ASPECTS OF NONLINEAR QUANTUM MECHANICS

Marek Czachor
Centrum Fizyki Teoretycznej
Polskiej Akademii Nauk
Warszawa

27 October 1993
Contents

1 INTRODUCTION 5

2 QUANTUM MECHANICS AS A CLASSICAL HAMILTONIAN SYSTEM 9

3 SOME GENERALIZATIONS 14
   3.1 States 14
   3.2 Observables 14
   3.3 Dynamics 21
   3.4 Results of Single Measurements and Probabilities 24

4 COMPOSITE SYSTEMS AND NONLOCALITY 33

5 A TWO-LEVEL ATOM IN NONLINEAR QM 42
   5.1 A Fully Quantum Approach 42
   5.2 A Semiclassical Approach 49

6 REMARKS ON THE MIND-BODY QUESTION IN NONLINEAR QM 50

7 ENTROPIC FRAMEWORK FOR NONLINEAR QM 55
   7.1 Elements of Information Theory — from Hartley to Rényi 56
   7.2 Poissonian Formulation of Quantum Mechanics 60
   7.3 Nambu-like Generalization of Quantum Mechanics 64
      7.3.1 The Jacobi Identity 66
      7.3.2 Composite Systems in the New Framework 68
7.3.3 Density Matrix Interpretation of Solutions
of the Evolution Equation ........................................ 69
7.3.4 Composition Problem for Subsystems
with Different Entropies .......................................... 73

8 APPENDICES .................................................. 76
8.1 Elements of Relativistic Formulation
— Hamiltonian Formulation
of the Dirac Equation ............................................. 76
8.2 Schrödinger Form of the Harmonic
Oscillator Equations .............................................. 80
8.3 On Certain Associative Product
of Nonlinear Observables ........................................ 81
8.4 Polarizing Mach-Zehnder Interferometer
and Other Analyzers of Polarization ......................... 86
  8.4.1 Analyzer of circular polarizations ...................... 84
  8.4.2 Analyzers of linear polarizations ....................... 85
  8.4.3 Polarizing Mach-Zehnder Interferometer .......... 86
...Of even greater interest it seems to me is the question of whether quantum mechanics is necessarily true. Quantum mechanics has had phenomenal successes in explaining the properties of particles and atoms and molecules, so we know that it is a very good approximation to the truth. The question then is whether there is some other logically possible theory whose predictions are very close but not quite the same as those of quantum mechanics. It is easy to think of ways of changing most physical theories in small ways. For instance, the Newtonian law of gravitation, that the gravitational force between two particles decreases as the inverse square of the distance, could be changed a little by supposing that the force decreases with some other power of the distance, close to but not precisely the same as the inverse square. To test Newton's theory experimentally we might compare observations of the solar system with what would be expected for a force that falls off as some unknown power of the distance and in that way put a limit on how far from an inverse square this power of distance can be. Even general relativity could be changed a little, for instance by including more complicated small terms in the field equations or by introducing weakly interacting new fields into the theory. It is striking that it has so far not been possible to find a logically consistent theory that is close to quantum mechanics, other than quantum mechanics itself.

I tried to construct such a theory a few years ago. My purpose was not seriously to propose an alternative to quantum mechanics but only to have some theory whose predictions would be close to but not quite the same as those of quantum mechanics, to serve as a foil that might be tested experimentally. I was trying in this way to give experimental physicists an idea of the sort of experiment that might provide interesting quantitative tests of the validity of quantum mechanics. One wants to test quantum mechanics itself, and not any particular quantum-mechanical theory like the standard model, so to distinguish experimentally between quantum mechanics and its alternatives one must check some general feature of any possible quantum-mechanical theory. In inventing an alternative to quantum mechanics I fastened on the one general feature of quantum mechanics that has always seemed somewhat more arbitrary than others, its linearity...

After some work I came up with a slightly nonlinear alternative to quantum mechanics that seemed to make physical sense and could be easily tested to very high accuracy by checking a general consequence of linearity, that the frequencies of oscillation of any sort of linear system do not depend on how the oscillations are excited. For instance, Galileo noticed that the frequency with which a pendulum goes back and forth does not depend on how far the pendulum swings. This is because, as long as the magnitude of the oscillation is small enough, the pendulum is a linear system; the rates of change of its displacement and momentum are proportional to its momentum and its displacement, respectively. All clocks are based on this feature of oscillations of linear systems, whether pendulums or springs or quartz crystals. A few years ago after a conversation with David Wineland of the National Bureau of Standards, I realized that the spinning nuclei that were used by the bureau to set time standards provided a wonderful test of the linearity of quantum mechanics; in my slightly nonlinear alternative to quantum mechanics the frequency with which the spin axis of the nucleus precesses around a magnetic field would depend very weakly on the angle between the spin axis and the magnetic field. The fact that no such effect had been seen at the Bureau of Standards told me immediately that any nonlinear effects in the nucleus studied (an isotope of beryllium) could contribute no more than one part in a billion billion billion to the energy of the nucleus. Since this work, Wineland
and several other experimentalists at Harvard, Princeton, and other laboratories have improved these measurements, so that we now know that nonlinear effects would have to be even smaller than this. The linearity of quantum mechanics, if only approximate, is a rather good approximation after all.

None of this was particularly surprising. Even if there are small nonlinear corrections to quantum mechanics, there was no reason to believe that these corrections should be just large enough to show up in the first round of experiments designed to search for them. What I did find disappointing was that this nonlinear alternative to quantum mechanics turned out to have purely theoretical internal difficulties. For one thing, I could not find any way to extend the nonlinear version of quantum mechanics to theories based on Einstein’s special theory of relativity. Then, after my work was published, both N. Gisin in Geneva and my colleague Joseph Polchinski at the University of Texas independently pointed out that in the Einstein-Podolsky-Rosen thought experiment (...) the nonlinearities of the generalized theory could be used to send signals instantaneously over large distances, a result forbidden by special relativity. At least for the present I have given up the problem; I simply do not know how to change quantum mechanics by a small amount without wrecking it altogether.

This theoretical failure to find a plausible alternative to quantum mechanics, even more than the precise experimental verification of linearity, suggests to me that quantum mechanics is the way it is because any small change in quantum mechanics would lead to logical absurdities. If this is true, quantum mechanics may be a permanent part of physics. Indeed, quantum mechanics may survive not merely as an approximation to a deeper truth, in the way that Newton’s theory of gravitation survives as an approximation to Einstein’s general theory of relativity, but as a precisely valid feature of the final theory.

Steven Weinberg in Dreams of a Final Theory [1]
Chapter 1

INTRODUCTION

A reader of this work may feel a bit confused by its title. “Aspects of what?” — one may ask — “And why only aspects?” And, indeed, the name “nonlinear quantum mechanics” suggests that we are dealing with an existing theory, whereas what we have in mind is a collection of various attempts of constructing a generalization of quantum mechanics (QM).

The adjective “nonlinear” corresponds usually to a nonlinear generalization of the Schrödinger equation. Much less often it is associated with the linearity of the Liouville-von Neumann equation, or the linearity of the space of states, or the linearity of the space of observables. Each of these elements constitutes a specific level of the linearity of ordinary QM, and therefore there are several places where the putative nonlinearity may be introduced.

The linearity of the Schrödinger (wave) equation results in the quantum mechanical superposition principle. The principle means mathematically that a linear combination of any number of solutions of the evolution equation is again a solution of this equation; and vice versa — any solution can be decomposed into a linear combination of other solutions. Physically the first part of the principle means that waves of probability can interfere in a linear way. But anybody who observed waves produced by a ship in a harbour knows that real waves on water interfere in a nonlinear way: There exist various feedback influences of the produced waves on their sources. Only in an idealized case of very weak waves the linear approximation describes the behaviour of the ship in a correct manner. This property is shared by practically all oscillations observed at the macroscopic level. It follows that the observed phenomena of quantum interferences cannot be considered as an evidence of the fundamental linearity of quantum evolution.

The other part of the superposition principle seems more essential and is related to the projection postulate. The postulate states that a measurement of a physical quantity is equivalent to a projection of an initial state vector, describing a pure state of the system in question before the measurement, in some direction in the Hilbert space of pure states. The projected state describes the state of the system after the measurement. According to the Copenhagen interpretation, the ontological status of the projection is the same as in ordinary probability calculus, where a measurement is represented by a projection on a subset of some probability space (the so-called Bayes rule for conditional probabilities). It is obvious that at the dynamical level the projection postulate can be consistent provided
the projection “maps solutions into solutions”, which is typical only of linear evolutions. The projection postulate is useful for calculation of probabilities but for obvious reasons it does not follow from the Schrödinger dynamics. It raises also interpretational questions which shall be discussed in Chapter 6.

The linearity of the Liouville-von Neumann equation reflects another principle: the convexity of the “figure of states”. Convexity means that a convex combination of two pure states is also a state — a mixed state. One of the properties of QM state figures is the fact that they are not simplexes, which means that a mixed state can be decomposed in a number of equivalent, physically indistinguishable ways. The pure states (projectors or propositions) form a boundary of the figure. The projection postulate can be also formulated for mixed states: The measurement projects the convex combination into one of the projectors belonging to the boundary.

The superposition and convexity principles are logically independent. To understand that, it is sufficient to note that a superposition of two pure states is a pure state, while a convex combination of two projectors is not a projector. In classical mechanics the two aspects are present as well — this can be illustrated by the usual nonlinearity of the Hamilton equations of motion and the linearity of the Liouville equation.

The putative nonlinearity can be introduced here in three ways. Either one generalizes the Schrödinger equation and maintains the linearity of the Liouville-von Neumann equation (like in classical mechanics), or generalizes both of them, or keeps the evolution of pure states linear but modifies the evolution of mixed states. In the approaches of Kibble [2, 3], Bialynicki-Birula et al. [4, 5, 6, 7], or Weinberg [8, 9, 10] one concentrates on the evolution of pure states. The question how to describe mixed states is left open in those works (one of possibilities — the probability measure approach — is discussed in this context by Bugajski [11], cf. also [12]). In the approach of Mielnik [12, 14, 15], see also [17], a basic principle is the convexity of the figure of states, but pure states can evolve in a nonlinear way, so there is no ordinary superposition principle. The convexity is, in this formulation, an important constituent of the probability interpretation of the theory. The third possibility, which introduces nonlinear evolution only for mixed states, is discussed here in Chapter 7.

Now, as we know more or less what can be “deformed” in the theory, it remains to find out how to do it. There are again numerous possibilities.

In the first place, it can be shown that the dynamics of states in QM is a classical Hamiltonian dynamics for pure states (Chapter 2), and Lie-Poissonian dynamics for mixed ones (Chapter 7). Accordingly, it turns out that it is sufficient to extend the set of QM observables to make the dynamics nonlinear. Such a modification was proposed by Kibble, Weinberg, Polchinski [18], Jordan [19] and perhaps others. Also the approach of Bialynicki-Birula and Mycielski can be regarded as belonging to this class of theories. The proposal of Kibble additionally initiated a different direction of investigations, where one tampers with the manifolds of states. This direction was further explored by the Milano group [21, 22, 23, 24, 25, 26] who discussed dynamical systems on general Kähler manifolds. Also all attempts of unifying general relativity with QM can be included in this class of theories [27]. It can be shown that a nonlinear Schrödinger equation appears naturally if a gravitational field of an isolated particle is taken into account [28].

All these attitudes to nonlinear generalization of QM can be collected under the name “nonlinear quantum mechanics”. Still, the reader must be hereby warned that none of the above proposals is a complete alternative even to the nonrelativistic linear QM.
So what aspects of the above formulations will be discussed in this work?

First of all, I will concentrate on Hilbert space models. This restriction, significantly simplifying the discussion, will be sufficiently general to illustrate various difficulties that should be met in any nonlinear QM. In majority of papers on nonlinear QM their authors concentrated on dynamics of states but left intact the problem of probability interpretation (the important exceptions are the works of Mielnik and Weinberg). Hence, one of the main themes returning in this work in several places will be the probability interpretation of states and observables (alternative definitions of eigenvalues and probabilities).

The other important point which, in my opinion, is a source of prejudices about nonlinear QM, is the problem of composite systems. The question is related to the existence of “faster-than-light telegraphs” (in separated systems, Chapter 4), “telekinetic phenomena” (the role of observers, Chapters 6 and 7), and even description of Rabi oscillations in two-level atoms (Chapter 5).

A separate chapter will be devoted to the two-level atom in Weinberg’s nonlinear QM (Chapter 5). The two-level atom will serve as an example helping to understand many arbitrary elements “smuggled” in usual discussions of possible experimental implications of generalized QM.

In Chapters 6 and 7 we shall attempt to develop the suggestion of Wigner who discussed a possible relationship between conscious observation and linearity of quantum evolution. We will formulate a hypothesis concerning a role of information for evolution of quantum states and define “observers” in information theoretic terms. In this way we shall outline a new program for nonlinear QM. The new framework will be free of the main difficulties of the ordinary approaches but, as expected, will lead to open problems of its own. In the final part we shall formulate the Dirac equation in a Hamiltonian way especially suitable for a relativistic extension of our formalism.

A few words have to be said about the topics that have not been included in this work. In particular, the reader will not find here any historical review of previous attempted generalizations. I am aware of the fact that an inclusion of such a review would make the work much more comprehensive but, simultaneously, much longer. Also a review of experiments made in search of putative nonlinearities would be in place here. I hope, however, that after having read the work the readers will understand the reasons for omitting the question. Even the experiments with two- or four-level systems in ion traps include such a number of arbitrary factors (to mention only the infinite number of possible descriptions of the “atom+field” composite system) that their serious discussion would make the work twice bigger. On the other hand, the interferometric experiments designed as tests of the Bialynicki-Birula–Mycielski equation applied only to quantum systems in pure states and hence are unrelated to my own proposal.

Finally, I think I should explain that I do not stick to the framework I propose as something physically “real”. I simply think that I have managed to show a direction for further investigations, the direction that has not been fully explored as yet. This may be important, since a large number of authors tried to find for nonlinear QM some general non-existence theorems, and many claim they have found them. My own viewpoint is almost the opposite, as the reader will see later.

Most of the results presented in this work appear for the first time in a written form. This concerns first of all the analysis of probabilities and eigenvalues. The examples were usually invented as counter examples illustrating points that were omitted or presented in
an unclear way in other papers. The quantum description of the two level atom, especially
the questions related to the description of the joint “atom+field” system and the two
level approximation cannot be found in literature. I have also included my earlier results
on composite systems, but here presented in a more systematic way. The description in
terms of density matrix is a generalization of the earlier works of Polchinski, Jordan, and
Bialynicki-Birula and Morrison. The latter work was an inspiration for my own proposal
of nonlinear QM.
Chapter 2

QUANTUM MECHANICS AS A CLASSICAL HAMILTONIAN SYSTEM

A departure point for the generalizations of ordinary, linear QM discussed below is the observation that quantum theory can be regarded as a particular classical Hamiltonian theory. This statement is in apparent conflict with what one usually reads in quantum mechanical textbooks. A student who has passed an exam on quantum mechanics will probably object and formulate two main quantum features of QM:

1. The algebra of quantum observables is non-Abelian as opposed to the Abelian algebra of functions on a phase space. In particular, the energy is represented in quantum theories by Hamiltonian operators in contrast to Hamiltonian functions of classical theories.

2. The evolution equation of a pure state of a conservative system is, in quantum theories, given by the Schrödinger equation in opposition to the Hamilton equations of classical theories.

It is one of the purposes of the studies on generalized quantum theories to understand what is actually the element of the theory that makes quantum theory quantum. We shall see that the two differences described above are not, in fact, actual. Therefore the real difference lies somewhere else. It will be argued that the difference is closely related to the associativity of the algebra of quantum observables, that is, to the fact that the quantum observables possess a richer algebraic structure than the classical ones.

Let us begin with some Hilbert space $\mathcal{H}$ of square integrable functions $M \to \mathbb{C}^n$ defined on a finite dimensional manifold $M$. The manifold $M$ may be, for example, a spacelike hyperplane in the Minkowski space or a mass hyperboloid. Let $\alpha \in M$ and $\psi_A(\alpha)$ be a value of an $A$-th component of the function $\psi \in \mathcal{H}$ taken at $\alpha$. Since $\mathbb{C}^n$ has a natural structure of a symplectic manifold with the symplectic form being the imaginary part of the scalar product, it follows that $\mathcal{H}$ has a natural structure of a “phase bundle” [29] whose fibers are the finite dimensional Hilbert spaces $\mathcal{H}_\alpha = \mathbb{C}^n$ with coordinates
\{ \varphi_A(\alpha) \}. In the approach used in this work the Hilbert space and its corresponding projective space \( P(\mathcal{H}) \) will be treated as real manifolds. This can be achieved by a realification of the model spaces. (The realification of a complex linear space is obtained by restricting the field of scalars to real numbers.) Since the manifolds are modelled on complex Hilbert spaces there exist complex valued scalar products in their tangent spaces. The scalar products lead naturally to symplectic and Riemannian structures on the (realificated) manifolds. Let \( \langle \cdot | \cdot \rangle_\psi \) be a scalar product in \( T_\psi \mathcal{H} \simeq \mathcal{H} \) or \( T_{\psi'}P(\mathcal{H}) \), for some \( \psi \) \[30\]. The symplectic form is

\[
\omega_\psi(\psi, \phi) = 2\nu \text{Im}(\langle \psi | \phi \rangle)_\psi
\]

and the Riemannian metric tensor

\[
g_\psi(\psi, \phi) = 2\nu \text{Re}(\langle \psi | \phi \rangle)_\psi.
\]

The constant \( \nu \) depends on conventions (say, whether the scalar product is linear in the first or in the second argument) or physical applications (for \( P(\mathcal{H}) \) \( \nu \) is related to the holomorphic sectional curvature; in the symplectic formulation of quantum mechanics it can be shown to be \( \hbar \) — here we will put \( \nu = 1 \) which is equivalent to the units with \( \hbar = 1 \) (cf. \[31, 32, 30, 21, 22, 23, 24, 25, 26, 2\]). Pure states \( \psi \in \mathcal{H} \) are sections of the bundle, in general defined on its dense subsets (manifold domains \[32\]). The manifold of states is an infinite dimensional manifold of the sections; this manifold (a phase space) will be denoted by \( P \).

A basis in a space tangent to a point in \( \mathcal{H}_\alpha \) (treated as a real manifold) is

\[
\frac{\partial}{\partial \varphi_A(\alpha)} \quad \text{and} \quad \frac{\partial}{\partial \varphi_{A'}(\alpha)};
\]

here \( \alpha \) is fixed and summation can be only over the discrete indices. The corresponding basis in \( T_\varphi P \) is denoted by

\[
\frac{\delta}{\delta \varphi_A(\alpha)} \quad \text{and} \quad \frac{\delta}{\delta \varphi_{A'}(\alpha)}
\]

where one can sum over discrete indices and integrate over the continuous ones (the integration involves some invariant, or quasi-invariant measure on \( M \)). The basis in the dual to \( T_\varphi P \) is denoted by

\[
d\varphi_A(\alpha) \quad \text{and} \quad d\varphi_{A'}(\alpha)
\]

where “dual” means that

\[
d\varphi_A(\alpha) \left( \frac{\delta}{\delta \varphi_B(\beta)} \right) = \delta_A^B \delta(\alpha - \beta)
\]

\[
d\varphi_{A'}(\alpha) \left( \frac{\delta}{\delta \varphi_{B'}(\beta)} \right) = \delta_{A'}^B \delta(\alpha - \beta).
\]

The symplectic form and the metric tensor in \( T_\varphi P \) are defined as

\[
\omega = i\delta_A^A' d\varphi_A(\alpha) \land d\varphi_{A'}(\alpha) = i\delta_A^A' \left( d\varphi_A(\alpha) \otimes d\varphi_{A'}(\alpha) - d\varphi_A(\alpha) \otimes d\varphi_{A'}(\alpha) \right)
\]

and

\[
g = \delta_A^A' d\varphi_A(\alpha) \lor d\varphi_{A'}(\alpha) = \delta_A^A' \left( d\varphi_A(\alpha) \otimes d\varphi_{A'}(\alpha) + d\varphi_{A'}(\alpha) \otimes d\varphi_A(\alpha) \right)
\]
where we have introduced the following summation convention: Summation is over repeated \( \text{Roman indices} \) and integration is over the repeated \( \text{Greek ones} \). This convention will be often applied. For two tangent vectors in \( T_{\varphi} P \)

\[
\psi = \psi_A(\alpha) \frac{\delta}{\delta \varphi_A(\alpha)} + \bar{\psi}_{A'}(\alpha') \frac{\delta}{\delta \bar{\varphi}_{A'}(\alpha')}
\]  

(2.10)

and

\[
\phi = \phi_A(\alpha) \frac{\delta}{\delta \varphi_A(\alpha)} + \bar{\phi}_{A'}(\alpha') \frac{\delta}{\delta \bar{\varphi}_{A'}(\alpha')}
\]  

(2.11)

one finds indeed

\[
\omega(\psi, \phi) = 2 \text{Im} \langle \psi | \phi \rangle
\]  

(2.12)

\[
g(\psi, \phi) = 2 \text{Re} \langle \psi | \phi \rangle
\]  

(2.13)

where the scalar product is anti-linear in the first argument.

The two tensors can be combined into one Hermitian form

\[
K = \frac{1}{2}(g + i \omega) = \delta^{AA'} d \varphi_A(\alpha) \otimes d \bar{\varphi}_{A'}(\alpha).
\]  

(2.14)

The components of \( K \) can be derived from the potential

\[
f[\varphi, \bar{\varphi}] = \delta^{AA'} \varphi_A(\alpha) \bar{\varphi}_{A'}(\alpha)
\]  

(2.15)

by

\[
K^{AA'}(\alpha, \alpha') = \frac{\delta^2 f}{\delta \varphi_A(\alpha) \delta \bar{\varphi}_{A'}(\alpha')}. 
\]  

(2.16)

The symplectic form is closed as independent of \( \varphi \), or as derivable from the potential by

\[
\omega = i \partial \bar{\partial} f
\]  

(2.17)

(cf. \cite{31, 33})

A analogous language can be formulated for the projective space \( P(\mathcal{H}) \). Let \( \Psi = [\psi] \) be a class of vectors complex proportional to a nonvanishing \( \psi \in \mathcal{H} \) and choose some \( \psi_K(\kappa) \neq 0 \). We define the following local projective coordinates

\[
\Psi^K_A(\alpha) = \psi_A(\alpha)/\psi_K(\kappa)
\]  

(2.18)

\[
\bar{\Psi}^{K_A'}(\alpha') = \bar{\psi}_{K_A'}(\alpha').
\]  

(2.19)

Let

\[
F[\Psi, \bar{\Psi}] = \ln(1 + \delta^{AA'} \Psi^K_A(\alpha) \bar{\Psi}^{K_A'}(\alpha'))
\]  

(2.20)

The symplectic form and the metric tensor for \( P(\mathcal{H}) \) can be derived from \( F \) by

\[
K^{AA'} = \frac{\delta^2 F}{\delta \Psi^K_A(\alpha) \delta \bar{\Psi}^{K_A'}(\alpha')}
\]  

\[
= \frac{\delta^{AA'} \delta(\alpha - \alpha')}{1 + \delta^{BB'} \Psi^K_B(\beta) \Psi^{K_B'}(\beta)} - \frac{\delta^{AC'} \delta^{C'A'} \bar{\psi}_{K_C'}(\alpha) \psi^{K_C}(\alpha')}{(1 + \delta^{BB'} \Psi^K_B(\beta) \Psi^{K_B'}(\beta))}.
\]  

(2.21)
In analogy to the Hilbertian formulation
\[
\bar{\omega} = iK^{AA'}_{K\kappa}(\alpha, \alpha')d\Psi^{K\kappa}_{A}(\alpha) \wedge d\bar{\Psi}^{K\kappa}_{A}(\alpha'),
\]
(2.22)
\[
\bar{g} = K^{AA'}_{K\kappa}(\alpha, \alpha')d\Psi^{K\kappa}_{A}(\alpha) \vee d\bar{\Psi}^{K\kappa}_{A}(\alpha').
\]
(2.23)
The metric (2.23) is an infinite dimensional version of the Fubini-Study metric, and plays an important role for geometric phases, such as the Berry phase, (cf. [34] and references therein). It is convenient to define also the “inverse” of \(K\) defined by
\[
I^{AA'}_{K\kappa}(\alpha, \alpha') = \left( 1 + \delta^{BB'}\psi^{K\kappa}_{B}(\beta)\bar{\psi}^{K\kappa}_{B}(\beta) \right) \left( \delta_{AA'}\delta(\alpha - \alpha') + \psi^{K\kappa}_{A}(\alpha)\bar{\psi}^{K\kappa}_{A}(\alpha') \right)
\]
(2.24)
where by the inverse it is meant that
\[
K^{AA'}_{K\kappa}(\alpha, \alpha')I^{BB'}_{K\kappa}(\alpha, \alpha') = \delta_{A'B}' \delta(\alpha' - \beta')
\]
(2.25)
\[
K^{AA'}_{K\kappa}(\alpha, \alpha')I^{BA'}_{K\kappa}(\beta, \alpha') = \delta_{AB} \delta(\alpha - \beta).
\]
(2.26)
Consider now two functionals: The matrix element
\[
H[\psi, \bar{\psi}] = \bar{\psi}_{A'}^{A}(\alpha')H^{AA'}(\alpha', \alpha)\psi_{A}(\alpha)
\]
(2.27)
of a self-adjoint operator \(\hat{H}\), and its average
\[
\bar{H}[\psi, \bar{\psi}] = \bar{\psi}_{A'}^{A}(\alpha')\bar{H}^{AA'}(\alpha', \alpha)\psi_{A}(\alpha)
\]
(2.28)
The latter is also a functional defined on \(P(\mathcal{H})\):
\[
\bar{H}[\psi, \bar{\psi}] = \bar{H}[\Psi, \bar{\Psi}]
\]
(2.29)
although in the projective coordinates \(\bar{H}\) (also playing a role of a Hamiltonian function) is not a bilinear functional of the projective coordinates.

A straightforward calculation shows [21, 22, 23, 24, 25, 26] that the Hamilton equations
\[
\iota_{X_{\mathcal{H}}}\omega = dH
\]
(2.30)
and
\[
\iota_{X_{\mathcal{H}}}\bar{\omega} = d\bar{H}
\]
(2.31)
are equivalent to the Schrödinger equation
\[
\frac{d}{dt}\psi = \hat{H}\psi
\]
(2.32)
and its adjoint expressed either in the Hilbertian or projective coordinates (by the way, the projective form of the Schrödinger equation is nonlinear). The Poisson bracket of observables (averages or matrix elements of some self-adjoint operators) written in terms of \(I\) and resulting from the Hamilton equations is in one-to-one relationship with the commutator of the operators. It follows that the Schrödinger equation is equivalent to the Hamilton equations for an infinite-dimensional Hamiltonian system, and the non-commutability of observables is of the same nature as the classical non-commutability.
of functions on a phase space with respect to the Poisson bracket (various explicit expressions of the bracket will be discussed in several places in this work). The only subtlety that is involved in such calculations is related to the differentiability of averages of unbounded operators. These functions are everywhere discontinuous and the differentiability must be understood in a weak sense [32, 21, 22, 23, 24, 25, 26]: we will usually use functional derivatives and all functionals will be assumed to be functionally differentiable. The ordinary canonical coordinates “q” and “p” correspond to real and imaginary parts of the state vector’s components.

The above conclusions may suggest that quantum mechanics is just an ordinary classical theory, perhaps more sophisticated from the dynamical viewpoint. On the other hand we know that the most profound conceptual difficulties of quantum mechanics (the EPR-Bohm and Schrödinger’s cat paradoxes, the Bell theorem, the measurement problem) can be formulated even in two dimensional Hilbert spaces. Such systems, being dynamically equivalent to a two dimensional harmonic oscillator, are dynamically trivial. It follows that there must exist some additional, “non-Hamiltonian”, properties of quantum mechanics that lead to its physical nontriviality. We will show later that an important property of linear QM is the additional, associative structure of the algebra of observables. This property is very restrictive and is closely related to the probabilistic nature of the theory.
Chapter 3

SOME GENERALIZATIONS

The observation that quantum dynamics is in fact a classical one immediately opens some possibilities of generalizations. One can either try to modify the Hamiltonian formulation itself or simply try to explore general possibilities immanent in the Hamiltonian framework.

In this chapter we shall follow the latter direction in a way analogous to this proposed by Kibble [2] and developed and extended by Weinberg [10], Polchinski [18] and Jordan [19].

3.1 States

We will discuss theories where states are represented in the same way as in ordinary quantum mechanics (the so-called normal states): by vectors or rays in a Hilbert space or by density matrices. We know that even in the ordinary statistical quantum theory there exist states that are not normal, but we shall not include such subtleties here. We should also bear in mind that possible extensions of QM may involve different manifolds of states [2], for example general Kähler manifolds [21, 22, 23, 24, 25, 26]. A necessity of some extension of that kind is suggested by Penrose [35]. We will not include stochastic generalizations either, cf. [37].

3.2 Observables

We have shown that the observables in QM belong to a restricted class of functionals defined on a phase space. In the Hilbertian formulation they take the form $\langle \psi | \hat{A} | \psi \rangle$ (diagonal matrix elements of a self-adjoint operator) where $|\psi\rangle$ is not normalized. It follows that such observables are not equal to the measurable quantities, and a relation with these latter is established in two equivalent ways: Either one divides the value of the observable evaluated at $|\psi\rangle$ by the squared norm of $|\psi\rangle$ or takes its value in the normalized $|\psi\rangle$. In the projective space formulation the observables are equal to measurable quantities (averages).
The fact that quantum observables are so specifically chosen leads to a special, complex Hilbertian type of the quantum mechanical probability calculus. It seems it was Mielnik who noticed [12, 14, 15, 16] that there exist many different and inequivalent probability models and tried to relate them to some nonlinear quantum theories. Such a perspective seems quite natural and suggests that one should try to investigate theories where observables, like in classical mechanics, belong to a more general set of functionals.

In the Kibble’s approach the set of observables consists of differentiable functions defined on a projective space. In the Weinberg formulation [10] one assumes that the observables are defined on a Hilbert space but satisfy the (1,1)-homogeneity condition

$$A(\lambda \psi, \tilde{\psi}) = A(\psi, \lambda \tilde{\psi}) = \lambda A(\psi, \tilde{\psi}).$$

(3.1)

Averages are defined in [10] like in ordinary quantum mechanics. The (1,1)-homogeneity and the definition of averages introduces the projective structure into the set of states and, effectively, into the dynamics. This is the way the Weinberg approach can be regarded as a variant of the Kibble’s one. (We shall see later that Weinberg’s theory contains also some elements that are new with respect to the Kibble’s proposal.)

