[x_{ij}\right]$ with its eigenvalues x_j as possible outcomes. Let us write the spectral decomposition of the matrix $\hat{X}$:

$$\hat{X} = \sum_{j=1}^{n} x_j \hat{E}_j, \tag{2.6}$$

where E_j is the orthogonal projector onto the subspace of eigenvectors corresponding to the eigenvalue x_j. The family $\hat{\boldsymbol{E}} = \{\hat{E}_j\}$ forms an orthogonal resolution of identity in $\mathcal{H}$:

$$\sum_{j=1}^{n} \hat{E}_j = \hat{I}; \quad \hat{E}_j \hat{E}_k = 0 \quad \text{if } j \neq k; \quad \hat{E}_j^2 = \hat{E}_j,$$

where $\hat{I}$ is the unit matrix. The space $\mathcal{H}$ is then decomposed into the direct orthogonal sum of the subspaces $\hat{E}_j(\mathcal{H})$.

According to the rules of linear algebra, one has for a real function f

$$f(\hat{X}) = \sum_{j=1}^{n} f(x_j)\hat{E}_j. \tag{2.7}$$

The numbers $f(x_j)$ are not all necessarily different, so that some terms in (2.7) may be joined in groups. Similarly to the classical picture, it is convenient from the beginning to consider representations of an observable in a form (2.6) where x_j are not necessarily different. Then the resolution of the identity $\boldsymbol{E}$ may be defined not uniquely and can be more or

less "detailed". The spectral representation gives the unique least detailed resolution of the identity. By making more and more detailed decompositions we finally arrive to a "maximal" decomposition which is defined by a basis of eigenvectors of $\hat{X}$. Such a maximal resolution of the identity will be nonunique if $\hat{X}$ has eigenvalues of multiplicity greater than 1.

By the statistical postulate, measurement of the observable $\hat{X}$ in the state $\hat{S}$ gives the outcome x_j with the probability

$$\mu_{\hat{S}}^{\hat{E}}(x_j) = \mathrm{Tr}\hat{S}\hat{E}_j. \tag{2.8}$$

This and (2.6) imply that the mean value of $\hat{X}$ in the state $\hat{S}$ is equal to $\mathrm{Tr}\hat{S}\hat{X}$. For a pure state $\hat{S}_\psi$ this is equal to $\langle\psi|\hat{X}\psi\rangle$.

Thus the standard form of quantum mechanics is described by the statistical model $(\mathfrak{S}(\mathcal{H}), \mathfrak{M}(\mathcal{H}))$, where $\mathfrak{M}(\mathcal{H})$ is the collection of all resolutions of the identity in $\mathcal{H}$ corresponding to measurements of quantum observables. The reader have probably noticed the analogy between the relations (2.1), (2.2), (2.3) of the Kolmogorov model and the relations (2.6), (2.7), (2.8). It is also natural to consider the quantum analog of the Wald model in which the measurements are described by arbitrary (non-orthogonal) resolutions of the identity in $\mathcal{H}$, *i.e.* the families of matrices $\hat{\boldsymbol{M}} = \{\hat{M}_j\}$ satisfying

$$\sum_j \hat{M}_j = \hat{I}, \quad \hat{M}_j \geq 0.$$

While the mathematical apparatus of quantum mechanics contained all the necessary prerequisites for that, the role of non-orthogonal resolutions of the identity was recognized only in 1970-s with the advances in the statistical approach. Formally this extension of the notion of quantum observable is similar to introduction of randomized procedures in the classical picture. From this point of view the orthogonal resolutions of the identity are similar to the classical deterministic procedures. However this analogy is not complete in one crucial respect: while in the classical picture the measurement procedures that are optimal in the sense of the ultimate accuracy and maximal informativity are usually deterministic, in quantum mechanics the statistically and informationally optimal measurement procedures are often described by non-orthogonal resolutions of the identity. As it was shown in Section 2, a classical randomized measurement can be reduced to observation over a composite system including both the object and the generator of random numbers. It is intuitively clear that such a procedure cannot bring more information about the state of the classical object than direct observation of the system. However in quantum statistics, paradoxically, the use of a "quantum roulette" allows in several cases to improve the data concerning the

state of the system. The profound reason for this is the "nonseparability" of quantum-mechanical description and the specific "entanglement" between the parts of a composite system to be discussed below. It is noteworthy that the recognition of this remarkable fact was stimulated by formulation and solution of applied problems in quantum communication theory (see Chapters 3, 4 of the main text).

Since the main objective of our further discussion will be the problem of hidden variables which arose in the standard framework of quantum mechanics, we shall proceed within this framework. However the main conclusions apply also to the extended statistical model which uses arbitrary resolutions of the identity for the description of quantum measurements.

2.5. Compatibility and complementarity

Subordination of observables defines partial order in the set of measurements $\mathfrak{M}$ which has a direct statistical meaning: if M_1 is subordinated to M_2, then M_2 is a more detailed, informative measurement than M_1. If M_1 and M_2 are mutually subordinated then they are *equivalent* from the point of view of statistical information. An example is given by measurements performed with one device with differently calibrated scales U_1, U_2.

Clearly, most important are the maximal elements of the set $\mathfrak{M}$ which describe the most informative measurements admitted in the given statistical model. Mathematically, existence of such maximal measurements may be a nontrivial problem. We simply assume that every measurement in $\mathfrak{M}$ is subordinated to a maximal one. In general, there can be many inequivalent maximal measurements. Uniqueness of the maximal measurement up to equivalence turns out to be a characteristic property of the classical model.

To explain this we introduce the important definition: the measurements M_1, M_2 are called *compatible* if they are both subordinated to some measurement M. In other words, outcomes of compatible measurements can be obtained as a result of post-processings in a single-measurement experiment M. Assume now that a separable statistical model $(\mathfrak{S}, \mathfrak{M})$ is such that all the measurements are compatible. Then the maximal measurement M^* is unique up to equivalence. Let $M^* : S \to \mu_S^*$ be a representative and let Ω be the set of its outcomes. Since every measurement M is subordinated to M^*, there is a function $f_M : \Omega \to U$, where U is the set of outcomes for M such that

$$\mu_S^M(B) = \mu_S^*(f_M^{-1}(B)), \quad B \subseteq U.$$

Due to assumed separatedness of the model the map $S \to \mu_S^*$ is one-to-one affine correspondence between the set $\mathfrak{S}$ and a convex subset of the

simplex $\mathfrak{S}(\Omega)$, so that states can be identified with probability distributions on the set Ω and μ_S^M is the distribution of the observable f_M in the classical state μ_S^*.

Thus a separated statistical model in which all the measurements are compatible essentially reduces to the classical picture, the role of the phase space is played by the set of outcomes of the maximal measurement. In this connection it is useful to observe that Kolmogorov's extension theorem [78] may be interpreted as a statement about existence of the maximal measurement for an infinite projective family of mutually compatible measurements.

On the other hand, if $\mathfrak{M}$ contains incompatible measurements then the maximal measurement cannot be unique. Let us return to the model of quantum mechanics. Let $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{F}}$ be the measurements described by orthogonal resolutions of the identity $\{\hat{E}_j\}, \{\hat{F}_j\}$ which commute in the sense that

$$\hat{E}_j \hat{F}_k = \hat{F}_k \hat{E}_j \quad \text{for all} \quad j, k. \tag{2.9}$$

Then the relation $\hat{G}_{jk} = \hat{E}_j \hat{F}_k$ defines a measurement with respect to which $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{F}}$ are subordinated since $\hat{E}_j = \sum_k \hat{G}_{jk}, \quad \hat{F}_k = \sum_j \hat{G}_{jk}$. The condition (2.9) is not only sufficient but also necessary for the compatibility of $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{F}}$. Observables $\hat{X}$ and $\hat{Y}$ are called *compatible* if there exist compatible measurements for them and this turns out to be equivalent to the condition $\hat{X}\hat{Y} = \hat{Y}\hat{X}$. In this way one derives the definition of compatibility from the standard formulation of quantum mechanics.

Since there are many incompatible observables described by noncommuting matrices, there are many inequivalent maximal measurements defined by different orthonormal bases in the space $\mathcal{H}$. These are the orthogonal resolutions of the identity which can not be further split into more detailed ones. In the infinite dimensional space the situation is more complicated since "continuous" maximal orthogonal resolutions of the identity appear. Notice however, that even in finite dimensional case there exist continuous maximal nonorthogonal resolutions of the identity. They arise as "overcomplete" systems of vectors obtained by projecting to $\mathcal{H}$ of the maximal orthogonal resolutions of the identity in some larger Hilbert space $\mathcal{H}'$. A typical example of an overcomplete system is the system of coherent vectors well known in quantum optics. In many cases the maximal information about quantum state is carried by the measurements described by such overcomplete systems (see Chapters 3, 4 of the main text).

There exist physical systems which are in a sense intermediate between classical and quantum such as systems with superselection rules. Let us call *center* of the statistical model $(\mathfrak{S}, \mathfrak{M})$ the collection of all measure-

ments compatible with all measurements in $\mathfrak{M}$. With the center one can associate its spectrum reflecting the classical properties of the statistical model. For a classical model the center coincides with the phase space. If the center is trivial *i.e.* consists only of the constants, as it is the case for quantum mechanics, then the model is *irreducible*. In general one can expect a decomposition of rather arbitrary statistical model into direct sum or integral of irreducible models. Such a structural theory is completely elaborated in the framework of algebraic approach and there are its generalizations to general convex sets of states [4].

Understanding the phenomenon of incompatibility was a decisive step in the creation of the "orthodox" interpretation of quantum mechanics. Incompatibility of measurements in quantum mechanics stems from the fact that physical measurements are performed with the macroscopic experimental setups, each one assuming complex specific organization of the space-time environment. Apparently two different ways of such organization can be mutually exclusive. "In quantum physics evidence about atomic objects obtained by different experimental arrangements exhibits a novel kind of peculiar complementary relationship" [155, page 4]. The classical mechanics relies upon idealization admitting theoretical compatibility of all measurement procedures which is justified in so far as it deals with macroscopic objects the interactions of which with the measurement devices can be as weak as one desires.

The surrounding world which is accessible to direct human's perception is macroscopic "by the definition". Therefore is it so difficult to give an accessible image for complementarity. However Niels Bohr advocated the idea that the principle of complementarity is quite general and especially important for biology and for subtle humanitarian relations which are so difficult to model mathematically. "The integrity of living organisms and the characteristics of conscious individuals and human cultures present features of wholeness, the account of which implies a typical complementary mode of description" [155, page 7]. Nevertheless so far quantum theory remains a unique case of a mathematical theory of a segment of the reality which, in its domain, gives an exact quantitative expression of the dialectic principle of complementarity.

2.6. Classical and nonclassical models

Consider statistical model $(\mathfrak{S}, \mathfrak{M})$ which is "classical" in the sense that the states and the measurements are described in terms of some phase space as in Section 2. However we do not require that the state space $\mathfrak{S}$ contains all the probability distributions as well as the set $\mathfrak{M}$ – measurements of all observables; so we admit that $\mathfrak{S}$ and $\mathfrak{M}$ can be defined by

some apriori restrictions (like those arising from complementarity). For this reason the model $(\mathfrak{S}, \mathfrak{M})$ need not be separated: there can be different probability distributions P_1, P_2 such that $\mu^M_{P_1} = \mu^M_{P_2}$ for all $M \in \mathfrak{M}$, *i.e.* indistinguishable from the point of view of all the measurement statistics. Similarly, there can be indistinguishable measurements. By identifying such indistinguishable states and measurements into equivalence classes, we obtain the new, already separated, statistical model.

The description of states and measurements in this new model is "compressed" just to the extent sufficient to preserve the measurements statistics in the initial classical model. As we have seen in Chapter 1, "gluing" the classical states geometrically amounts to projection of the state space $\mathfrak{S}$ onto a subset $\hat{\mathfrak{S}}$ in less dimensions. Then even if the initial model admits all possible probability distributions on Ω so that $\mathfrak{S}$ is just the simplex $\mathfrak{S}(\Omega)$, its projection $\hat{\mathfrak{S}}$ in the compressed model can be rather arbitrary convex set. Its form is determined by the set $\mathfrak{M}$, *i.e.* by the restrictions upon the classical measurements.

In statistical mechanics an important role is played by reduced description related to the probabilistic notion of partial observability. Assume that in the whole variety of variables related to a classical object with the phase space Ω one can observe only the random variables $X_1, \ldots, X_n$, as well as measurable functions of them. In this case $\mathfrak{M}$ consists of the measurements of observables of the form $f(X_1, \ldots, X_n)$. Then the classical states P_1, P_2 are indistinguishable if the corresponding expectations coincide: $\mathsf{M}_{P_1} f(X_1, \ldots, X_n) = \mathsf{M}_{P_2} f(X_1, \ldots, X_n)$ for all measurable functions f (this means that the restrictions of P_1, P_2 onto $\sigma-$subalgebra $\mathcal{B}$ generated by $X_1, \ldots, X_n$ coincide). Then the equivalence classes are represented by probability distributions $P(dx_1 \ldots dx_n)$ on the space $\hat{\Omega}$ of the values of the variables $X_1, \ldots, X_n$, which can thus be taken for the phase space of the reduced description. One has $\hat{\mathfrak{S}} = \mathfrak{S}(\hat{\Omega})$ *i.e.* the simplex $\mathfrak{S}(\Omega)$ is projected onto the simplex $\mathfrak{S}(\hat{\Omega})$ and the classical nature of the description is preserved. In this case the simplex is projected "along its faces" and the geometrical picture is preserved. Thus the reduced description is an important particular case of the state space compression.