The homogeneity condition (3.1) can be expressed by means of the Euler conditions as follows

$$\psi_n \frac{\partial A}{\partial \psi_n} = \bar{\psi}_n \frac{\partial A}{\partial \bar{\psi}_n} = A.$$

(3.2)

(In this section we will often apply finite dimensional conventions; in the full, infinite dimensional framework functional derivatives must be substituted for the partial ones and the Euler homogeneity condition would read, for example,

$$\psi_n(\alpha) \frac{\delta A}{\delta \psi_n(\alpha)} = \bar{\psi}_n(\alpha) \frac{\delta A}{\delta \bar{\psi}_n(\alpha)} = A).$$

(3.3)

With the help of (3.2) we can deduce some general properties of the observables. Applying (3.2) twice we find that

$$A = \frac{\partial A}{\partial \psi_n} \psi_n = \bar{\psi}_m \frac{\partial^2 A}{\partial \bar{\psi}_m \partial \psi_n} \psi_n = \bar{\psi}_m \hat{A}_{mn} \psi_n = \langle \psi | \hat{A}(\psi, \tilde{\psi}) | \psi \rangle$$

(3.4)

and

$$\frac{\partial A}{\partial \bar{\psi}_m} = \frac{\partial^2 A}{\partial \psi_n \partial \bar{\psi}_m} \psi_n = (\hat{A}(\psi, \tilde{\psi}) | \psi \rangle)_m.$$  

(3.5)

We can see that the homogeneity condition leads naturally to a nonlinear operator algebra. A matrix multiplication of such nonlinear operators can be easily expressed in terms of first partial derivatives of observables since

$$A * B := \frac{\partial A}{\partial \psi_n} \frac{\partial B}{\partial \bar{\psi}_m} = \langle \psi | \hat{A}(\psi, \tilde{\psi}) \hat{B}(\psi, \tilde{\psi}) | \psi \rangle.$$  

(3.6)

This $*$-product conserves the (1,1)-homogeneity of observables because the nonlinear matrices defined in (3.4) are (0,0)-homogeneous. The (0,0)-homogeneity implies, on the other hand, that the matrices are, in certain sense, “almost constant”. Indeed,

$$\left( \frac{\partial \hat{A}}{\partial \psi_k} | \psi \rangle \right)_l = \frac{\partial}{\partial \psi_k} \left( \frac{\partial^2 A}{\partial \psi_l \partial \psi_m} \right) \psi_m = \psi_m \frac{\partial}{\partial \psi_m} \frac{\partial^2 A}{\partial \psi_l \partial \psi_k} = 0$$

(3.7)
by virtue of the 0-homogeneity Euler condition. Consider now a differentiable function $t \rightarrow \psi_n(t)$. Eq. (3.7) and its complex conjugate imply that

$$\langle \psi | \frac{d}{dt} \hat{A}(\psi, \bar{\psi}) | \psi \rangle = 0$$

(3.8)

for any observable $A$. The algebra of observables equipped with the above $*$-product possesses a natural left and right unit element $n(\psi, \bar{\psi}) = \langle \psi | \bar{\psi} \rangle$.

Despite some similarities to the ordinary linear operator algebra our algebra of observables is associative if and only if the operators are linear. This leads to several differences with respect to ordinary quantum mechanics. For example, $*$-powers of $A$ will not, in general, $*$-commute with $A$, a fact that will influence the integrability of the theory.

In linear QM the $*$-powers of an observable are related to higher moments of random variables measured in experiments, hence lead to the probabilistic interpretation of the theory. In nonlinear theory the product (3.6) leads to the first non-unique element of the formalism we shall meet later.

For let us consider an observable $A$. We know that it can be written in a form of a matrix element $\langle \psi | \hat{A} | \psi \rangle$ where $\hat{A}$ is the nonlinear operator defined in (3.4) (for simplicity we shall, from now on, omit in our notation the dependence of $\hat{A}$ on $\psi$). The $*$-square of $A$ can be defined in two ways: Either as $\frac{\partial A}{\partial \rho_{nm}} \frac{\partial A}{\partial \rho_{mn}}$ or as $\langle \psi | \hat{A} \hat{A} | \psi \rangle$ and the two expressions are equal. If we want to define the third power we encounter several possibilities. The apparently most natural choice of $(A * A) * A$ or $A * (A * A)$ is not very reasonable as they are in general complex and non equal even if $A$ is real. The reality condition leads uniquely to $\frac{1}{2}(A * (A * A) + (A * A) * A)$ but this uniqueness will be lost for higher powers. In Section 3.4 we shall see also that $*$ does not lead to a consistent probability interpretation of $A$. The second definition can be chosen as $\langle \psi | \hat{A} \hat{A} \hat{A} | \psi \rangle := A \bar{*} A \bar{*} A$. We see that once we have defined $\hat{A}$, which is unique, all its $\bar{*}$-powers are unique as well. Notice that since $A * B = A \bar{*} B$ both products lead to the same Poisson bracket hence generate the same evolution of observables.

Both products can be defined also in terms of density matrices. Quantum observables are 1-homogeneous in density matrix i.e. $A(\lambda \rho) = \lambda A(\rho)$ or $\rho_{nm} \frac{\partial A}{\partial \rho_{mn}} = A$. The nonlinear operators are obtained by

$$\hat{A}_{nm} = \frac{\partial A}{\partial \rho_{nm}}$$

(3.9)

(notice the reversed order of the indices). The products are

$$A \bar{*} B = \rho_{mn} \frac{\partial A}{\partial \rho_{kn}} \frac{\partial B}{\partial \rho_{mk}} = A \bar{*} B$$

(3.10)

and

$$A \bar{*} B \bar{*} C = \text{Tr} \rho \hat{A} \hat{B} \hat{C}.$$  

(3.11)

These properties of $\bar{*}$ hint at the possibility of defining the probability interpretation of nonlinear QM in terms of $\bar{*}$ instead of $*$. We shall discuss this possibility in a more detailed way in the context of eigenvalues and dynamics.

Another way of viewing the difficulties with $*$ can be illustrated as follows. An observable $A$ can be defined as a scalar function on a manifold of states. All one-parameter
groups of canonical transformations generated by some observables are homogeneity preserving. Writing a canonical transformation in terms of components

$$\psi'_n(\psi, \bar{\psi}) = \frac{\partial \psi'_n}{\partial \psi_m} \psi_m + \frac{\partial \psi'_m}{\partial \psi_m} \bar{\psi}_m,$$

we find the second term vanishing by virtue of the (1,0)-homogeneity Euler condition for \(\psi'_n(\psi, \bar{\psi})\). The canonical transformations (3.12) look holomorphic, just like a change of a chart on a complex manifold. The “holomorphy” is however only apparent as the transformation coefficients depend on both \(\psi\) and \(\bar{\psi}\) even though the second term in (3.12) vanishes. Consider now the “gradient”

$$\frac{\partial A}{\partial \psi_m}.$$ (3.13)

We know that, by definition,

$$\psi_n \frac{\partial A}{\partial \psi_n} = A.$$ (3.14)

which may suggest that (3.13) transforms “contravariantly” with respect to the “covariant” transformation rule (3.12). This is not the case, however. We find

$$\frac{\partial A}{\partial \psi_m} = \frac{\partial \psi'_n}{\partial \psi_m} \frac{\partial A}{\partial \psi'_n} \frac{\partial A}{\partial \psi'_m} \frac{\partial \psi'_n}{\partial \psi_m} \frac{\partial A}{\partial \psi'_n}$$ (3.15)

with the last term, in general, nonvanishing. Still, its contraction with \(\psi_m\) does vanish as a result of the (0,1)-homogeneity of \(\psi'_m(\psi, \bar{\psi})\). The \(\ast\) product is defined as a “scalar product” of “gradients” \(\frac{\partial A}{\partial \psi_m}\) and \(\frac{\partial B}{\partial \psi_m}\) and as such does not represent a scalar. \(A \ast B\) is invariant under a change of coordinates provided the transformation \(\psi \rightarrow \psi'\) is holomorphic. Such a class is much narrower than the class of canonical transformations. As a particular consequence, \(\ast\)-products of two integrals of motion are not, in general, integrals of motion (an integral of motion is a “scalar” with respect to the time evolution understood as a canonical transformation).

One of the most important questions of nonlinear quantum mechanics concerns a definition of observables that correspond to subsystems of a larger system. Let us begin the discussion with recalling the ways we do it in ordinary QM.

From a physical viewpoint any system (with the exception of the Universe itself) is a subsystem of something greater. If we say that \(H(\psi_{\text{subsystem}}, \bar{\psi}_{\text{subsystem}})\) is a Hamiltonian function of a subsystem we mean that \(H(\psi_{\text{subsystem}}, \bar{\psi}_{\text{subsystem}})\) describes a subsystem which is noninteracting and noncorrelated with the rest of the Universe. Quantum mechanically this means that the state of the Universe is \(|\psi\rangle = |\psi_{\text{subsystem}}\rangle \otimes |\psi_{\text{rest}}\rangle\) (or \(\rho = \rho_{\text{subsystem}} \otimes \rho_{\text{rest}}\)).

Subsystem observables in linear QM can be written as

$$A(\psi, \bar{\psi}) = \langle \psi | \hat{A} \otimes 1_{\text{rest}} | \psi \rangle = \text{Tr} \rho_{\text{subsystem}} \hat{A}.$$ (3.16)

Let now \(\{|r\}\) denote a basis in the Hilbert space of the “rest” and \(\{|a\}\) be the one of the subsystem. A general form of the state of the whole system (here we assume for
inessential simplicity that the whole system is in a pure state) is
\[ |ψ⟩ = \sum_{a,r} \psi_{a,r} |a⟩|r⟩ = \sum_r |Φ^{(r)}⟩|r⟩. \]  

(3.17)

An average of \( A \) in the state \( |ψ⟩ \) can be expressed by means of the decomposition (3.17) as follows
\[ \mathcal{A}(ψ, ˘ψ) = \langle ψ| \hat{A} \otimes 1_{\text{rest}} |ψ⟩ = \sum_r \frac{⟨Φ^{(r)}| \hat{A} |Φ^{(r)}⟩}{⟨Φ^{(r)}⟩} \frac{⟨Φ^{(r)}⟩}{⟨ψ|ψ⟩} \]
\[ = \sum_r \frac{⟨Φ^{(r)}| \hat{A} |Φ^{(r)}⟩}{⟨Φ^{(r)}⟩} \frac{⟨ψ|1 \otimes |r⟩⟨r|ψ⟩}{⟨ψ|ψ⟩}. \]  

(3.18)

The physical meaning of (3.18) is obvious: The whole ensemble of subsystems is decomposed into subensembles corresponding to different results of measurements of an observable whose spectral family is given by \( \{|r⟩⟨r|\} \), and the average of \( A \) in the subsystem is calculated for each subensemble separately. The form (3.18) is useful as an illustration of the “collapse” of a state vector phenomenon. Indeed, the average is calculated as though the ensemble of subsystems consisted of subensembles collapsed by an external observer and, in certain sense, the ensemble is treated as the one composed of individual objects each of them possessing some “property” measured (via correlations) by the external observer.

The description in terms of the reduced density matrix \( ρ_{\text{subsystem}} \) suggests a slightly different interpretation of the average. If we believe that density matrices represent ensembles and not individual systems the form
\[ \mathcal{A}(ψ, ˘ψ) = \text{Tr} ρ_{\text{subsystem}} \hat{A} \]  

(3.19)

means that we do not care about specific possible decompositions of the ensemble, but treat it as a whole. The presence of the reduced density matrix in (3.19) does not mean, of course, that we cannot use the state vector’s components — it simply restricts the form of observables to only some functions of \( ψ_{a,r} \).

Quantum mechanics when viewed from the first perspective looks as a theory of individual systems but with some form of fundamental limitation of knowledge about states of such individuals. The second possibility turns QM into a kind of a “mean field theory”. Still, since both forms of description are physically indistinguishable such discussions within the linear framework seem purely academic.

In nonlinear theories the situation is drastically changed. The nonlinearity is supposed to be fundamental, hence applying to each individual even if the individual is well isolated from other physical systems and no mean-field approximation is justified.

The problem is practically the following. Assuming that we know a form of an observable describing a given subsystem itself (that is, if no interactions or correlations with the “rest” exist) what is the form of the observable if there do exist some interactions or correlations with the “rest”. A reversed problem is simpler: If we know a description of the “whole” system and the observable in question is additive then the observable representing the subsystem is this part of the whole one that refers to the subsystem. (Note that non-additive observables, like higher order correlation functions, can be defined only
in terms of the “whole” system, so, by definition, are excluded from our considerations here.)

The first, explicitly non-mean-field-theoretic, guess for the subsystem’s (additive) observable is suggested by the “collapse-like” form (3.18) of the average, i.e.

\[ A_{\text{subsystem}}(\psi, \bar{\psi}) = \sum_r A(\Phi(r), \bar{\Phi}(r)) \]  \hspace{1cm} (3.20)

where \( A \) is a given function defined on a Hilbert space of the subsystem and the other definitions are like in (3.17). It is this definition that was chosen by Weinberg in his formulation of nonlinear QM. For non-bilinear \( A(\Phi(r), \bar{\Phi}(r)) \) the form and value of (3.20) depend on the choice of the basis \( \{ |r \rangle \} \).

**Example 3.1** Consider the following non-bilinear observable

\[ A(\psi, \bar{\psi}) = \frac{\langle \psi | 3 \sigma_3 | \psi \rangle^2}{\langle \psi | \psi \rangle} = \frac{(|\psi_1|^2 - |\psi_2|^2)^2}{|\psi_1|^2 + |\psi_2|^2} \]  \hspace{1cm} (3.21)

where \( \sigma_3 \) is the Pauli matrix and \( |\psi\rangle \) belongs to a two-dimensional Hilbert space. Let the whole system be described by \( |\psi\rangle = \sum_{a=1}^2 \sum_r \psi_{a,r} |a \rangle |r \rangle \). Definition (3.20) yields

\[ A_{\text{subsystem}}(\psi, \bar{\psi}) = \sum_r \frac{(|\psi_{1,r}|^2 - |\psi_{2,r}|^2)^2}{|\psi_{1,r}|^2 + |\psi_{2,r}|^2} \]  \hspace{1cm} (3.22)

Let us choose now

\[ |\psi\rangle = \sqrt{\frac{\bar{a}_1}{2}} |11\rangle_1 |1\rangle_2 + \sqrt{\frac{\bar{a}_2}{2}} |22\rangle_1 |2\rangle_2. \]  \hspace{1cm} (3.23)

In this particular state

\[ A_{\text{subsystem}}(\psi, \bar{\psi}) = |\psi_{11}|^2 + |\psi_{22}|^2 \]  \hspace{1cm} (3.24)

so that each “collapsed” subensemble contributes to the observable separately. Let us choose now a different basis \( \{ |r \rangle \} \) in the Hilbert space of the “rest”, such that \( |1\rangle_2 = \frac{1}{\sqrt{2}} (|1\rangle_2 + |2\rangle_2), |2\rangle_2 = \frac{1}{\sqrt{2}} (|1\rangle_2 - |2\rangle_2) \). Now

\[ |\psi\rangle = \sqrt{\frac{\bar{a}_1}{2}} (|\psi_{11}|_1 + |\psi_{22}|_1) |1\rangle_2 + \sqrt{\frac{\bar{a}_2}{2}} (|\psi_{11}|_1 - |\psi_{22}|_1) |2\rangle_2. \]  \hspace{1cm} (3.25)

Substitution of \( |\Phi^{(1')}\rangle \) and \( |\Phi^{(2')}\rangle \) into (3.20) gives

\[ A_{\text{subsystem}}(\psi, \bar{\psi}) = \frac{1}{2} \frac{(|\psi_{11}|^2 - |\psi_{22}|^2)^2}{|\psi_{11}|^2 + |\psi_{22}|^2} + \frac{1}{2} \frac{(|\psi_{11}|^2 - |\psi_{22}|^2)^2}{|\psi_{11}|^2 + |\psi_{22}|^2} \]

\[ = \frac{(|\psi_{11}|^2 - |\psi_{22}|^2)^2}{|\psi_{11}|^2 + |\psi_{22}|^2} \]  \hspace{1cm} (3.26)

which is also a sum of contributions from two subensembles, but now “collapsed to linear polarizations”. \( \Box \)
The above example reveals two properties of (3.20):

1. This form is basis-dependent so one should rather write $A_{\text{subsystem}}(\psi, \bar{\psi}) = A_{\text{subsystem}}(\psi, \bar{\psi}, \{|r\}\}$;

2. If the “rest” is described by linear QM then no change of $\{|r\}\}$ can influence observables that correspond to “the rest” itself. In particular, if one considers a total energy of the whole system then a change of $\{|r\}\}$ will influence its value (we change the energy of the subsystem and maintain the energy of the “rest”). It follows that the form (3.20) can apply only to systems that, as a whole, are open (that is, there must exist also some another nonlinear system not contained in the “whole” one and compensating the changes of energy due to the changes of $\{|r\}\}$).

3. The “subsystem+compensating subsystem+rest” system must be described in a basis-independent way.

The last remark indicates that in order to describe systems with “isolated” nonlinearities, that is with the “rest” linear, one has to describe subsystems in a way different from Weinberg’s. The following modification was proposed by Polchinski [18].

Consider an observable $A(\psi, \bar{\psi})$ corresponding to some isolated physical system, and assume that the observable can be expressed as a function of the density matrix $\rho = |\psi\rangle\langle\psi|$, i.e. $A(\psi, \bar{\psi}) = A(\rho)$. For example, the observable from the previous example could be written in either of the forms

\[
\frac{(\text{Tr} \rho \sigma_3)^2}{\text{Tr} \rho}, \tag{3.27}
\]

\[
\frac{\text{Tr} (\rho \sigma_3 \rho \sigma_3)}{\text{Tr} \rho}, \tag{3.28}
\]

\[
\frac{(\text{Tr} \rho \sigma_3)^2 \text{Tr} (\rho^2)}{(\text{Tr} \rho)^3}, \tag{3.29}
\]

or

\[
\frac{(\text{Tr} (\rho \sigma_3 \rho \sigma_3))^{n+1}}{\text{Tr} \rho (\text{Tr} \rho \sigma_3)^{2n}}, \tag{3.30}
\]

and so on. For $\rho = |\psi\rangle\langle\psi|$ all of them, and all their convex combinations, reduce to (3.21). Polchinski’s idea was to start with definitions in terms $\rho$ where $\rho$ is the density matrix of the subsystem. For $|\psi\rangle$ describing the whole system the observables of the subsystem would be functions of $\rho_{\text{subsystem}} = \text{Tr}_{\text{rest}}|\psi\rangle\langle\psi|$.

**Example 3.2** Let $|\psi\rangle = \psi_{11}|1\rangle_1|1\rangle_2 + \psi_{22}|2\rangle_1|2\rangle_2$. The reduced density matrix representing the subsystem is

\[
\rho = |\psi_{11}|^2|1\rangle_1\langle 1|_1 + |\psi_{22}|^2|2\rangle_1\langle 2|_1. \tag{3.31}
\]

The subsystem’s observables are now

\[
\frac{(\text{Tr} \rho \sigma_3)^2}{\text{Tr} \rho} = \frac{|\psi_{11}|^2 - |\psi_{22}|^2|^2}{|\psi_{11}|^2 + |\psi_{22}|^2}, \tag{3.32}
\]
\[
\frac{\text{Tr} (\rho \sigma^3 \rho \sigma^3)}{\text{Tr} \rho} = \frac{|\psi_{11}|^4 + |\psi_{22}|^4}{|\psi_{11}|^2 + |\psi_{22}|^2},
\]  
(3.33)

\[
\frac{\left(\text{Tr} (\rho \sigma^3 \rho \sigma^3)\right)^2 \text{Tr} (\rho^2)}{(\text{Tr} \rho)^3} = \frac{|\psi_{11}|^2 - |\psi_{22}|^2|^2 (|\psi_{11}|^4 + |\psi_{22}|^4)}{(|\psi_{11}|^2 + |\psi_{22}|^2)^3},
\]  
(3.34)

and

\[
\frac{\left(\text{Tr} (\rho \sigma^3 \rho \sigma^3)\right)^{n+1} \text{Tr} \rho (\text{Tr} \rho)^{2n}}{(\text{Tr} \rho)^3} = \frac{|\psi_{11}|^4 + |\psi_{22}|^4)^{n+1}}{(|\psi_{11}|^2 + |\psi_{22}|^2)(|\psi_{11}|^2 - |\psi_{22}|^2)^2}. \]  
(3.35)

As we can see the observables are completely different. If \( |\psi_{11}| = |\psi_{22}| \) (like in the spin-1/2 singlet state) the average of the first and third observables vanishes, is non-vanishing but finite in the second case, and infinite in the last one.

Expression (3.32) seems, at first glance, the most natural one since can be rewritten as

\[
\frac{\langle \psi | \sigma_3 \otimes 1_{\text{rest}} | \psi \rangle^2}{\langle \psi | \psi \rangle}.
\]  
(3.36)

Kibble in [2] did not explicitly define his way of describing subsystems, but the examples he discussed suggest this type of description. In the next chapter we will discuss problems with nonlocality of nonlinear QM and see that all the descriptions by means of reduced density matrices are more safe than the proposal of Weinberg. However, if some correlations between subsystems actually exist we indeed can “collapse” a state of a remote individual. We may expect that an average value of this “remote” observable should be composed of the discussed averages of subensembles. The description by means of (3.32) does not seem to possess this property: It is the “average spin” that contributes to the observable (3.32). So we first calculate the “average” of \( \sigma_3 \) and then square it. In the Weinberg description we first square the “spin” and then take the “average”. (The quotation marks are necessary because we have not defined yet what we mean by eigenvalues, or values of single measurements; there exists also another subtle point that has been ignored in the definition of (3.20) — in nonlinear QM \( \frac{\langle \psi | r \rangle \langle r | \psi \rangle}{\langle \psi | \psi \rangle} \) may not have an interpretation in terms of probability.)

It is interesting that the “collapse”-like property of the average can be seen also in (3.33) where both values of “spin” are averaged separately. Finally, let us notice that the last observable (3.35) is unbounded, so that the values of single measurement cannot be identified with eigenvalues of \( \sigma_3 \).

3.3 Dynamics

Both Kibble and Weinberg assumed that the dynamics of pure states is given by Hamilton equations with Hamiltonian functions belonging to the algebra of generalized observables discussed in Section 3.2. More generally, all one parameter flows of canonical transformations are assumed to be integral curves of the Hamilton equations of motion

\[
\epsilon_X \omega = dH
\]  
(3.37)

which in the Hilbertian formulation can be written as

\[
\frac{d\psi_A(\alpha)}{dt} = \frac{\delta H}{\delta \psi_A(\alpha)},
\]  
(3.38)
and
\[-i \frac{d\hat{\psi}_A(\alpha)}{dt} = \delta_{AA} \frac{\delta H}{\delta \psi_A(\alpha)}. \tag{3.39}\]

Since the evolution is, in general, nonlinear no dynamical separation of states and observables is possible (the ordinary Heisenberg and Schrödinger pictures can be formulated only if the dynamics is linear). Instead, the evolution of observables is governed by the Poisson bracket resulting from the Hamilton equations (the summation convention)
\[\frac{d}{dt} F = \{F, H\} = i\delta_{AA} \left( \frac{\delta F}{\delta \psi_A(\alpha)} \frac{\delta H}{\delta \psi_A'(\alpha)} - \frac{\delta H}{\delta \psi_A(\alpha)} \frac{\delta F}{\delta \psi_A'(\alpha)} \right). \tag{3.40}\]

In more general case, where observables are functionals of a density matrix, the evolution can be postulated in a form of the “generalized Nambu mechanics” discussed in detail in Chapter 7
\[\frac{d}{dt} F = \{F, H, S\} = \Omega_{abc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \frac{\delta F}{\delta \rho_a(\tilde{\alpha})} \frac{\delta H}{\delta \rho_b(\tilde{\beta})} \frac{\delta S}{\delta \rho_c(\tilde{\gamma})} \tag{3.41}\]
where
\[S = \frac{1}{2} \text{Tr} (\rho^2)\]
is called, in the Bialynicki-Birula–Morrison [39] terminology the entropy. The entropy is not an observable in linear QM (observables are linear in \(\rho\)); in Weinberg’s nonlinear QM \(S\) can be related to the observable \(S/\text{Tr} \rho\) which is \(1\)-homogeneous in \(\rho\).

The formulation of nonlinear QM in terms of the bracket (3.41) has several advantages, as shown by Jordan in [19] (see also Chapter 7).

The Hamilton equations (3.38) in Kibble-Weinberg (KW) theory can be written in a Schrödinger-like form
\[i \frac{d}{dt} |\psi\rangle = \hat{H}(\psi, \tilde{\psi}) |\psi\rangle \tag{3.42}\]
where \(\hat{H}(\psi, \tilde{\psi})\) is Hermitian. All such equations conserve \(\langle \psi | \psi \rangle\) because
\[\frac{d}{dt} \langle \psi | \psi \rangle = i \langle \psi | \hat{H}^\dagger(\psi, \tilde{\psi}) - \hat{H}(\psi, \tilde{\psi}) |\psi\rangle. \tag{3.43}\]
(The Hamiltonians arising from the KW equations are (0,0)-homogeneous, but the above statement concerns also equations with non-homogeneous nonlinearities.) There exists also a larger class of nonlinear equations [40] with non-Hermitian Hamiltonians but also conserving \(\langle \psi | \psi \rangle\), namely
\[i \frac{d}{dt} |\psi\rangle = \left( \hat{H}(\psi, \tilde{\psi}) + \left( 1 - \frac{|\psi\rangle \langle \psi|}{\langle \psi | \psi \rangle} \right) U \right) |\psi\rangle := \hat{H}' |\psi\rangle \tag{3.44}\]
where \(U\) does not commute with \(|\psi\rangle \langle \psi|\). The term containing \(U\) does not occur in averages:
\[\langle \psi | \hat{H}(\psi, \tilde{\psi}) |\psi\rangle = \langle \psi | \hat{H}'(\psi, \tilde{\psi}) |\psi\rangle. \tag{3.45}\]
This, by the way, indicates that \(\langle \psi | \hat{H}'(\psi, \tilde{\psi}) |\psi\rangle\) cannot be regarded as a Hamiltonian function for (3.44). Another class of nonlinear Schrödinger equations which, for analogous reasons, are not equivalent to KW equations although the Hamiltonians are (0,0)-homogeneous is given by equations of the form
\[\frac{d}{dt} |\psi\rangle = \left( \hat{H}_0 + \frac{\langle \psi | \hat{A} |\psi\rangle}{\langle \psi | \psi \rangle} \hat{B} - \frac{\langle \psi | \hat{B} |\psi\rangle}{\langle \psi | \psi \rangle} \hat{A} \right) |\psi\rangle := \hat{H} |\psi\rangle. \tag{3.46}\]
In such a case all the nonlinear terms involving \( \hat{A} \) and \( \hat{B} \) cancel out in
\[
\langle \psi | \hat{H} | \psi \rangle = \langle \psi | H_0 | \psi \rangle.
\]

**Example 3.3** Consider a two-level atom irradiated by an external classical light and assume that the Hamiltonian function of the atom takes the We inberg form
\[
\langle \psi | \hat{H}_L + \hat{H}_{NL} | \psi \rangle = \langle \psi | \hat{H}_0 | \psi \rangle.
\]
where the nonlinear \( 2 \times 2 \) matrix is
\[
\hat{H}_{NLmn} = \partial^2 H_{NL} / \partial \psi_m \partial \psi_n = (h_0 + \tilde{h}_{NL} \tilde{s})_{mn}.
\]

Let \( \tilde{s}(\psi, \bar{\psi}) = \langle \psi | \tilde{\sigma} | \psi \rangle \). We find
\[
\dot{\tilde{s}} = \{ \tilde{s}, \hat{H} \} = -i \langle \psi | [\tilde{\sigma}, \hat{H}_L + \hat{H}_{NL}] | \psi \rangle = 2(\tilde{h}_L + \tilde{h}_{NL}) \times \tilde{s} = (\tilde{\omega}_L + \tilde{\omega}_{NL}) \times \tilde{s}.
\]

Defining
\[
\begin{pmatrix}
\tilde{\omega}_1 \\
\tilde{\omega}_2 \\
\tilde{\omega}_3
\end{pmatrix} =
\begin{pmatrix}
\cos \omega t & \sin \omega t & 0 \\
-\sin \omega t & \cos \omega t & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\omega_{NL1} \\
\omega_{NL2} \\
\omega_{NL3}
\end{pmatrix},
\]
with \( \omega \) being the frequency of light, we obtain after RWA the following nonlinear Bloch equations
\[
\begin{align*}
\dot{u} &= -\Delta v - \tilde{\omega}_2 v + \tilde{\omega}_2 w, \\
\dot{v} &= \Delta u + \Omega w + \tilde{\omega}_2 u - \tilde{\omega}_1 w, \\
\dot{w} &= -\Omega v - \tilde{\omega}_2 u + \tilde{\omega}_1 v.
\end{align*}
\]

Choosing \( (\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3) = (-A/2)v, (A/2)u, 2\epsilon w) \) we get
\[
\begin{align*}
\dot{u} &= -\Delta v - 2\epsilon uv + A/2 w, \\
\dot{v} &= \Delta u + \Omega w + 2\epsilon vu - A/2 w, \\
\dot{w} &= -\Omega v - A/2 u^2 - A/2 v^2.
\end{align*}
\]

Eqs. (3.51) are identical to Jaynes’s neoclassical Bloch equations [41] where \( \epsilon \) is the neoclassical Lamb shift and \( A \) is equal to Einstein’s coefficient of spontaneous emission. So let us check what kind of the KW-Schrödinger equation can lead to the neoclassical description.

Returning to the non-rotated reference frame we obtain
\[
(\omega_{NL1}, \omega_{NL2}, \omega_{NL3}) = (-A/2)s_2, (A/2)s_1, 2\epsilon s_3)
\]
and the nonlinear Schrödinger equation we are looking for is
\[
\frac{d}{dt} |\psi\rangle = \left( \hat{H}(\psi, \bar{\psi}) \right) |\psi\rangle = \hat{H}(\psi, \bar{\psi}) |\psi\rangle.
\]
If $\tilde{H}$ is some $(0,0)$-homogeneous function then the Hamiltonian is also $(0,0)$-homogeneous. However

$$\langle \psi | \tilde{H}(\psi, \bar{\psi}) | \psi \rangle = \langle \psi | \left( \tilde{H}(\psi, \bar{\psi}) 1 + \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle}{\langle \psi | \psi \rangle} \right) | \psi \rangle$$

(3.54)

cannot play the role of the Hamiltonian function for (3.53) with $A \neq 0$. For $A = 0$ the relevant KW Hamiltonian function is

$$E \langle \psi | \psi \rangle + \frac{1}{2} \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle^2}{\langle \psi | \psi \rangle}$$

(3.55)

and

$$\tilde{H}(\psi, \bar{\psi}) = E - \frac{1}{2} \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle^2}{\langle \psi | \psi \rangle^2}.$$  

(3.56)

\[ \square \]

### 3.4 Results of Single Measurements and Probabilities

Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\hat{H}$ a Hermitian operator. If $H = \langle \psi | \hat{H} | \psi \rangle$ is the associated observable, the values of single measurements of $H$ can be defined in at least three equivalent ways. Since in nonlinear QM the three options will lead to different results we will use here different names for each of them.

a) **Eigenvalue** $E$ of $H = \langle \psi | \hat{H} | \psi \rangle$ is the number satisfying for some *eigenstate* the equation

$$\hat{H} | \psi \rangle = E | \psi \rangle$$

(3.57)

or, in the “classical style”,

$$\frac{\partial H}{\partial \bar{\psi}_m} = E \bar{\psi}_m.$$  

(3.58)

b) **Diagonal values** are solutions of

$$\det(\hat{H} - E \mathbf{1}) = 0$$

(3.59)

or

$$\det\left( \frac{\partial^2 H}{\partial \bar{\psi}_m \partial \bar{\psi}_n} - E \delta_{mn} \right) = 0.$$  

(3.60)

c) Since eigenstates form a complete orthogonal set of vectors in $\mathcal{H}$ any solution of the Schrödinger equation $i \frac{d}{dt} | \psi \rangle = \hat{H} | \psi \rangle$ can be expressed as

$$\begin{pmatrix} \psi_1(t) \\ \vdots \\ \psi_N(t) \end{pmatrix} = \begin{pmatrix} \psi_1(0) e^{-i \omega_1 t} \\ \vdots \\ \psi_N(0) e^{-i \omega_N t} \end{pmatrix}$$

(3.61)

and the frequencies $\omega_n$ can be termed the *eigenfrequencies*. The three possibilities can be used for definitions of the results of single measurements also in nonlinear QM.

In nonlinear QM observables are not represented by linear operators so that we have to use the partial derivatives forms of the definitions. Let us begin the comparison of the various properties of the definitions with the following example.
Example 3.4 The so-called “simplest nonlinearity” considered in experiments designed as tests of Weinberg’s nonlinear QM corresponds to the following Hamiltonian function

\[ \langle \psi | \hat{H}_0 | \psi \rangle + \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle^2}{\langle \psi | \psi \rangle} \]  

(3.62)

where \( \hat{H}_0 = (E_0, 0) \). Denote \( \psi_1 = \sqrt{A_1} e^{i\alpha_1}, \psi_2 = \sqrt{A_2} e^{i\alpha_2} \). The eigenvalue condition reads

\[ \begin{align*}
E_1 + \epsilon \left[ 2 \frac{A_1 - A_2}{A_1 + A_2} - \left( \frac{A_1 - A_2}{A_1 + A_2} \right)^2 \right] \sqrt{A_1} &= \lambda \sqrt{A_1} \quad (3.63) \\
E_2 + \epsilon \left[ -2 \frac{A_1 - A_2}{A_1 + A_2} - \left( \frac{A_1 - A_2}{A_1 + A_2} \right)^2 \right] \sqrt{A_2} &= \lambda \sqrt{A_2}. \quad (3.64)
\end{align*} \]

For \( A_2 = 0 \) we find \( \lambda_+ = E_1 + \epsilon \), for \( A_1 = 0 \) \( \lambda_- = E_2 + \epsilon \). If \( A_1 \neq 0 \) and \( A_2 \neq 0 \) we get, for \( \epsilon \neq 0 \),

\[ \frac{A_1 - A_2}{A_1 + A_2} = \frac{E_2 - E_1}{4\epsilon} \]

which implies (\( A \) and \( B \) are positive)

\[ |E_2 - E_1| < 4|\epsilon| \]  

(3.65)

and

\[ \lambda_0 = \frac{1}{2} \left( E_1 + E_2 - \frac{1}{8\epsilon} (E_1 - E_2)^2 \right) \]

with the eigenvector

\[ |\psi\rangle = \left( \frac{1 + \frac{E_1 - E_2}{4\epsilon}}{1 - \frac{E_1 - E_2}{4\epsilon}} \right)^{-1/2} e^{i\alpha_1} \]

(3.66)

We find three eigenvalues although the dimension of the Hilbert space is two. The third eigenstate exists only for \( \epsilon \) satisfying the condition (3.65).

Of course, all the eigenstates are stationary. The eigenstate corresponding to \( \lambda_0 \) is not orthogonal to the remaining two eigenstates, which excludes the ordinary probability interpretation in terms of projectors.

Let us now calculate the diagonal values of this Hamiltonian function. The nonlinear Hamiltonian is

\[ \hat{H} = \begin{pmatrix} E_1 + \epsilon(8p^3 - 20p^2 + 16p - 3) & 8\epsilon |\psi_1|^2 |\psi_2|^2 \bar{\psi}_1 \psi_2 \\ 8\epsilon |\psi_1|^2 |\psi_2|^2 \bar{\psi}_1 \psi_2 & E_2 + \epsilon(-p^3 + 4p^2 + 1) \end{pmatrix} \]  

(3.67)

where the state is assumed normalized and \( p = |\psi|^2 \). The matrix is Hermitian and its eigenvalues (=diagonal values of \( \hat{H} \)) are

\[ E_\pm = \frac{1}{2} \left( E_1 + E_2 - 2\epsilon(8p^2 - 8p + 1) \right) \pm \left[ (E_1 - E_2)^2 - 8\epsilon \left\{ (E_1 - E_2)(4p^3 - 6p^2 + 4p - 1) \\ + 2\epsilon(20p^4 - 40p^3 + 28p^2 - 8p + 1) \right\} \right]^{1/2}. \]  

(3.68)
The diagonal values are in the nonlinear case \textit{functions} as opposed to eigenvalues which are numbers and the number of them is always equal to the dimension of the suitable Hilbert space.

The solution of the respective nonlinear Schrödinger equation is

\[
\begin{pmatrix}
\psi_1(t) \\
\psi_2(t)
\end{pmatrix} = \begin{pmatrix}
\psi_1(0)e^{-i(E_1 + 2\epsilon\langle \sigma_3 \rangle - \epsilon(\sigma_3)^2)t} \\
\psi_2(0)e^{-i(E_2 - 2\epsilon\langle \sigma_3 \rangle - \epsilon(\sigma_3)^2)t}
\end{pmatrix}
\]  

(3.69)

where the averages are integrals of motion. It follows that the eigenfrequencies are also state dependent functions but differ from the diagonal values.

Fig. 3.1 illustrates the mutual relations between the three definitions.

In linear QM mechanics observables correspond to averages and, at experimental level, to random variables. With any random variable one can associate its higher moments. The higher moments, on the other hand, can be used to deduce a probability interpretation of the theory. Therefore, one of the essential points of any generalized algebra of observables is the question of the representability of higher moments corresponding to some given observable. In linear QM the problem is solved by the spectral theorem.

The first interesting result concerning the nonlinear \textit{eigenvalues}, noticed by Weinberg in [10], is the following

Lemma 3.1 Let \( F \) and \( G \) be two \((1,1)\)-homogeneous observables possessing a common eigenstate \(|\psi\rangle\), i.e.

\[
\delta_{AA'} \frac{\delta F}{\delta \psi_A(\alpha)} = f\psi_A(\alpha) \quad \text{and c.c.,}
\]

(3.70)

\[
\delta_{AA'} \frac{\delta G}{\delta \psi_A(\alpha)} = g\psi_A(\alpha) \quad \text{and c.c.,}
\]

(3.71)

and let

\[
F \ast G = \delta_{AA'} \frac{\delta F}{\delta \psi_A(\alpha)} \frac{\delta G}{\delta \psi_A'(\alpha)}.
\]

(3.72)

Then

\[
\delta_{AA'} \frac{\delta(F \ast G)}{\delta \psi_A'(\alpha)} = fg\psi_A(\alpha) \quad \text{and c.c.}
\]

(3.73)

Proof:

\[
\delta_{BB'} \frac{\delta(F \ast G)}{\delta \psi_B'(\beta)} = \delta_{BB'} \frac{\delta}{\delta \psi_B'(\beta)} \left( \delta_{AA'} \frac{\delta F}{\delta \psi_A(\alpha)} \frac{\delta G}{\delta \psi_A'(\alpha)} \right)
\]

\[= \delta_{AA'} \delta_{BB'} \left( \frac{\delta^2 F}{\delta \psi_B'(\beta) \delta \psi_A(\alpha)} \frac{\delta^2 G}{\delta \psi_A'(\alpha)} \right)
\]

\[+ \delta_{BB'} \psi'(\beta) \left( \frac{\delta^2 F}{\delta \psi_B'(\beta) \delta \psi_A(\alpha)} \right)
\]

\[= g\delta_{BB'} \frac{\delta^2 F}{\delta \psi_B'(\beta) \delta \psi_A(\alpha)} \psi_A(\alpha)
\]

\[+ \delta_{BB'} \psi_A'(\alpha) \frac{\delta^2 G}{\delta \psi_B'(\beta) \delta \psi_A'(\alpha)}
\]

\[= fg\psi_B(\beta)
\]

26
where in the last step we have applied the (1,0)-homogeneity Euler condition for the
derivatives over $\bar{\psi}$.

The homogeneity implies that an average of $F$ in an eigenstate is equal to the respective eigenvalue. For finite dimensional Hilbert spaces we know also that a number of eigenvalues is not smaller than the dimension of the space (a result from the Morse theory). Moreover, the eigenvalues are critical points of averages. For averages defined on the whole Hilbert/projective space their maxima and minima must be found at critical points (averages defined on a finite dimensional projective space are smooth functions defined on a compact set). These facts suggest that the eigenvalues are correct candidates for the results of single measurements of the generalized observables. However, the following examples show that the problem is not that simple.

Example 3.5 The observable

$$A = \frac{\langle \psi | \psi \rangle^2}{\langle \psi | \sigma_3 | \psi \rangle}$$

satisfies the KW requirements. Its eigenvalues are $\pm 1$ so do not bound the averages. It follows that such singular observables must be excluded if we want to have the probability interpretation in terms of the eigenvalues. The algebra of observables would have to be restricted to (1,1)-homogeneous smooth functions defined on the whole Hilbert/projective space. This requirement is not very restrictive. Notice that no difficulties will appear if we will apply to $A$ the interpretation in terms of diagonal values or eigenfrequencies.

Example 3.6 In QM we often encounter observables whose eigenstates are degenerate. Consider now

$$H = E \langle \psi | \psi \rangle + \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle^3}{\langle \psi | \psi \rangle^2}.$$  

Its eigenvalues are $E \pm \epsilon$ and $E$. The number of them is hence greater than the dimension of the Hilbert space no matter how small $\epsilon$ is. Again, no problems appear if we apply the diagonal values or eigenfrequencies interpretation. Had we substituted the first term with some $(E_1,0,0,E_2)$, where $E_1 \neq E_2$, we would have obtained the third eigenvalue provided $|E_1 - E_2| \leq 6|\epsilon|$ so that the eigenvalue will not appear for sufficiently small $\epsilon$. This phenomenon seems to be related to the Kolmogorov-Arnold-Moser theorem where a dimension of an invariant torus may not change with nonlinear perturbation if a non-perturbed Hamiltonian system is nondegenerate and the perturbation is not too large [30].

Example 3.7 Let

$$H = \langle \psi | \hat{H}_0 | \psi \rangle + \epsilon \frac{\langle \psi | \sigma_3 | \psi \rangle^{2N}}{\langle \psi | \psi \rangle^{2N-1}}$$

where $\hat{H}_0 = (E_1,0,E_2)$. Two eigenvalues corresponding to the eigenstates $|\pm\rangle$ of $\sigma_3$ are $E_+ = E_1 + \epsilon$ and $E_- = E_2 + \epsilon$ so are shifted by the same amount. The third eigenstate occurs if

$$|\epsilon| > \frac{|E_1 - E_2|}{4N}.$$  

The third eigenvalue

$$E_0 = \frac{1}{2}(E_1 + E_2) + \epsilon (1 - 2N) \left( \frac{E_2 - E_1}{4N \epsilon} \right)^{\frac{1}{2}}.$$
tends to $E_1$ with $N \to \infty$. Although such nonlinearities are not “simplest”, they are sufficiently simple to be borne in mind in experiments whose ambition is to eliminate all nonlinear corrections to QM. □

Example 3.8 Consider

$$ H = E\langle\psi|\psi\rangle + \epsilon\frac{(\psi|\sigma_3|\psi)^2}{\langle\psi|\psi\rangle}.$$  \hfill (3.79)

The eigenvalues are $E$ and $E + \epsilon$. Probabilities calculated by means of

$$\langle H \rangle = (E + \epsilon)p_{E+\epsilon} + Ep_E$$  \hfill (3.80)

and the normalization of probability are

$$p_{E+\epsilon} = \langle \sigma_3 \rangle^2 \quad p_E = 1 - \langle \sigma_3 \rangle^2.$$  \hfill (3.81)

We know that $H * H$ has eigenvalues $(E + \epsilon)^2$ and $E^2$, so we can calculate the probabilities by means of $(H * H) = (E + \epsilon)^2p_{E+\epsilon} + E^2p_E$ and the normalization condition. Since

$$\langle H * H \rangle = E^2 + (4\epsilon^2 + 2E\epsilon)(\langle \sigma_3 \rangle)^2 - 3\epsilon^2(\langle \sigma_3 \rangle)^4$$  \hfill (3.82)

we obtain

$$p_{E+\epsilon} = \frac{(4\epsilon^2 + 2E\epsilon)(\langle \sigma_3 \rangle)^2 - 3\epsilon^2(\langle \sigma_3 \rangle)^4}{2E\epsilon + \epsilon^2}$$  \hfill (3.83)

which is, of course, inconsistent with the previous result. The troubles arise because of the nonassociativity of *. An analogous result was derived by Jordan [20], who showed that propositions cannot be defined consistently within the *-algebraic approach to Weinberg’s observables. We conclude that * cannot be applied for a definition of higher moments and the above lemma is useless from the viewpoint of the probability interpretation. □

Next candidate for a single result of a measurement of some $H$ is the diagonal value. We have already seen that the diagonal values are state dependent functions. There is no general guarantee that a diagonal value of an integral of motion is itself an integral of motion. In fact, as the KW equations are not necessarily integrable for $\dim\mathcal{H} > 2$, whereas the number of the diagonal values is equal to the dimension of $\mathcal{H}$, the diagonal values, for $N > 2$, can be time independent only for integrable systems. In addition, as diagonal values are roots of algebraic equations, they even do not have to be differentiable functions of states. Therefore, we cannot find for them any general equation of motion.

In spite of these disadvantages the diagonal values are in some respects superior to eigenvalues. If for two observables $A$ and $B$ the commutator $\hat{A}\hat{B} - \hat{B}\hat{A} = 0$, then $A$ and $B$ can be diagonalized simultaneously and the product $\hat{A}\hat{B}$ has eigenvalues values being products of the eigenvalues of $A$ and $B$. This leads to the following unique probability interpretation. Let $U(\psi, \bar{\psi})$ be the unitary transformation diagonalizing $\hat{H}$. Then

$$H = \sum_n E_n(\psi, \bar{\psi})|(U\psi)_n|^2.$$  \hfill (3.84)

The functions $p_n = |(U\psi)_n|^2/\langle\psi|\psi\rangle$ are probabilities resulting from the higher moments procedure.
Let us recall that for two observables \( F \ast G = F\bar{\ast}G \). The nonassociativity of \( \ast \) is reflected in the \( \bar{\ast} \)-algebra of observables in the following non-uniqueness of \( \bar{\ast} \). Although always

\[
(\psi|H\ldots\bar{\ast}H)|\psi\rangle = (\psi|\hat{H}\ldots\hat{H})\psi\rangle
\]

(3.85)

but for non-bilinear observables

\[
(H\bar{\ast}\ldots\bar{\ast}H) \neq \hat{H}\ldots\hat{H}.
\]

(3.86)

The eigenfrequencies approach can be applied only to systems that are close to integrable \[30\]. For finite dimensional Hamiltonian systems some general KAM theorems are known. It is known that components of state vectors exhibit in such a case, unless some resonant initial conditions occur, a quasi periodicity described by \[10\]

\[
\psi_k(t) = \sum_{n_1\ldots n_N} c_k(n_1\ldots n_N)e^{-i\sum_{\nu} n_{\nu}\omega_{\nu}t}
\]

(3.87)

where both the amplitudes and the frequencies are dependent on initial conditions, as shown in our first example in this subsection, and the sums run over all positive and negative integers. The form (3.87) of the solution is valid only if the frequencies are nonresonant, which means that \( \sum_{\nu} n_{\nu}\omega_{\nu} \neq 0 \). Now, following Weinberg, we note that both

\[
\langle\psi|\psi\rangle = \sum_k \sum_{n_1\ldots n_N} \sum_{n_1'\ldots n_N'} c_k(n_1\ldots n_N)\bar{c}_k(n_1'\ldots n_N')\exp\left(-i\sum_{\nu}(n_{\nu} - n_{\nu}')\omega_{\nu}t\right)
\]

(3.88)

and

\[
H = \sum_k \psi_k^* \frac{\partial H}{\partial \psi_k} = i \sum_k \psi_k \frac{d}{dt}\bar{\psi}_k
\]

\[
= \sum_k \sum_{n_1\ldots n_N} \sum_{n_1'\ldots n_N'} c_k(n_1\ldots n_N)\bar{c}_k(n_1'\ldots n_N')
\]

\[
\times \left(\sum_{\nu} n_{\nu}\omega_{\nu}\right)\exp\left(-i\sum_{\nu}(n_{\nu} - n_{\nu}')\omega_{\nu}t\right)
\]

are, on general grounds, time independent. The condition of nonresonance means that the only time independent terms that can contribute to \( \langle\psi|\psi\rangle \) and \( H \) are those with \( n_{\nu} = n_{\nu}' \). Thus

\[
\langle\psi|\psi\rangle = \sum_k \sum_{n_1\ldots} |c_k(n_1\ldots)|^2
\]

(3.89)

and

\[
H = \sum_k \sum_{n_1\ldots} |c_k(n_1\ldots)|^2 \sum_{\nu} n_{\nu}\omega_{\nu}.
\]

(3.90)

These formulas lead to the natural interpretation of the eigenfrequencies \( \sum_{\nu} n_{\nu}\omega_{\nu} \) as the results of single measurements and the respective normalized coefficients as probabilities. The interpretation can be naturally extended to other observables by assuming that all observables generate one parameter groups of canonical transformations \( \text{via} \) the Hamilton equations of motion. The same result would be also obtained by means of the “standard”
theory of measurements where we assume that the so-called pre-measurement is described by an interaction Hamiltonian function which is proportional to the observable measured, and that all the measuring procedures are based on measurements of observables that are bilinear like in linear QM. To be precise, we must remark that this procedure (chosen finally by Weinberg) is not unique either because there is no unique description of the composite “object+observer” system even if the observer is assumed to be linear (see remarks in the section devoted to observables).

To better understand the mutual relations between eigenvalues and eigenfrequencies it is interesting to compare the two notions in situations where a number of eigenvalues is different from this of eigenfrequencies.

Example 3.9 We know that

$$H = E\langle\psi|\psi\rangle + \epsilon\frac{\langle\psi|\sigma_3|\psi\rangle^3}{\langle\psi|\psi\rangle^2}. \quad (3.91)$$

has three eigenvalues $E \pm \epsilon$ and $E$. The eigenfrequencies are

$$E_1(\psi, \bar{\psi}) = E + \epsilon(3\langle\sigma_3\rangle^2 - 2\langle\sigma_3\rangle^3) \quad (3.92)$$

$$E_2(\psi, \bar{\psi}) = E + \epsilon(-3\langle\sigma_3\rangle^2 - 2\langle\sigma_3\rangle^3) \quad (3.93)$$

The three eigenvalues correspond to eigenvectors satisfying

$$\langle\sigma_3\rangle = \begin{cases} 0 & \text{for } \psi^0 \\ 1 & \text{for } \psi^1 \\ -1 & \text{for } \psi^{-1} \end{cases}$$

In this notation

$$E_1(\psi^0, \bar{\psi}^0) = E \quad \text{with probability } p_1(\psi^0, \bar{\psi}^0) = 1/2$$

$$E_2(\psi^0, \bar{\psi}^0) = E \quad \text{with probability } p_2(\psi^0, \bar{\psi}^0) = 1/2$$

$$E_1(\psi^1, \bar{\psi}^1) = E + \epsilon \quad \text{with probability } p_1(\psi^1, \bar{\psi}^1) = 1$$

$$E_2(\psi^1, \bar{\psi}^1) = E - 5\epsilon \quad \text{with probability } p_2(\psi^1, \bar{\psi}^1) = 0$$

$$E_1(\psi^{-1}, \bar{\psi}^{-1}) = E + 5\epsilon \quad \text{with probability } p_1(\psi^{-1}, \bar{\psi}^{-1}) = 0$$

$$E_2(\psi^{-1}, \bar{\psi}^{-1}) = E - \epsilon \quad \text{with probability } p_2(\psi^{-1}, \bar{\psi}^{-1}) = 1.$$ 

As expected even in an eigenstate there are two eigenfrequencies. The ones which are not equal to the eigenvalues occur with probabilities 0. □

A reader of the main Weinberg’s paper may be a little bit confused with what he finally understands as a result of a single measurement. A half of the paper suggests that this role will be played by eigenvalues, then he defines probabilities in terms of eigenfrequencies and, finally, in his analysis of a two-level atom he treats difference of eigenvalues as the energy difference while the difference of the eigenfrequencies is treated as the frequency of the emitted photon. It seems that from the viewpoint of the analysis above the eigenfrequency difference should be treated as the energy difference of atomic levels.

I think that each of the three possibilities lacks elegance and generality. Therefore in the last chapter of this work we will attempt to develop a version of nonlinear QM where observables will remain like in ordinary QM.
Finally, to close the variations on probability interpretation, let us outline some other general possibilities of the relations between averages, probabilities and values of single measurements.

We have stressed several times that in the Hilbertian formulation of QM the Hamiltonian function is not equal to the energy because the states are not assumed normalized. This freedom may be useful. Notice that observables in physics in general are identified by means of their algebraic properties or the way they change under an action of some symmetry group. It is reasonable to expect that a modification \( H \) of some bilinear

\[
H_L = \langle \psi | \hat{H}_0 | \psi \rangle
\]

will have the same transformation properties as \( H_L \). If we add some nonlinear term to \( H_L \) we will generally change its transformation properties. Therefore it seems we should consider the generalized observables of the form

\[
H = I(\psi, \bar{\psi})(\psi | \hat{H}_0 | \psi)
\]

where \( I \) is \((0,0)\)-homogeneous and depends on \( \psi \) and \( \bar{\psi} \) only via averages of Casimirs of the symmetry group and/or, perhaps, superselection operators like charge. It can be proved easily that \( I \) is in such a case an integral of motion, and we can define average values by

\[
\langle H \rangle = \frac{H(\psi, \bar{\psi})}{I(\psi, \bar{\psi})}\langle \psi | \psi \rangle.
\]

The evolution in each eigensubspace (“superselection sector”) of the Casimirs and the superselection operators would be linear, but with different “fundamental constants”. The linearity of the evolution would be violated only for states that are initially superpositions of states belonging to different sectors. It is obvious that the definition of averages would eliminate any difficulties with probabilities and eigenvalues.

Another general line of extensions would be to consider observables whose nonlinear parts vanish on normalized states. For instance the following modification of the Bialynicki-Birula–Mycielski equation

\[
\frac{d}{dt} |\psi\rangle = (\hat{H}_0 + b \ln \langle \psi | \psi \rangle) |\psi\rangle
\]

leads to two inequivalent definitions of averages. One is

\[
\frac{\langle \psi | \hat{H}_0 + b \ln \langle \psi | \psi \rangle | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{H(\psi, \bar{\psi})}{\langle \psi | \psi \rangle} = \frac{\langle \psi | \hat{H}_0 | \psi \rangle}{\langle \psi | \psi \rangle} + b \ln \langle \psi | \psi \rangle
\]

and the other is

\[
H \left( \frac{\psi}{||\psi||}, \frac{\bar{\psi}}{||\bar{\psi}||} \right) = \frac{\langle \psi | \hat{H}_0 | \psi \rangle}{\langle \psi | \psi \rangle}.
\]

The distinction between the generator of evolution and the observable energy may be one of the ways of introducing nonlinearities without affecting the linearity of averages. Such a distinction between Hamiltonian and energy appears, for example, in some versions of symplectic formulation of field theories [29], and the dynamics so derived would have to be treated as a dynamics with constraints [38].
Figure 3.1: Eigenvalues, eigenfrequencies and diagonal values are in the nonlinear case completely different concepts.
Chapter 4

COMPOSITE SYSTEMS AND NONLOCALITY

One of the most often quoted “impossibility theorems” about nonlinear QM is that any such theory must imply faster than light communications. We will argue in this and last chapters that this statement is too strong. In fact, we will show that there exists a rich class of theories where the mentioned phenomenon does not occur. As we shall see the theories must satisfy two conditions.

1. Their interpretation must not be based on the “collapse of a state vector” postulate. It means that any reasoning based on this postulate has to be regarded as unphysical. (In linear QM there exist such limitations. For example, all “counterfactual” problems like EPR paradox or Bell theorem can be eliminated trivially by rejecting reasonings involving alternative measurements.) Interpretations of QM that do not introduce the collapse (projection) postulate exist, to mention the many worlds interpretation [42].

2. Observables corresponding to subsystems must be functionals depending on density matrices of those subsystems.

In the last chapter we shall prove some general, quite strong theorems related to the latter condition. Here, we shall limit ourselves to several simple examples showing different ways of making quantum mechanical nonlocality “malignant”, to use the marvelous phrase of Bogdan Mielnik.

It seems it was Nicolas Gisin who noted for the first time that a nonlinear evolution can lead to a faster-than-light communication between two separated systems. His argument was the following [43]. Let $\mathcal{H}$ be a finite dimensional Hilbert space, $|\psi_i\rangle, |\phi_j\rangle \in \mathcal{H}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $\langle\psi_i|\psi_j\rangle = \langle\phi_i|\phi_j\rangle = \delta_{ij}$. Then the following lemma holds.

Lemma 4.1 If for some nonvanishing probabilities $x_i, y_j$,

$$\sum_i x_i |\psi_i\rangle \langle \psi_i| = \sum_j y_j |\phi_j\rangle \langle \phi_j|$$

(4.1)
then there exist two orthonormal bases \{\{\alpha_i\}\}, \{\{\beta_j\}\} in some Hilbert space \mathcal{H}' and the state
\[
|\chi\rangle = \sum_i \sqrt{x_i} |\psi_i\rangle \otimes |\alpha_i\rangle = \sum_j \sqrt{y_j} |\phi_j\rangle \otimes |\beta_j\rangle.
\]
(4.2)

\[\square\]

The proof can be found in [43]. The meaning of the lemma is that the two decompositions of the density matrix can be obtained by means of the EPR correlations. Indeed, we can take the density matrix \(|\chi\rangle\langle\chi|\) and trace out \(\mathcal{H}'\). The resulting density matrices are these appearing in the lemma. Assume now that we have a nonlinear evolution of pure states
\[
|\psi_i\rangle\langle\psi_i| \mapsto g_t(|\psi_i\rangle\langle\psi_i|).
\]
(4.3)

Then, in general,
\[
\sum_i x_i g_t(|\psi_i\rangle\langle\psi_i|) \neq \sum_j y_j g_t(|\phi_j\rangle\langle\phi_j|)
\]
(4.4)
even if initially the two decompositions were equal. The important assumption leading to the faster-than-light communication is that each of the pure state sub-ensembles evolves according to (4.3), even if the whole state is given by their convex combination. This
seems reasonable if we assume that the ensemble consists of the collapsed sub-ensembles. If we apply some “no-collapse” interpretation then the argument cannot be consistently formulated. However, it has to be stressed that the form of the evolution appearing in (4.4) can be derived from quite general assumptions. It has been shown in a rigorous way in the language of the theory of categories by Posiewnik [44] that this form is implied by Mielnik’s definition of mixed states as probability measures on the set of pure states. This remark has quite nontrivial consequences: The description of mixed states in terms of probability measures (Mielnik’s “convex approach”) combined with putative nonlinearity of the Schrödinger equation leads to faster that light communication. If one wants to get rid of such difficulties one cannot keep the figure of states convex.

Returning to Gisin, in his second paper [45] he considered an example of a nonlinear evolution taken from the Weinberg paper [10]. Consider an ensemble of pairs of spin-1/2 particles in the singlet state and assume that in one arm of the experiment the evolution is given by the Hamiltonian function
\[
\langle\psi|\sigma_3|\psi\rangle^2
\]
\[
\langle\psi|\psi\rangle.
\]
(4.5)

An experimenter in the other arm chooses between two settings of his Stern-Gerlach device and decomposes, by means of the EPR correlations, the ensemble in the other arm into subensembles corresponding to, say, spins up or down in the z direction, or spins up or down in some u direction tilted at 45° with respect to the z-axis. The evolution equation, up to an overall phase, is given by
\[
\frac{d}{dt} |\psi\rangle = 2\langle\sigma_3\rangle \sigma_3 |\psi\rangle.
\]
(4.6)

If the initial state is either up or down in the z direction the evolution is stationary and the average of \(\sigma_2\) is always zero. If the state is either up or down in the u direction the spin is precessing around the z axis but the sense of the rotation is opposite for the up
and down states. After a time of a quarter of the “Larmor period” the spin will have the same positive value of \(\langle \sigma_2 \rangle\). And this result can be detected by an observer in the arm with the nonlinearity.

As we can see Gisin in both of his proofs has completely ignored details of the description and the evolution of the “large” system. To be precise he should rather define a Hamiltonian function of the two arms, then solve the nonlinear Schrödinger equation for the whole system or calculate the evolution of the average of \(\sigma_2\) in one arm. This is the crucial point. We know already that the description of composite systems is not unique as long as we know only the evolution of pure states of subsystems. It will be shown below that the Gisin’s telegraph will never work if we describe the whole system in a correct way. It will be shown also that his proof is valid for the specific choice of the description proposed by Weinberg.

But before we shall pass on to the details of the calculations let us consider another “general proof” which was formulated by myself independently and simultaneously with this of Gisin [46]. Consider the situation depicted in Fig. 4.1 (for a detailed description of the polarizing Mach-Zehnder interferometer (PMZI) see Appendix 8.4.3). Kets representing, respectively, right- or left-hand polarized photons travelling to right along the horizontal “x” axis, evolve in the interferometer in the following way (the index \(y\) means that the photon travels along the vertical “y” axis):

\[
|+,x\rangle \xrightarrow{BS_1} i|+,y\rangle \xrightarrow{M_1} i|+,x\rangle \xrightarrow{\alpha} i e^{i\alpha}|+,x\rangle \\
|-,x\rangle \xrightarrow{BS_1} |-,x\rangle \xrightarrow{\lambda/2} |+,x\rangle \xrightarrow{M_2} |+,y\rangle \\
|-,x\rangle \xrightarrow{BS_2} \frac{1}{\sqrt{2}}|+,x\rangle + |+,y\rangle.
\] (4.7)

Here \(\alpha\) denotes a phase shifter which introduces a phase shift \(\alpha\), \(\lambda/2\) is a half-wave plate, \(BS_1\) is a polarizing beam splitter, \(BS_2\) a non-polarizing one, and \(M_1\) and \(M_2\) are mirrors. A transmitted beam is always phase-shifted by \(\pi/2\) with respect to the reflected one [47].

Let us consider now the following two cases:

1. A source produces a linear polarization state, say

\[
|\psi\rangle = \frac{1}{\sqrt{2}}(|+,x\rangle + |-,x\rangle).
\] (4.9)

The M-Z interferometer acts by means of (4.8) thus the whole state transforms into

\[
|\psi'\rangle = \frac{1}{2}(i(e^{i\alpha} + 1)|+,x\rangle + (1 - e^{i\alpha})|+,y\rangle).
\] (4.10)

We observe an interference between the two outgoing channels. In particular, for \(\alpha = 0\) the whole beam goes through the \(x\) channel.

2. A source produces a singlet state [47]

\[
|\phi\rangle = \frac{1}{\sqrt{2}}(|+,x\rangle|+,x\rangle + |-,x\rangle|-,x\rangle),
\] (4.11)
where the photons in a pair travel along the $x$ axis but in opposite directions. Now the whole system consists of two separated subsystems, $I$ and $II$, where in $I$ the photons enter the M-Z interferometer. This subsystem is described by a reduced density matrix

$$\rho_1 = \text{Tr}_2|\phi\rangle \langle \phi| = \frac{1}{2}|+x\rangle\langle +x| + \frac{1}{2}|−x\rangle\langle −x|$$  \hspace{1cm} (4.12)

and is therefore in a mixed state. No interference should be observed, as there is no coherence between the incoming right- or left-handed photons. Indeed, the transformation (4.8) transforms the density matrix into

$$\rho'_1 = \frac{1}{2}|+x\rangle\langle +x| + \frac{1}{2}|+y\rangle\langle +y|$$  \hspace{1cm} (4.13)

hence the intensities of beams in the outgoing channels are equal and independent of the optical retarder.

This property of the singlet state was considered by Pykacz and Žukowski [47] as the proof that a superluminal signalisation is impossible in this experiment. This is obviously true in linear quantum mechanics.

Let us, however, recalculate the problem in a different way.

We must obtain an identical result if we first transform the whole singlet state according to (4.8) then form a density matrix of the whole system, and, finally, take the trace over $II$.