Just to present a different picture consider the following modification of partial observability: assume that one can observe only the variables of the form $f_1(X_1), \ldots, f_n(X_n)$, where $f_1, \ldots, f_n$ are arbitrary functions. Then the classical states P_1, P_2 will be indistinguishable if

$$\mathsf{M}_{P_1} f_i(X_i) = \mathsf{M}_{P_2} f_i(X_i)$$

for all measurable functions $f_i; i = 1, \ldots, n$ (*i.e.* the restrictions of P_1, P_2 onto $\sigma-$subalgebras $\mathcal{B}_1, \ldots, \mathcal{B}_n$ generated by corresponding random variables $X_1, \ldots, X_n$ coincide). Then the equivalence classes are

represented by collections $\hat{P} = (P_1(dx_1), \dots P_n(dx_n))$ of probability distributions on the spaces of values of the variables $X_1, \dots, X_n$. In that case $\hat{\mathfrak{S}}$ is a direct product of n simplexes; in particular, if $X_1, \dots, X_n$ are two-valued then $\hat{P}$ runs through $n-$dimensional hypercube. Thus introducing apriori restrictions can radically change the convex structure of the compressed state space.

As we have seen in Chapter 1, for any sufficiently regular separated statistical model $(\hat{\mathfrak{S}}, \hat{\mathfrak{M}})$ there is a classical model for which $(\hat{\mathfrak{S}}, \hat{\mathfrak{M}})$ is the compression in the sense described above. Does this mean a possibility of introducing hidden variables, in particular, for quantum mechanics? To answer this question we should analyze the requirements which has to be satisfied by a hidden variable model. Such requirements can be conventionally split into two classes. The first class is constituted by the "minimal" requirements which refer only to the general properties of statistical description for a single system. These will be considered in the first place since almost all attempts of mathematical proofs of nonexistence pretended to appeal only to such general requirements. We shall see that in fact they were based on additional assumptions lacking physical motivation. Moreover, we shall demonstrate the classical description for a single quantum system satisfying all the general requirements of statistical description.

Another class is constituted by requirements which refer to composite quantum systems and it is these requirements that present the main obstacle for a hidden variable theory.

3. The problem of hidden variables

3.1. "No-go" proofs and the minimal statistical requirements on hidden variables

A hidden variable theory pretends to give an explanation of the randomness of the experimental results through fluctuations of the values of certain variables ω describing the "real" properties of the object. Therefore attempts of the "no-go" proofs usually started with associating probability distributions $S(d\omega)$ with quantum states $\hat{S}$ and random variables $X(\omega)$ with quantum observables $\hat{X}$, where $\omega \in \Omega$ – a hypothetical phase space. Thus it was usually assumed that there are one-to-one correspondences $\hat{S} \leftrightarrow S$ and $\hat{X} \leftrightarrow X$.

However the discussion of the compressed statistical description in the previous section leads to an idea that a hidden variable theory should admit "gluing" statistically equivalent states and observables. Therefore we accept from the beginning that a classical description of a quantum system consists of the phase space Ω and a pair of mappings: $S \to \hat{S}$ from

the simplex $\mathfrak{S}(\Omega)$ onto the set of quantum states $\hat{\mathfrak{S}}$ and $X \to \hat{X}$ from the set of classical observables $\mathfrak{O}(\Omega)$ onto the set of quantum observables $\hat{\mathfrak{O}}$.

The domain of definition of the first mapping need not be the whole $\mathfrak{S}(\Omega)$ as well as that of the other – the whole $\mathfrak{O}(\Omega)$. Thus the classical description brings into correspondence some classical states $S(d\omega)$ with the density operators $\hat{S}$ and some classical observables $X(\omega)$ with Hermitian operators $\hat{X}$ in the Hilbert space $\mathcal{H}$. One and the same quantum state $\hat{S}$ can be described by several different probability distributions $S(d\omega)$ interpreted as different ways of preparation of the quantum ensemble $\hat{S}$, and one and the same quantum observable $\hat{X}$ can be described by different functions $X(\omega)$ interpreted as different ways of observation of $\hat{X}$. To denote possible non-uniqueness of the classical description Bell used the term "contextuality". The contextuality in the description of a quantum state displays in particular in the fact that one and the same density operator $\hat{S}$, depending on the context of a preparation procedure, can be obtained as completely different mixtures of pure states. Similarly, one and the same projector $\hat{P}$, depending on the context of a measurement procedure, can arise as a result of coarse-graining from different orthonormal bases.

Now let us consider requirements which where imposed on hidden variable theories. In view of the preceding discussion the following assumptions are the restrictions to be stated explicitly:

(S.0) *the mapping* $S \to \hat{S}$ *is one-to-one*;

(X.0) *the mapping* $X \to \hat{X}$ *is one-to-one.*

These assumptions which may seem "technical" from a mathematical viewpoint are in fact of crucial importance. At the same time their physical motivation may be questioned.

The requirement that all the statistical predictions of the hidden variable theory coincide in all respects with the quantum mechanical ones can be expressed as equality between the expectations:

(E.1) $\mathrm{Tr}\hat{S}\hat{X} = \int_\Omega S(d\omega)X(\omega) \quad$ for all X, S.

If this requirement does not hold, then there is a disagreement between statistical predictions of the theories which in principle can be detected experimentally. This would lead to physical questions which are beyond the scope of the present mathematical consideration.

The next group of requirements concerns the properties of the mapping $X \to \hat{X}$. The relation (2.7) shows that the outcomes of observable $f(\hat{X})$ can be obtained from the outcomes of $\hat{X}$ by mere computation $x_i \to f(x_i)$. But in a hidden variable theory a measurement of

$\hat{X}$ is reduced to observation of some $X(\omega)$. The subsequent computation $x \to f(x)$ is equivalent to a direct observation of $f(X(\omega))$. Thus in the hidden variable theory observation of $f(X)$ represents a way of measuring the quantum observable $f(\hat{X})$. This motivates the following functional condition:

(X.1) *if* $X \to \hat{X}$ *then* $f(X) \to f(\hat{X})$.

This is closely related to the following two conditions:

(X.2) *if* $X \to \hat{X}$ *then any value* $X(\omega)$ *belongs to the spectrum* $\{x_i\}$ *of quantum observable* $\hat{X}$.

The meaning of this *spectral condition* is clear: a hidden variable description should preserve "objective values" of observables.

(X.3) *for any pair of compatible quantum observables* $\hat{X}, \hat{Y}$ *there exist the corresponding classical observables* X, Y *such that* $X \to \hat{X}, Y \to \hat{Y}$ *and* $X + Y \to \hat{X} + \hat{Y}$.

Compatibility of quantum observables means that there is a measuring device which produces the outcomes of measurement of both $\hat{X}$ and $\hat{Y}$. Complementing this with a summator we obtain a device for measuring $\hat{X} + \hat{Y}$. The *finite sum rule* (X.3) reflects this possibility in the hidden variable theory. It can be replaced with the following *finite product rule*:

(X.4) *for any pair of compatible quantum observables* $\hat{X}, \hat{Y}$ *there exist the corresponding classical observables* X, Y *such that* $X \to \hat{X}, Y \to \hat{Y}$ *and* $XY \to \hat{X}\hat{Y}$.

Lemma 3.1. *The functional condition* (X.1) *implies the finite sum rule* (X.3) *and the finite product rule* (X.4). *Under* (X.0)*, the condition* (X.1) *implies also the spectral condition* (X.2).

Proof. If $\hat{X}, \hat{Y}$ are compatible then there exists $\hat{Z}$ such that $\hat{X} = f(\hat{Z}), \hat{Y} = g(\hat{Z})$. Let $Z \to \hat{Z}$ be the corresponding classical observable, then by (X.1) $X = f(Z) \to \hat{X}$ and $Y = g(Z) \to \hat{Y}$. Therefore $X + Y = (f + g)(Z) \to (f + g)(\hat{Z}) = \hat{X} + \hat{Y}$. This proves (X.3), and (X.4) is proved similarly.

Let the correspondence $X \to \hat{X}$ be one-to-one so that we can write $X \leftrightarrow \hat{X}$. Consider the function $f_0(x) \equiv 0$. Applying (X.1) to this function, we have $0 \leftrightarrow \hat{0}$, where $\hat{0}$ is the zero operator and 0 is the classical observable identically equal to zero. Let $P(x)$ be the characteristic polynomial of an Hermitian operator $\hat{X}$, so that $P(\hat{X}) = \hat{0}$. If $X \leftrightarrow \hat{X}$ then by (X.1) $P(X) \leftrightarrow P(\hat{X})$ hence $P(X(\omega)) \equiv 0$ *i.e.* any value $X(\omega)$ belongs to the spectrum of quantum observable $\hat{X}$. □

The conditions (E.1) and (X.1) imply the following amplification of the property of statistical correspondence:

(E.2) $\mathrm{Tr}\hat{S}f(\hat{X}) = \int_\Omega S(d\omega)f(X(\omega))$ for all X, S, f.

Notice that while (E.2) trivially implies (E.1), the condition (X.1) follows only under some additional assumption. A subset $\mathfrak{S}_0 \subseteq \mathfrak{S}(\Omega)$ will be called *separating* if the equality

$$\int_\Omega S(d\omega)X_1(\omega) = \int_\Omega S(d\omega)X_2(\omega) \quad \text{for all } S \in \mathfrak{S}_0$$

implies $X_1 = X_2$. If the set of probability distributions which correspond to all possible quantum states in the given classical description (let us denote it $\mathfrak{S}_0$) is separating then the condition (E.1) implies (X.0) and (E.2) implies (X.1). Indeed, let $Y \to f(\hat{X})$ then by (E.2)

$$\int_\Omega S(d\omega)Y(\omega) = \mathrm{Tr}\hat{S}f(\hat{X}) = \int_\Omega S(d\omega)f(X(\omega)) \quad \text{for all } S \in \mathfrak{S}_0$$

whence $f(X) = Y \to f(\hat{X})$.

Finally let us discuss the linearity condition:

(X.5) $\widehat{(\lambda X + \mu Y)} = \lambda\hat{X} + \mu\hat{Y}$ for all X, Y and real λ, μ .

This mathematically innocent condition is not justified physically as it was mentioned already in the von Neumann's book [138]. If $\hat{X}$ and $\hat{Y}$ are incompatible then measurements of observables $\hat{X}$,$\hat{Y}$ and $\hat{X}+\hat{Y}$ may have nothing in common apart from equality of the mean values. This last relation was used by von Neumann to motivate the condition (X.5). Let us analyze the corresponding argument as given in[2]: "In quantum mechanics the mean values satisfy the relation $\langle\hat{X}+\hat{Y}\rangle = \langle\hat{X}\rangle + \langle\hat{Y}\rangle$. Therefore in a hidden variable model it should hold

$$\int_\Omega S(d\omega)(X+Y)(\omega) = \int_\Omega S(d\omega)X(\omega) + \int_\Omega S(d\omega)Y(\omega). \qquad (3.10)$$

Since to different density matrices $\hat{S}$ correspond different probability distributions $S(d\omega)$, it is natural to require that

$$(X+Y)(\omega) = X(\omega) + Y(\omega) \qquad (3.11)$$

[2] A.I. Akhiezer, R.V. Polovin, Why it is impossible to introduce hidden variables in quantum mechanics? UFN 107(3) (1972) 463-487.

for arbitrary classical observables X, Y which may correspond to commuting as well as noncommuting operators $\hat{X}, \hat{Y}$."

However (3.11) follows from (3.10) only under additional assumption that the set of classical states $\mathfrak{S}_0$ in the hidden variable theory is separating. Is such an assumption "natural"? A very simple (and generic) example of nonseparating set is the subset of probability distributions $S(d\omega)$ on a product $\Omega' \times \Omega''$ which have the form $S(d\omega) = S'(d\omega')P(d\omega'')$, where $S'(d\omega')$ is arbitrary distribution on Ω' and $P(d\omega'')$ is a fixed distribution on Ω''. This is just the case in the explicit hidden variable models. The distribution $P(d\omega'')$ plays a role of the equilibrium state of the "hidden" subsystem which provides stochasticity to the measurement outcomes [10].

Under the condition of one-to-one correspondence (X.0) the conditions (X.1)-(X.5) turn into the requirements which were used in the "no-go" proofs for hidden variables. Let us give here the most significant results in this direction in the formulation adapted to the proposed general classification.

The first statement is close to what is called the von Neumann theorem.

Proposition 3.2. *There is no classical description satisfying the conditions* (X.0), (X.2), (X.5).

From the discussion above we see that this can be also formulated as follows: *there is no classical description with separating set of classical states, satisfying the condition* (E.2).