The whole state transforms as follows:

$$|\phi'\rangle = \frac{1}{2}(ie^{i\alpha} |+x\rangle |+x\rangle - e^{i\alpha} |−x\rangle |+y\rangle + i |−x\rangle |+x\rangle + |−x\rangle |+y\rangle).$$  \hspace{1cm} (4.14)

Now we can explicitly see why there cannot appear interference in $I$: The linear dependence of the states which interfered in the linear polarization case is destroyed by orthogonality of their singlet state partners. The lack of the interference in $I$ is therefore a nonlocal phenomenon. The two explanations of the lack of interference are, unfortunately, experimentally indistinguishable. As long as ordinary linear devices are applied to the subsystem $II$ no possibility of a signalisation between $I$ and $II$ is possible, because the kets will always evolve according to some unitary transformation and remain orthogonal.

We know that nonlinear evolutions in a Hilbert space do not conserve scalar products (the “mobility phenomenon” [16]) and states that are initially orthogonal may lose their orthogonality during the course of the evolution. It is clear that if we will violate in $II$ the orthogonality of $|+x\rangle$ and $|−x\rangle$ because of some nonlinearity then the photons in $I$ will start to interfere and the interference will be the stronger the more violation of the orthogonality has been obtained. Putting it differently, tracing over a space where the evolution does not conserve scalar products may result in some “remains” of the traced out system in the reduced density matrix describing the remote, separated system.

This argument looks general but a careful reader has probably noticed that one additional assumption has been smuggled here: It is implicitly assumed that a nonlinear evolution in $\mathcal{H}$ can be extended to $\mathcal{H} \otimes \mathcal{H}'$ in such a way that a solution describing the composite system is of the form

$$|\phi\rangle \otimes |+\rangle + |\phi'\rangle \otimes |−\rangle$$  \hspace{1cm} (4.15)
where $|\phi\rangle$ and $|\phi'\rangle$ are solutions of the subsystem’s nonlinear evolution equation that are “in mobility”, that is whose scalar product is not conserved. This assumption is a strong limitation. In fact, it never occurs if separated systems are described in a correct way. However, the proposal of Weinberg does allow for such pathological solutions.

In what follows we shall formulate the arguments of Gisin and myself in a precise way within the framework of Weinberg’s approach.

Let us assume that we have two separated systems, I and II, described by Hamiltonian functions

$$H_1(\varphi, \bar{\varphi}) = E_1(\varphi), \quad H_2(\chi, \bar{\chi}) = E_2(\chi) + \epsilon \frac{(\chi|\sigma_3|\chi)^2}{\langle \chi|\chi \rangle}. \quad \text{(4.16)}$$

According to Weinberg, the Hamiltonian function of the whole I+II system is given by

$$H(\psi, \bar{\psi}) = \sum_l H_1(\varphi_{(l)}, \bar{\varphi}_{(l)}) + \sum_k H_2(\chi_{(k)}, \bar{\chi}_{(k)}), \quad \text{(4.17)}$$

where $\varphi_{k(l)} = \chi_{l(k)} = \psi_{kl}$.

A general solution of

$$\frac{d\psi_k}{dt} = \frac{\partial H}{\partial \psi_k} \quad \text{(4.18)}$$

with $H$ given by (4.17), in a basis diagonalizing $\sigma_3$, is

$$\psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} = \begin{pmatrix} \alpha x & \alpha y \\ \alpha x' & \alpha y' \end{pmatrix} = |\xi_1\rangle \otimes |\phi\rangle + |\xi_2\rangle \otimes |\psi\rangle \quad \text{(4.19)}$$

where $|\xi_1\rangle = \begin{pmatrix} \alpha \\xi_1 \end{pmatrix}$, $|\xi_2\rangle = \begin{pmatrix} \alpha \xi_2 \end{pmatrix}$, $|\phi\rangle = \begin{pmatrix} \xi_x \xi_y \end{pmatrix}$, $|\psi\rangle = \begin{pmatrix} \xi'_x \xi'_y \end{pmatrix}$ are some solutions

$$|\psi(t)\rangle = \begin{pmatrix} \frac{\sqrt{A}e^{-i(E_2+(A-B)(A+B)^{-2}(A+3B))t+i\delta}}{\sqrt{B}e^{-i(E_2-(A-B)(A+B)^{-2}(3A+B))t}} \end{pmatrix} \quad \text{(4.20)}$$

of

$$i\frac{d}{dt}|\psi\rangle = E_2\psi + \epsilon \left( 2\frac{(|\psi|\sigma_1|\psi\rangle|\sigma_3\psi\rangle}{(|\psi\rangle^2} - \frac{(|\psi|\sigma_3|\psi\rangle|^2}{(|\psi\rangle^2} \right) |\psi\rangle \quad \text{(4.21)}$$

resulting, via the Hamilton equations, from the Hamiltonian function

$$H_2(\psi, \bar{\psi}) = E_2(|\psi\rangle^2 + \epsilon \frac{|\psi|\sigma_3|\psi\rangle|^2}{(|\psi\rangle^2}. \quad \text{(4.22)}$$

Like in ordinary quantum mechanics, the state evolves here by means of separate evolutions of the “collapsed states”.

Now, if $|\psi\rangle$ is a solution of (4.21), then $|\phi\rangle = i\sigma_2|\psi\rangle$ is another solution corresponding to the same mean value of the energy, and we get

$$\langle \psi|\phi\rangle = 2i\sqrt{AB} \sin(4\epsilon(A-B)(A+B)^{-1}t + \delta). \quad \text{(4.23)}$$

For $\epsilon \neq 0$, $AB \neq 0$, $\delta = 0$, and $A \neq B$ the states are orthogonal for $t = 0$, but later lose their orthogonality, although the norm of each of them is conserved. This is an explicit example of the mobility phenomenon.
The Hamiltonian functions

\[ H_1(\psi, \bar{\psi}) = \sum_l H_1(\varphi(l), \bar{\varphi}(l)) \quad \text{and} \quad H_2(\psi, \bar{\psi}) = \sum_k H_2(\chi(k), \bar{\chi}(k)) \]  

(4.24)

commute because \( H_1(\psi, \bar{\psi}) = E_1(\psi, \bar{\psi}) \). Keeping in mind that there is no interaction part in \( H \), we conclude that there is no flow of energy between \( I \) and \( II \).

Let us now calculate the reduced density matrix of the linear subsystem \( I \). With the notation of (4.19), we find (we take normalized states)

\[ \rho_1 = \frac{1}{2}(\langle \xi_1 | \xi_1 \rangle + \langle \xi_2 | \xi_2 \rangle + \langle \psi | \phi \rangle |\xi_1 \rangle \langle \xi_2 | + \langle \phi | \psi \rangle |\xi_2 \rangle \langle \xi_1 |). \]  

(4.25)

The off-diagonal elements vanish in the linear theory (\( \epsilon = 0 \)) if initially the states are orthogonal, and we get a “fully mixed” state. For \( \epsilon \neq 0 \) these coherences oscillate with the mobility frequency given by (4.23). There therefore exist observables whose average values oscillate in this way. For example, the components of spin satisfy

\[ \langle \sigma_1 \rangle = \langle \sigma_3 \rangle = 0, \quad \text{but} \quad \langle \sigma_2 \rangle = \text{Im} \langle \phi | \psi \rangle = 2\sqrt{AB} \sin(4\epsilon(A - B)(A + B)^{-1}t + \delta). \]  

(4.27)

The result that has been derived is in agreement with our general analysis. Let us return to the example with the interference and assume that we introduce in \( II \) an optical device that induces the nonlinear evolution discussed above.

According to the superposition principle, the source emits pairs which are superpositions of the states corresponding to results (+, +) or (−, −). Once the “left” photon enters the non-linearity region, the superposition starts to evolve like the state (4.19). The right- and left-handed states start to rotate around each other because of the mobility phenomenon. Their partners do not interfere as long as the states are orthogonal, but the “less orthogonal” they are, the stronger the interference of their partners. It follows that the intensities in both of the outgoing channels of the M-Z interferometer in \( I \) start to change. In the beginning of the process the intensities were equal. At the moment of time in which the orthogonality is broken in the strongest way (let us accept, for inessential simplicity, that the two states are then proportional to each other; for reversible evolutions, the states can be arbitrarily close, but never proportional) the interference is strongest and, for \( \alpha = 0 \), the intensity in the \( y \) channel is 0, while in the \( x \) one it is maximal. After this extreme point, the mobility starts to make the states “more and more orthogonal”, and the interference becomes weaker and weaker. At the moment when the states are again orthogonal the intensities are again equal. And so on. The outgoing intensities must oscillate with the mobility frequency which agrees with the result of the calculation.

Another way of viewing this phenomenon can be the following. We know that the interference does not take place if one knows a path of a particle (say, a photon) in an interferometer. This observation inclined many people to propose the so-called delayed-choice experiments. It is sometimes believed that this interpretation means that the photons in one arm of the interferometer can interfere until a remote observer in the other arm makes an experiment which determines their paths in the interferometer. Such an indirect measurement could be possible if there were a one-to-one relationship
between a trajectory in the interferometer and the results of the remote measurements. This, on the other hand, means that the state of the whole system takes the form

$$\psi = |yes\rangle|\text{one route}\rangle + |no\rangle|\text{another route}\rangle,$$

where the kets $|yes\rangle$ and $|no\rangle$ must be orthogonal since correspond to different results of measurements made by the remote observer. But our earlier analysis proves that there is no interference in such a case. Therefore the question whether the experiment is delayed or not, is inessential. It is important that in principle one can make such an experiment.

Now, what happens if the photons traverse some non-linear device which introduces the mobility of initially orthogonal states $|yes\rangle$ and $|no\rangle$ before they reach the remote observer? The stronger the violation of their orthogonality, the less information one gets about the trajectories in question. In the extreme case of the mobility, say $|yes\rangle \to |yes\rangle, |no\rangle \to |yes\rangle$, the result “yes” tells us nothing about the trajectories in the other arm, and then one obtains the strongest interference. Again, the “more orthogonal” the states are, the greater the probability of a correct prediction of the route of the photons in the interferometer and the weaker should be the interference. By the way, this analysis helps to unify the well known “Feynman rules” for calculation of probabilities in QM: We add amplitudes in case we do not know trajectories, and probabilities in case we know them. The above analysis shows that we always add amplitudes, but the knowledge (in principle or actual) of trajectories is possible if and only if the states corresponding to different trajectories are correlated with mutually orthogonal states and then there is no interference terms in the probabilities because of the same reasons as there is no interference in our Mach-Zehnder interferometer.

Let us complete the analysis of this part with two comments.

1. The telegraph is based on the fact that an average of an observable in the linear system $I$ depends on parameters of the Hamiltonian function of the nonlinear system $II$. Recalling the form of the Poisson bracket equation for observables $A_1$ related to $I$

$$\dot{A}_1 = \{A_1, H_1 + H_2\}$$

we can see that such a dependence is possible if and only if

$$\{A_1, H_2\} \neq 0.$$  

It is an easy exercise to check that this condition is met indeed in the case of $A = \langle\psi|\sigma_2|\psi\rangle$ and $H_2$ from (4.17). It follows that the necessary an sufficient condition for elimination of telegraphs based on the mobility phenomenon is that any observables corresponding to separated systems must be in involution with respect to the Poisson bracket generating the evolution of observables. The observation that such observables are not necessarily in involution in the Weinberg approach is due to J. Polchinski [18].

2. If one of the systems is linear, then the telegraph based on the mobility phenomenon can be used for sending information only from the nonlinear system to the linear one. In Gisin’s telegraph one utilizes the remote preparation of mixtures entering the nonlinear system. Accordingly, this kind of telegraph works in the opposite direction hence cannot be equivalent to this based on the mobility phenomenon.
Let us describe now the Gisin’s telegraph [48]. Consider again the same Hamiltonian functions of the subsystems and the Weinberg’s description of the whole one (Eqs. (4.16) and (4.17)).

The sender is free to choose a basis in his Hilbert space by a rotation of his Stern-Gerlach device. Let the unitary matrix with unit determinant

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \bar{\alpha} \end{pmatrix}$$

describe the freedom in the choice of bases in the linear subsystem and the relation between the components of the whole state in the chosen basis and in the spin up-down one is given by

$$\begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha & \beta \\ -\beta & \bar{\alpha} \end{pmatrix} \begin{pmatrix} \psi_{++} & \psi_{+-} \\ \psi_{-+} & \psi_{--} \end{pmatrix}.$$  \hspace{1cm} (4.32)

The state is initially the singlet, which means that

$$\begin{pmatrix} \psi_{11}(0) & \psi_{12}(0) \\ \psi_{21}(0) & \psi_{22}(0) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha & \beta \\ -\beta & \bar{\alpha} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -\beta & \alpha \\ -\bar{\alpha} & -\bar{\beta} \end{pmatrix}.$$ \hspace{1cm} (4.33)

Finally the solution for the whole system is

$$|\psi\rangle = \frac{1}{\sqrt{2}} e^{-i(E_1+2\epsilon X^2)t} \begin{pmatrix} -\beta e^{-i2\epsilon Xt} & \alpha e^{i2\epsilon Xt} \\ -\bar{\alpha} e^{i2\epsilon Xt} & -\bar{\beta} e^{-i2\epsilon Xt} \end{pmatrix}$$ \hspace{1cm} (4.34)

where $X = |\beta|^2 - |\alpha|^2$. The reduced density matrix of the nonlinear system II reads

$$\rho^{II} = \frac{1}{2} + \text{Re}(\bar{\alpha}\beta) \sin(4\epsilon(|\alpha|^2 - |\beta|^2)t) \sigma_2 + \text{Im}(\bar{\alpha}\beta) \sin(4\epsilon(|\alpha|^2 - |\beta|^2)t) \sigma_1.$$ \hspace{1cm} (4.35)

The average of $\sigma_2$ in the nonlinear system is

$$\langle \sigma_2 \rangle = 2\text{Re}(\bar{\alpha}\beta) \sin(4\epsilon(|\alpha|^2 - |\beta|^2)t)$$ \hspace{1cm} (4.36)

hence depends on the choice of basis made in the linear one.

Notice that since the Hamiltonian function of the linear subsystem is proportional to $\langle \psi|\psi \rangle$ it is in involution with any observable. In particular

$$\{ \langle \sigma_2 \rangle, H_1 \} = 0$$ \hspace{1cm} (4.37)

which means that the commutability of observables corresponding to separated systems is not a sufficient condition for the nonexistence of faster-than-light telegraphs. The dependence of $\langle \sigma_2 \rangle$ on $\alpha$ and $\beta$ follows from the dependence of $H_2$ on these parameters. We have remarked already in the chapter dealing with observables that the Weinberg’s choice is basis dependent: By a change of basis in I we can change a value of energy in II.

It becomes clear now under what conditions this kind of pathology can be eliminated. Consider a density matrix $\rho$ describing a “large” system. A change of basis in a subsystem I is represented by the unitary transformation

$$\rho \mapsto U_I \otimes 1_{II} \rho U_I^{-1} \otimes 1_{II}.$$ \hspace{1cm} (4.38)
Any function of $\rho$ that can be written in a form of a series

$$\hat{f}(\rho) = \sum_k s_k \rho^k$$

transforms in the same way. It follows that all expressions like

$$\text{Tr} (\hat{f}(\rho) 1_I \otimes \hat{A}_{II})$$

are $U_I$-independent. Also all functionals depending on reduced density matrices of subsystems are independent of changes of bases outside of those subsystems. In the last chapter we shall prove a general theorem stating that if two observables depend on reduced density matrices of different subsystems then they are in involution with respect to a large class of brackets. In such a case both kinds of telegraphs will be eliminated.

The last point that has to be explained is the question of uniqueness of the description of subsystems. We have shown already that a knowledge of observables on pure states of subsystems does not determine their form if the subsystems are correlated with something else. Here we will show that observables that differ only on mixed states will, in general, generate different evolutions of subsystems.

Consider our “canonical” example (4.16) but rewritten in a form including density matrices. For the sake of clarity let us also substitute a more general matrix

$$\hat{\epsilon} = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix}$$

for $\sqrt{\epsilon}\sigma_3$.

We have now an infinite number of possibilities. The simplest nontrivial ones are

$$H_1(\rho_1) = E_1 \text{Tr } \rho_1, \quad H_2(\rho_2) = E_2 \text{Tr } \rho_2 + \frac{\text{Tr } (\rho_2 \hat{\epsilon})^2}{\text{Tr } \rho_2}$$

and

$$H_1(\rho_1) = E_1 \text{Tr } \rho_1, \quad H_2(\rho_2) = E_2 \text{Tr } \rho_2 + \frac{\text{Tr } (\rho_2 \hat{\epsilon})^2}{\text{Tr } \rho_2} \frac{\text{Tr } (\rho_2^2)}{(\text{Tr } \rho_2)^2}.$$  (4.42)

Let $\rho_{kl} = \sum_m \bar{\psi}_{mk} \psi_{ml}$ denote the components of the reduced density matrix $\rho_2$. The Hamiltonian function of the whole system is

$$H(\rho) = H_1(\rho_1) + H_2(\rho_2).$$

The two forms of $H_2$ lead, respectively, to the following two evolution equations

$$\frac{d}{dt} \rho_{kl} = -2i \frac{\text{Tr } \rho_2 \hat{\epsilon}}{\text{Tr } \rho_2} \rho_{kl} (\epsilon_k - \epsilon_l)$$

and

$$\frac{d}{dt} \rho_{kl} = -2i \frac{\text{Tr } \rho_2 \hat{\epsilon} \text{Tr } (\rho_2^2)}{\text{Tr } \rho_2 (\text{Tr } \rho_2)^2} \rho_{kl} (\epsilon_k - \epsilon_l)$$

where the expressions involving traces are integrals of motion. The equations are different.

There is only one way out of the above dilemma: The description must from the outset be given in terms of density matrices. This remark will be the departure point for the new version of nonlinear QM proposed in the last chapter.
Chapter 5

A TWO-LEVEL ATOM IN NONLINEAR QM

Since the present-day ion trap experimental techniques enable unprecedentedly precise measurements of quantum optical properties of atoms, the question of a correct description of a two-level system in nonlinear QM has acquired particular importance. The aim of the analysis below is to clarify some elementary features of the nonlinear formalism involved in calculations of such problems. Therefore, the content of this chapter should not be understood as a complete, unique solution or systematization of all the questions encountered. Still, I hope that I have managed to point out some elements essential for correct calculations in practical, experimental situations.

5.1 A Fully Quantum Approach

In linear QM a two-level system is mathematically equivalent to a spin-1/2 nonrelativistic particle. All the examples discussed in this work were based on a two-dimensional Hilbert space and the reader may have a feeling that they could be applied equally well to both spin-1/2 particles and two-level atoms. This kind of conviction has been shared by all the authors (including Weinberg himself) dealing with theoretical and experimental aspects of nonlinear QM. The main result of this chapter, as we shall see later, is that, paraphrasing G. Orwell’s words, in nonlinear QM all two-level systems are two-level but some of them are more two-level than others.

We shall assume that the supposed nonlinearity is of purely atomic origin. This means that we shall consider the atom as a nonlinear subsystem of the larger “atom+field” system where both the field and interaction Hamiltonian functions are linear in density matrix. (The papers discussing the problem can be divided into two groups: Either the authors do not care about the description of the “atom+field” composite system or treat it in the way proposed by Weinberg. We know that none of them can be correct unless the atom and the field are in a product state, which is typical for semiclassical treatments. It follows that no really quantum description of the problem has been given as yet.) We shall assume also the dipole and rotating wave approximations. To clarify the role of the latter we shall treat both $\Delta m = 0$ and $\Delta m = \pm 1$ cases.
We begin with the form of the atomic Hamiltonian function. The simplest one (at least from the point of view of simplicity of calculations) is

\[ H_{AT}[\rho] = \text{Tr} \rho \hat{H}_L + \frac{(\text{Tr} \rho \hat{\epsilon})^2}{\text{Tr} \rho} \]  

(5.1)

where \( \hat{H}_L \) is the linear Hamiltonian of the atom and \( \hat{\epsilon} \) is an operator commuting with \( \hat{H}_L \). Assuming that we consider the atom in a pure state \( \rho = |\psi\rangle\langle\psi| \) we find that a general solution of the resulting nonlinear Schrödinger equation is (in this chapter we shall use the ordinary units with \( \hbar \neq 1 \))

\[ \psi_k(t) = \psi_k(0) \exp \left[ -\frac{i}{\hbar} \left( E_k + 2\langle \hat{\epsilon} \rangle \epsilon_k - \langle \hat{\epsilon} \rangle^2 \right) t \right]. \]  

(5.2)

The averages of \( \hat{\epsilon} \) are integrals of motion and, of course, depend on all nonvanishing components of \( |\psi\rangle \). This is an important point. In the analysis of a coupling between the atom and an external electromagnetic field we meet two difficulties. First of all we have to decide which states will be involved in the absorption-emission process. In linear QM the situation is simple: We take two stationary states of the noninteracting atom. In nonlinear QM the atomic nonlinearity may lead to stationary states that are not orthogonal between one another. Atomic creation and annihilation operators corresponding to such levels cannot satisfy ordinary anticommutation relations.

To avoid such complications the interaction term we shall choose will be defined in the ordinary way, that is, in terms of creation and annihilation operators corresponding to the levels of the linear Hamiltonian.

Second, we cannot a priori restrict the atomic Hilbert space to two dimensions because a \( k \)-th eigenfrequency depends on amplitudes of all other components of \( |\psi\rangle \) and the evolution cannot be naturally “cut into \( N \)-dimensional pieces”. Of course, we cannot also assume that only these \( \epsilon_k \) are nonvanishing which we are interested in (a probability that in a given case we will find just those only nonvanishing ones is 0). Therefore, to be perfectly consistent we should resign from the two level approximation in the interaction term. We shall not, however, consider such complications although this approximation will further restrict the physical validity of the calculations presented below.

Let \( b_k, b_k^\dagger \) be the \( k \)-th level atomic annihilation and creation operators satisfying the fermionic algebra \( [b_k, b_l^\dagger]_+ = \delta_{kl} \) and \( a, a^\dagger \) be the annihilation and creation operators of a monochromatic photon field whose frequency is \( \omega \). The choice of the creation-annihilation operators language leads naturally to the following Hamiltonian function of the whole “atom+field” composite system

\[ H_{A+F}(\psi, \bar{\psi}) = \langle \psi | \sum_k \hbar \omega_k b_k^\dagger b_k + \hbar \omega a^\dagger a + \frac{i\hbar}{2} (b_k^\dagger b_{k+1} - b_{k-1} b_k^\dagger) | \psi \rangle \]  

\[ + \frac{\langle \psi | \sum_k \epsilon_k b_k^\dagger b_k | \psi \rangle^2}{\langle \psi | \psi \rangle}. \]  

(5.3)

The state in the Fock basis is \( |\psi\rangle = \sum_{kn} \psi_{kn} |k\rangle |n\rangle \). The nonlinear term is therefore equivalent to

\[ \frac{\langle \psi | \hat{\epsilon} \otimes 1_F | \psi \rangle^2}{\langle \psi | \psi \rangle} = \frac{(\text{Tr} \rho_{AT} \hat{\epsilon})^2}{\text{Tr} \rho_{AT}} \]  

(5.4)
which seems natural but, as we have seen before, is only one out of the whole variety of inequivalent possibilities. Such a description is consistent with the definition of the atomic Hamiltonian function (5.1) (in this sense it is unique) and free from any “malignant” nonlocalities. On the other hand it is in a mean-field style; we have remarked in the section on observables that

\[
\frac{\text{Tr} \rho A \hat{\rho} A \hat{\rho}}{\text{Tr} \rho A} \tag{5.5}
\]

would look more “fundamental”. Anyway, although the conclusions drawn on a basis of such an analysis are limited in their generality, some choice has to be made and calculations with (5.5) would be more complicated.

The Hamiltonian resulting from (5.3) is

\[
\hat{H}_{A+F} = \sum_k \hbar \omega_k \hat{b}_k^\dagger \hat{b}_k + \hbar \omega a^\dagger a + \frac{i\hbar q}{2} \left( b_{11}^\dagger a - a^\dagger b_{11}^\dagger \right)
\]

\[+ 2 \frac{\langle \psi \mid \sum_k \epsilon_k \hat{b}_k^\dagger \hat{b}_k \mid \psi \rangle}{\langle \psi \mid \psi \rangle} \sum_k \epsilon_k \hat{b}_k^\dagger \hat{b}_k - \frac{\langle \psi \mid \sum_k \epsilon_k \hat{b}_k^\dagger \hat{b}_k \mid \psi \rangle^2}{\langle \psi \mid \psi \rangle^2} \tag{5.6}
\]

and the nonlinear Schrödinger equation is

\[i \hbar \dot{\psi}_1 = \left( \hbar \omega_1 + 2 \sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2 \sum_{lm} \mid \psi_{lm} \mid^2 - \epsilon_1 - \left( \frac{\sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2}{\sum_{lm} \mid \psi_{lm} \mid^2} \right)^2 + n \hbar \omega \right) \psi_1 \]

\[= \frac{i}{2} \hbar q \sqrt{\alpha} \psi_{2,n-1} \]

\[i \hbar \dot{\psi}_2 = \left( \hbar \omega_2 + 2 \sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2 \sum_{lm} \mid \psi_{lm} \mid^2 - \epsilon_2 - \left( \frac{\sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2}{\sum_{lm} \mid \psi_{lm} \mid^2} \right)^2 + n \hbar \omega \right) \psi_2 \]

\[= \frac{i}{2} \hbar q \sqrt{\alpha} \psi_{2,n+1} \]

\[\vdots \]

\[i \hbar \dot{\psi}_k = \left( \hbar \omega_k + 2 \sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2 \sum_{lm} \mid \psi_{lm} \mid^2 - \epsilon_k - \left( \frac{\sum_{lm} \epsilon_l \mid \psi_{lm} \mid^2}{\sum_{lm} \mid \psi_{lm} \mid^2} \right)^2 + n \hbar \omega \right) \psi_k \]

for \( k > 2. \) \tag{5.7}

Writing \( \psi_n = \hat{A}_n \exp(-i \alpha_n / \hbar) \) we find that for \( k > 2 \) \( \hat{A}_n = \text{const} \) for all \( n. \) Since also \( \| \psi \| \) is time independent (hereafter we put \( \| \psi \| = 1 \)) it follows that for \( k > 2 \) the exponents depend on time also via \( \langle \psi | \hat{e}_{12} | \psi \rangle = \sum_n (\epsilon_1 \mid \psi_{1n} \mid^2 + \epsilon_2 \mid \psi_{2n} \mid^2) \) whose explicit form has to be determined. Decomposing

\[
\langle \psi | \epsilon_{12} | \psi \rangle = \frac{1}{2} \langle \psi | \epsilon_1 + \epsilon_2 | (b_1^\dagger b_1 + b_2^\dagger b_2) | \psi \rangle + \frac{1}{2} \langle \psi | \epsilon_1 - \epsilon_2 | (b_1^\dagger b_1 - b_2^\dagger b_2) | \psi \rangle
\]

\[= (\epsilon_1 + \epsilon_2) \langle \psi | R_1 | \psi \rangle + (\epsilon_2 - \epsilon_1) \langle \psi | R_3 | \psi \rangle, \tag{5.8}
\]

where the first expression is an integral of motion, we see that the problem reduces to calculating \( \langle \psi | R_3 | \psi \rangle \) which is one half the atomic inversion. Denoting \( \epsilon = \epsilon \otimes 1_F, \) the total Hamiltonian can be decomposed now into two parts

\[
\hat{H}_1 = \hbar (\omega_2 - \omega_1) R_3 + 2(\epsilon_2 - \epsilon_1) \langle \psi | \epsilon | \psi \rangle R_3 + \hbar \omega a^\dagger a
\]
we finally obtain
\[ A \]
where \( A \) is a constant depending on initial conditions. Denoting further
\[ \langle \psi | e | \psi \rangle = 2 \epsilon_0 \langle \psi | a^\dagger a | \psi \rangle + \hbar \epsilon_0 a^\dagger a + \frac{i \hbar q}{2} (R_+ a - a^\dagger R_-), \]
where \( \omega_0 = \omega_2 - \omega_1 \) and \( \epsilon_0 = \epsilon_2 - \epsilon_1 \). To explicitly distinguish between initial conditions and dynamical objects we shall decompose \( \langle \psi | e | \psi \rangle \) as follows
\[ \langle \psi | e | \psi \rangle = \sum_{k > 2, n} \epsilon_k | \psi_{kn} \rangle \langle \psi_{kn} | + (\epsilon_1 + \epsilon_2) \langle \psi | R_0 | \psi \rangle + \epsilon_0 \langle \psi | R_3 | \psi \rangle := A + \epsilon_0 \langle \psi | R_3 | \psi \rangle, \]
where \( A \) is a constant depending on initial conditions. Denoting further
\[ \hbar \omega_0 + 2 \epsilon_0 A = \hbar \omega_0', \quad 2 \epsilon_0 = \hbar \epsilon \]
we finally obtain
\[ \hat{H}_1 = \hbar \omega_0' R_3 + \hbar \epsilon \langle \psi | R_3 | \psi \rangle R_3 + \hbar \epsilon_0 a^\dagger a + \frac{i \hbar q}{2} (R_+ a - a^\dagger R_-). \]
It is clear now what is the actual meaning of the two-level approximation in nonlinear QM. The evolution of the atomic operators \( R_j \) is generated by a two-dimensional Hamiltonian, in analogy to the linear case, but the parameters of \( \hat{H}_1 \) depend on components of the wave function corresponding to the levels being outside of the two-dimensional Hilbert space. On the other hand, the phases of the remaining components depend on the average of the atomic inversion of the two levels.

(It has to be stressed here that one can consider a quantum system whose Hamiltonian function is a sum of a linear term and some nonlinearity which involves only spinor components of the wave function. Formally, such a case would be equivalent to a composite system whose constituents are a scalar particle that evolves according to the laws of the linear QM and some \textit{nonlinear} two-level system noninteracting with the particle. We know that various solutions of the nonlinear Schrödinger equation, including the product one, will exist and no components other than the spinor ones will be involved in the nonlinear part of the evolution. Systems with nonlinearities of this kind would be “truly two-level” as opposed to the systems which are two-level in the sense specified above.)

Our next task is to find and solve an equation for the atomic inversion. The Poisson bracket evolution equation reads
\[ i \hbar \frac{d}{dt} \langle \psi | R_3 | \psi \rangle = \langle \psi | [R_3, H_1] | \psi \rangle = \frac{q}{2} \langle \psi | R_+ a + a^\dagger R_- | \psi \rangle \]
so is just like in linear QM [49]. Following the notation of [49] the second derivative is found equal

\[
\frac{d^2}{dt^2} \langle \psi | R_3 | \psi \rangle = \frac{1}{2} \frac{q}{\hbar} \langle \psi | [R_+ a + a^\dagger R_-] H_1 | \psi \rangle \\
= -q^2 \langle \psi | R_3 \left( \hat{N} + \frac{1}{2} \right) | \psi \rangle \\
+ \frac{iq}{2} \left( \Delta' + \epsilon \langle \psi | R_3 | \psi \rangle \right) \langle \psi | R_+ a - a^\dagger R_- | \psi \rangle 
\]

(5.15)

where \( \hat{N} = R_+ + a^\dagger a \) and \( \Delta' = \omega_0 - \omega \). Define

\[
B = \sum_{k>2, n} \hbar \omega_k |\psi_k n|^2 + \hbar (\omega_1 + \omega_2) \langle \psi | R_0 | \psi \rangle .
\]

(5.16)

\( \langle \psi | \hat{N} | \psi \rangle \), like in the linear case, is constant. In order to get rid of the average in the last row of (5.15) we rewrite the whole Hamiltonian function as follows

\[
H_{A+F} = \hbar \omega_{A+F} = \hat{A}^2 + B + \hbar \omega \langle \psi | \hat{N} | \psi \rangle \\
+ \hbar \Delta' \langle \psi | R_3 | \psi \rangle + \epsilon^2_0 \langle \psi | R_3 | \psi \rangle^2 + \frac{i\hbar q}{2} \langle \psi | R_+ a - a^\dagger R_- | \psi \rangle .
\]

(5.17)

Denoting \( w = 2 \langle \psi | R_3 | \psi \rangle \), \( \omega_A = \hat{A}^2 / \hbar \), \( \omega_B = B / \hbar \), \( \omega_{RW A} = \omega_{A+F} - \omega_B \) we finally get

\[
\ddot{w} = 2\Delta' (\omega_{RW A} - \omega \langle \hat{N} \rangle - \omega_A) \\
+ \left( \epsilon (\omega_{RW A} - \omega \langle \hat{N} \rangle - \omega_A) - \Delta' \right) w - q^2 \langle \psi | R_3 \left( \hat{N} + \frac{1}{2} \right) | \psi \rangle \\
- \frac{3}{4} \epsilon \Delta' w^2 - \frac{\epsilon^2}{8} w^3 .
\]

(5.18)

We have met here the characteristic inconvenience of the Poisson bracket formalism of nonlinear QM: The nonexistence of the Heisenberg picture. In the linear case we can solve the Jaynes-Cummings problem completely, independently of any particular initial conditions for states. Here the term \( \langle \psi | R_3 \left( \hat{N} + \frac{1}{2} \right) | \psi \rangle \) involves correlations between the atom and the field and I have not managed to express it solely in terms of constants and \( w \) unless the state is an eigenstate of \( \hat{N} \), or a semiclassical decorrelation is assumed. So let initially the state of the system be a common eigenstate of \( R_3 \) and \( a^\dagger a \) with respective eigenvalues \( n' \) and \( n \). The atomic inversion satisfies then the general elliptic equation [50]

\[
\ddot{w} = 2\Delta' (\Delta' n' + \frac{1}{8} \epsilon) + \left( \epsilon (\Delta' n' + \frac{1}{8} \epsilon) - \Delta' - \frac{q^2}{2} (N + \frac{1}{2}) \right) w - \frac{3}{4} \epsilon \Delta' w^2 - \frac{\epsilon^2}{8} w^3 .
\]

(5.19)

The general method of solving equations of the form

\[
\ddot{y} = A + By + Cy^2 + Dy^3
\]

(5.20)

is the following. After multiplying (5.20) by \( \dot{y} \), integrating over \( t \), and simplifying the resulting expression we obtain

\[
\dot{y}^2 = a + 2Ay + By^2 + \frac{2}{3} Cy^3 + \frac{1}{2} Dy^4
\]

(5.21)
where $a$ is an arbitrary constant, which has to be determined from initial conditions. We rewrite the equation in the more convenient form

$$
\dot{y}^2 = a + by + cy^2 + dy^3 + ey^4
$$

and seek a transformation $z = z(y)$ by means of which the equation reduces to the standard elliptic form

$$
z^2 = (1 - z^2)(1 - k^2 z^2) := F^2(z).
$$

Writing

$$
\dot{y}^2 = h^2(y - \alpha)(y - \beta)(y - \gamma)(y - \delta) := G^2(y)
$$

and defining

$$
z^2 = \frac{(\beta - \delta)(y - \alpha)}{(\alpha - \delta)(y - \beta)},
$$

$$
k^2 = \frac{(\beta - \gamma)(\alpha - \delta)}{(\alpha - \gamma)(\beta - \delta)},
$$

$$
M^2 = \frac{(\beta - \delta)(\alpha - \gamma)}{4}
$$

we obtain

$$
\frac{1}{F(z)} \frac{dz}{dt} = \frac{M}{G(z)} \frac{dy}{dt} = hM
$$

hence

$$
z = sn(hM(t - t_0), k).
$$

The applicability of the method is up to problems with vanishing of denominators in the above fractions. In such a case there exist other transformations analogous to these given above [50].

Although we could try to find a general expression for the atomic inversion following from (5.19) it seems more instructive to make here some simplifying assumptions. First of all we can take the “two-level initial conditions”, that is, assume that the initial state of the system is such that the only nonvanishing components of the wave function are these with $k = 1, 2$. Then $h\Delta' = h\Delta + \epsilon_2^2 - \epsilon_1^2$ where $\Delta = \omega_0 - \omega$. (Let me remark here that in most of the papers dealing with two-level systems (cf. [41, 8]) their authors assumed for the “simplest” nonlinearities $\epsilon_2 = 0$ which seemed to suggest that the nonlinearity must shift the resonant frequency. As we can see, the more symmetric “$\sigma_3$” choice does not change the detuning.) Further, choosing the “detuning” $\Delta' = 0$ and denoting $\epsilon^2/8 = 2\varsigma^2$ we get

$$
\ddot{w} = (2\varsigma^2 - \Omega^2)w - 2\varsigma^2 w^3.
$$

This equation can be solved immediately. For example, with the initial condition $w(0) = -1$ we find

$$
w(t) = \begin{cases} 
-cn(\Omega t, \varsigma/\Omega) & \text{for } \Omega > \varsigma \\
-sech(\Omega t) & \text{for } \Omega = \varsigma \\
-dn(\varsigma t, \Omega/\varsigma) & \text{for } \Omega < \varsigma
\end{cases}
$$

The result is analogous to this of Wódkiewicz and Scully [41] who chose the Bloch equations approach.
Simplifying the fractions we obtain

\[ H_{A+F}(\psi, \bar{\psi}) = \langle \psi \rangle \sum_k h \omega_k b_k^\dagger b_k + h \omega a^\dagger a + \frac{ihq}{2} (b_2^\dagger b_1 a - a^\dagger b_2) \psi + \sum_n \frac{\langle \psi | \sum_k \epsilon_k b_k^\dagger \hat{P}_n | \psi \rangle^2}{\langle \psi | \hat{P}_n | \psi \rangle}. \]

where \( \hat{P}_n \) project on \( n \)-photon states. Assume now that initially the state of the whole system has only one nonvanishing component \( \psi_{11} \) (one photon and the atom in the ground state). It follows that only \( \psi_{11} \) and \( \psi_{20} \) will appear in the nonlinear Schrödinger equation. Our “nonmalignant” choice leads to the Schrödinger equation of the form

\[ i\hbar \dot{\psi}_{11} = \left\{ h \omega_1 + 2 \frac{\epsilon_1 |\psi_{11}|^2 + \epsilon_2 |\psi_{20}|^2}{|\psi_{11}|^2 + |\psi_{20}|^2} \epsilon_1 - \left( \frac{\epsilon_1 |\psi_{11}|^2 + \epsilon_2 |\psi_{20}|^2}{|\psi_{11}|^2 + |\psi_{20}|^2} \right)^2 + h \omega \right\} \psi_{11} - \frac{i}{2} h q \psi_{20} \]

\[ i\hbar \dot{\psi}_{20} = \left\{ h \omega_2 + 2 \frac{\epsilon_1 |\psi_{11}|^2 + \epsilon_2 |\psi_{20}|^2}{|\psi_{11}|^2 + |\psi_{20}|^2} \epsilon_2 - \left( \frac{\epsilon_1 |\psi_{11}|^2 + \epsilon_2 |\psi_{20}|^2}{|\psi_{11}|^2 + |\psi_{20}|^2} \right)^2 \right\} \psi_{20} + \frac{i}{2} h q \psi_{11}. \]

while the Weinberg’s one leads to

\[ i\hbar \dot{\psi}_{11} = \left\{ h \omega_1 + 2 \frac{\epsilon_1 |\psi_{11}|^2}{|\psi_{11}|^2} \epsilon_1 - \left( \frac{\epsilon_1 |\psi_{11}|^2}{|\psi_{11}|^2} \right)^2 + h \omega \right\} \psi_{11} - \frac{i}{2} h q \psi_{20} \]

\[ i\hbar \dot{\psi}_{20} = \left\{ h \omega_2 + 2 \frac{\epsilon_2 |\psi_{20}|^2}{|\psi_{20}|^2} \epsilon_2 - \left( \frac{\epsilon_2 |\psi_{20}|^2}{|\psi_{20}|^2} \right)^2 \right\} \psi_{20} + \frac{i}{2} h q \psi_{11}. \]

Simplifying the fractions we obtain

\[ i\hbar \dot{\psi}_{11} = (h \omega_1 + \epsilon_1^2) \psi_{11} - \frac{i}{2} h q \psi_{20} \]

\[ i\hbar \dot{\psi}_{20} = (h \omega_2 + \epsilon_2^2) \psi_{20} + \frac{i}{2} h q \psi_{11} \]

which are linear and the only modification with respect to ordinary QM is that the energy levels of the atom are the nonlinear eigenvalues corresponding to the atom in the absence of radiation. In addition, the “\( \sigma_3 \)” nonlinearity satisfies \( \epsilon_1^2 = \epsilon_2^2 \) so that for this choice of the “simplest” nonlinearity neither the energy difference nor the shape of the inversion’s oscillation would be affected, whereas we know already that the correct description leads to elliptic oscillations even for \( w(0) = -1 \) and \( N + \frac{1}{2} = 1 \).
In the calculations in this section we have not needed any assumption about the “smallness” of the nonlinearity. Moreover, as we have shown before, it is not evident what is actually meant by a small nonlinearity. This lack of uniqueness is related to the fact that there exist singular nonlinearities that are negligible in the lack of correlations, but can become dominant if the nonlinear system in question correlates with something else. Anyway, disregarding these subtle points it seems reasonable to expect that a physical nonlinearity, if any, should be in ordinary situations small in some sense. Therefore, it becomes interesting to understand in what respect the solutions we have found depend on approximations. In particular the role of RWA should be clarified.

The easiest way of doing that is to consider transitions with the selection rules $\Delta m = \pm 1$ involving circularly polarized light. It can be shown easily that the only difference with respect to the $\Delta m = 0$ transitions discussed above is the necessity of substituting $q\bar{q}$ for $q^2$ in the equations for $w$, so that no qualitative change in the time dependence of $w$ will appear.

5.2 A Semiclassical Approach

The derivation of the nonlinear Bloch equations has been given shortly in the example with the neoclassical Jaynes theory. We have seen there that Weinberg’s theory cannot, in general, give predictions equivalent to those based on the neoclassical approach. Here we will briefly follow the Wódkiewicz and Scully paper to compare their semiclassical results with those given in the previous subsection.

The nonlinearity chosen by Wódkiewicz and Scully corresponds in our notation to $\epsilon_2 = 0$ and $\epsilon_1 = \sqrt{2\hbar\varsigma}$. We know that this choice leads to the “detuning” $\Delta' = \Delta - 2\varsigma$. If no spontaneous emission is introduced phenomenologically into the Bloch equations, they take the form

$$\dot{u} = -\Delta' v - 2\varsigma wv$$
$$\dot{v} = \Delta' u + \Omega w + 2\varsigma wu$$
$$\dot{w} = -\Omega v.$$  \hspace{1cm} (5.36)

Let now $\Delta' = 0$. Then

$$\dot{u} = -2\varsigma wv$$
$$\dot{v} = \Omega w + 2\varsigma wu$$
$$\dot{w} = -\Omega v.$$  \hspace{1cm} (5.38)

These equations have an additional conserved quantity

$$u^2 + v^2 + \frac{\Omega}{\epsilon}w$$  \hspace{1cm} (5.40)

which we shall put equal 0. Differentiating the last Bloch equation we get

$$\ddot{w} = -2\varsigma \Omega w - \Omega^2 w$$
$$= (2\varsigma^2 - \Omega^2)w - 2\varsigma^2 w^3$$  \hspace{1cm} (5.41)

which is identical to our result (here, of course, $\Omega$ is the semiclassical Rabi frequency).
Chapter 6

REMARKS ON THE MIND-BODY QUESTION IN NONLINEAR QM

The present writer has no other qualification to offer his views than has any other physicist and he believes that most of his colleagues would present similar opinions on the subject, if pressed.

E. P. Wigner in Remarks on the Mind-Body Question

The problem of measurement seems to be the most profound difficulty of quantum mechanics. A very clear and concise presentation of the problem was given by Wigner in his philosophical paper [51], a fragment of which has been chosen as the motto to this chapter. The essential part of his argumentation goes as follows.

Let us consider (...) an initial state of the object which is a linear combination $\alpha \psi_1 + \beta \psi_2$ of the two states $\psi_1$ and $\psi_2$. It then follows from the linear nature of the quantum mechanical equations of motion that the state of the object plus observer is, after the interaction, $\alpha \psi_1 \otimes \chi_1 + \beta \psi_2 \otimes \chi_2$. If I now ask the observer whether he saw a flash, he will with a probability $|\alpha|^2$ say that he did, and in this case the object will also give me the responses as if it were in the state $\psi_1$. If the observer answers “No” — the probability for this is $|\beta|^2$ — the object’s responses from then on will correspond to a wave function $\psi_2$. The probability is zero that the observer will say “Yes”, but the object gives the response which $\psi_2$ would give because the wave function (...) of the joint system has no $\psi_2 \otimes \chi_1$ component. (...) All this is quite satisfactory: the theory of measurement, direct or indirect, is logically consistent so long as I maintain my privileged position as ultimate observer.

However, if after having completed the whole experiment I ask my friend, “What did you feel about the flash before I asked you?” he will answer, “I told
you already, I did [did not] see a flash”, as the case may be. In other words, the question whether he did or did not see the flash was already decided in his mind, before I asked him. If we accept this, we are driven to the conclusion that the proper wave function immediately after the interaction of friend and object was already either $\psi_1 \otimes \chi_1$ or $\psi_2 \otimes \chi_2$ and not the linear combination $\alpha \psi_1 \otimes \chi_1 + \beta \psi_2 \otimes \chi_2$. This is a contradiction, because the state described by the wave function $\alpha \psi_1 \otimes \chi_1 + \beta \psi_2 \otimes \chi_2$ describes a state which has properties which neither $\psi_1 \otimes \chi_1$ nor $\psi_2 \otimes \chi_2$ has. If we substitute for “friend” some simple physical apparatus, such as an atom which may or may not be excited by the light-flash, this difference has observable effects and there is no doubt that $\alpha \psi_1 \otimes \chi_1 + \beta \psi_2 \otimes \chi_2$ describes the properties of the joint system correctly, the assumption that the wave function is either $\psi_1 \otimes \chi_1$ or $\psi_2 \otimes \chi_2$ does not. If the atom is replaced by a conscious being, the wave function $\alpha \psi_1 \otimes \chi_1 + \beta \psi_2 \otimes \chi_2$ (which also follows from the linearity of the equations) appears absurd because it implies that my friend was in a state of suspended animation before he answered my question.

It follows that the being with consciousness must have a different role in quantum mechanics than the inanimate measuring device: the atom considered above. In particular, the quantum mechanical equations of motion cannot be linear if the preceding argument is accepted... ([51], p. 179–180)

The last sentence was the main motivation for the inclusion of the measurement problem in the work on aspects of nonlinear QM.

It seems that an unquestionable element of all the known approaches to the measurement problem is the so-called pre-measurement procedure. The pre-measurement is the stage before the moment the friend of Wigner “starts to realize” that he has [has not] seen the flash of light. The pre-measurement prepares the initial conditions for the forthcoming “actualisation” of the result of the measurement and, assuming that the supposed nonlinearity of the evolution is “localised in the observer”, we return to the theme tune of our work — the description of composite systems.

I have argued several times that the “collapse-like” description proposed by Weinberg cannot be correct in usual situations, first of all because it is basis dependent. Nevertheless, we must bear in mind that so long as we are dealing with observers the linear superposition principle is no longer valid for the observers themselves. This difficulty, outlined by Wigner in the above quotation, is known as the pointer basis problem [52].

Putting aside the various theoretical proposals of its solution, we remark that once some way of inducing the “pointer superselection rules” is given, no freedom in the choice of the observer’s basis is left, and the Weinberg approach may be motivated. The point is that results of single measurements are evidently, at least operationally, precisely defined for an observer who notices the flashes of light. A technical difficulty in a nonlinear description of the collapse (a projection of a solution is not necessarily a solution) is intelligently passed round by Weinberg who introduces the collapse implicitly at the level of the Hamiltonian function and not by means of some additional discontinuous evolution law.

A modification of the observed (linear) system can, of course, exert an influence on the observer; the Gisin’s telegraph belongs to such a class of phenomena, so may not be very pathological in the context of measurements. On the other hand, the telegraph based on the mobility phenomenon would lead to “telekinetic phenomena”. Indeed, consider a
pre-measurement that produces an entangled state of a system and an observer. Let the observed system be a two-level atom and an electromagnetic field, initially in the state
\[ |1\rangle|1\rangle + |2\rangle|0\rangle = |11\rangle + |20\rangle \]
(6.1)
where the notation is like in the previous chapter. Assume that a Hilbert space of the observer is also spanned by two states \(|\pm\rangle\). After the pre-measurement the state of the joint system is
\[ |11\rangle + |20\rangle \]
(6.2)
If now the observer’s state space undergoes a nonlinear evolution with the “\(\sigma_3\)” nonlinearity like in (4.16) then two possibilities occur. First, if only the \(|\pm\rangle\) start to rotate with the mobility frequency then the atomic reduced density matrix is like in linear QM. However, if the nonlinearity violates the orthogonality of the states \(|1\rangle|+\rangle := |1+\rangle\) and \(|0\rangle|−\rangle := |0−\rangle\) (i.e. when the observer makes both his own consciousness and the interacting photons evolve in a nonlinear way) then the atom starts to “feel” it. It follows that, at least in principle, a suitable form of his “own” nonlinearity can enable the observer influence in a statistically observable way (compare (4.27)) a behaviour of a random generator just by watching it! Notice that there is also some limitation on the possibility of the influence. For consider a situation where there is a number of intermediate states between the random generator and the observer, so that the entangled state takes the form
\[ |1\rangle|+\rangle|+\rangle \ldots |+\rangle|+\rangle + |2\rangle|−\rangle|−\rangle \ldots |−\rangle|−\rangle. \]
The random generator will “feel” the mobility only provided the mobility will involve the underbraced states
\[ |1\rangle|+\rangle|+\rangle \ldots |+\rangle|+\rangle + |2\rangle|−\rangle|−\rangle \ldots |−\rangle|−\rangle. \]
In case the nonlinearity is more localized, say
\[ |1\rangle|+\rangle|+\rangle \ldots |+\rangle|+\rangle + |2\rangle|−\rangle|−\rangle \ldots |−\rangle|−\rangle, \]
then the generator’s density matrix does not contain the “malignant” terms (one cannot influence the generator by watching it on TV; or, it is much easier to influence with my thoughts my own finger than someone else’s).

The above phenomenon is present only if we assume the kind of description à la Weinberg which is neither very elegant nor easy in practical applications. One may hope that at least in the “correct” description no ways of exerting observer’s influence on external world by “thoughts” or “intentions” exist. The following surprising example is due to J. Polchinski [18] (the version presented here is slightly modified with respect to the original one).

Consider a process involving four steps.

1. A spin-1/2 ion enters a Stern-Gerlach device, which couples to the linear spin \(\sigma_3\) component and the beam splits into two, “+1” and “−1” sub-beams.

2. The “+1” beam evolves freely; in the path of the “−1” beam a macroscopic observer (or a random generator) takes one of two actions: (a) nothing (say, with probability \(\lambda_1\)), or (b) rotates the spin into the “1” direction with a magnetic field coupled to \(\sigma_2\) (with probability \(\lambda_2\)).
3. The two beams are rejoined and the ion enters a region of field coupled to the nonlinear observable

$$f \frac{(\text{Tr } \rho \sigma_1)^2}{\text{Tr } \rho}. \quad (6.3)$$

4. The observer again measures the spin with a Stern-Gerlach device coupled to $\sigma_3$.

The steps 1 and 2 prepare the initial condition for the nonlinear evolution, and the reduced density matrix of the ion after step 2 is

$$\rho_0 = \lambda_1 \left( \begin{array}{cc} 1/2 & 0 \\ 0 & 1/2 \end{array} \right) + \lambda_2 \left( \begin{array}{cc} 3/4 & 1/4 \\ 1/4 & 1/4 \end{array} \right). \quad (6.4)$$

The ion’s nonlinear Hamiltonian for the step 3 is

$$\hat{H} = 2f \frac{\text{Tr } \rho \sigma_1 \sigma_1}{\text{Tr } \rho} \quad (6.5)$$

where we have assumed the “correct” form of the evolution, i.e. that it is generated by the Hamiltonian function depending only on the reduced density matrix of the ion. The reduced density matrix satisfies

$$i \dot{\rho} = 2f \frac{\text{Tr } \rho \sigma_1}{\text{Tr } \rho} [\sigma_1, \rho] \quad (6.6)$$

whose solution is

$$\rho(t) = e^{-2if(\sigma_3)\sigma_1 t \rho(0)}e^{2if(\sigma_3)\sigma_1 t}$$

$$= \frac{\lambda_1}{2} \mathbf{1} + \frac{\lambda_2}{4} \left( 2 \mathbf{1} + \sigma_1 + \sigma_3 \cos 2\lambda_2 ft + \sigma_2 \sin 2\lambda_2 ft \right). \quad (6.7)$$

In the analogous manner we can calculate the evolution of the projector $P_\pm = \frac{1}{2}(\mathbf{1} \pm \sigma_3)$. We find in the “Heisenberg picture” (i.e. we solve the Heisenberg equations of motion with the nonlinear Hamiltonian)

$$P_\pm(t) = \frac{1}{2} \left( \mathbf{1} \pm \sigma_3 \cos 2\lambda_2 ft \mp \sigma_3 \sin 2\lambda_2 ft \right). \quad (6.8)$$

The linear case, where in step three the ion couples to the linear $\sigma_1$, would yield

$$P_\pm(t) = \frac{1}{2} \left( \mathbf{1} \pm \sigma_3 \cos 2ft \mp \sigma_3 \sin 2ft \right). \quad (6.9)$$

Assuming that the time of the interaction during the third step satisfies $2ft = \pi$ we get, in the linear case,

$$P_\pm(t) = \frac{1}{2} \left( \mathbf{1} \mp \sigma_3 \right) = P_\mp(0), \quad (6.10)$$

which means that the spin changes its sign during the evolution. In the nonlinear case, however,

$$P_\pm(t) = \frac{1}{2} \left( \mathbf{1} \pm \sigma_3 \cos \lambda_2 \pi \mp \sigma_3 \sin \lambda_2 \pi \right) \quad (6.11)$$

hence, in particular, the evolution of $P_\pi$ depends on $\lambda_2$ — the probability of one of the two actions taken by the observer or the random generator in case the spin turned out
to be $-1$. The result means that the evolution of the ion depends on “intentions” of the observer concerning his possible actions he would have undertaken had the spin of the ion turned out to be $-1$ — even in case the spin is $+1$ and the observer is passive! The point is that we assume that the ion evolves linearly after having left the nonlinearity region, so that we can apply the ordinary interpretation of results of single measurements. For example, for $\lambda_2 = 0$, that is when the observer is decided not to take any actions, the spin state of the ion would remain unchanged. In the opposite case, $\lambda_2 = 1$, the evolution would be like in linear QM.

I cannot find any mistake in this reasoning. Notice also that since the interpretational background must be some “no collapse” interpretation of QM, the effect means that different (Everett) branches of the Universe can somehow influence results of single measurements, which is not so bizarre if one accepts the spirit of the many-worlds interpretation.

It seems that the essential point of this argumentation is the assumption that a single member of a beam of ions is described by the same density matrix as the whole ensemble. This is exactly opposite to the reasoning leading to Gisin’s telegraph. And I think this example indicates one of the most fundamental conceptual, or practical, difficulties of nonlinear QM: The fact that we do not really know how to treat beams of single objects. Intuitively, weak beams should be “linear” while strong ones could be, perhaps, “nonlinear” in some mean-field sense. The second hint for further generalizations of linear QM is a possibility, suggested by Wigner, that the only domain of fundamental nonlinearities could be the consciousness of observer. Then the density matrix representing the consciousness could evolve nonlinearly and no decompositions of the observer into sub-ensembles would make any sense. Still, the Polchinski’s phenomenon can describe something like intuition: A perception of a single event depends on the overall property of an ensemble of such events since the whole density matrix is involved.

A question that arises immediately is how can we distinguish between systems that are “conscious” (observers) and “non-conscious”. The only natural characterisation that comes to my mind is the way they behave from the point of view of information gain — the notion defined in theory of information.

This viewpoint will become the starting point for my own proposal of nonlinear QM presented in the next chapter. It seems it is the first version of the theory free of all the main difficulties presented until now.
Chapter 7

ENTROPIC FRAMEWORK FOR NONLINEAR QM

This chapter is devoted to a new proposal of a general framework for nonlinear QM. A fundamental drawback of the Hamiltonian generalizations à la Kibble or Weinberg is, as we have seen, the fact that one has to modify observables in order to modify dynamics. A modification of observables will, in general, lead to the known problems with results of single measurements and probabilities. These, on the other hand, form the fundamental physical and mathematical body of the theory. Physically, averages are inherently linear concepts and higher order moments of observable quantities (random variables) are related to the associativity of the algebra of observables. Mathematically, the spectral theory of self-adjoint linear operators leads to the required structures in a very natural way and, last but not least, is a beautiful and elaborated theory.

The Hamiltonian form of the quantum mechanical equations of motion, with the averages playing effectively the role of generators of canonical transformations, indicates that quantum dynamics is deeply rooted in probability. In fact, not only the probabilities — via average values of energy — generate the evolution of quantum systems, but the evolving being is again... the probability, since the canonical coordinates for a quantum system are the real and imaginary parts of probability amplitudes. Let me stress it again: The evolution of states is generated by objects that describe statistical regularities of quantum ensembles.

Another important, or even essential, physical characterization of a statistical physical theory is provided by the notion of entropy. Now, is the entropy of quantum systems an observable quantity or not? Can we take a single member of an ensemble and measure its entropy in analogy to the measurement of energy? No, we cannot. Why? Because, logically, the entropy is not a linear sum of contributions from the constituents of the ensemble, but — rather — describes an overall property of the ensemble. It follows that there does not exist a linear operator of entropy whose average value could describe the (average) entropy of the ensemble while its eigenvalues would be the elementary entropies of single systems. This property of entropy can be seen in its standard definition (the von Neumann entropy)

\[ S[\rho] = -\text{Tr}(\rho \ln \rho) \] (7.1)
or, in a form of averages,

$$S[\rho] = -\langle \ln \rho \rangle.$$  \hspace{1cm} (7.2)

Sometimes another, similar expression is considered [55, 56]

$$S[\rho] = - \ln \langle \rho \rangle = - \ln \text{Tr} (\rho^2)$$  \hspace{1cm} (7.3)

which we shall term the Stueckelberg entropy , as it seems it was he who discussed it for the first time in his nowadays almost forgotten papers [57].

Now, what is the role of entropy for quantum evolution? A first look at general structures of quantum mechanics suggests that there is no role whatever. What we can say generally is that both entropies are conserved by unitary evolutions, because unitarity implies conservation of Tr ($\rho^n$) for any $n$, a property apparently typical only for linear evolutions.

Although it seems there is no physical principle relating evolution and entropy (with the notable exception of the second law of thermodynamics, but this concerns a different level of discussion), one of the main purposes of this chapter is an attempt to propose such a principle. A dynamics resulting from the principle is rather the Nambu [39] than the Hamilton or Poisson dynamics, although it contains the Hamiltonian and Poissonian frameworks as particular cases. It is surprising, as we shall see, that most of the properties of QM that are usually attributed to the linearity are typical for a large class of theories (which are not necessarily linear!) resulting from the generalized formalism.

A departure point of the new construction will be a generalization of the ordinary Poisson bracket to a triple bracket involving an additional functional $S$ which, following Białyńcki-Birula and Morrison (BBM) [39], will be interpreted as a measure of entropy. Since the generalized framework will be based on different measures of entropy (or information), it is best to begin the discussion with an introduction of some elementary concepts from information theory.

7.1 Elements of Information Theory  
— from Hartley to Rényi

A logarithmic measure of information was introduced by R. V. Hartley in 1928 [58]. According to him, to characterize an element of a set of size $N$ we need $\log_2 N$ units of information. It follows that a unit of information (1 bit) is the amount of information necessary for a characterization of a pair. Of course, one can choose also other units such that the unit is the amount of information necessary for a characterization of a set with $0 < k \in \mathbb{N}$ elements, or even with $0 < r \in \mathbb{R}$ elements in average. The respective measures of information in arbitrary units $a$ are $\log_a N$. The most important feature of the logarithmic information measure is its additivity: If a set $E$ is a disjoint union of $M$ $N$-tuples $E_1, \ldots, E_M$, then we can specify an element of this $MN$-element set $E$ in two steps: First we need $\log_a M$ units of information to describe which $E_k$ of the sets $E_1, \ldots, E_M$ contains the element, then we need $\log_a N$ further units to tell which element of this $E_k$ is the considered one. The information necessary for a characterization of an element of $E$ is the sum of the partial informations: $\log_a MN = \log_a M + \log_a N$. Next step in the development of the measures of information was done independently by C. E. Shannon [59] and N. Wiener [60] in 1948 who derived a formula analogous
to Boltzmann’s entropy. Their formula has the following heuristic motivation. Let $E$ be the disjoint union of the sets $E_1, \ldots, E_n$ having $N_1, \ldots, N_n$ elements respectively ($\sum_{k=1}^n N_k = N$). Let us suppose that we are interested only in knowing the subset $E_k$. (This is typical for classical statistical problems in physics: Statistical quantities depend on classes of microscopic conditions and not on single microscopic properties.) The information characterizing an element of $E$ consists of two parts: The first specifies the subset $E_k$ containing this particular element and the second locates it within $E_k$. The amount of the second piece of information is, by Hartley formula, $\log_a N_k$ thus depends on the index $k$. On the other hand, to specify an element of $E$ we need $\log_a N$ units of information. The amount necessary for the specification of the set $E_k$ is therefore

$$I_k = \log_a N - \log_a N_k = \log_a \frac{N}{N_k} = \log_a \frac{1}{p_k}.$$  \hspace{1cm} (7.4)

It follows that the amount of information received by learning that a single event of probability $p$ took place equals

$$I(p) = \log_a \frac{1}{p}.$$ \hspace{1cm} (7.5)

In statistical situations measured quantities correspond to averages of random variables. Therefore the average information is

$$I = \sum_k p_k \log_a \frac{1}{p_k}.$$ \hspace{1cm} (7.6)

This is the Shannon’s formula and $I$ is called the entropy of the probability distribution $\{p_1, \ldots, p_n\}$. If all the probabilities are equal $1/N$ then the Shannon’s formula is equal to the Hartley’s one. The mean we have applied is the so-called linear mean. Rényi observed that there exist information theoretic problems where the measures of information are those obtained by more general ways of averaging — the Kolmogorov–Nagumo function approach [61]. Let $\varphi$ be a monotonic function on real numbers. The Kolmogorov–Nagumo average information can be defined by means of $\varphi$ as

$$I = \varphi^{-1} \left( \sum_k p_k \varphi \left( \log_a \frac{1}{p_k} \right) \right).$$ \hspace{1cm} (7.7)

If the generalized information measure is to satisfy the postulate of additivity, the function $\varphi$ cannot be arbitrary. Indeed, let a chance experiment be a union of two independent experiments. Let us suppose that we obtain $I_k$ units of information with probability $p_k$ in the first experiment and $J_l$ units of information with probability $q_l$ in the second one. Thus we receive $I_k + J_l$ units of information with probability $p_k q_l$. If the average amount of information obtained from the joint experiment is the sum of the average amounts of information obtained from both experiments, then

$$\varphi^{-1} \left( \sum_{k,l} p_k q_l \varphi (I_k + J_l) \right) = \varphi^{-1} \left( \sum_k p_k \varphi (I_k) \right) + \varphi^{-1} \left( \sum_l q_l \varphi (J_l) \right)$$ \hspace{1cm} (7.8)

must hold for any probability distributions $\{p_k\}$ and $\{q_l\}$. If $J_l = J$ for all $l$ then

$$\varphi^{-1} \left( \sum_{k,l} p_k q_l \varphi (I_k + J_l) \right) = \varphi^{-1} \left( \sum_k p_k \varphi (I_k) \right) + J$$ \hspace{1cm} (7.9)
which can hold if and only if \( \varphi \) is linear or exponential function. The linear function corresponds to Shannon’s information. The exponential functions provide a large class of new measures of information. Consider a function \( \varphi(x) = a^{(1-\alpha)x} \). We can always choose the units of information in such a way that

\[
I = \varphi^{-1} \left( \sum_k p_k \varphi \left( \log_a \frac{1}{p_k} \right) \right) = \frac{1}{1-\alpha} \log_a \left( \sum_k p_k^\alpha \right) = \log_a \left( \left( \sum_k p_k^{\alpha} \right)^{1/(1-\alpha)} \right). \tag{7.10}
\]

For \( p_k = 1/N \) we obtain again the Hartley formula. Formula (7.10) describes Rényi’s \( \alpha \)-entropy which, from now on, will be denoted \( I_\alpha(P) \), where \( P \) denotes the probability distribution. We see that the essential part of the definition is played by

\[
I_\alpha^*(P) = a^{I_\alpha(P)} = \left( \sum_k p_k^\alpha \right)^{1/(1-\alpha)}. \tag{7.11}
\]

which is independent of the choice of the unit \( a \). (The power \( 1/(1-\alpha) \) could be also replaced by \( K/(1-\alpha) \), where \( K \) is a positive constant. The entropy would be then multiplied by \( K \); we shall need this freedom later.) To distinguish between \( \alpha \)-entropy and \( I_\alpha^*(P) \) we shall call the latter \( \alpha^* \)-entropy (* will remind us that this quantity is multiplicative in opposition to the additivity of \( I_\alpha(P) \)). (The observation that what is in fact informationally fundamental in \( I_\alpha(P) \) is \( I_\alpha^*(P) \) is strengthened by Daróczy’s definition of entropy of order \( \alpha \) [62] defined as

\[
(2^{1-\alpha} - 1)^{-1} \left( \sum_k p_k^{\alpha} - 1 \right). \tag{7.12}
\]

This expression possesses many ordinary properties of the entropy and in the limit \( \alpha \to 1 \) becomes, the so-called Shannon’s information function.)