Proof. Assume that a classical description with the properties (X.0), (X.2), (X.5) exists. By (X.0) there is unique X such that $X \to \hat{X}$. Fix a point ω_0 of the phase space and consider the functional on operators given by $F(\hat{X}) = X(\omega_0)$. By (X.5) this functional is linear. In the finite dimensional case we are considering, it is almost obvious that every such functional has the form $F(\hat{X}) = \mathrm{Tr}\hat{P}\hat{X}$, where $\hat{P} \in \hat{\mathfrak{O}}$. Let $\hat{X}$ run over all possible projections in $\mathcal{H}$ so that its eigenvalues are $x_i = 0, 1$. Then by (X.2) the quantity $X(\omega_0) = \mathrm{Tr}\hat{P}\hat{X}$ assumes only the values 0, 1 which is apparently impossible for any choice of $\hat{P}$. □

Of historical interest is the following remark of Wigner to his article on hidden variables [198]: "The discussion of Von Neumann, most commonly quoted, is that contained in his book, Sections 4.1 and 4.2. As an old friend of Von Neumann, and in order to preserve historical accuracy, the present writer may be permitted the observation that the proof contained in this book was not the one which was principally responsible for Von Neumann's conviction of the inadequacy of hidden variable

theories. . . "[3] However independently of the meaning given to these arguments by von Neumann, published in his monographs they were considered as decisive argument against hidden variables for the years until 1950-s with the appearance of papers of Bohm (1952) and Wiener and Siegel (1953), which contained explicit (although not completely transparent) constructions of hidden variable models. On the other hand, several improvements of the von Neumann "no-go theorem" appeared. The situation was radically clarified with the works of Bell (1966) and also Kochen and Specker (see [11, 77]).

In the year 1957 Gleason [168] proved the following quite nontrivial theorem which gave an answer to a question of Mackey:

Let $F(\hat{E})$ be a probability measure on projections in the Hilbert space $\mathcal{H}$ of dimensionality ≥ 3, i.e. the real function satisfying the conditions: 1) $F(\hat{E}) \geq 0$; 2) *for any orthogonal resolution of the identity $\{\hat{E}_i\}$ it holds $\sum_i F(\hat{E}_i) = 1$. Then $F(\hat{E}) = \mathrm{Tr}\hat{S}\hat{E}$, where $\hat{S}$ is a density operator in $\mathcal{H}$.*

This implies

Proposition 3.3. *If* $\dim \mathcal{H} \geq 3$ *then there is no classical description satisfying the conditions* (X.0), (X.2), (X.3).

Indeed, if such a description exists then the function $F(\hat{E}) = E(\omega_0)$, where $E \leftrightarrow \hat{E}$, satisfies the conditions of Gleason's theorem. Then $F(\hat{E}) = \mathrm{Tr}\hat{S}\hat{E}$ and we come to the contradiction as in the proof of Proposition 1.

The proof of Gleason's theorem remains difficult even after a number of subsequent simplifications. Gleason himself did not apply his theorem to the hidden variable problem. This possibility was noticed by Bell. Moreover he extracted the geometrical idea behind Gleason's proof which is essential from the viewpoint of hidden variables issue and gave a short direct proof of the following statement which in our presentation follows from Proposition 3.3 and the Lemma 3.1:

Proposition 3.4. *If* $\dim \mathcal{H} \geq 3$ *then there is no classical description satisfying the conditions* (X.0) *and* (X.1).

Similar result was independently and in a quite different way obtained by Kochen and Specker [77] who gave an explicit construction of 117 (later more "economical" constructions were designed) unit vectors in three-dimensional space on which one cannot define a measure with the

[3] This unpublished argument of von Neumann will be considered in the next section.

properties 1), 2) from Gleason's theorem and assuming only the values 0 or 1. In other words, the geometry of this system is such that the conditions defining such a measure become self-contradictory. Later it was observed that the proof of Bell in fact amounts to construction of a certain system of 13 vectors [10].

The case $\dim \mathcal{H} = 2$ corresponding to spin-1/2 particle is special. In the papers of Bell and of Kochen-Specker explicit hidden variable constructions were given for this case. One more construction follows from the general model which will be given in the next section. These constructions clearly demonstrate insufficiency of the "no-go" proofs of the type of Proposition 3.2.

Notice that the Propositions 3.3 and 3.4 do not use at all the condition of statistical correspondence (E.1) which in fact should be central for a hidden variable theory. These results do not concern statistics and speak only of impossibility of the one-to-one correspondence between the "quantum logic" of projections and the Boolean algebra of classical events preserving the algebraic relations between the compatible variables. The key condition here is (X.0) as it was understood in the work of Bell.

These mathematical results do not contradict explicit hidden variable models such as Bohm's model and Wiener-Siegel model: the assumption which is not fulfilled in these models is the condition (X.0) of one-to-one correspondence. We already noticed that a reasonable assumption would be to admit that one and the same quantum observable can be measured in many different ways (which in particular is reflected in possible nonuniqueness of the resolution of the identity entering the representation (2.6)). Propositions 3.3, 3.4 imply that in a hidden variable theory satisfying functional conditions of the type (X.1)-(X.4) such a nonuniqueness is unavoidable.

A different approach to "no-go" proofs is related to the ideas of Wigner and Blokhintsev on the impossibility of defining a joint distribution for incompatible quantum observables and is based on the analysis of the properties of the mapping $S \to \hat{S}$.

Consider the classical state S and the corresponding quantum state $\hat{S}$. Then the classical ensemble defined by the probability distribution $S(d\omega)$ replaces the quantum ensemble representing $\hat{S}$. The mixture $pS_1 + (1 - p)S_2$ of the classical ensembles represents the quantum ensemble $p\hat{S}_1 + (1 - p)\hat{S}_2$. This leads to the following affinity condition:

(S.1) *for* $S_1 \to \hat{S}_1$, $S_2 \to \hat{S}_2$ *and a real* p, *satisfying* $0 < p < 1$ *it holds* $pS_1 + (1 - p)S_2 \to p\hat{S}_1 + (1 - p)\hat{S}_2$.

Proposition 3.5. *There is no classical description satisfying the conditions* (E.2), (S.0), (S.1).

Proof. Assume that such a description exists and consider two arbitrary quantum observables $\hat{X}, \hat{Y}$. Let X, Y be some corresponding classical observables. Due to the condition (S.0), there is one-to-one affine correspondence $S \leftrightarrow \hat{S}$ and the formula

$$\mu_{\hat{S}}(B) = S(\omega : (X(\omega), Y(\omega)) \in B)$$

defines a probability distribution on the plane $\mathbb{R}^2$. Due to (S.1) the mapping $\hat{S} \to \mu_{\hat{S}}$ is affine and hence defines a generalized quantum measurement with outcomes (x, y). By using the condition (E.2) we obtain that distributions of observables $\hat{X}, \hat{Y}$ are marginals of the distribution $\mu_{\hat{S}}(dx, dy)$. Hence we come to an absurd conclusion that any two quantum observables $\hat{X}, \hat{Y}$ are compatible. □

This argument follows essentially the paper of Srinivas [195], where it is stressed that the key condition here is (S.0). Any quantum state $\hat{S}$ can be represented in many different ways as the mixture of pure states. If a hidden variable theory satisfying the conditions (E.2), (S.1) is possible then it should necessarily admit such a nonuniqueness. In this respect the condition (S.0) is similar to (X.0).

Thus among the requirements which were proposed for hidden variable theories one should distinguish (E.1), (X.1), (S.1) (as well as the related conditions (E.2), (X.2)-(X.4)) which have sound statistical motivation. Essentially, these conditions require that the classical description should preserve the main structural properties of statistical model expressed by the axioms (A.1)-(A.3). The "no-go" theorems in fact do not forbid such theories. A mathematically and physically motivated conclusion is that a classical description satisfying these requirements should have necessarily non-unique nature admitting possibility of compression in transition to quantum theory.

3.2. A hidden variable model for a single quantum system

In the work of Kochen and Specker [77] the authors gave a "trivial" hidden variable model satisfying the statistical condition (E.1) but not preserving the structure of functional dependences in quantum mechanics. The idea was straightforward and amounts to introducing a personal "hidden variable" for each observable $\hat{X}$ producing stochasticity in outcomes of measurement of $\hat{X}$. The totality of all such hidden variables is then the phase space variable ω of the model. The purpose of this construction was to demonstrate insufficiency of the statistical condition (E.1) and

the power of the functional condition (X.1). However we will show that basing on a similar idea one can construct a hidden variable model satisfying all the minimal requirements (E.1), (X.1), (S.1). The key feature is of course the non-uniqueness of the classical description.

Denote by Ω' the set of all pure quantum states, so that for any $\omega' \in \Omega'$ there is a unit vector $\psi_{\omega'} \in \mathcal{H}$ such that $\mathrm{Tr}\omega'\hat{X} = \langle\psi_{\omega'}|\hat{X}\psi_{\omega'}\rangle$. Every quantum state can be written as a mixture of pure states

$$\hat{S} = \int_{\Omega'} \omega' S'(d\omega'), \tag{3.12}$$

where $S'(d\omega')$ is a probability distribution on Ω', and this representation is of course not unique. The relation (3.12) defines affine mapping $S' \to \hat{S}$ of the simplex $\mathfrak{S}(\Omega')$ onto the convex set $\hat{\mathfrak{S}}$.

Let $\hat{\boldsymbol{E}} = \{\hat{E}_i\}$ be an orthogonal resolution of the identity in $\mathcal{H}$ describing a quantum measurement. It suffices to restrict to the maximal measurements for which $\hat{E}_i$ are projections onto the vectors e_i of an orthonormal basis. Put

$$M_i(\omega') = \mathrm{Tr}\omega'\hat{E}_i = |\langle\psi_{\omega'}|e_i\rangle|^2, \tag{3.13}$$

then $\hat{\boldsymbol{M}} = \{M_i(\omega')\}$ will be a classical randomized measurement on Ω'. The relations (3.12), (3.13) imply

$$\mathrm{Tr}\hat{S}\hat{E}_i = \int_{\Omega'} S'(d\omega')M_i(\omega'). \tag{3.14}$$

Thus we succeeded to build a classical model in which states are described by probability distributions on the "phase space" Ω' and the (maximal) quantum measurements – by randomized classical measurements M, with the statistical correspondence condition (3.14) fulfilled. The quantum theory is the compressed description of this model in the sense of Section 2.6 (*cf.* also Theorem 1.7.1).

The next step is to realize every classical randomized measurement (3.13) with the help of randomizing probability space $\left(\Lambda_{\hat{\boldsymbol{E}}}, d\lambda_{\hat{\boldsymbol{E}}}\right)$ for example as it is described at the end of Section 2.2. One can say that to every maximal quantum measurement corresponds a "roulette" $\left(\Lambda_{\hat{\boldsymbol{E}}}, d\lambda_{\hat{\boldsymbol{E}}}\right)$ which models stochasticity of the measurement outcomes in any pure state ω'. If the state $\hat{S}$ is mixed, then the probability distribution of the measurement outcomes is given by the corresponding mixture

$$\mathrm{Tr}\hat{S}\hat{E}_i = \int_{\Omega'}\int_{\Lambda_{\hat{\boldsymbol{E}}}} S'(d\omega')d\lambda_{\hat{\boldsymbol{E}}}E_i\left(\omega', \lambda_{\hat{\boldsymbol{E}}}\right), \tag{3.15}$$

as it is seen from (3.14) and (2.5). To embrace the totality of the maximal quantum measurements, let us introduce the product of probability spaces

$\left(\Omega'', P''\right) = \prod_{\hat{E}} \left(\Lambda_{\hat{E}}, d\lambda_{\hat{E}}\right)$. Thus every $\omega'' \in \Omega''$ is a collection $\omega'' = \prod_{\hat{E}} \lambda_{\hat{E}}$ of independent "roulettes" corresponding to all possible maximal quantum measurements so that $P(d\omega'') = \prod_{\hat{E}} d\lambda_{\hat{E}}$.