The limit \( \alpha \to 1 \) is interesting also for \( \alpha \)-entropies. It can be shown that \( I_1 = \lim_{\alpha \to 1} I_\alpha \) equals Shannon’s entropy.

\( I_\alpha(P) \) is a monotonic, decreasing function of \( \alpha \). For negative \( \alpha \) \( I_\alpha(P) \) tends to infinity if one of \( p_k \) tends to zero. This property excludes \( \alpha < 0 \) because adding a new event of probability 0 to a probability distribution, what does not change the probability distribution, turns \( I_\alpha(P) \) into infinity.

A fundamental notion in information theory is the gain of information. Consider an experiment whose results are \( A_1, \ldots, A_n \) having probabilities \( p_k = P(A = A_k) \). We observe an event \( B \) related to the experiment and obtain a result \( B = B_i \). Now the conditional probabilities are \( p_{ki} = P(A = A_k | B = B_i) \). Consider now a system (an “observer”) whose information is measured by some \( \alpha \)-entropy. How much information about the random variable \( A \) has he received by observation of \( B = B_i \)? The amount of information he would have obtained by observing \( A = A_k \) would be equal to

\[
\log_a \frac{1}{p_k} \tag{7.13}
\]

if he had not measured \( B \). After having observed \( B = B_i \) the amount of information he would have obtained by observing \( A = A_k \) would be

\[
\log_a \frac{1}{p_{ki}}. \tag{7.14}
\]
It follows that the measurement of \( B = B_l \) has given him already

\[
\log_a \frac{1}{p_k} - \log_a \frac{1}{p_{kl}} = \log_a \frac{p_k}{p_{kl}} \tag{7.15}
\]

units of information about \( A \). The expression (7.15) is called the decrease of uncertainty about \( A = A_k \) by observing \( B = B_l \). We define the gain of information about \( A \), obtained when the probability distribution \( \{p_k\} \) is replaced by \( \{p_{kl}\} \), by

\[
\varphi^{-1} \left( \sum_k p_{kl} \varphi \left( \log_a \frac{p_k}{p_{kl}} \right) \right) = \frac{1}{1 - \alpha} \log_a \left( \sum_k \frac{p_{kl}^{2 - \alpha}}{p_k^{1 - \alpha}} \right). \tag{7.16}
\]

If we define the increase of the uncertainty by minus decrease of uncertainty we can calculate the average “loss of information” defined by

\[
\varphi^{-1} \left( \sum_k p_{kl} \varphi \left( \log_a \frac{p_{kl}}{p_k} \right) \right) = \frac{1}{1 - \alpha} \log_a \left( \sum_k \frac{p_{kl}^{2 - \alpha}}{p_k^{1 - \alpha}} \right). \tag{7.17}
\]

For Shannon’s entropy the gain is minus the loss. For \( \alpha \)-entropies the two concepts are inequivalent.

The gain of information defined by (7.16) for \( \alpha > 2 \) has the same pathological properties as \( I_\alpha \) for \( \alpha < 0 \) so, it seems, cannot be consistently applied. This is the reason why Rényi defined the gain of information as minus the loss. From the viewpoint of our quantum mechanical applications the situation is not so clear, however.

It should be emphasized that when we speak about information, what we have in mind is not the subjective “information” possessed by a particular, animate observer. In reality the information contained in an observation is a quantity independent of the fact whether it does or does not reach the perception of the observer (be it a man, some registering device, a computer, or some other physical system). On the other hand, different kinds of entropies introduced above may be characteristic for different systems. The entropy (information) is objective in the same sense as probability, and in the same sense it is reasonable to expect that there are classical and quantum informations, as there are classical and quantum probabilities.

A fundamental fact about physical interpretation of the statistical nature of QM is that the probabilities do not result from our lack of knowledge about some deeper level of quantum systems but are, in certain sense, fundamental. It means that there does not exist (in principle!) any deeper level of description. When we think about a classical statistical system we can imagine that its internal order, or disorder, can change due to interactions with the rest of the world. Putting it more formally, we can say that an information characterizing a classical system should allow different gains of information in different situations. A contemplation of the quantum case suggests that a quantum information might be of such a kind that its corresponding gain of information is zero under all circumstances. It is tempting to develop this hypothesis a little and find whether a measure of information possessing this property exists.

The Shannon’s information gain is given by

\[
- \sum_k p_{kl} \log_a \frac{p_{kl}}{p_k} \tag{7.18}
\]
and vanishes only if $A$ and $B$ are independent. So this case can be excluded because we want the gain of information to be 0 for all probability distributions. For $\alpha$-entropies we find that the vanishing of (7.16) implies

$$
\sum_k \frac{p_k^{2-\alpha}}{p_k} = \sum_k p_k(\frac{2-\alpha}{p_k^{\alpha-2}}) = 1
$$

(7.19)

which can hold for all $p_k$ and $p_{kl}$ if and only if $\alpha = 2$. It follows that the correct candidate for the quantum entropy is the Rényi’s $2$-entropy which reads

$$
-\log_a \left( \sum_k p_k^2 \right).
$$

(7.20)

Expressing the probabilities by means of a density matrix and choosing the unit of information with $a = e$ we obtain

$$
I_2[\rho] = -\ln \text{Tr} (\rho^2)
$$

(7.21)

hence the entropy of Stueckelberg! A possible interpretation of this result is that $\alpha$-entropies for $1 \leq \alpha \leq 2$ represent a continuous transition between classical and quantum entropies.

### 7.2 Poissonian Formulation of Quantum Mechanics

Let $\mathcal{H}$ be a Hilbert space. Consider the Hamilton equations

$$
\omega^{AA'}(\alpha,\alpha') \frac{d\psi_A(\alpha)}{dt} = \frac{\delta H}{\delta\bar{\psi}_{A'}(\alpha')}
$$

(7.22)

and

$$
\bar{\omega}^{AA'}(\alpha,\alpha') \frac{d\bar{\psi}_{A'}(\alpha')}{dt} = \frac{\delta H}{\delta \psi_A(\alpha)}
$$

(7.23)

where the bars denote complex conjugations and the conventions concerning primed and unprimed indices are assumed like in the spinor abstract index calculus [36]. The summation convention is like in Chapter 2: We sum over repeated Roman indices and integrate over repeated Greek ones. In QM the “symplectic form” is given by the delta distribution

$$
\omega^{AA'}(\alpha,\alpha') = i\delta^{AA'}\delta(\alpha - \alpha') = \omega^{AA'}\delta(\alpha - \alpha')
$$

(7.24)

$$
\delta^{AB'} = \delta_{AB} = 1 \text{ if } A = B' \text{ and } 0 \text{ for } A \neq B'.
$$

The inverse of $\omega^{AA'}(\alpha,\alpha')$ is

$$
I_{AA'}(\alpha,\alpha') = -i\delta_{AA'}\delta(\alpha - \alpha') = I_{AA'}\delta(\alpha - \alpha')
$$

(7.25)

where by the inverse we understand that

$$
\omega^{AA'}(\alpha,\alpha') I_{BA'}(\beta,\alpha') = \delta_B^A\delta(\alpha - \beta)
$$

(7.26)

$$
\omega^{AA'}(\alpha,\alpha') I_{AB'}(\alpha,\beta') = \delta_B^A\delta(\alpha' - \beta').
$$

(7.27)
Accordingly, the form of the Hamilton equations we shall use is

$$\frac{d\psi_A(\alpha)}{dt} = I_{AA'} \frac{\delta H}{\delta \psi_A(\alpha)}$$  \hspace{1cm} (7.28)

and

$$\frac{d\psi_A(\alpha)}{dt} = -I_{AA} \frac{\delta H}{\delta \psi_A(\alpha)}.$$  \hspace{1cm} (7.29)

(7.28) and (7.29) describe a quantum evolution of pure states. Assume now that all observables of the theory depend on $|\psi\rangle$ and $\langle\psi|$ by the density matrix $\rho = |\psi\rangle\langle\psi|$. Let $F$ and $G$ be two such observables, that is $F[\psi, \bar{\psi}] = F[\rho]$ and $G[\psi, \bar{\psi}] = G[\rho]$. The Poisson bracket resulting from the Hamilton equations is

$$\{F, G\} = I_{AA'} \left( \frac{\delta F}{\delta \psi_A(\alpha)} \right) \frac{\delta G}{\delta \psi_A'(\alpha')} - \left( \frac{\delta G}{\delta \psi_A(\alpha)} \right) \frac{\delta F}{\delta \psi_A'(\alpha')}.$$  \hspace{1cm} (7.30)

Applying the chain rule to the components of the pure state density matrix

$$\rho_{AA'}^{\alpha, \alpha'} = \psi_A(\alpha)\bar{\psi}_{A'}(\alpha')$$  \hspace{1cm} (7.31)

we find that

$$\{F, G\} = I_{AA'} \left( \frac{\delta F}{\delta \rho_{AB}^{\alpha, \beta}} \right) \frac{\delta G}{\delta \rho_{AB}^{\gamma, \delta}} - \left( \frac{\delta G}{\delta \rho_{AB}^{\alpha, \beta}} \right) \frac{\delta F}{\delta \rho_{AB}^{\gamma, \delta}} = (F \leftrightarrow G).$$  \hspace{1cm} (7.32)

So long as the density matrix in (7.32) is given by (7.31) the bracket is equivalent to the Poisson bracket (7.30). Jordan, in a context of the Weinberg’s theory [19] and for a finite dimensional Hilbert space, investigated properties of the bracket (7.32) with $\rho$ being an arbitrary density matrix. For reasons that will be explained below I will term such a general bracket the Białynicki-Birula–Morrison–Jordan (BBMJ) bracket.

We will now show that (7.32), for a general $\rho$, can be written in a form of a generalized Nambu bracket. Let $\rho$ be arbitrary. The BBMJ bracket can be rewritten as

$$\{F, G\} = \rho_{AA'}^{\alpha, \alpha'} \Omega^{AA'}_{BB'CC'}(\alpha, \alpha', \beta, \beta', \gamma, \gamma') \frac{\delta F}{\delta \rho_{BB}^{\alpha, \beta}} \frac{\delta G}{\delta \rho_{CC}^{\gamma, \delta}}$$  \hspace{1cm} (7.33)

with

$$\Omega^{AA'}_{BB'CC'}(\alpha, \alpha', \beta, \beta', \gamma, \gamma') = \delta^A_{B'}\delta^A_{C'}\delta^{\alpha - \beta}_\gamma\delta^{\alpha' - \beta'}_\delta - \delta^A_{B} \delta^A_{C} \delta^{\alpha - \beta}_\gamma \delta^{\alpha' - \beta'}_\delta = \Omega^a_{bc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$$  \hspace{1cm} (7.34)

where, in analogy to the spinor calculus, we have clumped together the respective pairs of indices into composite indices ($a = AA'$, $(\alpha, \alpha') = \tilde{\alpha}$, etc.).

The “structure kernels” $\Omega^a_{bc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ satisfy conditions characteristic for Lie-algebraic structure constants:

$$\Omega^a_{bc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = -\Omega^a_{bc}(\tilde{\beta}, \tilde{\alpha}, \tilde{\gamma})$$  \hspace{1cm} (7.35)

and

$$\Omega^a_{bc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})\Omega^b_{de}(\tilde{\gamma}, \tilde{\delta}, \tilde{\epsilon}) + \Omega^c_{de}(\tilde{\alpha}, \tilde{\delta}, \tilde{\epsilon})\Omega^a_{bc}(\tilde{\gamma}, \tilde{\beta}, \tilde{\delta}) + \Omega^a_{dc}(\tilde{\alpha}, \tilde{\delta}, \tilde{\gamma})\Omega^c_{eb}(\tilde{\gamma}, \tilde{\epsilon}, \tilde{\delta}) = 0.$$  \hspace{1cm} (7.36)
These two conditions imply the Jacobi identity. The composite index form of the BBMJ bracket

$$\{F, G\} = \rho_\alpha(\tilde{\alpha})\Omega^a_{bc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \frac{\delta F}{\delta \rho_\alpha(\tilde{\beta})} \frac{\delta G}{\delta \rho_\alpha(\tilde{\gamma})}$$

(7.37)

shows that it takes the same form as the generalized BBM-Nambu bracket written in terms of the Wigner function for a scalar field [39]. As a matter of fact, the BBMJ bracket is simply a different representation of the BBM bracket. The formula (7.37) looks much the same as the Poisson bracket related to the Kirillov form on coadjoint representations of Lie groups [30]. In fact, such brackets for general structure constants are called the Lie-Poisson brackets (cf. [63]). As shown in [39], the structure kernels appearing in the BBMJ bracket correspond to the Weyl-Heisenberg Lie algebra.

It remains to find out how to formulate the explicit triple bracket equivalent to (7.37). In order to do this we first have to define a “metric tensor” to lower the upper index in the structure kernels. The apparently natural guess

$$g_{ab}(\tilde{\alpha}, \tilde{\beta}) = \Omega^c_{ad}(\tilde{\gamma}, \tilde{\alpha}, \tilde{\delta})\Omega^d_{bc}(\tilde{\delta}, \tilde{\beta}, \tilde{\gamma})$$

(7.38)

is incorrect as (7.38) involves expressions like $\delta(0)$ which are not distributions in the Schwartz sense.

The correct definitions are

$$g_{ab}(\tilde{\alpha}, \tilde{\beta}) = -I_{AB'}I_{BA'}\delta(\alpha - \beta')\delta(\beta - \alpha')$$

(7.39)

$$g^{ab}(\tilde{\alpha}, \tilde{\beta}) = -\omega^{AB'}\omega^{BA'}\delta(\alpha - \beta')\delta(\beta - \alpha').$$

(7.40)

The metric tensor is symmetric

$$g_{ab}(\tilde{\alpha}, \tilde{\beta}) = g_{ba}(\tilde{\beta}, \tilde{\alpha})$$

(7.41)

and satisfies the invertibility conditions

$$g^{ab}(\tilde{\alpha}, \tilde{\beta})g_{bc}(\tilde{\beta}, \tilde{\gamma}) = g_{ab}(\tilde{\alpha}, \tilde{\beta})g^{ba}(\tilde{\beta}, \tilde{\gamma})$$

$$= \delta^A\delta^A\delta(\alpha - \gamma)\delta(\alpha' - \gamma')$$

(7.42)

$$= \delta^\beta\delta^\beta\delta(\alpha - \gamma).$$

(7.43)

The metric tensor is a useful tool. Consider for example a $\rho$-independent $F_b(\tilde{\beta}) = F_{BB'}(\beta, \beta')$. Then

$$F[\rho] = g^{ab}(\tilde{\alpha}, \tilde{\beta})\rho_\alpha(\tilde{\alpha})F_b(\tilde{\beta})$$

$$= \delta^{AB'}\delta^{BA'}\delta(\alpha - \beta')\delta(\beta - \alpha')\rho_{AA'}(\alpha, \alpha')F_{BB'}(\beta, \beta')$$

$$= \rho^{B' A'}(\beta', \alpha')F^A_{B'}(\alpha', \beta') = \text{Tr}\rho\hat{F}$$

(7.45)

and we see that bilinear observables can be naturally expressed with the help of (7.40). This example is important also as an illustration of the convention concerning lowering and raising of indices (following from the use of the metric tensor). For notice that

$$\frac{\delta F}{\delta \rho^{B' A'}(\beta', \alpha')} = F^A_{B'}(\alpha', \beta')$$

(7.46)

although the staggering of indices like $F^A_{B'}(\beta', \alpha')$ might seem more natural.
Example 7.1 To show explicitly how the “kernel formalism” works consider the non-relativistic kinetic Hamiltonian function

\[ H[\rho] = \text{Tr} \hat{H} \hat{\rho} = \int d^3x d^3y \delta(\vec{x} - \vec{y}) \frac{-\Delta}{2m} \rho(\vec{x}, \vec{y}) \]

\[ = \int d^3x \int d^3p/(2\pi)^3 \frac{p^2}{2m} e^{-i\vec{p} \cdot \vec{x}} \rho(\vec{p}, \vec{x}) \]

\[ = \int d^3x \int d^3p/(2\pi)^3 \frac{p^2}{2m} e^{-i\vec{p} \cdot \vec{x}} \int d^3y e^{i\vec{p} \cdot \vec{y}} \rho(\vec{y}, \vec{x}) \]

The latter two equations are consistent with the definition of the functional derivative in terms of the directional derivative:

\[ \nabla f H[\rho] = \lim_{\lambda \to 0} \frac{H[\rho + \lambda f] - H[\rho]}{\lambda} = \int d^3x d^3y \frac{\delta H[\rho]}{\delta \rho(\vec{x}, \vec{y})} f(\vec{x}, \vec{y}). \]

Eq. (7.47) can be also written as

\[ \int d^3x d^3y H(\vec{y}, \vec{x}) \rho(\vec{x}, \vec{y}) \]

\[ = \int d^3x d^3y d' d'y' \delta(\vec{x} - \vec{y}) \delta(\vec{y} - \vec{x}') \rho(\vec{y}, \vec{x}) \]

\[ = g(\vec{x}, \vec{y}) H(\vec{x}) \rho(\vec{y}) = g^{\text{vy}} H_\text{x} \rho_\text{y}. \]

The fully covariant form of the structure kernels is

\[ \Omega_{abc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = g_{ad}(\tilde{\alpha}, \tilde{\delta}) \Omega^d_{bc}(\tilde{\delta}, \tilde{\beta}, \tilde{\gamma}) \]

\[ = -I_{AB} I_{CA} I_{BC} \delta(\alpha - \beta') \delta(\beta' - \gamma') \]

\[ \times (\delta^D_B \delta^D_C) \Omega^D_{AB} \delta(\beta - \delta') \delta(\delta' - \gamma') \delta(\gamma - \beta') \]

\[ = -I_{AC} I_{BA} I_{CB} \delta(\alpha - \beta) \delta(\gamma - \alpha') \delta(\beta - \gamma') \]

\[ - I_{AC} I_{BA} I_{CB} \delta(\alpha - \gamma) \delta(\beta - \alpha') \delta(\gamma - \beta'). \]

One easily verifies that \( \Omega_{abc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \) is totally antisymmetric (where the permutation of indices is understood in the sense of (7.35)).
Following Bialynicki-Birula and Morrison let us introduce the functional
\[ S_2 = \frac{1}{2} g^{ab}(\bar{\alpha}, \bar{\beta}) \rho_{a}(\bar{\alpha}) \rho_{b}(\bar{\beta}) = \frac{1}{2} \delta^{AA'} \delta^{B'B} \rho_{AA'}(\alpha, \beta) \rho_{B'B}(\beta, \alpha) = \frac{1}{2} \text{Tr}(\rho^2). \tag{7.51} \]
We can see that \( S_2 \) is one half of the inverse of Rényi’s \( 2^* \)-entropy and for this reason it will be termed just entropy.

The BBMJ bracket is now equal to the following triple bracket
\[ \{F, G \} = [F, G, S_2] = \Omega_{\alpha \beta \gamma}(\bar{\alpha}, \bar{\beta}, \bar{\gamma}) \frac{\delta F}{\delta \rho_{a}(\bar{\alpha})} \frac{\delta G}{\delta \rho_{b}(\bar{\beta})} \frac{\delta S_2}{\delta \rho_{c}(\bar{\gamma})}. \tag{7.52} \]
The antisymmetry of the triple bracket means that \( S_2 \) is the Casimir for the BBMJ bracket Lie algebra of observables. Another Casimir is
\[ \text{Tr} \rho = \delta^{AA'} \delta(\alpha - \alpha') \rho_{AA'}(\alpha, \alpha') \tag{7.53} \]
because
\[
\{ \text{Tr} \rho, G \} = -i \delta^{AA'} \left( \delta^{AB'} \delta(\alpha - \beta') \rho_{CB'}(\gamma, \beta') \frac{\delta G}{\delta \rho_{CA'}(\gamma, \alpha)} \right)
- \delta^{G} \frac{\delta G}{\delta \rho_{AB'}(\alpha, \beta')} \rho_{CB'}(\gamma, \beta') \delta^{CA'} \delta(\gamma - \alpha) = 0. \tag{7.54}
\]
Notice that until now we have not made any specific assumptions about the form of the observables (including the Hamiltonian function in (7.28), (7.29)) like, say, linearity or 1-homogeneity in \( \rho \). The results are therefore general. The wave functions have been eliminated from the dynamical equations, but the Hilbert space background is implicitly present in the structure kernels and the metric tensor which are defined in terms of \( \omega \) and \( I \), and in the very notion of the density matrix which acts in the Hilbert space.

Components of the pure state density matrix satisfy
\[ \frac{d}{dt} \rho_{AA'}(\alpha, \alpha') = \{ \rho_{AA'}(\alpha, \alpha'), H \}. \tag{7.55} \]
which holds also for general density matrices as can be seen from the familiar, operator version of the Liouville–von Neumann equation. It follows that the density matrices form a Poisson manifold, as opposed to state vectors that form a phase space.

### 7.3 Nambu-like Generalization of Quantum Mechanics

At a first glance it is rather surprising that in linear QM the triple bracket \([F, G, S_2]\), whose antisymmetric form does not favour any of the three functions, involves observables \( F \) and \( G \) which are necessarily linear in \( \rho \), and \( S_2 \) which is not linear in \( \rho \). A closer look
shows, however, that the bracket with $S_2$ makes $[F, G, S_2]$ linear in $\rho$, so that a time derivative of an observable is also an observable. It is clear that this property holds only for entropies quadratic in $\rho$, which suggests that although $S_2$ itself is not an observable, generalizations of $S_2$ to some more general $S$ make no sense for physical reasons. If we replace the linearity with the weaker homogeneity requirement, like in the Kibble-Weinberg approach, the triple bracket will map the set of observables into itself provided the entropy will have the same homogeneity as $S_2$.

From the viewpoint of nonlinear generalizations of QM it is important to understand whether it is indeed physically necessary to have a bracket which maps observables into themselves. To fix our attention let us consider the nonrelativistic position operator. An average velocity of an ensemble of particles can be calculated either by first calculating an average position and then taking its time derivative, or by first measuring the velocity of each single particle and then taking the average. We can say that the first procedure is a calculation of the time derivative of an average, whereas the latter is taking the average of the time derivative. The situation can be described symbolically by the equation

$$\frac{d}{dt} \langle \vec{q} \rangle = \langle \frac{d}{dt} \vec{q} \rangle.$$  \hspace{1cm} (7.56)

Still, it is an important property of the QM formalism that we cannot realize the two procedures simultaneously, as $\vec{v} = \vec{p}/m$ and $\vec{q}$ are complementary. It follows that in a concrete experiment we have to decide which way of measuring to choose. In this meaning if we can measure $\vec{q}$, we cannot measure $\frac{d}{dt} \vec{q}$, and vice versa. To express it differently, if $\vec{q}$ is observable (not an observable!) then $\frac{d}{dt} \vec{q}$ is not. This restriction is very important in QM because it can be used to eliminate the paradoxes of the Einstein-Podolsky-Rosen variety.

In the version of nonlinear QM discussed below we will meet the following problem. The observables will be defined as quantities that are in one-to-one relationship to some experimentally measured random variables. Two observables will be said to be complementary if there does not exist a physical situation where the two respective random variables can be measured simultaneously, that is in a single run of an experiment. We will therefore define observables as functionals linear in $\rho$ but the nonlinearity will be introduced through the entropy by means of the triple bracket evolution equation. We will see that in this formulation a derivative of an observable will not be linear in $\rho$ hence will not be considered as an observable, unless the derivative is 0, of course. For example, if a position of some system is given by $\vec{q} \rho = \text{Tr} \hat{q} \rho$, where $\hat{q}$ is a linear operator, then there will not exist, in general, any linear $\frac{d}{dt} \hat{q}$ such that

$$\frac{d}{dt} \text{Tr} \hat{q} \rho = \text{Tr} \frac{d}{dt} \hat{q} \rho.$$ \hspace{1cm} (7.57)

It will be assumed, however, that there does exist an operator $\frac{\partial}{\partial \rho} \hat{q}$ with the interpretation of velocity, which can be measured in some (complementary) experiment and whose average will correspond to $\text{Tr} \frac{d}{dt} \hat{q} \rho$ but which will not have to satisfy (7.57). To put it differently, each observable can be measured in a standard frequency-like way but derivatives of observables cannot be considered simultaneously with the observables themselves. And vice versa, if we measure an observable which is the time derivative of some other observable then the latter cannot be considered simultaneously with its derivative. This
is a somewhat stronger version of the complementarity principle which has to be accepted in the nonlinear framework.

After these remarks I can begin with my own proposal. Let us replace $S_2$ with some arbitrary $S$ and introduce the “$S$-bracket”

$$\{F,G\}_S = [F,G,S]. \quad (7.58)$$

where now for different $S$ we will have different brackets.

The main postulate of the generalized framework is that the evolution equation for density matrices is the triple bracket analogue of the Liouville-von Neumann equation, i.e.

$$\frac{d}{dt} \rho_{aa'}(\alpha,\alpha') = \{\rho_{aa'}(\alpha,\alpha'),H\}_S \quad (7.59)$$

where $S$ has the same homogeneity as $S_2$. This implies that, like in ordinary QM, the scaling by a constant $\rho \rightarrow \lambda \rho$ is a symmetry of the dynamics. We will assume that observables are linear in $\rho$, so can maintain the associative structure of random variables just like in the linear theory.

The first step in this new program is to understand to what degree the structure of ordinary QM depends on the form of $S_2$.

We have explained already that only for $S = S_2$ the set of linear observables is closed with respect to $\{\cdot,\cdot\}_S$. Next question concerns the Jacobi identity, or under what conditions the manifold of states can be regarded as a Poisson manifold. We shall see that the identity holds for all $S$ that are differentiable functions of $f_2[\rho] = \operatorname{Tr}(\rho^2)$, i.e. $S[\rho] = S(f_2[\rho])$. It seems that in more general cases the Jacobi identity will not have to hold (I am working on this point presently but there still exist some unclear elements in the proof of a more general theorem). However, even if the manifold of states will no longer be a Poisson manifold and the dynamics will not be Poissonian, the dynamical system will have many properties characteristic for ordinary QM. We will find a large class of conserved quantities (including energy) and show that, at least for a large class of initial conditions, the solution of the evolution equation is a density matrix with time independent eigenvalues.

### 7.3.1 The Jacobi Identity

Let $F$, $G$, $H$ and $S$ be arbitrary twice functionally differentiable functionals. Let

$$\{F,G\}_S = \Omega_{abc}(\tilde{\alpha},\tilde{\beta},\tilde{\gamma}) \frac{\delta F}{\delta \rho_a(\tilde{\alpha})} \frac{\delta G}{\delta \rho_b(\tilde{\beta})} \frac{\delta S}{\delta \rho_c(\tilde{\gamma})}$$

$$= \Omega_{abc} \frac{\delta F}{\delta \rho_a} \frac{\delta G}{\delta \rho_b} \frac{\delta S}{\delta \rho_c} \quad (7.60)$$

where in the second line the discrete and continuous indices have been clumped into single ones. (We could do this in majority of calculations. The fact that I have decided to use normally the more complicated form involving both Roman and Greek letters follows from fear of losing control over the convention; here the mathematical operations are reduced mainly to permutations of indices.) We consider the expression


66
\[ \frac{\delta^2 S}{\delta \rho_a \delta \rho_f} = g^{af} \text{ for } S = S_2 \text{ and (7.63) vanishes in virtue of (7.36). For more general } S = S(f_2[\rho]) \text{ we find} \]

\[ \frac{\delta S}{\delta \rho_c} = 2 \frac{\partial S}{\partial f_2} \rho^c \]  
\[ \frac{\delta^2 S}{\delta \rho_a \delta \rho_f} = 4 \frac{\partial^2 S}{\partial f_2^2} \rho^a \rho^f + 2 \frac{\partial S}{\partial f_2} g^{af}. \]

Inserting these expressions into (7.63) we obtain

\[ J = 8 \frac{\delta F}{\delta \rho_c} \frac{\partial^2 S}{\partial \rho_a \partial \rho_f} \frac{\partial^2 S}{\partial f_2^2} \rho^a \rho^f \frac{\partial H}{\partial \rho_c} \frac{\partial S}{\partial f_2} (\Omega_{def} \Omega_{abc} + \Omega_{def} \Omega_{abc} + \Omega_{def} \Omega_{abc}) = 0 \]

since \( \Omega_{abc} \rho^a \rho^c = 0 \). With this choice of \( S \) we obtain the dynamics given by

\[ \frac{d}{dt} \rho_{AA}(\alpha, \alpha') = \{ \rho_{AA}(\alpha, \alpha'), H \} \delta S[\rho] \]

where \( C[\rho] = 2 \frac{\delta S}{\delta f_2} = C(f_2[\rho]) \) is an integral of motion, as we shall see later. The only difference with respect to ordinary QM would be in a \( \rho \)-dependent rescaling of time.