Now define the phase space of the sought classical description as $\Omega = \Omega' \times \Omega''$, so that $\omega = \left(\omega', \omega''\right)$. The classical states will be given by the probability distributions on Ω having the form $S(d\omega) = S'(d\omega')P(d\omega'')$, where S' is an arbitrary probability distribution on Ω'. The mapping $S \to \hat{S}$ is defined as

$$\hat{S} = \int_{\Omega} \omega' S'(d\omega) = \int_{\Omega'} \omega' S'(d\omega'). \tag{3.16}$$

To every maximal quantum measurement $\hat{\boldsymbol{E}}$ corresponds uniquely the deterministic classical measurement $\boldsymbol{E} = \{E_i(\omega)\}$, where $E_i(\omega) = E_i\left(\omega', \lambda_{\hat{E}}\right)$, and $\lambda_{\hat{E}} = \pi_{\hat{E}}(\omega'')$ is the coordinate projection of the point ω''. From (3.15) it follows that

$$\mathrm{Tr}\hat{S}\hat{E}_i = \int_{\Omega} S(d\omega)E_i(\omega). \tag{3.17}$$

It remains to establish correspondence between quantum observables and random variables on Ω. Let $\hat{X}$ be a quantum observable and $\hat{\boldsymbol{E}} = \{\hat{E}_i\}$ - one of the corresponding *maximal* measurements, so that $\hat{X} = \sum_i x_i \hat{E}_i$. Let $\boldsymbol{E} = \{E_i(\omega)\}$ be the corresponding classical deterministic measurement. Consider the random variable $X(\omega) = \sum_i x_i E_i(\omega)$. Since $\hat{\boldsymbol{E}}$ is in general nonunique, we obtain a collection of random variables $X(\omega) \to \hat{X}$ corresponding to different $\hat{\boldsymbol{E}}$. It is important that $X(\omega)$ depends on ω'' only via the coordinate projection $\lambda_{\hat{E}} = \pi_{\hat{E}}(\omega'')$. If $X(\omega) \neq \text{const}$ this allows to reconstruct $\boldsymbol{E}$ uniquely given $X(\omega)$. The values x_i are also reconstructed by X. Then from $\boldsymbol{E}$ one recovers $\hat{\boldsymbol{E}}$ and hence $\hat{X} = \sum_i x_i \hat{E}_i$. Therefore the mapping $X \to \hat{X}$ is well defined for $X(\omega) \neq \text{const}$. In case $X(\omega) \equiv \lambda$ we have $x_i \equiv \lambda$, hence $\hat{X} = \lambda \hat{I}$ so the mapping is unambiguously defined also in this case.

From (3.17) it follows that the statistical correspondence (E.1) holds. The mapping $S \to \hat{S}$ is affine, so that (S.1) also holds. Let us check the functional condition (X.1). If X is a random variable, $X \to \hat{X}$, and f a function, then $f(X(\omega)) = \sum_i f(x_i)E_i(\omega)$. Without loss of generality we can assume both X and f nonconstant. Then $\hat{\boldsymbol{E}}$ and hence $\hat{X}$ is recovered from X uniquely, so that $\widehat{f(X)} = \sum_i f(x_i)\hat{E}_i = f(\hat{X})$.

Notice that in the case $\dim \mathcal{H} = 2$ every nontrivial resolution of the identity in $\mathcal{H}$ is maximal, therefore our construction in this case satisfies even the uniqueness condition (X.0). This explains the restriction $\dim \mathcal{H} \geq 3$ in Propositions 3.3, 3.4.

In brief, the proposed classical description is constructed so as to make distinguishable all the mixtures giving the same quantum state as well as all the measurements giving the same quantum observable. The correspondence $S \to \hat{S}$ is one-to-one only for the pure states $\hat{S}$ and the correspondence $X \to \hat{X}$ – only for the maximal observables with simple spectrum.

Certainly this model does not pretend to replace the mathematical apparatus of quantum mechanics. It is also extremely wasteful from the viewpoint of representing the measurement statistics as it contains a lot of irrelevant "detail". However it is relevant in making explicit the structural properties of quantum theory which can be preserved in classical description and in showing that "no-go" proof for hidden variables cannot be based solely on the minimal statistical requirements (E.1), (S.1), (X.1). Complementarity does not prevent from a classical description as soon as the classical-quantum correspondence is not required to be one-to-one. Similar conclusions hold for rather arbitrary separated statistical model since the minimal statistical requirements concern only the common properties of quantum mechanical description. Thus consideration of the hidden variable hypothesis should involve more specific properties of the quantum mechanical description.

3.3. Hidden variables and evolutions of quantum system

Here we discuss the question: can a classical description reproduce temporal quantum evolutions including a) the reversible dynamics as determined by the Schrödinger equation; b) state changes (reductions) due to repeated quantum measurements.

The quantum dynamics of a single system is translated into the classical description proposed in the previous section without great difficulties. The complex unit state vector ψ defines the coordinates $\left[\psi_j, \bar{\psi}_j\right]$ on the variety of pure states Ω', and the Schrödinger equation $\frac{d\psi}{dt} = -iH\psi$ with the Hamiltonian H generates a flow $\left\{T'_t\right\}$ on Ω' which can be written in the coordinates $\left[\psi_j, \bar{\psi}_j\right]$ as a classical Hamiltonian system

$$\frac{d\psi_j}{dt} = -i\frac{\partial}{\partial\bar{\psi}_j}\Gamma\left(\psi, \bar{\psi}\right),$$
$$\frac{d\bar{\psi}_j}{dt} = i\frac{\partial}{\partial\psi_j}\Gamma\left(\psi, \bar{\psi}\right); \quad j = 1, \ldots, n,$$

with the Hamiltonian function $\Gamma\left(\psi, \bar{\psi}\right) = (\psi|H\psi)$. Putting $T''_t\omega'' = \omega''$, we obtain a flow $T_t = T'_t \times T''_t$ defining the dynamics of the corresponding classical model.

The question of describing repeated measurements and reductions is more complicated. A genuine hidden variable model should completely reduce measurements to observations and in this way resolve the enigmatic "measurement problem". A lot of attention was paid to this problem in physical and philosophical literature (see, *e.g.* [71, 138, 171, 186, 198]) while our remarks will be rather sketchy.

First we describe repeated measurements in quantum mechanics. Let a measurement $\hat{\boldsymbol{E}} = \{\hat{E}_i\}$ and then a measurement $\hat{\boldsymbol{F}} = \{\hat{F}_j\}$ be performed over a quantum system in the initial state $\hat{S}$. The formula (2.8) is insufficient in this case and an additional assumption is required. The "repeatability hypothesis" of von Neumann leads to the following expression for the joint probability distribution of the repeated quantum measurements

$$\mu_{\hat{S}}^{\hat{\boldsymbol{E}},\hat{\boldsymbol{F}}}(i,\, j) = \mathrm{Tr}\hat{S}\hat{E}_i\hat{F}_j\hat{E}_i. \tag{3.18}$$

If $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{F}}$ are compatible, then these probabilities do not depend on the order of measurements and are equal to the probabilities for the joint measurement $\mathrm{Tr}\hat{S}\hat{E}_i\hat{F}_j$. If $\hat{\boldsymbol{E}} = \hat{\boldsymbol{F}}$ then $\mu_{\hat{S}}^{\hat{\boldsymbol{E}},\hat{\boldsymbol{F}}}(i,\, j) = 0$ for $i \neq j$, which explains the term "repeatability hypothesis" . In the general case of incompatible measurements $\mu_{\hat{S}}^{\hat{\boldsymbol{E}},\hat{F}} \neq \mu_{\hat{S}}^{\hat{\boldsymbol{F}},\hat{\boldsymbol{E}}}$, reflecting impossibility to give an objective meaning to the joint probability distribution. Notice also that for a given observable $\hat{X}$ the distribution (3.18) in general depends on the choice of its measurement $\hat{\boldsymbol{E}}$.

The relation (3.18) can be directly generalized to the case of arbitrary number of repeated measurements. For example, for three measurements

$$\mu_{\hat{S}}^{\hat{\boldsymbol{E}},\hat{\boldsymbol{F}},\hat{\boldsymbol{G}}}(i,\, j,\, k) = \mathrm{Tr}\hat{S}\hat{E}_i\hat{F}_j\hat{G}_k\hat{F}_j\hat{E}_i. \tag{3.19}$$

The unpublished von Neumann's argument mentioned in the remark of Wigner was just about the possibility of reproducing the statistics of repeated quantum measurements. That remark concerns the case of spin-1/2 particle which is described by two-dimensional Hilbert space $\mathcal{H}$ and the spin components are $2 \times 2-$matrices.

"... Rather, Von Neumann often discussed the measurement of the spin component of a spin-J particle in various directions. Clearly, the probabilities for the two possible outcomes of a single such measurement can be easily accounted for by hidden variables (see, *e.g.*, the rest of the present section or the more specific discussion on page 448 of Bell's article, Reference 2). However, Von Neumann felt that this is not the case for many consecutive measurements of the spin component in various different directions. The outcome of the first such measurement restricts

the range of values which the hidden parameters must have had before that first measurement was undertaken. The restriction will be present also after the measurement so that the probability distribution of the hidden variables characterizing the spin will be different for particles for which the measurement gave a positive result from that of the particles for which the measurement gave a negative result. The range of the hidden variables will be further restricted in the particles for which a second measurement of the spin component, in a different direction, also gave a positive result. A great number of consecutive measurements will select particles the hidden variables of which are all so closely alike that the spin component has, with a high probability, a definite sign in all directions. However, according to quantum mechanical theory, no such state is possible. Schrödinger raised the objection against this argument that the measurement of a spin component in one direction, while possibly specifying some hidden variables, may restore a random distribution of some other hidden variables. It is this writer's impression that Von Neumann did not accept Schrödinger's objection. His point was that the objection presupposed hidden variables in the apparatus used for the measurement. Von Neumann's argument needs to assume only two apparata, with perpendicular magnetic fields, and a succession of measurements alternating between the two apparata. Eventually, even the hidden variables of both apparata will be fixed by the outcomes of many subsequent measurements of the spin component in their respective directions so that the whole system's hidden variables will be fixed. Von Neumann did not publish this apparent refutation of Schrödinger's objection."

Let us see how one should change the state in a hidden variable model of Section 2.2 to reproduce the outcomes of the repeated measurements. For simplicity we restrict to pure states and maximal measurements. If in the state ω' one performs the measurement $\hat{\boldsymbol{E}}$ and then the measurement $\hat{\boldsymbol{F}}$ then according to (3.18) the probability of the outcome (i, j) is equal to

$$\mu_{\omega'}^{\hat{\boldsymbol{E}},\hat{\boldsymbol{F}}}(i,\, j) = |\langle \psi_{\omega'} | e_i \rangle|^2 \left| \langle e_i | f_j \rangle \right|^2 ,$$

where $\{e_i\}$, $\left\{f_j\right\}$ are the bases determining the measurements $\hat{\boldsymbol{E}}$, $\hat{\boldsymbol{F}}$. One can see that such values of the probabilities will be ensured if after the first measurement the initial point $\omega = \left(\omega', \omega''\right)$ of the phase space will go into the point $(\hat{E}_i, \omega'')$ under the condition that the outcome of the first measurement was i, *i.e.* if $\lambda_{\hat{\boldsymbol{E}}} = \pi_{\hat{\boldsymbol{E}}}(\omega'')$ satisfies the inequality

$$\sum_{k=1}^{i-1} |\langle \psi_{\omega'} | e_i \rangle|^2 < \lambda_{\hat{\boldsymbol{E}}} \leq \sum_{k=1}^{i} |\langle \psi_{\omega'} | e_i \rangle|^2 .$$

The change of the component ω' is thus a controlled Markov process. We also have to describe the change $\omega'' \to \tilde{\omega}''$. If the second measurement is the same as the first one, then (3.18) implies that its outcome should coincide with the outcome of the first measurement. To ensure this it should be $\pi_{\hat{E}}(\tilde{\omega}'') = \pi_{\hat{E}}(\omega'')$ *i.e.* the "roulette" corresponding to the measurement $\hat{E}$ should preserve its state after the first measurement. It is not important what then happens with the other "roulettes" unless they preserve the uniform distribution.

Consider, however, three repeated measurements $\hat{E}, \hat{F}, \hat{E}$. According to (3.19) the probability of the outcome (i, j, k) is equal to

$$\mu_{\omega'}^{\hat{E},\hat{F},\hat{E}}(i, j, k) = |\langle \psi_{\omega'}|e_i\rangle|^2 \left|\langle e_i|f_j\rangle\right|^2 \left|\langle f_j|e_k\rangle\right|^2 .$$

In order to obtain such an expression in the classical model, it is necessary that after the second measurement the probability distribution of the "roulette" corresponding to the measurement $\hat{E}$ should be completely renewed, *i.e.* $\pi_{\hat{E}}(\omega'')$ should become a random variable independent of the previous values. Since $\hat{E}$ is arbitrary, this leads to the following rule of state change for the roulettes: after a measurement $\hat{F}$ the states of all the roulettes $\lambda_{\hat{E}} = \pi_{\hat{E}}(\omega'')$ with $\hat{E} \neq \hat{F}$ are completely renewed while $\lambda_{\hat{F}} = \pi_{\hat{F}}(\omega'')$ preserves its value. This completely corresponds to the Schrödinger's remark. One can check that such a prescription allows to reproduce probabilities for all possible repeated maximal measurements. To include not necessarily maximal measurements one has to extend further the collection of "roulettes" in the classical descriptions.

The hidden variable ω have the two components, the first of which ω' can be considered as the characteristic of the system itself. The question – to what corresponds ω'' – properly to the system or to the measuring devices is very interesting although it does not give direct arguments contradicting to Schrödinger's prescription. On one hand it appears natural to consider $\lambda_{\hat{E}} = \pi_{\hat{E}}(\omega'')$ as parameters of the measuring device $\hat{E}$. Interaction of $\hat{E}$ with the system fixes the parameters of the device and $\lambda_{\hat{E}}$ does not change after the repeated measurement $\hat{E}$. On the other hand, the totality of all possible measurements $\hat{E}$ which can be performed over the system and the corresponding collection of the parameters $\omega'' = \prod_{\hat{E}} \lambda_{\hat{E}}$ can be regarded as the characteristic of the whole system, taking into account that measurement $\hat{F}$ affects all the parameters $\lambda_{\hat{E}}, \hat{E} \neq \hat{F}$. Here it is appropriate to remind of Bohr's saying that in quantum physics "the interaction between the measuring instruments and the objects forms an integral part of the phenomena" [155, page 4]. Anyhow we have seen that there exists a way to define the stochastic rule of state change of a classi-

cal system which allows to reproduce probabilities of repeated quantum measurements.