For general \( S \) the work on Jacobi identity is in progress but, as of now, we have to accept the possibility that the (mixed) states in the generalized QM do not form a Poisson manifold. Accordingly, the observables will not form a Lie algebra which may suggest that the generalized framework cannot be formulated in a Poissoné covariant way. In Appendix 8.1 I present a formulation of the Dirac equation à la Tomonaga-Schwinger [53, 54] which can be used for a generalization of the Dirac equation along the lines proposed in this work.
7.3.2 Composite Systems in the New Framework

Let the Hilbert space in question and the density matrix of some composite system be \( \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \) and

\[
\rho_\alpha(\tilde{\alpha}) = \rho_{AA'}(\alpha, \alpha') = \rho_{A_1 A_2 A'_1 A'_2}(\alpha_1, \alpha_2, \alpha'_1, \alpha'_2).
\] (7.68)

The same doubling of indices concerns

\[
I_{AA'}(\alpha, \alpha') = -i \delta_{A_1 A'_1} \delta_{A_2 A'_2} \delta(\alpha_1 - \alpha'_1) \delta(\alpha_2 - \alpha'_2).
\] (7.69)

Reduced density matrices of the two subsystems are

\[
\rho_{A_1 A'_1}(\alpha_1, \alpha'_1) = \delta^{A_2 A'_2} \delta(\alpha_2 - \alpha'_2) \rho_{A_1 A_2 A'_1 A'_2}(\alpha_1, \alpha_2, \alpha'_1, \alpha'_2)
\] (7.70)

\[
\rho_{A_2 A'_2}(\alpha_2, \alpha'_2) = \delta^{A_1 A'_1} \delta(\alpha_1 - \alpha'_1) \rho_{A_1 A_2 A'_1 A'_2}(\alpha_1, \alpha_2, \alpha'_1, \alpha'_2)
\] (7.71)

and satisfy

\[
\frac{\delta \rho_{A_1 A'_1}(\alpha_1, \alpha'_1)}{\delta \rho_{B_1 B'_1 B_2^1 B'_2^1}(\beta_1, \beta_2, \beta'_1, \beta'_2)} = \delta^{B_1 B'_1} \delta_{A_1 A'_1} \delta_{B_2 B'_2} \delta_{A_2 A'_2} \delta(\beta_1 - \alpha_1) \delta(\beta_2 - \beta'_1) \delta(\beta'_2 - \alpha'_2)
\] (7.72)

and

\[
\frac{\delta \rho_{A_2 A'_2}(\alpha_2, \alpha'_2)}{\delta \rho_{B_1 B'_1 B_2^1 B'_2^1}(\beta_1, \beta_2, \beta'_1, \beta'_2)} = \delta^{B_2 B'_2} \delta_{A_2 A'_2} \delta_{B_1 B'_1} \delta_{A_1 A'_1} \delta(\beta_2 - \alpha_2) \delta(\beta_1 - \alpha'_1) \delta(\beta'_2 - \alpha'_2).
\] (7.73)

The structure kernels for the composite system are

\[
\Omega_{abc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = \Omega_{a_1 a_2 b_1 b_2 c_1 c_2}(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\gamma}_1, \tilde{\gamma}_2)
\]

\[
= -i \bigg( \delta_{A_1 B'_1} \delta_{C_1 A'_1} \delta_{B_1 C'_1} \delta_{A_2 B'_2} \delta_{C_2 A'_2} \delta_{B_2 C'_2} \times \delta(\alpha_1 - \beta_1) \delta(\gamma_1 - \alpha'_1) \delta(\beta_1 - \gamma'_1) \delta(\alpha_2 - \beta'_2) \delta(\gamma_2 - \alpha'_2) \delta(\beta_2 - \gamma'_2)
\]

\[
- \delta_{A_1 C'_1} \delta_{B_1 A'_1} \delta_{C_1 B'_1} \delta_{A_2 C'_2} \delta_{B_2 A'_2} \delta_{C_2 B'_2} \times \delta(\alpha_1 - \gamma'_1) \delta(\beta_1 - \alpha'_1) \delta(\gamma_1 - \beta'_1) \delta(\alpha_2 - \gamma'_2) \delta(\beta_2 - \alpha'_2) \delta(\gamma_2 - \beta'_2) \bigg).
\] (7.74)

The following two results solve generally the question of faster-than-light telegraphs in both Hamiltonian and triple bracket frameworks.

**Lemma 7.1**

\[
\Omega_{abc}(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \frac{\delta \rho_{A'_1}(\tilde{\delta}_1)}{\delta \rho_{A}(\tilde{\alpha})} \frac{\delta \rho_{A'_1}(\tilde{\delta}_2)}{\delta \rho_{B}(\tilde{\beta})} = 0.
\] (7.75)

**Proof:** It is sufficient to contract (7.74) with (7.72) and (7.73). \( \Box \)

**Theorem 7.2** Let \( F = F[\rho'] \) and \( G = G[\rho'''] \), that is depend on \( \rho \) via (7.70) and (7.71),

then for any \( S \)

\[
\{F, G\}_S = 0.
\] (7.76)

68
Proof: By virtue of the lemma one has
\[ 0 = \Omega_{\alpha \beta \gamma}(\tilde{a}, \tilde{b}, \tilde{c}) \frac{\delta F_{\tilde{a}}(\tilde{a})}{\delta \rho_{\alpha}(a)} \frac{\delta F_{\tilde{b}}(\tilde{b})}{\delta \rho_{\beta}(\beta)} \frac{\delta F_{\tilde{c}}(\tilde{c})}{\delta \rho_{\gamma}(\gamma)} \delta S = \{F, G\}_S. \tag{7.77} \]
\[ \square \]
Notice that we have not assumed anything but differentiability not only about \( S \) but also about \( F \) and \( G \). So, in particular, for arbitrary (nonlinear) observables and \( S = S_2 \) we obtain the Polchinski-Jordan result for Weinberg’s nonlinear QM.

### 7.3.3 Density Matrix Interpretation of Solutions of the Evolution Equation

One of the essential questions we have to clarify concerns the density matrix interpretation of the solutions of (7.59). In fact, it seems there is no general \textit{a priori} guarantee that the generalized dynamics will conserve the positivity of \( \rho \). The next theorems will give a partial answer to this problem.

In order to attack the question we have to make the language of the \( S \)-brackets more readable. Consider the triple bracket \( [F, G, H] \) of arbitrary functionals \( F \), \( G \) and \( H \). We find that
\[ i[F, G, H] = \frac{\delta F}{\delta \rho^{A'}(\beta')} \frac{\delta G}{\delta \rho^{C'}(\gamma')} \frac{\delta H}{\delta \rho^{B'}(\alpha')} \delta \rho^{A'}(\alpha') \delta \rho^{B'}(\beta') \delta \rho^{C'}(\gamma'). \tag{7.78} \]

Applying the notation of (7.46) (where now the “operator” kernels are in general \( \rho \)-dependent) we transform (7.78) into
\[ F^{A'}_{B'}(\alpha', \beta')G^{B'}_{C'}(\beta', \gamma')H^{C'}_{A'}(\gamma', \alpha') - F^{A'}_{C'}(\alpha', \gamma')G^{B'}_{A'}(\beta', \alpha')H^{C'}_{B'}(\gamma', \beta') = \text{Tr} (\hat{F} - \hat{G} - \hat{H}). \tag{7.79} \]

In the last line we have introduced an abbreviated convention based on the assignment to any functional \( F \) an operator
\[ \hat{F} = \frac{\delta F}{\delta \rho} \tag{7.80} \]
which is defined by the kernel form used in (7.79). For example
\[ \rho = \frac{\delta S}{\delta \rho} \tag{7.81} \]
and
\[ \frac{\delta \text{Tr} (\rho^n)}{\delta \rho} = n \rho^{n-1}, \tag{7.82} \]
the latter being the shortened form of
\[ \frac{\delta \text{Tr} (\rho^n)}{\delta \rho^{A'}(\alpha', \alpha')} = n \delta^{B_1}_{A'} \delta^{B_2}_{A'} \delta^{B_3}_{A'} \ldots \delta^{B_{n-1}}_{A'} \delta^{B_n}_{A'} \times \delta(\beta'_n - \alpha) \delta(\alpha' - \beta_2) \delta(\beta'_2 - \beta_3) \ldots \delta(\beta'_{n-1} - \beta_n) \times \rho^{B_2}_{B_1}(\beta_2, \beta_3) \ldots \rho^{B_n}_{B_{n-1}}(\beta_n, \beta_n). \tag{7.83} \]
The first of these implies the known result

\[ [F,G,S_2] = -i \text{Tr} (\rho [\hat{F}, \hat{G}]). \quad (7.84) \]

Consider now a functional \( S \) (differentiable in \( f_k \))

\[ S[\rho] = S(f_1[\rho], \ldots, f_n[\rho], \ldots) \quad (7.85) \]

where \( f_k[\rho] = \text{Tr} (\rho^k) \).

**Theorem 7.3** For any \( m \in \mathbb{N} \), and any \( G \), if \( S \) satisfies (7.85) then

\[ [f_m, G, S] = 0. \quad (7.86) \]

**Proof:**

\[
[f_m, G, S] = \sum_n [\text{Tr} (\rho^m), G, f_n] \frac{\partial S}{\partial f_n} = -im \sum_n n \text{Tr} (\rho^{m-1}, \hat{G} (\rho^{n-1}) \frac{\partial S}{\partial f_n}

= -im \sum_n n \text{Tr} (\hat{G} (\rho^{n-1}, \rho^{n-1}) \frac{\partial S}{\partial f_n} = 0. \quad (7.87)
\]

\( \square \)

This remarkable result covers many interesting and nontrivial generalizations of \( S_2 \). As a by-product it shows also that the same property holds for the Polchinski-Jordan version of Weinberg’s nonlinear QM because we have not assumed that \( G \) is linear in \( \rho \) (moreover, it includes other theories where observables do not satisfy any homogeneity condition). The particular case \( m = 1 \) implies that \( \text{Tr} \rho \) is conserved by all evolutions, a fact important for a definition of averages. For pure states \( \text{Tr} (\rho^m) = (\text{Tr} \rho)^m \) so that the integrals \( f_m \) are not necessarily independent, but for all \( m, n \) \( f_m \) and \( f_n \) are in involution with respect to \( \{\cdot, \cdot\}_S \). Jordan proved in [19] by an explicit calculation that in his formulation of Weinberg’s nonlinear QM \( \text{Tr} \rho \) and \( \text{Tr} \rho^2 \) are conserved — our theorem considerably generalizes this result.

**Theorem 7.4** Let \( S \) satisfy (7.85) and \( \rho_0 \) be a self-adjoint solution of (7.59). If \( \rho_0 \) is positive and has a finite number of nonvanishing eigenvalues \( p_k(0), 0 < p_k(0) \leq 1 \), then the eigenvalues of \( \rho_t \) are integrals of motion, and the evolution conserves positivity of \( \rho_t \).

**Proof:** Since the nonvanishing eigenvalues of \( \rho_0 \) satisfy \( 0 < p_k(0) \leq 1 < 2 \), it follows that for any \( \alpha p_k(0)^\alpha \) can be written in a form of a convergent Taylor series. By virtue of the spectral theorem the same holds for \( p_k^0 \) and \( \text{Tr} (\rho_k^0) \). Each element of the Taylor expansion of \( \text{Tr} (\rho_k^0) \) is proportional to \( f_n[\rho_0] \), for some \( n \). But \( f_n[\rho_0] = f_n[\rho_t] \) hence

\[
\text{Tr} (\rho_k^0) = \text{Tr} (\rho_k^t) = \sum_k p_k(0)^\alpha = \sum_k p_k(t)^\alpha \quad (7.88)
\]

for all real \( \alpha \). Since all \( p_k(0) \) are known (the initial condition), we know also \( \sum_k p_k(0)^\alpha = \sum_k p_k(t)^\alpha \) for any \( \alpha \). We can now use the known result used in the information theory [61] stating that the knowledge of \( \sum_k p_k(t)^\alpha \) for all \( \alpha \) uniquely determines \( p_k(t) \). The continuity in \( t \) implies that \( p_k(t) = p_k(0) \). \( \square \)
The spectral decomposition of the density matrix

\[ \rho_t = \sum_k p_k |k, t\rangle \langle k, t|, \]  

(7.89)

where \( t \mapsto |k, t\rangle \) defines a one-parameter continuous family of orthonormal vectors, leads to the unitary (although \( \rho \)-dependent) transformation \( |k, t\rangle = U(\rho_t, \rho_0)|k, 0\rangle \). The density matrix evolves then as follows

\[ \rho_t = U(\rho_t, \rho_0)\rho_0 U(\rho_t, \rho_0)^{-1}. \]  

(7.90)

**Example 7.2** Let us consider the simplest nontrivial case of a two dimensional Hilbert space. For \( S = S_{n+1} \) discussed below (cf. (7.97)) the evolution equation (up to constant factors) is given by

\[ i \frac{d}{dt} \rho = [\hat{H}, \rho^n]. \]  

(7.91)

Take, for simplicity, a natural even \( n = 2m \). Let

\[ \rho = \mu (1 + \vec{\nu} \vec{\sigma}), \]  

(7.92)

then

\[ \rho^n = \mu^n \left( \sum_{2k=0}^{n} \nu^{2k} 1 + \sum_{2k+1=1}^{n-1} \nu^{2k+1} \vec{n} \vec{\sigma} \right). \]  

(7.93)

Finally, for \( \hat{H} = H_0 1 + \vec{H} \vec{\sigma} \),

\[ \dot{\mu} = 0, \]
\[ \dot{\nu} = 0, \]
\[ \dot{\vec{\nu}} = 2 \mu^n \sum_{2k+1=1}^{n-1} \nu^{2k} \vec{H} \times \vec{\nu}, \]  

(7.94)

hence we get Bloch equations with the frequencies depending on constant functions of initial conditions. Traces of all powers of \( \rho \) depend only on \( \mu \) and \( \nu \) so are integrals of motion. It follows that in such a case the nonlinearity of the equation of motion does not lead us too far from the linear evolution of ordinary QM. The Bloch vector is rotating but the angles of rotation depend nonlinearly on initial conditions.

The question whether the same holds good for \( \rho_0 \) having an infinite number of nonvanishing eigenvalues will be left open here. In any case, it seems that the above theorem is sufficient at least “for all practical purposes”.

The fact that for any \( S \) the evolution conserves \( \text{Tr}(\rho^2) \) suggests that the generalization of \( S_2 \) to (7.85) may be in a sense trivial. To see that this is not the case we have to make our proposal a little bit more concrete — we have to choose some explicit “physical” class of \( S \) — and here the information theoretic introduction may be helpful.

We have noted that the functional \( \text{Tr}(\rho^2) \) is a natural measure of quantum entropy — it describes a system that cannot “learn” or, more formally, whose gain of information is 0 under all circumstances. Let us assume that the appearance of \( S_2 \) in the BBMJ bracket is not just a coincidence but reflects the fact that the linear theory is described by the 2-entropy of Rényi. We formulate therefore the following
Hypothesis 7.1  The fundamental law of quantum physics is that the evolution of density matrices is governed by

\[ \frac{d}{dt} \rho_{AA'}(\alpha, \alpha') = \{ \rho_{AA'}(\alpha, \alpha'), H \}_S \]  

(7.95)

where \( S \) is a measure of entropy and \( H = \text{Tr} \rho \hat{H} \) is a measure of average energy. Averages of observables are defined by

\[ \text{Tr} \rho \hat{F} = \text{Tr} \rho (7.96) \]

which implies that \( \rho \to \lambda \rho \), for any constant \( \lambda \in \mathbb{C} \), must be a symmetry of (7.95).

The suggestion of Wigner that a natural area for nonlinear generalizations of the linear formalism of QM is the domain of observations leads to investigation of systems that can gain information hence are described by \( \alpha \neq 2 \) entropies. A homogeneity preserving generalization of \( S_2 \) for other \( \alpha \)-entropies can be, for instance,

\[ S_\alpha[\rho] = \left( 1 - \frac{1}{\alpha} \right) \frac{\left( \text{Tr} (\rho^\alpha) \right)^{1/(\alpha-1)}}{(\text{Tr} \rho)^{1/(\alpha-1)-1}}. \]  

(7.97)

The choice of the denominator is important only from the point of view of the homogeneity of the evolution equation. The multiplier \( 1 - 1/\alpha \) guarantees that the evolution of pure states is the same, hence linear, for all \( \alpha \) (this is reasonable as pure states have the same, vanishing \( \alpha \)-entropies). The generalized Liouville-von Neumann equation following from (7.97) is

\[ i \frac{d}{dt} \rho = \left( \frac{\left( \text{Tr} (\rho^\alpha) \right)^{1/(\alpha-1)-1}}{(\text{Tr} \rho)^{1/(\alpha-1)-1}} \right) [ \hat{H}, \rho^{\alpha-1} ]. \]  

(7.98)

For pure states and \( \text{Tr} \rho = 1, \rho^n = \rho \) and the equation reduces to the ordinary, linear one; for mixed states the evolution is nonlinear unless the states are “so mixed” that \( \rho \) is proportional to the unit operator (which makes sense in finite dimensional cases, of course) and all \( \alpha \)-entropies reduce to the Hartley formula.

The evolution of (now linear) observables is governed by

\[ i \frac{d}{dt} \hat{F} = \left( \frac{\left( \text{Tr} (\rho^\alpha) \right)^{1/(\alpha-1)-1}}{(\text{Tr} \rho)^{1/(\alpha-1)-1}} \right) \text{Tr} (\rho^{\alpha-1} [\hat{F}, \hat{H}]). \]  

(7.99)

which shows that for the generalized \( S \) the time derivative of an observable is not linear in the density matrix. For \( \alpha = 2 \) the equations reduce again to the ordinary linear equations.

It seems that the following choice of \( S_\alpha \) is also interesting:

\[ S_\alpha[\rho] = \frac{1}{2} \frac{\left( \text{Tr} (\rho^\alpha) \right)^{1/(\alpha-1)-1}}{(\text{Tr} \rho)^{1/(\alpha-1)-1}}. \]  

(7.100)

For pure states the expression reduces to the linear form \( \frac{1}{2} \langle \psi | \psi \rangle^2 = \frac{1}{2} \text{Tr} (\rho^2) \). The density matrix would satisfy then the equation

\[ i \frac{d}{dt} \rho = \frac{1}{2} \frac{\alpha}{\alpha - 1} \frac{\left( \text{Tr} (\rho^\alpha) \right)^{1/(\alpha-1)-1}}{(\text{Tr} \rho)^{1/(\alpha-1)-1}} [\hat{H}, \rho^{\alpha-1}] \]  

(7.101)
which for pure states and normalized $\rho$ would become

$$2 \frac{\alpha - 1}{\alpha} \frac{d}{dt} \rho = [\hat{H}, \rho]$$

(7.102)

and the “Boltzmann-Shannon classical limit” $\alpha \to 1$ of the Rényi entropy is analogous to the $\hbar \to 0$ classical limit of QM.

## 7.3.4 Composition Problem for Subsystems with Different Entropies

Of course, even assuming that the ideology advocated here is meaningful we should bear in mind that the regions of our real world where some nonlinearities reside may be rather few and far between. So one of the first things we have to understand is again the composition problem: What is the form of the equations of motion of a composite system whose parts are described by different entropies?

I think it is best to approach the question again in an information theoretic way. To begin with, let us consider a system whose entropy is $I_\alpha$, and whose subsystems have entropies of the same kind: for all $k$ a $k$-th system’s entropy satisfies $I_{\alpha_k} = I_\alpha$. Let the $k$-th subsystem be described by a reduced density matrix $\rho_k$. The entropy of the “large” system should be defined, as usual in information theory, as the average entropy of the subsystems. The overall entropy of the large systems should not depend on the way we decompose it into subsystems. Therefore the average cannot have the apparently natural form

$$I_{1...n} [\rho] = \sum_k p_k I_\alpha [\rho_k],$$

(7.103)

where $p_k$ are some weights, because the LHS is sensitive to correlations between the subsystems whereas the RHS is not, so that the entropy would be sensitive to the decompositions which are arbitrary. It seems we have to assume that in such a case the composition takes the trivial form

$$I_{1...n} [\rho] = \sum_k p_k I_{\alpha_k} [\rho_k] = \sum_k p_k I_\alpha [\rho] = I_\alpha [\rho].$$

(7.104)

Consider now a situation where the different subsystems have different entropies, say, $I_{\alpha_k}$. The average entropy of the composite system is now defined in analogy to (7.104) as

$$I_{\alpha_1...\alpha_n} [\rho] = \sum_{\alpha_k} p_k I_{\alpha_k} [\rho]$$

(7.105)

where the probabilities $p_k$ are weights describing the “percentage” of each of the entropies in the overall entropy of the system. We do not know how to determine the weights — they can play a role of parameters characterizing the system.

The above definitions imply that the $\alpha^*$-entropy of the large system is

$$I_{\alpha_1...\alpha_n}^* [\rho] = \prod_{\alpha_k} I_{\alpha_k}^* [\rho_k]^{p_k},$$

(7.106)

so it is natural to define

$$S_{\alpha_1...\alpha_n} [\rho] = \prod_{\alpha_k} S_{\alpha_k} [\rho_k]^{p_k}.$$
Denoting the latter expression by $S$, we obtain the evolution of the density matrix of the whole system governed by

$$
\frac{d}{dt} \rho_b(\beta) = [\rho_b(\beta), H, S]
$$

$$
= \sum_{\alpha_k} [\rho_b(\beta), H, S_{\alpha_k}] S_{\alpha_1}^{p_k} \cdots S_{\alpha_{k-1}}^{p_k} \cdots S_{\alpha_n}^{p_n}
$$

$$
= \sum_{\alpha_k} p_b(\beta), H, S_{\alpha_k}] S_{\alpha_1}^{p_k} \cdots S_{\alpha_{k-1}}^{p_k} \cdots S_{\alpha_n}^{p_n}
$$

$$
= \sum_{\alpha_k} p_b(\beta), H, S_{\alpha_k}] S S_{\alpha_k}
$$

(7.108)

Consider again a system which consists of subsystems equipped with the entropy of the same kind. Then $S_{\alpha_k} = S_{\alpha_l} = S$ for all $k$ and $l$ and the system evolves according to

$$
\frac{d}{dt} \rho_b(\beta) = [\rho_b(\beta), H, S]
$$

(7.109)

as expected.

Next possibility is that the entropies that sum to the overall entropy are again sums of other entropies. The description of the whole system should not depend on the order in which the partial entropies are summed up. So consider two entropies $S^I$ and $S^{II}$, with appropriate weights $\lambda^I$ and $\lambda^{II}$, and let the entropies $S^I$ and $S^{II}$ consist of some other entropies $S^I_{\alpha_k}$ and $S^{II}_{\alpha_l}$ appearing with weights $\{p^I_{\alpha_k}\}_{k=1}^{n}$ and $\{p^{II}_{\alpha_l}\}_{l=1}^{n}$, respectively. Then

$$
\frac{d}{dt} \rho_b(\beta) = \{\rho_b(\beta), H\}_S = \lambda^I \{\rho_b(\beta), H\}_{S^I} S^{II} S_{S^I} + \lambda^{II} \{\rho_b(\beta), H\}_{S^{II}} S_{S^I} S_{S^{II}}
$$

$$
= \sum_{k} \lambda^I p^I_{\alpha_k} \{\rho_b(\beta), H\}_{S_k} S^{II} S_{S^I} + \sum_{l} \lambda^{II} p^{II}_{\alpha_l} \{\rho_b(\beta), H\}_{S_l} S_{S^I} S_{S^{II}}
$$

$$
= \sum_{k} \lambda^I p^I_{\alpha_k} \{\rho_b(\beta), H\}_{S_k} S^{II} S_{S^I} + \sum_{l} \lambda^{II} p^{II}_{\alpha_l} \{\rho_b(\beta), H\}_{S_l} S_{S^I} S_{S^{II}}
$$

(7.110)

which shows that the evolution can be indeed consistently composed of “sub-entropies”.

If all $S_{\alpha_k}[\rho]$ depend on $\rho$ only via $f_m[\rho]$, like in our definitions (7.97) and (7.101), we know that on general grounds they are integrals of motion. Consider a subsystem described by $\rho_k$ and which is noninteracting with the subsystem “where the nonlinearity resides”. In such a case the overall Hamiltonian function is

$$
H[\rho] = \sum_k H_k[\rho_k]
$$

(7.111)

and

$$
\frac{d}{dt} \rho_{k} b(\beta) = \sum_l p_l[\rho_{k} b(\beta), H_k[\rho_k], S_{\alpha_l}] S_{S_{\alpha_l}}
$$

(7.112)

A system described by $\rho_k$ and $S_{\alpha_k}$ can be totally isolated from the “rest of the Universe”, if for $k \neq l$,

$$
[\rho_{k} b(\beta), H_k[\rho_k], S_{\alpha_l}] = 0
$$

(7.113)
and
\[ p_k = \frac{S_{\alpha_k}}{S}. \] (7.114)

However, even in such a case the global properties of the large system leave their mark on the local properties of all the subsystems as
\[ p_k \frac{S}{S_{\alpha_k}} \] (7.115)
is at most an integral of motion hence depends on initial conditions. Let, for example,
\[ S[p] = S_{\alpha_1 \ldots \alpha_n}[p] = \prod_{\alpha_k} S_{\alpha_k}[p_k]. \] (7.116)

where the different subsystems have different entropies. Each of the sub-entropies satisfies
\[ \frac{d}{dt} S_{\alpha_k}[\rho_k] = [S_{\alpha_k}[\rho_k], H, S] \]
\[ = \sum_{\alpha_l} \rho_{l}[S_{\alpha_k}[\rho_k], H, S_{\alpha_l}(\rho_l)] \frac{S}{S_{\alpha_l}} = 0 \] (7.117)
in virtue of the theorem 7.2 (we have used here the fact that \( \{F, H\}_S = -\{F, S\}_H \)), and the antisymmetry of the triple bracket. Therefore not only the overall entropy, but also the sub-entropies are integrals of motion even if the Hamiltonian function \( H \) contains interaction terms.

Consider now the reduced density matrix \( \rho_k \) of one of the subsystems. Using the same theorem we find that
\[ \frac{d}{dt} \rho_{k b}(\tilde{\beta}) = p_k[\rho_{k b}(\tilde{\beta}), H, S_{\alpha_k}] \frac{S}{S_{\alpha_k}}. \] (7.118)

where \( S/S_{\alpha_k} \) is an integral of motion but its value depends on initial conditions. If the change of the initial conditions does not affect the reduced density matrices in (7.116), the integral of motion is also unchanged. Therefore in order to change this quantity we have to change correlations between the subsystems. In particular, a time dependence of a linear system which is noninteracting with the nonlinear one is insensitive to changes of initial conditions within the linear system if the particular form (7.116) holds. For global entropies different from (7.116) some kind of sensitivity appears but the influences between the subsystems cannot propagate faster than light unless we introduce the projection postulate.

Such a trace of nonlinearity observed in some linear system might be used to detect the nonlinearity. Following Santilli [64] we can expect that an evolution of an inside part of a hadron may be nonlinear (like in hadronic mechanics). In such a case correlations between a hadron (say, a proton) and some linear system (say, an electron) could be observed in a form of a \( \rho \)-dependent rescaling of time in the electron’s evolution.
Chapter 8
APPENDICES

8.1 Elements of Relativistic Formulation — Hamiltonian Formulation of the Dirac Equation

Consider the Dirac equation

\[ (i\gamma^a \nabla_a - m)\psi = 0 \] (8.1)

where \( \nabla_a = \partial_a + ie\Phi_a \) and \( \Phi_a \) is an electromagnetic potential world-vector. We are going to rewrite the equation in a form of the “proper-time” covariant Hamilton equations of motion. The “proper time” will be defined in terms of spacelike hyperplanes constructed as follows. Let \( \sigma_\tau(x(\tau)) = 0 \) be an equation defining a family of spacelike hyperplanes. The field of timelike, future-pointing, normalized vectors \( n^a_\tau(x) \propto \partial_a \sigma_\tau(x) \), satisfying the continuity equation \( \partial_\tau n^a_\tau(x) = 0 \), defines the field of “proper time” directions. Integral curves \( \tau \mapsto x^a(\tau) \) of \( n^a_\tau(x) \), where \( \tau \) is the parameter of the family \( \{ \sigma_\tau \} \), play the role of the world-lines. We shall need the continuity equation to guarantee the reality of the Hamiltonian function. Notice that this condition eliminates some physically meaningful hyperplanes, like the proper-time hyperboloid \( \sigma_\tau(x) = x^a x_a - \tau^2 = 0 \), but admits simultaneity hyperplanes \( \sigma_\tau(x) = n^a x_a - \tau = 0 \). The “proper time” following from the construction should not, for this reason, be identified with the ordinary proper time of the electron. The “proper time” derivative at \( x \) is defined as

\[ \frac{d}{d\tau} = n^a_\tau(x) \partial_a. \] (8.2)

Let \( \phi_A(x) \) and \( \chi^{A'}(x) \) be the spinor components of a bispinor. Our aim is to rewrite the Dirac equation in the Hamiltonian form

\[ \omega^{AA'} \frac{d}{d\tau} \phi_A(x) = \frac{\delta H}{\delta \phi_{A'}(x)} \] (8.3)

\[ \omega_{AA'} \frac{d}{d\tau} \chi^{A'}(x) = \frac{\delta H}{\delta \chi^A(x)} \] (8.4)

\[ \bar{\omega}^{AA'} \frac{d}{d\tau} \bar{\phi}_A(x) = \frac{\delta H}{\delta \bar{\phi}_{A'}(x)} \] (8.5)
\[ \hat{\omega}_{AA'} \frac{d}{dt} \chi^{A'}(x) = \frac{\delta H}{\delta \chi^A(x)}. \]  

We cannot do this directly in terms of (8.1) because the derivatives are already contracted over the world-vector indices. So let us transform (8.1) into new, rather unusual form. Multiplying (8.1) from left by the Dirac matrices we obtain

\[ 0 = (i \gamma^a \gamma^b \nabla_b - m \gamma^a) \psi = (i \gamma^a + \sigma^{ab} \nabla_b - m \gamma^a) \psi. \]  

Writing the four-potential explicitly in

\[ i \partial^a \psi = (\sigma^{ab} \partial_b + e \Phi^a + i e \sigma^{ab} \Phi_b + m \gamma^a) \psi \]  

and contracting with \( n^a(x) \) we get

\[ i \frac{d}{dt} \psi(x) = \left( -\sigma^{ab} n^b(x) \partial_b + e n^a(x) \Phi_a(x) + i e \sigma_{ab} n^a(x) \Phi^b(x) \right. \]  

\[ \left. + \ m \ n^a(x) \gamma_a \right) \psi(x) \]  

\[ = \hat{H} \psi(x). \]  

The Hamiltonian operator is here a four-scalar, as opposed to the ordinary Dirac Hamiltonian which transforms as \( p_0 \). The Dirac equation (8.8) can be written in a more familiar form if, for \( \Phi^a = 0 \), we introduce the Pauli-Lubanski vector

\[ \frac{i}{2} \sigma^{ab} p_b = -\frac{1}{2} \sigma^{ab} \partial_b = \gamma^5 S^a. \]  

The equation

\[ \frac{1}{2} p^a = \gamma^5 S^a + \frac{m}{2} \gamma^a \]  

expresses the relationship between the four-velocity operator \( \gamma^a \) and the four-momentum \( p^a \). For \( m = 0 \) it shows that spin is helicity \( \times \) momentum. It can be shown also that spinor equations for any spin and \( m = 0 \) (including the source-free Maxwell equations) always take this form.