In the case of spin-1/2 system this was observed by Clauser [159] in a discussion which followed the publication of Wigner's paper. Having considered the Bell's model, Clauser proposed a simple rule for state change due to measurement which in fact uses Schrödinger's prescription and reproduces probabilities of all repeated measurements. In his answer Wigner raised the following objection which is applicable also to our model: a hidden variable model should give explanation for state changes in "mechanistic" *i.e.* deterministic but not in stochastic terms. A permanent flow of randomness which is necessary for partial renewal of the distribution of ω'' after each measurement is in the conflict with Liouville type theorem concerning conservation of the phase volume. To describe such a permanent renewal, a radical increase of dimensionality is required. Let us explain this on the simplest example of a sequence of independent random variables $\{X_i\}$ with values in $\mathbb{R}$. To describe it as a dynamical system with an invariant measure one has to pass to the trajectory space $\mathbb{R}^{\mathbb{Z}}$. The stochastic renewal of the state of one dimensional system can be represented as a mechanical evolution only in the infinite dimensional space of sequences.

The comment of Wigner is concluded with explanation that "all schemes of hidden parameters which either Von Neumann himself, or anyone else whom he knew, could think of and which reproduced the probabilities of outcomes of several successive measurements of the spin directions, had some feature which made it unattractive, in fact unreasonable" and "this was the true reason for his (von Neumann's) conviction of the inadequacy of the theories of hidden variables".

Thus the requirement of reproducing the results of repeated measurements in a hidden variable theory seems to lead to unappealing constructions which evoke negative emotions both from physicists and mathematicians. However no definite results which would close search in this direction were obtained. After publication of Bell's works the edge of investigations shifted to a different aspect of the hidden variable problem related to the description of composite systems[4].

3.4. Composite systems, EPR paradox and the Bell inequality

The most important physical applications of the mathematical apparatus of quantum mechanics concern the special features of *interactions* of

[4] Recently there was a revival of interest to noncontextual hidden variables due to discovery of tests for noncontextuality, including experimental work (see [169, 183] and the references therein).

microobjects which have no place in the classical physics. Nevertheless the first step in the quantum mechanical description interaction follows the classical pattern. Namely, one starts from a collection of noninteracting "single" components and then the interaction is described in terms of this composite system. Thus speaking about interaction one deals not so much with a model for single system, but rather with a category of these models equipped with the operation of product defining the rule for making a composite system. In classical mechanics one considers (symplectic) phase spaces with the operation of Cartesian product, while in quantum mechanics – the Hilbert spaces with the operation of tensor product.

In our discussion of this operation, for simplicity, we will not take into account additional complications related to possible indistinguishability of particles. One calls by tensor product of the vector $\psi_1 \in \mathcal{H}_1$ with the components $[\psi_1^i]$ and the vector $\psi_2 \in \mathcal{H}_2$ with the components $[\psi_2^j]$ the vector $\psi_1 \otimes \psi_2$ with the components $[\psi_1^i \psi_2^j]$ which is conveniently represented by the matrix. The space $\mathcal{H}_1 \otimes \mathcal{H}_2$ consists of all possible linear combinations (superpositions) of vectors of the form $\psi_1 \otimes \psi_2$, *i.e.* of all complex matrices $[\psi^{ij}]$. Consider a pure state of the composite system defined by a unit vector $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$. There is a lot of vectors which cannot be written in the form $\psi_1 \otimes \psi_2$ corresponding to the case where the first and the second components are in the uniquely defined pure states. Most of the vectors ψ are superpositions of such vectors. For a superposition it is not possible to separate uniquely the first and the second components of the composite system. These unfactorizable *entangled* states represent a holistic entity in which the components exists, as one is accustomed to say, virtually. This reflects the property of quantum *nonseparability*.

At a first glance it looks unclear how such a merging of the components could happen before the interaction of the subsystems. The explanation is that preparation of an entangled pure state of a composite system assumes preliminary interaction between the components. Indeed, any vector ψ can be obtained from factorizable one $\psi_1 \otimes \psi_2$ by action of a unitary operator U in $\mathcal{H}_1 \otimes \mathcal{H}_2$ so that $\psi = U(\psi_1 \otimes \psi_2)$. Then the preparation consists of the interaction described by the operator U. As distinct from the classical mechanics where state preparation can be described in purely kinematical terms, preparation of many quantum mechanical states requires a dynamical interaction.

Consider from this point of view the hidden variable model of Section 2.2. Since the set of pure states of the composite system Ω' is larger than Cartesian product $\Omega'_1 \times \Omega'_2$, where Ω'_j is the set of pure states of the $j-$th component, the phase space of the classical description of the

composite system will be larger than the product of phase spaces for the components: $\Omega_1 \times \Omega_2 \subsetneq \Omega$. Therefore this classical description is not a correspondence between the categories of classical and quantum system preserving the operation of forming the composite systems.

Moreover, it appears that there is no way to establish such a correspondence. In any classical description of a composite quantum system the variables corresponding to observables of the components are necessarily entangled in the way unusual for classical subsystems. To give exact formulation, notice that observables corresponding to the first and the second components have the form $\hat{X} = \hat{X}_1 \otimes \hat{I}_2$, $\hat{Y} = \hat{I}_1 \otimes \hat{Y}_2$, where $\hat{X}_1$ is operator in $\mathcal{H}_1$, $\hat{Y}_2$ operator in $\mathcal{H}_2$ and $\hat{I}_j$ are the unit operators in $\mathcal{H}_j$. Certainly $\hat{X}\hat{Y} = \hat{Y}\hat{X}$ so that $\hat{X}, \hat{Y}$ are compatible, but they have even stronger property of algebraic independence: if $f(\hat{X}) = g(\hat{Y})$ then $f(\hat{X}) = g(\hat{Y}) = c\hat{I}$ (such a property is not satisfied *e.g.*by the commuting variables $\hat{X}$ and $\hat{X}^2$ with $\hat{X}^2 \neq c\hat{I}$).

Proposition 3.6. *There is no classical description for a composite quantum system in the space $\mathcal{H}_1 \otimes \mathcal{H}_2$ satisfying* (E.1), (X.2) *and the following separability condition:*

(X.6) *for any observables $\hat{X}_1, \dots, \hat{X}_n$ of the first subsystem and $\hat{Y}_1, \dots, \hat{Y}_m$ of the second subsystem there are random variables $X_1, \dots, X_n$; $Y_1, \dots, Y_m$ such that $X_i \to \hat{X}_i$, $Y_j \to \hat{Y}_j$ and $X_i Y_j \to \hat{X}_i \hat{Y}_j$.*

It will be shown that there is no such a description already for $n = m = 2$. Notice that the conditions (X.0) or (S.0) of one-to-one correspondence are not required here. The proposition means that while for any pair $\hat{X}, \hat{Y}$, where $\hat{X}$ (respectively $\hat{Y}$) refers to the first (respectively to the second) system, it is always possible to find X, Y such that $X \to \hat{X}$, $Y \to \hat{Y}$ and $XY \to \hat{X}\hat{Y}$ (this follows from the possibility to satisfy the product rule (X.4)), it is impossible to do it in the way that Y would be the same for all choices of $\hat{X}$ and X – the same for all $\hat{Y}$. The expression $\hat{X}\hat{Y}$ enters into the correlation

$$\left\langle \hat{X}\hat{Y} \right\rangle = \mathrm{Tr}\hat{S}\hat{X}\hat{Y} \tag{3.20}$$

of the outcomes of joint measurements of $\hat{X}$ and $\hat{Y}$. Thus to reproduce the quantum mechanical correlations between the subsystems, a hidden variable theory should possess the following strange property: the observation method for an observable of the second subsystem $\hat{Y}$ should necessarily depend on which $\hat{X}$ is observed over the first subsystem.

Proof. The proof is based on *Clauser-Horne-Shimony-Holt inequality* which is a version of the famous Bell inequality [160]. Assume that one

can satisfy the condition (X.6) with $n = m = 2$ and let us prove that for all observables $\hat{X}_1$, $\hat{X}_2$ of the first component and all observables $\hat{Y}_1$, $\hat{Y}_2$ of the second, with the outcomes in $[-1, 1]$, the correlations (3.20) satisfy

$$\left|\left\langle \hat{X}_1\hat{Y}_1\right\rangle + \left\langle \hat{X}_1\hat{Y}_2\right\rangle + \left\langle \hat{X}_2\hat{Y}_1\right\rangle - \left\langle \hat{X}_2\hat{Y}_2\right\rangle\right| \leq 2. \tag{3.21}$$

By the assumptions (E.1), (X.6) it is sufficient to prove this for correlations of classical random variables $X_1(\omega)$, $X_2(\omega)$, $Y_1(\omega)$, $Y_2(\omega)$ given by

$$\langle XY\rangle = \int_\Omega S(d\omega)\, X(\omega)\, Y(\omega).$$

But taking into account the spectral condition (X.2) one has $|X_j| \leq 1$, $|Y_k| \leq 1$; $j, k = 1, 2$, which implies

$$|X_1Y_1+X_1Y_2+X_2Y_1-X_2Y_2| \leq |Y_1+Y_2|+|Y_1-Y_2| \leq 2\max\{|Y_1|,|Y_2|\} \leq 2,$$

hence $-2 \leq X_1Y_1 + X_1Y_2 + X_2Y_1 - X_2Y_2 \leq 2$, whence, taking the expectation, one obtains the required inequality.

It remains to show that in any composite quantum system one can find observables $\hat{X}_1$, $\hat{X}_2$; $\hat{Y}_1$, $\hat{Y}_2$ and the state $\hat{S}$ violating the inequality (3.21). For this consider first the system of two distinguishable spin-1/2 particles so that $\mathcal{H}_1$ and $\mathcal{H}_2$ are two-dimensional. Denote by $\hat{X}(\vec{a})$ the spin observable in the direction $\vec{a} = (a_x, a_y, a_z)$ for the first particle and by $\hat{Y}(\vec{b})$ – the spin observable in the direction $\vec{b}$ for the second particle. In the basis

$$|\uparrow\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |\downarrow\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

one has

$$\hat{X}(\vec{a}) = \begin{bmatrix} a_z & a_x - ia_y \\ a_x + ia_y & -a_z \end{bmatrix},$$

and similarly for $\hat{Y}(\vec{b})$. Consider the entangled pure state $\hat{S}_\psi$ of the composite system given by the vector

$$|\psi\rangle = \frac{1}{\sqrt{2}}\big[|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle\big]. \tag{3.22}$$

A simple calculation shows that

$$\langle\psi|\sigma(\vec{a}) \otimes \sigma(\vec{b})|\psi\rangle = -\vec{a}\cdot\vec{b}. \tag{3.23}$$

If one chooses the four vectors $\vec{a}_j, \vec{b}_k$, $(j, k = 1, 2)$ as shown on Fig. 6.3, then the correlations between $\hat{X}(\vec{a}_j)$ and $\hat{Y}(\vec{b}_k)$ give the value $2\sqrt{2}$ for the left side of (3.21), which breaks the inequality.

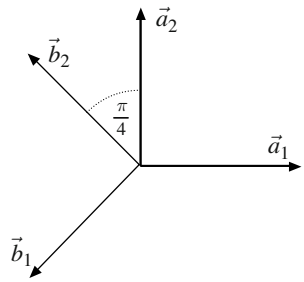

Figure 6.3. Choice of $\vec{a}_j$ and $\vec{b}_k$.

For an arbitrary composite quantum system $\mathcal{H}_1 \otimes \mathcal{H}_2$ one can always take the two-dimensional subspaces of $\mathcal{H}_1$ and $\mathcal{H}_2$ to build this construction which proves the Proposition in the general case. □

The underlying case of two spatially separated particles has quite important physical consequences. This was realized already in 1935 after publication of the paper of Einstein, Podolsky and Rosen (EPR) [162] in the subsequent discussion on completeness of quantum mechanics (considering entangled state of two spin-1/2 particles was proposed later by Bohm). EPR believed that their argument shows "incompleteness" of quantum mechanical description. The argument of Bell shifts the accent substantially: if the quantum mechanical description is correct then any attempt to "complete" it with hidden variables leads necessarily to contradiction with the physical principle of *locality*. Assume that $\mathcal{H}_1$ and $\mathcal{H}_2$ describe spin degrees of freedom of two spin-1/2 particles located in spatially separated domains. Quantum mechanics implies existence of states for such a system in which the spins are entangled as in (3.22). Moreover, such a state in principle can be realized experimentally as a product of decay of a spin-zero system.

Now assume that a joint measurement is performed for the spin of the first particle in the direction $\vec{a}$ and the spin of the second particle in the direction $\vec{b}$. After a long series of independent repetitions the experimenter computes statistical estimates for the correlation between $\hat{X}(\vec{a})$ and $\hat{Y}(\vec{b})$. Consider the following three statements:

I the correct values for the correlations are given by the quantum mechanical formula (3.20);

II there is a classical description of the composite quantum system satisfying the spectral condition *i.e.* preserving the "objective values" of observables;

III "the real factual situation of the system S_1 is independent of what is done with the system S_2, which is spatially separated from the former".

The last property is the *Einstein locality* [162] or *separability* [163] related to the principle "no instantaneous action at a distance". Notice that the property (III) has an unambiguous meaning only in the classical picture *i.e.* under the condition (II). The classical description satisfying the requirement (III) is called a local hidden variable theory or "local realism". The Bell argument shows that the properties (I)-(III) are incompatible *i.e. a local hidden variable theory reproducing the statistical predictions of quantum mechanics does not exist*. This can be considered as modern interpretation of the "EPR paradox" .

Moreover, the Bell type inequalities in principle open a possibility for experimental test: quantum mechanics versus local hidden variable theory. A practical implementation of such experiments is however aggravated by several possible loopholes and requires great efforts. The famous Aspect's experiment performed in 1981-1982 showed the agreement with quantum mechanical formula for correlations. After that there was a continuing series of experiments with similar conclusions while a couple of them claimed possible agreement with the Bell inequality [160, 194]. However the question is not closed because of the high price of rejecting the "realistic" description of the Nature [163].

We are not speaking of defenders of the naive realism who would be happy with a pictorial image of the microworld as something similar to the world directly accessible to the human perception where mechanistic idealizations like a material point still have sense. They continue attempts to find a gap in the argument leading to the conclusion that any local hidden variable theory cannot reproduce the statistical predictions of quantum theory concerning correlations between the parts of a composite quantum system. It should be noticed that while the Bell inequality as such is elementary, the logic of its application in the hidden variable problem is far from trivial; any critique of Bell's argument with its subsequent refinements, however it might seem sophisticated, sooner or later was found based on a misunderstanding. A critical survey of recent discussions and different opinions is given in the article [164] where in particular a proof is given of a Bell inequality taking into account finiteness of the sample and possibility of arbitrary local correlations between the subsequent experiments.

What is really disturbing is the apparent impossibility of a peaceful coexistence of quantum mechanics and "local realistic" theory such as general relativity (which was in fact the motivation of Einstein's concern).

Admitting a local hidden variable theory would mean that quantum mechanics is limited at least in some of its most basic predictions which would be quite an extraordinary development. Some compromise has to be found if one believes in the possibility of unified quantum theory of interactions including gravitation.

On the other hand, macroscopic or at least mesoscopic manifestations of the entanglement are critically important for the future technological implementations of the new quantum information processing protocols including quantum computation.

Therefore much attention is paid to close possible logical loopholes in the actual Bell-test experiments. The most significant one is the *detection loophole* which still exists after almost 40 years of experimental research. Soon after appearance of Bell's work it was recognized that the inequality can be violated in local hidden variable models with post-selection with a positive probability of failure for the particle detection if the "no detection" outcomes are just neglected. A simple explicit local hidden variable model for such correlation experiment with two spin-1/2 particles was demonstrated by N. Gisin and B. Gisin [167] implying in particular that experiments with the detector efficiency below 75% cannot be considered as decisive. Therefore researchers continue to improve experimental techniques and search for the more sensitive inequalities [187]. Nowadays theoretical study of the Bell-type inequalities providing quantitative boarders between "classical" and "quantum" is one of the topics in the modern quantum information theory with applications to quantum cryptography, entanglement detection, multipartite interactive proof systems, communication complexity etc., see [181] and the references therein. However these exciting subjects are already beyond the scope of the present essay.

References

[1] L. ACCARDI, A. FRIGERIO and J. T. LEWIS, "Quantum Stochastic Processes", Publ. RIMS Kioto University **18** (1982), 97-133.

[2] N. I. AKHIEZER and I. M. GLAZMAN, "Theory of Linear Operators in Hilbert Space", Ungar, New York, 1963; 2nd Russian ed. Nauka, Moscow, 1966.

[3] G. R. ALLCOCK, *The time of arrival in quantum mechanics*, Ann. Physics **53** (1969), 253–348.

[4] E. M. ALFSEN, "Compact Convex Sets and Boundary Integrals", Ergebnisse der Mathematik 57, Springer-Verlag, Berlin, 1971.

[5] W. O. AMREIN, *Localizability for particles of mass zero*, Helv. Phys. Acta **42** (1) (1969), 149–190.

[6] V. BERGMANN, *On unitary ray representations of continuous groups*, Ann. of Math., **59** (1) (1954), 1–46.

[7] V. BARGMANN, *On a Hilbert space of analytic functions and an associated integral transform*, I, Comm. Pure Appl. Math. **14** (3) (1961), 187–214.

[8] V. BARGMANN, *Note on Wigner's theorem on symmetry operations*, J. Math. Phys. **5** (1964), 862–868.

[9] H. BAUER, *Minimalstellen von Functionen und Extremalpunkte*, Arch. Math. **9** (1958), 389–393.

[10] F. J. BELINFANTE, "A Survery of Hidden Variables Theories", Pergamon Press, New York, 1973.

[11] J. BELL, *On the problem of hidden variables in quantum mechanics*, Rev. Modern Phys. **38** (1966), 447–452.

[12] F. A. BEREZIN, "The Method of Second Quantization", Nauka, Moscow, 1965, (R).

§ The year of the first edition is indicated in brackets for some monographs. (R) indicates a work in the Russian language.

[13] D. I. BLOKHINTZEV, "Fundamentals of Quantum Mechanics (Vysšaja shkola, Moscow, 1961, (R).
[14] D. I. BLOKHINTZEV, "Fundamental Questions of Quantum Mechanics, Nauka, Moscow, 1966, (R).
[15] N. N. BOGOLJUBOV, "Lectures on the Theory of Symmetry of Elementary Particles, Moscow State University, Moscow, 1966, (R).
[16] N. N. BOGOLJUBOV, A A. LOGUNOV and I. T. TODOROV, "Introduction to Axiomatic Quantum Field Theory, Benjamin, New York, 1975; Russian ed.: Nauka, Moscow, 1969.
[17] D. BOHM, "Quantum Theory", Prentice-Hall, Englewood Cliffs, NJ, 1951.
[18] N. BOHR, "Atomic Physics and Human Knowledge, Wiley, New York, 1958.
[19] V. B BRAGINSKIJ and YU. I. VORONTZOV, *Quantum-mechanical limitations in macroscopic experiments and contemporary experimental technique*, Uspehi Fiz. Nauk **114** (1974), 41–53 (R).
[20] P. CARRUTHERS and M. M. NIETO, *Phase and angle variables in quantum mechanics*, Rev. Modern Phys. **40** (2) (1968), 411–440.
[21] D. CASTLER, *The C^*-algebras of a free Boson field*, Comm. Math. Phys. **1** (1) (1965), 14–48.
[22] N. N. CHENTZOV, "Statistical Decision Rules and Optimal Inference", Translations of Mathematical Monographs; **53**, American Mathematical Society, 1982; (1972) (R).
[23] H. CRAMER, "Mathematical Methods of Statistics", Princeton Univ. Press, Princeton, NJ, 1946.
[24] C. D. CUSHEN and R. L. HUDSON, *A quantum mechanical central limit theorem*, J. Appl. Probab. **8** (1971), 454–469.
[25] A. DANERI, A. LOINGER and G. M. PROSPERI, *Quantum theory of measurement and the ergodicity conditions*, Nuclear Phys. **33** (1962), 297–319.
[26] E. B. DAVIES, *On repeated measurements of continuous observables in quantum mechanics*, J. Funct. Anal. **6** (1970), 318–346.
[27] E. B. DAVIES, "Quantum Theory of Open Systems", Academic Press, London, 1976.
[28] E. B. DAVIES, *Quantum communication systems*, IEEE Trans. Inform. Theory IT **23** (7) (1977), 530-534.
[29] E. B. DAVIES and J. T. LEWIS, *An operational approach to quantum probability*, Comm. Math. Phys. **17** (3) (1970), 239–260.
[30] P. A. M. DIRAC, "The Principles of Quantum Mechanics", 4th ed., Oxford Univ. Press, Oxford, 1958, (1930).

[31] N. DUNFORD and J. T. SCHWARTZ, "Linear Operators", part 1, Interscience, New York, 1958.

[32] H. EKSTEIN and A. J. F. SIEGERT, *On a reinterpretation of decay experiments*, Ann. Physics **68** (1971), 509–520.

[33] G. G. EMCH, "Algebraic Methods in Statistical Mechanics and Quantum Field Theory", Wiley-Interscience, New York, 1972.

[34] T. S. FERGUSON, "Mathematical Statistics: a Decision Theoretic Approach", Academic Press, New York, 1967.

[35] R. P. FEYNMANN, *Concept of probability in quantum mechanics*, Proc. 2nd Berkeley Symp. on Math. Stat. and Probab. Berkeley (1951), 533–541.

[36] R. P. FEYNMANN, R. B. LEIGHTON and M. SANDS, "Quantum Mechanics", Addison-Wesley, Reading, MA, 1965.

[37] V. A. FOCK, "Fundamentals of Quantum Mechanics, Nauka, Moscow, 1976, 1932, (R).

[38] V. A. FOCK, *On the interpretation of quantum mechanics*, Uspehi Fiz. Nauk **62** (4) (1957), 461–474, (R).

[39] D. GABOR, *Communication theory and physics*, Phil. Mag. **41** (1950), 1161-1197.

[40] I. M. GELFAND and N. YA. VILENKIN, "Some Applications of Harmonic Analysis, Fitzmatgiz, Moscow, 1961, (R).

[41] I. M. GELFAND, R. A. MINLOS and Z. YA. SHAPIRO, "Representations of the Rotation Group and the Lorentz Group", Fizmatgiz, Moscow, 1958, (R).

[42] R. J. GLAUBER, *The quantum theory of optical coherence*, Phys. Rev. **130** (1963), 2529–2539.

[43] B. V. GNEDENKO, "The Theory of Probability", Chelsea, New York, 1962; 5rh Russian ed. Nauka, Moscow, 1969.

[44] J. P. GORDON and W. LOUISELL, *Simultaneous measurements of non-commuting observables*, In: "Physics of Quantum electronics", P. Kelley, M. Lax and B. Tannenwold (eds.), McGraw-Hill, New York, 1966, 833–840.

[45] V. GORINI, A. FRIGERIO, M. VERRI, A. KOSSAKOWSKI and E. C. G. SUDARSHAN, *Properties of quantum Markovian equantions*, Rep. Math. Phys. **13** (3) (1978), 149–173.

[46] S. GUDDER, *Convex structures and operational quantum mechanics*, Comm. Math. Phys. **29** (3) (1973), 249–264.

[47] P. R. HALMOS, "Finite-Dimensional Vector Spaces", 2nd ed. Van Nostrand, Princeton, NJ, 1958.

[48] P. R. HALMOS, "A Hilbert Space Problem Book", Van Nostrand, Princeton, NJ, 1967.

[49] A. HARTKÄMPER, *Axiomatics of preparing and measuring processes*, In: "Foundations of Quantum Mechanics and Ordered Linear Spaces", Lecture Notes in Phys. **29** (1974), 107–115.

[50] C. W. HELSTROM, *Detection theory and quantum mechanics*, Inform. and Control **10** (3) (1967), 254–291.

[51] C. W. HELSTROM, *Minimum mean-square error estimation in quantum statistics*, Phys. Let. **25A** (1967), 101–102.

[52] C. W. HELSTROM, *Estimation of a displacement parameter of a quantum system*, Internat. J. Theoret. Phys. **11** (6) (1974), 357–378.

[53] C. W. HELSTROM, "Quantum Detection and Estimation Theory", Academic Press, New York, 1976.

[54] K. HEPP, *Quantum theory of measurement and macroscopic observables*, Helv. Phys. Acta **45** (1972), 237–248.

[55] A. S. HOLEVO, *On quantum characteristic functions*, Problemy Peredači Informacii **6** (4) (1970), 44–48, (R).

[56] A. S. HOLEVO, *Towards a mathematical theory of quantum communication channels*, Problemy Peredači Informacii **8** (1) (1972), 63–71, (R).

[57] A. S. HOLEVO, *An analog of the statistical decision theory in the noncommutative probability theory*, Trudy Moscow, Mat. Obšč,**26**(1972), 133–149, (R), AMS Trans. (1974).

[58] A. S. HOLEVO, *Statistical problems in quantum physics*, 2nd Japan-USSR Symp. on Probability Theory, Kyoto **1** (1972), 22–40.

[59] A. S. HOLEVO, *Statistical decision theory for quantum systems*, J. Multivariate Anal. **3** (4) (1973), 337–394.

[60] A. S. HOLEVO, *Information-theoretic aspects of quantum measurement*, Problemy Peredači Informacii **9** (2) (1973), 31–42, (R).

[61] A. S. HOLEVO, *Optimal quantum measurements*, Teoret. Mat. Fiz. **13** (3) (1973), 184–199, (R).

[62] A. S. HOLEVO, *Some statistical problems for quantum Gaussian states*, IEEE Trans. Inform. Theory IT **21** (5) (1975), 533–543.

[63] A. S. HOLEVO, *Noncommutative analogs of the Cramer-Rao inequality in the quantum measurement theory*, 3nd Soviet-Japan Symp. on Probability Theory, Trashkent, 1975, Lecture Notes in Math. **550** (1976), 194–222.

[64] A. S. HOLEVO, *Investigations in the general theory of statistical decision*, Proceedings of the Steklov Institute of Mathematics **124** (1976), (R), AMS Transl. **3** (1978).

[65] A. S. HOLEVO, *Commutation superoperator of a state and its applications in the noncommutative statistics*, Rep. Math. Phys. **12** (2) (1977), 251–271.

[66] A. S. HOLEVO, *Estimation of a shift parameter of a quantum state*, Rep. Math. Phys. **13** (3) (1978), 287–307.

[67] A. S. HOLEVO, *Covariant measurements and uncertainty relations*, Rep. Maht. Phys. **16** (3) (1979), 385–400.

[68] W. HUNZIKER, *A note on symmetry operations in quantum mechanics*, Helv. Phys. Acta **45** (2) (1972), 233–236.

[69] R. S. INGARDEN, *Quantum information theory*, Rep. Math. Phys. **10** (1) (1976), 43–72.

[70] E. INÖNU and E. P. WIGNER, *Representations of the Galilei group*, Nuovo Cimento **9** (1952), 705–718.

[71] J. M. JAUCH, "Foundations of Quantum Mechanics", Addison-Wesley, Reading, MA, 1968.

[72] J. M. JAUCH and C. PIRON, *Generalized localizability*, Helv. Phys. Acta **45** (1967), 559-570.

[73] D. JUDGE, *On the uncertainty relation for angle variables*, Nuovo Cimento **31** (1964), 332–340.

[74] R. V. KADISON, *Transformation of states in operator theory and dynamics*, Topology **3** (suppl.) (1965), 177–198.

[75] A. A. KIRILLOV, "Elements of the Representation Theory", Nauka, Moscow, 1978, (R).

[76] J. R. KLAUDER and E. C. G. SUDARSHAN, "Fundamentals of Quantum Optics", Benjamin, New York, 1968.

[77] S. KOCHEN and E. P. SPECKER, *The problem of hidden variables in quantum mechanics*, J. Math. Mech. **17** (1) (1967), 59–88.

[78] A. N. KOLMOGOROV, "Foundations of the Theory of Probability", Chelsea, New York, 1950, (1933).

[79] A. N. KOLMOGOROV, "Theory of the Information Transmission", USSR Academy of Sciences Press, Moscow, 1957, (R).

[80] K. KRAUS, *Operations and effects in the Hilbert space formulation of quantum theory*, In: "Foundations of quantum mechanics and ordered linear space", Lecture Notes in Phys. **29** (1974), 206–229.

[81] K. KRAUS and J. SCHRÖTER, *Expectation values of unbounded observables*, Internat. J. Theoret. Phys. **7** (6) (1973), 431–442.

[82] U. KRAUSE, *The inner orthogonality of convex sets in axiomatic quantum mechanics*, In: "Foundations of quantum mechanics and ordered linear spaces", Lecture Notes in Phys. **29** (1974), 269–280.

[83] N. S. KRYLOV and V. A. FOCK, *On the main two interpretations of the "time-energy" uncertainty relation*, JETP **17** (2) (1947), 93–107, (R).

[84] L. LANZ, L. A. LUGIATO and G. RAMELLA, *On the quantum mechanical treatment of decaying non-relativistic systems*, Internat. J. Theoret. Phys. **8** (5) (1973), 341–352.

[85] G. LINDBLAD, *Entropy, information and quantum measurements*, Comm. Math. Phys. **33** (4) (1973), 305–322.

[86] J. L. LOPES and M. PATY (eds.), "Quantum Mechanics: a Half Century Later", Reidel Publ. Company, Doredrecht-Holland, 1977.

[87] W. LOUISELL, *Amplitude and phase uncertainty relations*, Phys. Lett. **7** (1963), 60–61.

[88] W. LOUISELL, "Radiation and Noise in Quantum Electronics", McGraw-Hill, New York, 1964.

[89] G. LOUPIAS and S. MIRACLE-SOLÉ, *C^*-algèbre des systèmes canoniques*, Comm. Math. Phys. **2** (1) (1966), 31–42.

[90] G. LUDWIG, *Measuring and preparing process*, In: "Foundations of quantum mechanics and ordered linear spaces", Lecture Notes in Phys. **29** (1974), 122–161.

[91] G. W. MACKEY, *Imprimitivity for representations of locally compact groups*, Proc. Nat. Acad. Sci. U.S.A. **35** (4) (1949), 537–545.

[92] G. W. MACKEY, "Mathematical Foundations of Quantum Mechanics", Benjamin, New York, 1963.

[93] I. A MALKIN and V. I. MAN'KO, "Dynamical Symmetries and Coherent States of Quantum Systems", Nauka, Moscow, 1979, (R).

[94] A. I MALTZEV, "Fundamentals of Linear Algebra", Nauka, Moscow, 1970, (R).

[95] L. I. MANDELSTAM, "Lectures on Optics, Relativity and Quantum Mechanics", Nauka, Moscow, 1972, (1939), (R).

[96] L. I. MANDELSTAM and I. E. TAMM, *"Time-energy" uncertainty relation in the non-relativistic quantum mechanics*, Izv. Acad. Nauk SSSR Ser. Fiz. **9** (1-2) (1945), 122-128, (R).

[97] J. MANUCEAU and A. VERBEURE, *Quasifree states of the CCR*, Comm. Math. Phys. **9** (4) (1968), 293–302.

[98] K. V. MARDIA, "Statistics of Directional Data", Academic Press, New York, 1972.

[99] J. E. MOYAL, *Quantum mechanics as a statistical theory*, Proc. Camb. Phil. Soc. **45** (1949), 99–124.

[100] M. A. NAIMARK, *On the self-adjoint extensions of the second kind of a symmetric operator*, Izv. Acad. Nauk SSR ser. Mat. **4** (1) (1940), 53–104, (R).

[101] M. A. NAIMARK, *Spectral functions of a symmetric operator*, Izv. Acad. nauk SSSR Ser. Mat. **4** (3) (1940), 277–318 (R).

[102] E. NELSON, *Derivation of the Schrödinger equation from Newtonian mechanics*, Phys, Rev. **150** (4) (1966), 1079–1085.

[103] E. NELSON, "Dynamical Theories of Brownian Motion", Princeton, NJ, 1967.

[104] H. NEUMANN, *The structure of ordered Banach spaces in axiomatic quantum mechanics*, In: "Foundations of Quantum Mechanics and Ordered Linear spaces", Lecture Notes in Phys. **29** (1974), 116–121.

[105] T. D NEWTON and E. P. WIGNER, *Localized states for elementary systems*, Rev. Modern Phys. **21** (1949), 400–406.

[106] M. OZAWA, *Optimal measurements for general quantum systems*, Rep. Math. Phys. (1980).

[107] W. PAULI, "General Principles of Wave Mechanics", Encyclopaedia of Phys. v. 5/1, Spinger-Verlag, Berlin, 1965.

[108] A. M PERELOMOV, *Coherent states for arbitrary Lie groups*, Comm. Math. Phys. **26** (3) (1972), 222–236.

[109] A. M PERELOMOV, *Generalized coherent states and some of their applications*, Uspehi Fiz. Nauk **123** (1) (1977), 23–53, (R).

[110] C. PIRON, *Axiomatique quantique*, Helv. Phys. Acta **37** (4-5) (1964), 439–468.

[111] L. S. POLAK (ed.), "50 Years of Quantum Mechanics", Nauka, Moscow, 1979, (R).

[112] G. POLYA and G. SZEGÖ, "Aufgaben und Lehrsatze aus der Analysis", Band II, Springer-Verlag, Berlin, 1925.

[113] L. S. PONTRJAGIN, "Continuous Groups", Nauka, Moscow, 1970, (R).

[114] J. C. T POOL, *Mathematical aspects of Weyl correspondence*, J. Math. Phys. **7** (1) (1966), 66–76.

[115] E. PRUGOVEČKI, *On a theory of measurement of incompatible observables in quantum mechanics*, Canad. J. Phys. **45** (6) (1967), 2173–2219.

[116] C. R. PUTNAM, "Commutation Properties of Hilbert Space Operators and Related Topics", Springer-Verlag, Berlin, 1967.

[117] J. M. RADCLIFFE, *Some properties of coherent spin states*, J. Phys. **A4** (1971), 313–323.

[118] M. REED and B. SIMON, "Methods of Modern Mathematical Physics", Vol. 1, "Function Analysis", Academic Press, New York, 1972.
[119] F. RIESZ and B. SZ.-NAGY, "Lecons d'Analyse Fonctionnelle", Academiai Kiado, Budapest, 1972, 1952.
[120] H. P. ROBERTSON, *The uncertainty principle*, Phys. Rev. **34** (1) (1929), 163–164.
[121] R. T. ROCKAFELLAR, "Convex Analysis", Princeton University Press, Princeton, NJ, 1970.
[122] R. SCHATTEN, "Norm Ideals of Completely Continuous Operators, Springer-Verlag, Berlin, 1960.
[123] C. Y. SHE and H. HEFFNER, *Simultaneous measurements of noncommuting observables*, Phys. Rev. **152** (4) (1966), 1103–1110.
[124] YU. M. SHIROKOV, *Group-theoretic consideration of relativistic quantum mechanics*, JETP **5** (1957), 1196–1214, (R).
[125] YU. M. SHIROKOV, *Quantum and classical mechanics in the phase-space representation*, Fiz. Èlementarn. Častic i Atom. Jadra **10** (1) (1979), 5–50, (R).
[126] I. SEGAL, *A generating functional for the states of a linear Boson field*, Canad. J. Math. **13** (1) (1961), 1–18.
[127] I. SEGAL,"Mathematical Problems of Relativistic Physics", AMS, Providence, RI, 1963.
[128] B. SIMON, "Functional Integration and Quantum Physics", Academic Press, New York, 1979.
[129] M. H. STONE, "Linear Transformations in Hilbert Space and Their Applications to Analysis", AMS, Providence, RI, 1932.
[130] R. L. STRATONOVICH, *The quantum generalization of optimal statistical estimation and testing hypothesis*, J. Stoch. **1** (1973), 87–126.
[131] L. SUSSKIND and J. GLOGOWER, *Quantum mechanical phase and time operator*, Phys. **1** (1964), 49–61.
[132] M. TAKESAKI, *Tomita's theory of modular Hilbert algebras*, Lecture Notes in Math. **128** (1972).
[133] K. URBANIK, *Joint probability distribution of observables in quantum mechanics*, Studia Math. **21** (1) (1961), 117–133.
[134] F. VALENTINE, "Convex Sets", McGraw Hill, New York, 1964.
[135] V. S. VARADARAJAN, "Geometry of Quantum Theory", Van Nostrand, New York, 1968.
[136] V. S. VARADARAJAN,"Geometry of Quantum Theory", II, "Quantum Theory of Covariant Systems", Van Nostrand, New York, 1970.

[137] H. C. VOLKIN, *Phase operators and phase relations for photon states*, J. Math. Phys. **14** (12) (1973), 1965–1976.
[138] J. VON NEUMANN, "Mathematical Foundations of Quantum Mechanics", Princeton University Press, Princeton, NJ, 1955, (1932).
[139] J. VON NEUMANN, *Die Eidentigkeit der Schrödingerschen Operatoren*, Math. Ann. **104** (3) (1931), 570–578.
[140] A. WALD, "Statistical Decision Fucntions", Wiley, New York, Chapman and Hall, London, 1950.
[141] H. WEYL, "The Theory of Groups and Quantum Mechanics", Dover, New York, 1931, (1928).
[142] A. S. WIGHTMAN, *On the localizability of quantum mechanical systems*, Rev. Modern Phys. **34** (1962), 845–872.
[143] A. S. WIGHTMAN, *Hilbert's sixth problem: mathematical treatment of the axioms of physics*, Proc. Symp. in Pure Math. **28** (1) (1977), 147–240.
[144] E. P. WIGNER, *Unitary representations of the inhomogeneous Lorentz group*, Ann. Math. **40** (1) (1939), 149–204.
[145] E. P. WIGNER, "Group Theory", Academic Press, New York, 1959.
[146] E. P. WIGNER, *On the time-energy uncertainty relation*, In: "Aspects of Quantum Theory", A. Salam and E. P. Wigner (eds.), Cambridge University Press, Cambridge, 1972.
[147] H. P. H YUEN and M. LAX, *Multiple-parameter quantum estimation and measurement of non-selfadjoint observables*, IEEE Trans. **IT-19** (6) (1973), 740–750.
[148] D. P. ŽELOBENKO, "Compact Lie Groups and Their Representations", Nauka, Moscow, 1970, (R).
[149] U. ZWANZIGER, *Representations of the Lorentz group corresponding to unstable particles*, Phys. Rev. **131** (6) (1963), 2818–2819.

References added to the second edition

[150] A. BARCHIELLI and M. GREGORATTI, "Quantum Trajectories and Measurements in Continuous Time. The Diffusive Case", Lect. Notes Phys. 782, Springer, Berlin-Heidelberg, 2009.
[151] O. E. BARNDORFF-NIELSEN, R. D. GILL and P. E. JUPP, *On quantum statistical inference*, J. Royal Statist. Soc. B, **65** (2003), 1–31.
[152] I. BENGTSSON and K. ŻYCZKOWSKI, "Geometry of Quantum States: An Introduction to Quantum Entanglement", Cambridge University Press, 2006.

[153] G. BIRKHOFF and J. VON NEUMANN, *The logic of quantum mechanics*, Ann. Math. **37** (1936), 823–843.

[154] D. BOHM, *A suggested interpretation of quantum theory in terms of "hidden variables"*, Phys. Rev. **85** (1952), 166–174, 180–193.

[155] N. BOHR, "Essays 1958/1962 on Atomic Physics and Human Knowledge", Interscience Publishers, New York-London, 1963.

[156] S. L. BRAUNSTEIN, "Squeezing as an irreducible resource", e-print quant-ph/9904002.

[157] P. BUSCH, M. GRABOVSKI and P. J. LAHTI, "Operational quantum physics", Lect. Notes Phys. **m31**, Springer-Verlag: New York-Heidelberg-Berlin, 1997.

[158] N. N. CHENTSOV, "Selected works", Mathematics. Part IV, geometrostatistics and noncommutative probability, Moscow: Fizmatlit, 2001 (R).

[159] J. F. CLAUSER, *Von Neumann's informal hidden variable argument*, Amer. J. Phys. **39** (1971), 1095–1099.

[160] J. F. CLAUSER and A. SHIMONI, *Bell's theorem: experimental tests and implications*, Rep. Progr. Phys. **41** (1978), 1881–1928.

[161] G.M.D'ARIANO,*Homodyning as universal detection*, In: "Quantum Communication, Computing and Measurement" O. Hirota, A. S. Holevo, C. M. Caves (eds), New York: Plenum Press, 253–264, 1997; e-print quant-ph/9701011.

[162] A. EINSTEIN, B. PODOLSKY and N. ROSEN, *Can quantum mechanical description of physical reality be considered complete?*, Phys. Rev. **47** (1935), 777-780.

[163] B. D'ESPAGNAT, "Veiled Reality", an Analysis of Present-Day Quantum-Mechanical Concepts, Addison-Wesley, 1995.

[164] R. D. GILL. "Time, finite statistics and Bell's fifth position", Foundations of Probability and Physics – 2, Vaxjo Univ. Press, 2003, 179–206 (2003); e-print quant-ph/0301059.

[165] R. D. GILL and S. MASSAR, *State estimation for large ensembles*, Phys. Rev. A **61** (2000), 042312/1-16, e-print quant-ph/9902063.

[166] R. GILL and M. I. GUTA, *An invitation to quantum tomography*, e-print quant-ph/0303020.

[167] N. GISIN and B. GISIN, *A local hidden variable model of quantum correlation exploiting the detection loophole*, Phys.Lett. **A260** (1999), 323–327; arXiv:quant-ph/9905018.

[168] A. GLEASON, *Measures on the closed subspaces of a Hilbert space*, J. Rat. Mech. and Anal. **6** (1957), 885–894.

[169] O. GÜHNE, M. KLEINMANN, A. CABELLO, J.-A. LARSSON, G. KIRCHMAIR, F. ZÄHRINGER, R. GERRITSMA and C. F. ROOS,

Compatibility and noncontextuality for sequential measurements, Phys. Rev. **A81** (2010), 022121; arXiv:0912.4846 [quant-ph].

[170] M. GUTA and A. JENCOVA, *Local asymptotic normality in quantum statistics*, Communications in Mathematical Physics **276** (2007), 341–379, arXiv:quant-ph/0606213.

[171] A. HARTKÄMPER and H. NEUMANN (eds.), *Foundation of quantum mechanics and ordered linear spaces*, Advanced Study Institute, Marburg 1973, Lect. Notes in Phys. **29** (1974).

[172] M. HAYASHI, *Asymptotic estimation theory for a finite dimensional pure state model*., J. Phys. A **31** (1998), 4633–4655.

[173] M. Hayashi, *Two quantum analogues of Fisher information from a large deviation viewpoint of quantum estimation*, J. Phys. A Math. Gen. **35** (2002), 7689–7727, arXiv:quant-ph/0202003.

[174] M. HAYASHI, "Quantum Information: an Introduction", Springer, Berlin, 2006.

[175] F. HIAI and D. PETZ, *The proper formula for relative entropy and its asymptotics in quantum probability*, Commun. Math. Phys. **143** (1991), 99–114.

[176] D. HILBERT, J. VON NEUMANN and L. NORDHEIM, *Uber die grundlegen der quantenmechanik*, Math. Ann. **98** (1927), 1–30.

[177] A. S. HOLEVO, *Bounds for generalized uncertainty of shift parameter*, Lect. Notes Math. **1021** (1983), 243–251.

[178] A. S. HOLEVO, *Statistical structure of quantum theory*, Lect. Notes Phys. **m67** Springer, Berlin 2001.

[179] A. S. HOLEVO, "Quantum Systems, Channels, Information", Moscow Independent University: Moscow 2010, (R)

[180] P. JORDAN, J. VON NEUMANN and E. WIGNER, *On an algebraic generalization of the quantum mechanical formalism*, Ann. of Math. **35** (1) (1934), 29–64.

[181] M. JUNGE, C. PALAZUELOS, D. PEREZ-GARCIA, I. VILLANUEVA and M. M. WOLF, "Operator Space Theory: a Natural Framework for Bell Inequalities", arXiv:0912.1941[quant-ph].

[182] M. KEYL M. and R. F. WERNER, *Estimating the spectrum of a density operator*, Phys. Rev. A **64** (5) (2001), 052311.

[183] A. A. KLYACHKO, M. A. CAN, S. BINICIOGLU and A. S. SHUMOVSKY, *A simple test for hidden variables in spin-1 system*, Phys. Rev. Lett. **101** (2008), 020403; arXiv: 0706.0126 [quant-ph].

[184] U. LEONHARDT, "Measuring the Quantum State of Light", Cambridge University Press, 1997.

[185] A. LESNIEWSKI and M. B. RUSKAI, *Monotone Riemannian metrics and relative entropy on non-commutative probability spaces*, J. Math. Phys. **40** (1999), 5702–5724.
[186] G. LUDWIG, "Foundations of Quantum Mechanics", Springer - Verlag, Berlin, 1983.
[187] S. MASSAR, S. PIRONIO, J. ROLAND and B. GISIN, *A zoology of Bell inequalities resistant to detector inefficiency*, Phys. Rev. **A66** (2002), 052112; arXiv:quant-ph/0205130.
[188] S. MASSAR and S. POPESCU, *Optimal extraction of information from finite quantum ensembles*, Phys. Rev. Lett. **74** (1995), 1259–1263.
[189] T. OGAWA and H. NAGAOKA, *Strong converse and Stein's lemma in quantum hypothesis testing* IEEE Trans. Inf. Theory, **IT-46** (2000), 2428–2433.
[190] D. PETZ, *Monotone metrics on matrix spaces*, Lin. Alg. Appl. **244** (1996), 81–96.
[191] D. PETZ, "Quantum Information Theory and Quantum Statistics", Berlin: Springer, 2008.
[192] C. PIRON, "Foundations of Quantum Physics", W. A. Benjamin Inc., London, 1976.
[193] W. P. SCHLEIH and S. M. BARNETT (eds.), *Quantum phase and phase dependent measurements*, Phys. Scripta **T48**, (1993).
[194] F. SELLERI and G. TAROZZI, *Quantum mechanics, reality and separability*, La Rivista del Nuovo Cimento **2** (1981).
[195] M. D. SRINIVAS, *When is a hidden varible theory comparable with quantum mechanics?*, Pramana **19** (1982), 159–173.
[196] A. UHLMANN, *Density operators as an arena for differential geometry*, Rep. Math. Phys. **33** (1993), 253–263.
[197] N. WIENER et al., "Differential Space, Quantum Systems and Prediction", MIT Press, 1966.
[198] E. P. WIGNER, "Symmetries and Reflections", Indiana University Press, Bloomington-London, 1970.
[199] H. M. WISEMAN and R. B. KILLIP, *Adaptive single-shot phase measurements: the full quantum theory* Phys. Rev. A, **57** (3) (1998), 2169–2185.
[200] H. P. YUEN and J. H. SHAPIRO, *Quantum statistics of homodyne and heterodyne detection*, In: "Coherence and Quantum Optics" L. Mandel, E. Wolf (IV. eds.), Plenum Press, New York, 719–727, 1978.

Index

MONOGRAPHS

This series is a collection of monographs on advanced research topics of current interest in the fields of mathematics and physics.

Published volume

1. A. HOLEVO, *Probabilistic and Statistical Aspects of Quantum Theory*, 2010. ISBN 978-88-7642-375-8

Volumes published earlier

QUADERNI

E. DE GIORGI, F. COLOMBINI, L. C. PICCININI, *Frontiere orientate di misura minima e questioni collegate,* 1972.

C. MIRANDA, *Su alcuni problemi di geometria differenziale in grande per gli ovaloidi,* 1973.

G. PRODI, A. AMBROSETTI, *Analisi non lineare,* 1973.

C. MIRANDA, *Problemi di esistenza in analisi funzionale,* 1975 (out of print).

I. T. TODOROV, M. MINTCHEV, V. B. PETKOVA, *Conformal Invariance in Quantum Field Theory,* 1978.

A. ANDREOTTI, M. NACINOVICH, *Analytic Convexity and the Principle of Phragmén-Lindelöf,* 1980.

S. CAMPANATO, *Sistemi ellittici in forma divergenza. Regolarità all'interno,* 1980.

Topics in Functional Analysis 1980-81, Contributors: F. STROCCHI, E. ZARANTONELLO, E. DE GIORGI, G. DAL MASO, L. MODICA, 1981.

G. LETTA, *Martingales et intégration stochastique,* 1984.

Old and New Problems in Fundamental Physics, in honour of GIAN CARLO WICK, 1986.

Interaction of Radiation with Matter, in honour of ADRIANO GOZZINI, 1987.

M. MÉTIVIER, *Stochastic Partial Differential Equations in Infinite Dimensional Spaces,* 1988.

Symmetry in Nature, in honour of LUIGI A. RADICATI DI BROZOLO, 2 voll., 1989 (reprint 2005).

Nonlinear Analysis, in honour of GIOVANNI PRODI, 1991 (out of print).

C. LAURENT-THIÉBAUT, J. LEITERER, *Andreotti-Grauert Theory on Real Hypersurfaces,* 1995.

J. ZABCZYK, *Chance and Decision. Stochastic Control in Discrete Time,* 1996 (reprint 2000).

I. EKELAND, *Exterior Differential Calculus and Applications to Economic Theory,* 1998 (2000).

Electrons and Photons in Solids, A Volume in honour of FRANCO BASSANI, 2001.

J. ZABCZYK, *Topics in Stochastic Processes,* 2004.

N. TOUZI, *Stochastic Control Problems, Viscosity Solutions and Application to Finance,* 2004.

CATTEDRA GALILEIANA

P. L. LIONS, *On Euler Equations and Statistical Physics,* 1998.

T. BJÖRK, *A Geometric View of the Term Structure of Interest Rates,* 2000 (2001).

F. DELBAEN, *Coherent Risk Measures,* 2002.

W. SCHACHERMAYER, *Portfolio Optimization in Incomplete Financial Markets,* 2004.

D. DUFFIE, *Credit Risk Modeling with Affine Processes,* 2004.

H. M. SONER, *Stochastic Optimal Control in Finance,* 2004 (2005).

LEZIONI LAGRANGE

C. VOISIN, *Variations of Hodge Structure of Calabi-Yau Threefolds,* 1998.

LEZIONI FERMIANE

R. THOM, *Modèles mathématiques de la morphogénèse,* 1971.

S. AGMON, *Spectral Properties of Schrödinger Operators and Scattering Theory,* 1975.

M. F. ATIYAH, *Geometry of Yang-Mills Fields,* 1979.

KAC M., *Integration in Function Spaces and some of its Applications,* 1983 (out of print).

J. MOSER, *Integrable Hamiltonian Systems and Spectral Theory,* 1983.

T. KATO, *Abstract Differential Equations and Nonlinear Mixed Problems,* 1988.

W. H. FLEMING, *Controlled Markov Processes and Viscosity Solution of Nonlinear Evolution Equations,* 1988.

V. I. ARNOLD, *The Theory of Singularities and its Applications,* 1991.

J. P. OSTRIKER, *Development of Larger-Scale Structure in the Universe,* 1993.

S. P. NOVIKOV, *Solitons and Geometry,* 1994.

L. A. CAFFARELLI, *The Obstacle Problem,* 1999.