In what follows we will need the spinor version of (8.9). The Dirac equation will be written as

\[ i \nabla^{AA'} \phi_A = -\mu \chi^{A'} = i g_{aAA'} \nabla_a \phi_A \]  

\[ i \nabla_{AA'} \chi^{A'} = -\mu \phi_A = i g_{aAA'} \nabla_a \chi^{A'} \]  

where \( \mu = m/\sqrt{2} \) and \( g_{aAA'} \) are the Infeld-van der Waerden symbols [36]. We will need the identities [65]

\[ g^{a}_{AA'} g^{b}_{AY'} + g^{b}_{AY} g^{a}_{AY'} = g^{ab}_{XY} \]  

\[ g^{a}_{AA'} g^{b}_{AY'} - g^{b}_{AY} g^{a}_{AY'} = 4 \sigma^{ab}_{XY} \]  

\[ g^{a}_{AY} g^{b}_{AY'} - g^{b}_{AY'} g^{a}_{AY} = 4 \bar{\sigma}^{ab}_{XY} \]
where $\sigma^{ab} Y$ and $\tilde{\sigma}^{ab} Y'$ are generators of $(\frac{1}{2},0)$ and $(0, \frac{1}{2})$ representations of $SL(2, \mathbb{C})$. The generators written in a purely spinorial way are

$$\sigma_{AA'BB'XY} = \frac{1}{4} \varepsilon_{A'B'}(\varepsilon_{AX} \varepsilon_{BY} + \varepsilon_{BX} \varepsilon_{AY})$$

(8.18)

$$\tilde{\sigma}_{AA'BB'XY'} = \frac{1}{4} \varepsilon_{AB}(\varepsilon_{A'X} \varepsilon_{B'Y'} + \varepsilon_{B'X} \varepsilon_{A'Y'}).$$

(8.19)

Repeating essentially the same calculations as those leading from (8.1) to (8.8) we obtain

$$i \nabla_a \phi_X = -4i \sigma_{abX} Y \nabla^b \phi_Y + 2 \mu g_a X \chi^{X'}$$

(8.20)

$$i \nabla_a \chi^{X'} = 4i \sigma_{abY'} \nabla^b \chi^{Y'} - 2 \mu g_a X \phi_X$$

(8.21)

and the equations obtained by their complex conjugation. These equations are especially simple if we express generators and Infeld-van der Waerden symbols in purely spinorial terms. Remembering that $n^a_\tau n_{a\tau} = 1$ implies $n_{AA'\tau} n^{BA'}_{\tau} = \frac{1}{2} \epsilon_A^B$ we get after some calculations

$$i \frac{d}{dt} \phi_X(x) = i n^Y Y' (x) \nabla_X Y' \phi_Y(x) + \mu n_{\tau X} Y' (x) \chi^{Y'}(x)$$

$$+ e n^a_\nu(x) \Phi_a(x) \phi_X(x)$$

(8.22)

$$i \frac{d}{dt} \chi^{X'}(x) = i n_{\tau Y} Y' (x) \nabla_X Y' \chi^{Y'}(x) - \mu n^X Y' (x) \phi_X(x)$$

$$+ e n^a_\nu(x) \Phi_a(x) \chi^{X'}(x)$$

(8.23)

$$- i \frac{d}{dt} \phi_{X'}(x) = -i n^Y Y' (x) \nabla_X Y' \phi_Y(x) + \mu n_{\tau Y} Y' (x) \chi^{Y'}(x)$$

$$+ e n^a_\nu(x) \Phi_a(x) \phi_{X'}(x)$$

(8.24)

$$- i \frac{d}{dt} \chi^{X'}(x) = -i n_{\tau Y} X' (x) \nabla_X X' \chi^{Y'}(x) - \mu n^X X' (x) \phi_{X'}(x)$$

$$+ e n^a_\nu(x) \Phi_a(x) \chi^{X'}(x)$$

(8.25)

Let $d\sigma_\tau(x)$ be some invariant measure on the hyperplane $\sigma_\tau$. The equations can be derived from the Hamiltonian function

$$H[\psi, \bar{\psi}] = \langle \psi | \hat{H} | \psi \rangle$$

$$= \int_{\sigma_\tau} \left\{ i \bar{\phi}_X(x)n^X_{\tau}(x)n^Y_{\tau}(x)\nabla_X Y \phi_Y(x) - i \bar{\chi}^{X'}(x)n_{\tau X'}(x)n_{\tau Y'}(x)\nabla_X Y' \chi^{Y'}(x) + \frac{1}{2} \mu \left( \bar{\phi}_{X'}(x)\chi^{X'}(x) + \bar{\chi}^{X'}(x)\phi_X(x) \right) 
+ e n^a_\nu(x) \Phi_a(x)n^{X'}_{\tau}(x) \left( \phi_X(x)\bar{\phi}_{X'}(x) + \bar{\chi}_X(x)\chi_{X'}(x) \right) \right\} d\sigma_\tau(x)$$

(8.26)

provided

$$\partial^{X'} n_{\tau X'}(x) = \partial^{X'} n_{\tau X'}(x) = 0$$

(8.27)
and the wave functions vanish at boundaries of the hyperplane $\sigma\tau$. Reality of $H$ is guaranteed by the same conditions. The Hamiltonian function is not positive definite, which is correct since we are working here in first quantized formalism. Contraction of (8.27) over the remaining indices implies the continuity equation discussed above.

The explicit form of the Hamilton equations is

\begin{align}
 i n^X_{\tau}(x) \frac{d}{d\tau} \phi^X(x) &= \frac{\delta H}{\delta \phi^X(x)}, \\
 i n_{\tau X'}(x) \frac{d}{d\tau} \chi^X(x) &= \frac{\delta H}{\delta \chi^X(x)}, \\
 -i n^X_{\tau}(x) \frac{d}{d\tau} \phi^X(x) &= \frac{\delta H}{\delta \phi^X(x)}, \\
 -i n_{\tau X'}(x) \frac{d}{d\tau} \chi^X(x) &= \frac{\delta H}{\delta \chi^X(x)}.
\end{align}

or, in the Poissonian way,

\begin{align}
 i \frac{d}{d\tau} \phi^X(x) &= 2 n_{\tau X'}(x) \frac{\delta H}{\delta \phi^X(x)}, \\
 i \frac{d}{d\tau} \chi^X(x) &= 2 n^X_{\tau}(x) \frac{\delta H}{\delta \chi^X(x)}, \\
 -i \frac{d}{d\tau} \phi^X(x) &= 2 n_{\tau X'}(x) \frac{\delta H}{\delta \phi^X(x)}, \\
 -i \frac{d}{d\tau} \chi^X(x) &= 2 n^X_{\tau}(x) \frac{\delta H}{\delta \chi^X(x)}.
\end{align}

We can see that $i n_{\tau X'}(x) = \omega_{\tau X'}(x)$ are the components of the symplectic (since derivable from a Kähler potential) form on $\sigma\tau$ at point $x \in \sigma\tau$, and the Poissonian form $I_{\tau X'}(x) = -2i n_{\tau X'}(x)$.

Now, we can repeat the whole reasoning that has led us to the BBMJ bracket, its triple bracket generalization and the entropic framework. It seems there is no general obstacle for such a generalization. Note that now we can do that in a manifestly covariant way, so that the possible non-Poissonian structure of the manifold of states is unrelated to the relativistic covariance of the theory.

It is clear, on the other hand, that the representation of the covering of the Poincaré group we are dealing with acts in the Hilbert space of the bispinor $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation of $SL(2, \mathbb{C})$. The operators of the “space-time translations” which determine the origin of a Minkowski space coordinate system are, of course, linear operators in opposition to the generator of evolution which will not be a linear operator. This is a very interesting problem. A hint for its clarification is, perhaps, our observation that an evolution of the density matrix may be determined by a unitary $\rho$-dependent transformation acting in the Hilbert space.
8.2 Schrödinger Form of the Harmonic Oscillator Equations

Consider an arbitrary number of independent harmonic oscillators. Their Hamiltonian function is
\[ H(\vec{q}, \vec{p}) = \sum_k \left( \frac{p_k^2}{2m_k} + \frac{m_k\omega_k^2 q_k^2}{2} \right). \] (8.36)

Let us now introduce the complex coordinates
\[ z_k = \frac{p_k}{\sqrt{2m_k}} - i \sqrt{\frac{m_k}{2}} \omega_k q_k \] (8.37)
whose dimension is [energy]^{1/2}, and let \( \sqrt{\varepsilon_0} \) be a constant with the same dimension. It is natural to introduce the dimensionless coordinates \( \psi_k = z_k/\sqrt{\varepsilon_0} \). The Hamiltonian function is
\[ H(\psi, \bar{\psi}) = \sum_k z_k \bar{z}_k = \varepsilon_0 \sum_k \psi_k \bar{\psi}_k. \] (8.38)

Let \( \omega_0 \) be a constant with the dimension of frequency. The Hamilton equations of motion read
\[ i\varepsilon_0 \omega_0 \frac{d}{dt}\psi_k = \omega_k \omega_0 \frac{\partial H}{\partial \bar{\psi}_k} \] (8.39)
\[ -i\varepsilon_0 \omega_0 \frac{d}{dt}\bar{\psi}_k = \omega_k \omega_0 \frac{\partial H}{\partial \psi_k}. \] (8.40)

The same equations can be obtained from the Hamiltonian function
\[ H'(\psi, \bar{\psi}) = \sum_k \frac{\varepsilon_0}{\omega_0} \omega_k \psi_k \bar{\psi}_k. \] (8.41)

Introducing another constant \( \frac{\varepsilon_0}{\omega_0} := h \) we get the Schrödinger equation
\[ i\hbar \frac{d}{dt}\psi_k = \frac{\partial H'}{\partial \bar{\psi}_k} = h\omega_k \psi_k \] (8.42)
\[ -i\hbar \frac{d}{dt}\bar{\psi}_k = \frac{\partial H'}{\partial \psi_k} = h\omega_k \bar{\psi}_k \] (8.43)
which is nothing but a different representation of the ordinary harmonic oscillator equation for independent oscillators. It is surprising, however, that the Hamiltonian \( H' \) is not equal to energy. The energy \( H \) is, on the other hand, proportional to the “squared norm” \( \sum_k \psi_k \bar{\psi}_k \).

Consider now a three-dimensional harmonic oscillator oscillating with the frequency \( \omega \). There exists a conserved quantity \( \vec{J} = \vec{q} \times \vec{p} \). In the dimensionless coordinates
\[ \vec{J} = i\hbar \frac{\omega_0}{\omega} \vec{\psi} \times \bar{\vec{\psi}}. \] (8.44)

A \( k \)-th component of \( \vec{J} \) is
\[ J_k(\vec{\psi}, \bar{\vec{\psi}}) = -i\hbar \frac{\omega_0}{\omega} e_{klm} \bar{\psi}_l \psi_m \] (8.45)
hence an observable bilinear like in QM. Analogously to $H'$ we can introduce the conserved quantity $\vec{J}' = (\omega/\omega_0)\vec{J}$. Both $\vec{J}'$ and $\vec{J}$ can be represented in terms of matrices

$$ (J'_{km})_{lm} = -i\hbar\epsilon_{klm} \quad (8.46) $$

satisfying

$$ [J'_k, J'_l] = i\hbar\epsilon_{klm}J'_m. \quad (8.47) $$

The same Lie-algebraic property follows from the Poisson bracket resulting from the $H'$ version of the Hamilton equation. The bracket following from the $H$ version is analogous but involves a modified, $\omega$ dependent “$\hbar$”.

### 8.3 On Certain Associative Product of Nonlinear Observables

What is the role of Weinberg’s multiplication rule?...

Can it be used like that for algebraic relations of observable quantities? Or is its role only to express the Lie bracket as a commutator? Is there another multiplication rule to be used for algebraic relations of physical quantities? None has been proposed so far.

In linear QM the associativity of observables follows from their bilinearity in $\psi$ and $\bar{\psi}$. We know that this property makes possible two equivalent formulations of the algebra of observables: The operator and Weinberg’s $\ast$-product ones. The $\ast$-product appears naturally because the Poisson bracket (which is equivalent to the commutator of operators) can be written as

$$ \{F, G\} = F \ast G - G \ast F. \quad (8.48) $$

In what follows I will present another product having this property, but in addition being associative for any infinite differentiable functions defined on a Hilbert space (I must admit that initially I hoped that such products will play an important role for the probability interpretation of Weinberg’s nonlinear QM, but this program has finally led nowhere). The product is analogous (although not equivalent) to the Moyal product, and the proof of its associativity is identical to this for general $\ast_{\lambda}$-products investigated by Flato et al. [66]. The amount of typographical errors in [66] makes the proof practically unreadable, so let me repeat it. I am using the finite dimensional conventions but its extension to functional derivatives is straightforward.

The Weinberg product is denoted here by

$$ F \ast G = \delta_{AA'} \frac{\partial F}{\partial \psi_A} \frac{\partial G}{\partial \bar{\psi}_{A'}} := \partial^A F \partial_{A} \bar{G}. \quad (8.49) $$

We shall consider the product

$$ F \ast G = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_{A_1 A'_1} \cdots \delta_{A_k A'_k} \frac{\partial^k F}{\partial \psi_{A_1} \cdots \partial \psi_{A_k}} \frac{\partial^k G}{\partial \bar{\psi}_{A'_1} \cdots \partial \bar{\psi}_{A'_k}} $$

81
\[
\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \frac{\partial^k F}{\partial \psi_{A_1} \ldots \partial \psi_{A_k}} \frac{\partial^k G}{\partial \bar{\psi}_{A_1} \ldots \partial \bar{\psi}_{A_k}}
= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \partial A_1 \ldots \partial A_k F \partial A_1 \ldots \partial A_k G
= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} P^k(F,G)
\]

which, for bilinear observables, reduces only to the first two terms

\[
F \ast G = FG + \lambda F \ast G.
\]  

and the commutator

\[
F \ast G - G \ast F = \lambda^{-1} (F \ast G - G \ast F).
\]  

For \( \lambda = -1 \), and normalized states (8.51) is a correlation. Accordingly, \( \ast \) can be regarded as an associative generalization of correlations and standard deviations for generalized observables.

**Theorem 8.1** \( \ast \)-product is formally associative.

**Proof:** Since

\[
(F \ast G) \ast H = \sum_{t=0}^{\infty} \lambda^t \sum_{r+s=t} \frac{1}{r!s!} P^r(P^s(F,G),H)
\]
and

\[
F \ast (G \ast H) = \sum_{t=0}^{\infty} \lambda^t \sum_{r+s=t} \frac{1}{r!s!} P^r(F,P^s(G,H))
\]

the two expressions will be equal if

\[
T_t(F,G,H) = U_t(F,G,H)
\]

where

\[
T_t(F,G,H) = \sum_{r+s=t} \frac{1}{r!s!} P^r(P^s(F,G),H)
\]
and

\[
U_t(F,G,H) = \sum_{r+s=t} \frac{1}{r!s!} P^r(F,P^s(G,H)).
\]
Taking into account that
\[
P^r(F, P^s(G, H)) = \partial A_{1}...A_{r} F \sum_{r'=0}^{r} \left( \begin{array}{c} r \\ r' \end{array} \right) \tilde{\partial} A_{1}...A_{r'} \partial B_{1}...B_{s} G \tilde{\partial} A_{r'+1}...A_{r} B_{1}...B_{s} H \tag{8.58}
\]
we obtain
\[
U_t(F, G, H) = \sum_{r'+s \leq t} \frac{1}{r'! s!(t-r'-s)!} \partial A_{1}...A_{r'} B_{1}...B_{s} G \partial A_{r'+1}...A_{t-s} F \tag{8.59}
\]
\[
T_t(F, G, H) = \sum_{r'+s \leq t} \frac{1}{r'! s!(t-r'-s)!} \partial B_{1}...B_{r'} A_{1}...A_{s} G \partial A_{r'+1}...A_{t-s} B_{1}...B_{s} H \tag{8.60}
\]

The proof is completed by the exchange between \( r' \) an \( s \).

### 8.4 Polarizing Mach-Zehnder Interferometer and Other Analyzers of Polarization

In Chapter 4 we have illustrated the physical meaning of the “malignant” nonlocal behavior of separated systems in Weinberg’s nonlinear QM with the help of a specific polarizing Mach-Zehnder interferometer. A detailed description of the interferometer and other analyzers of polarizations is given below. The results presented in this section are a part of [67].

From a quantum field theoretic viewpoint an optical device is a scattering area and, as such, has to be treated in a quantum manner. Any scattering process is described by means of a unitary transformation, the \( S \)-matrix, which maps a system’s “in” Fock space at time \( t = -\infty \) into the “out” Fock space at \( t = \infty \). The limits \( t = \pm \infty \) play a role of a mathematical idealisation expressing the fact that at such times the “in” and “out” fields can be assumed free. It must be stressed that, even in the most elaborated axiomatic approaches, like the Haag-Ruelle scattering theory [68], one does not develop any physical interpretation of what is going on at the very scattering region and, instead, concentrates on relations between the asymptotic fields.

In practical applications in quantum optics the situation considerably simplifies with respect to the general scattering theory. All the scattering areas are localized within small regions of space and the free field “\( \pm \infty \)” asymptotics is achieved in a trivial way by simply assuming that the evolution is free outside of the device. Like in the more general case, one does not attempt to describe all the physical phenomena occurring during, say, a photon’s transmission through a half-wave plate.

In the Heisenberg picture the \( S \)-matrix transforms the creation operators according to
\[
S^\dagger a^\text{out} S = a^\text{in} \tag{8.62}
\]
The $S$-matrix at the LHS of Eq. (8.62) maps an input state $|\Psi_{\text{in}}\rangle$ into $|\Psi_{\text{out}}\rangle = S|\Psi_{\text{in}}\rangle$. In the majority of practical applications the linear mapping (8.62) can be written in a form of a unitary, finite dimensional matrix $U$ acting in some subspace of the whole “in” CCR algebra. Such a situation occurs, for example, if the whole linear transformation decomposes into irreducible parts corresponding to some given frequency, direction of propagation or polarization. The form of this matrix depends on the choice of the “in” and “out” bases in the Fock space. Therefore, in order to describe a given optical device it is often easier to find the form of $U$ than this of $S$, and once such a matrix is given the optical device is defined.

One of the simplest optical systems, a lossless, symmetric half-silvered mirror, is described by a $2 \times 2$ unitary matrix corresponding to two “in” and two “out” modes of the electromagnetic field. Inclusion of polarizations generally increases the number of the required modes to four.

### 8.4.1 Analyzer of circular polarizations

A lossless analyzer of circular polarizations (ACP) can be described by a $4 \times 4$ unitary matrix corresponding to two “in” ports with two polarization modes for each of them; operators with subscripts “+” and “−” correspond to, respectively, the right- and left-handed modes. The “in” and “out” annihilation operators must, by definition of ACP, satisfy the following relations

$$
\langle \Psi | S_C^{\dagger} a_{1+}^{\text{out}} | \Psi \rangle = \langle \Psi | a_{1+}^{\text{in}} | \Psi \rangle
$$

for any $|\Psi\rangle$ satisfying $a_{2+}^{\text{in}} |\Psi\rangle = a_{2-}^{\text{in}} |\Psi\rangle = 0$.

Conditions (8.63) are minimal in the sense that we assume the analyzer to analyze at least one input channel’s circular polarizations. We do not make here any assumptions about the possible output state of light had some photons been introduced also through the second input port.

In what follows we shall use the annihilation operators ordered in vectors of the following form (the superscript $T$ denotes a transpose of a vector)

$$
a^T = (a_1, a_2, a_3, a_4) = (a_{1+}, a_{1-}, a_{2+}, a_{2-}).
$$

Let us fix now the circular polarization “in” and “out” annihilation operators according to the ordering given by (8.64). The action of ACP can be characterized by the matrices $S_C$ and $U_C$ defined by $S_C^{\dagger} a_{k}^{\text{out}} S_C = U_C k |a_{k}^{\text{in}}\rangle$. The relations (4) supplemented with the unitarity condition lead to

$$
U_C = \begin{pmatrix}
\exp(\im \varphi) & 0 & 0 & 0 \\
0 & U_{23} & U_{24} & 0 \\
0 & 0 & U_{33} & U_{34} \\
0 & \exp(\im \chi) & 0 & 0
\end{pmatrix}
$$

(8.65)

where the matrix $U' = (U_{23}, U_{24})$ is unitary. The nonsymmetric form of (8.65) follows only from our conventions concerning the numbering of the annihilation operators; (8.65) is evidently unitarily equivalent to a block diagonal unitary matrix.
The difference $\varphi - \chi$ corresponds to a phase shift between the reflected and the transmitted beams and the freedom in the choice of $U'$ reflects the freedom of the choice of the possible behavior of the ACP with respect to the second input port. One can consider, for example, either symmetric or non-symmetric ACP. For the purposes of this paper the choice of $\varphi$ and $\chi$ is irrelevant and remains a question of convention as we will use the ACP only in the context of the polarizing Mach-Zehnder interferometer and, of course, all the phase shifts can be collected into an overall phase of the interferometer’s phase shifter. Therefore, in no way restricting the generality of the solution, we will later put $\chi = 0$. We shall also put $e^{i\varphi} = i$ in order to obtain an agreement with conventions appearing in other papers (cf. [46]).

One can check now that (8.65) represents indeed an analyzer of circular polarizations for the “1-in” port by explicitly writing the “out” circular polarization modes annihilation operators in terms of the “in” ones:

\begin{align*}
a^{\text{out}}_{1+} &= S_C e^{i\varphi} a^{\text{in}}_{1+} S_C^\dagger \\
a^{\text{out}}_{1-} &= S_C (U_{23} a^{\text{in}}_{2+} + U_{24} a^{\text{in}}_{2-}) S_C^\dagger \\
a^{\text{out}}_{2+} &= S_C (U_{33} a^{\text{in}}_{2+} + U_{34} a^{\text{in}}_{2-}) S_C^\dagger \\
a^{\text{out}}_{2-} &= S_C e^{i\chi} a^{\text{in}}_{1-} S_C^\dagger.
\end{align*}

(8.66) means that the lefthanded photons in the “1-out” port can arrive only from the “2-in” port. The same concerns the righthanded photons in “2-out”. On the other hand, all the righthanded photons coming from “1-in” are reflected into “1-out” and all the lefthanded ones entering through “1-in” are transmitted into “2-out”. By the way, we see from (8.66) that the device acts also as an analyzer of the orthogonal modes $U_{23} a^{\text{in}}_{2+} + U_{24} a^{\text{in}}_{2-}$ and $U_{33} a^{\text{in}}_{2+} + U_{34} a^{\text{in}}_{2-}$ of the second input port.

To complete this part of the discussion let us note that although the two inputs evolve independently of each other, we cannot use $2 \times 2$ matrices to describe the analyzer. This follows from the fact that the outputs do not actually correspond to 1-dimensional subspaces of the Fock space. We can always place another analyzer in one of the output ports of our analyzer and this device will analyze again some two orthogonal polarizations. Therefore the “vacuum input port” “2” must be introduced in order to make us capable of considering problems of the Malus law variety.

### 8.4.2 Analyzers of linear polarizations

The Wollaston quartz or calcite prisms, nicol, or Glan-Thomson and Glan-Foucault polarizers are analyzers of linear polarizations (ALP).

Let us introduce the following linear polarization “in” annihilation operators

\begin{align*}
a^{\text{in}}_{1\parallel\alpha} &= \frac{1}{\sqrt{2}} (e^{i\alpha} a^{\text{in}}_{1+} + a^{\text{in}}_{1-}) \\
a^{\text{in}}_{1\perp\alpha} &= \frac{1}{\sqrt{2}} (e^{i\alpha} a^{\text{in}}_{1+} + a^{\text{in}}_{1-}) \\
a^{\text{in}}_{2\parallel\alpha} &= \frac{1}{\sqrt{2}} (-e^{i\alpha} a^{\text{in}}_{2+} + a^{\text{in}}_{2-}) \\
a^{\text{in}}_{2\perp\alpha} &= \frac{1}{\sqrt{2}} (e^{i\alpha} a^{\text{in}}_{2+} + a^{\text{in}}_{2-}).
\end{align*}

(8.67)
The linear polarizations are parametrized by the relative phase between some fixed circular polarizations. Such a parametrization is “spinorial” in the sense that a change of $\alpha$ by $\pi$ rotates the polarization plane by $\pi/2$. Let now the “out” linear polarization modes be constructed from $a_{1\perp}^{\text{out}}$ and $a_{2\parallel}^{\text{out}}$ in the same way. The ALP is defined in the way analogous to (8.63), i.e. by

$$\langle \Psi | S_{L\alpha}^{\dagger} a_{1\perp}^{\text{out}} a_{1\perp} \Psi \rangle = \langle \Psi | a_{1\perp}^{\text{in}} a_{1\perp} | \Psi \rangle$$
$$\langle \Psi | S_{L\alpha}^{\dagger} a_{2\parallel}^{\text{out}} a_{2\parallel} \Psi \rangle = \langle \Psi | a_{2\parallel}^{\text{in}} a_{2\parallel} | \Psi \rangle$$
$$a_{1\parallel\alpha}^{\text{out}} S_{L\alpha} \Psi = a_{2\perp\alpha}^{\text{out}} S_{L\alpha} | \Psi \rangle = 0 \quad (8.68)$$

for any $| \Psi \rangle$ satisfying $a_{2\perp\alpha}^{\text{in}} | \Psi \rangle = a_{2\perp\alpha}^{\text{in}} | \Psi \rangle = 0$.

Repeating the calculations leading to (8.65) we will obtain the following general formula for the ALP

$$a_{1\perp\alpha}^{\text{out}} = S_{L\alpha} e^{i\varphi} a_{1\perp\alpha}^{\text{in}} S_{L\alpha}^{\dagger}$$
$$a_{1\parallel\alpha}^{\text{out}} = S_{L\alpha} (U_{23} a_{2\perp\alpha}^{\text{in}} + U_{24} a_{2\parallel\alpha}^{\text{in}}) S_{L\alpha}^{\dagger}$$
$$a_{2\parallel\alpha}^{\text{out}} = S_{L\alpha} (U_{33} a_{2\perp\alpha}^{\text{in}} + U_{34} a_{2\parallel\alpha}^{\text{in}}) S_{L\alpha}^{\dagger}$$
$$a_{2\perp\alpha}^{\text{out}} = S_{L\alpha} e^{i\chi} a_{1\parallel\alpha}^{\text{in}} S_{L\alpha}^{\dagger} \quad (8.69)$$

### 8.4.3 Polarizing Mach-Zehnder Interferometer

Let us consider PMZI depicted in Fig. 4.1. $S$ is a source of light, $BS_1$ analyzer of circular polarizations (one may think about some other orthogonal polarizations — the algebraic relations will in general remain the same), $M$ are fully reflecting mirrors, $\alpha$ an optical retarder, $\lambda/2$ a half-wave plate and $BS_2$ an ordinary symmetric beam splitter.

The description of $BS_1$ has been given in the subsection 8.4.1. In the rest of this paper we will always assume that a state of the electromagnetic input field is annihilated by the annihilation operators of the “2-in” port of $BS_1$. Therefore we can further simplify our calculations by putting $U' = I$ and this will not affect the physical results and their interpretations.

The remaining parts of PMZI are given by

$$U_{\lambda/2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (8.70)$$

$$U_{\alpha} = \begin{pmatrix} e^{i\alpha} & 0 & 0 & 0 \\ 0 & e^{i\alpha} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (8.71)$$

$$U_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} i & 0 & 1 & 0 \\ 0 & i & 0 & 1 \\ 1 & 0 & i & 0 \\ 0 & 1 & 0 & i \end{pmatrix} \quad (8.72)$$
The whole interferometer acts therefore as

\[ U_{I\alpha} = U_{2}U_{\alpha}U_{\lambda/2}U_{1} = \frac{1}{\sqrt{2}} \begin{pmatrix}
-\epsilon^{i\alpha} & 1 & 0 & 0 \\
0 & 0 & \epsilon^{i\alpha} & 1 \\
\epsilon^{i\alpha} & i & 0 & 0 \\
0 & 0 & \epsilon^{i\alpha} & i
\end{pmatrix} \]  

(8.73)

and the annihilation operators at the detector positions are given by

\[
\begin{align*}
a^{out}_{1+} &= S_{I\alpha} \frac{1}{\sqrt{2}} (-\epsilon^{i\alpha}a^{in}_{1+} + a^{in}_{1-}) S_{I\alpha}^d \\
a^{out}_{1-} &= S_{I\alpha} \frac{1}{\sqrt{2}} (\epsilon^{i\alpha}a^{in}_{2+} + a^{in}_{2-}) S_{I\alpha}^d \\
a^{out}_{2+} &= S_{I\alpha} \frac{1}{\sqrt{2}} (\epsilon^{i\alpha}a^{in}_{1+} + ia^{in}_{1-}) S_{I\alpha}^d \\
a^{out}_{2-} &= S_{I\alpha} \frac{1}{\sqrt{2}} (\epsilon^{i\alpha}a^{in}_{2+} + ia^{in}_{2-}) S_{I\alpha}^d.
\end{align*}
\]

(8.74)

Recalling our conclusions from Sec. 8.4.2 we understand that the interferometer acts as an analyzer of two orthogonal linear polarizations incoming through “1-in” and also of some two linear polarizations arriving from “2-in”. Indeed, rewriting (8.74) with the help of (8.67) we obtain

\[
\begin{align*}
a^{out}_{1+} &= S_{I\alpha} a^{in}_{1\parallel\alpha} S_{I\alpha}^d \\
a^{out}_{1-} &= S_{I\alpha} a^{in}_{2\perp(\alpha+\pi/2)} S_{I\alpha}^d \\
a^{out}_{2+} &= S_{I\alpha} ia^{in}_{1\perp\alpha} S_{I\alpha}^d \\
a^{out}_{2-} &= S_{I\alpha} ia^{in}_{2\parallel(\alpha+\pi/2)} S_{I\alpha}^d.
\end{align*}
\]

(8.75)

Let us note that this ALP belongs to the class of analyzers mentioned in the second remark after Eq. (8.68). We have here the required 1–1 relationship between the input polarization state and the relevant output channel, but no matter what the polarization of the input beam is, any light arriving from “1-in” leaves the PMZI in the righthanded state (and also any light arriving from “2-in” will leave the PMZI in the lefthanded state).

This property of the PMZI is easy to understand. Any state arriving from “1-in” is split into circular polarization components at BS1. The righthanded part goes through the “upper” path, is scattered on BS2 and does not change the polarization, whereas the lefthanded component changes its polarization in the half-wave plate and also reaches BS2 in the righthanded polarization state.

The unitarity of the transformation implies that any lefthanded “out” photon must arrive from “2-in”.

The linear polarization annihilation operators (8.67) satisfy the following formulas

\[
\begin{align*}
a^{in}_{1\parallel\alpha} &= \frac{1}{2} (1 + \epsilon^{i(\alpha-\beta)}) a^{in}_{1\parallel\beta} + \frac{1}{2} (1 - \epsilon^{i(\alpha-\beta)}) a^{in}_{1\perp\beta} \\
a^{in}_{1\perp\alpha} &= \frac{1}{2} (1 - \epsilon^{i(\alpha-\beta)}) a^{in}_{1\parallel\beta} + \frac{1}{2} (1 + \epsilon^{i(\alpha-\beta)}) a^{in}_{1\perp\beta}
\end{align*}
\]

(8.76)
(the Malus law). With the help of (14) one finds that

\[
S_{I\beta}a_{1\parallel\alpha}^{in}S_{I\beta}^\dagger = \frac{1}{2}\left(1 + e^{i(\alpha - \beta)}\right)a_{1+}^{out} - \frac{i}{2}\left(1 - e^{i(\alpha - \beta)}\right)a_{2+}^{out}
\]

\[
S_{I\beta}a_{1\perp\alpha}^{in}S_{I\beta}^\dagger = \frac{1}{2}\left(1 - e^{i(\alpha - \beta)}\right)a_{1+}^{out} - \frac{i}{2}\left(1 + e^{i(\alpha - \beta)}\right)a_{2+}^{out}.
\] (8.77)

The particular case \(\beta = \alpha + \pi\) reads

\[
S_{I(\alpha+\pi)}a_{1\parallel\alpha}^{in}S_{I(\alpha+\pi)}^\dagger = -ia_{2+}^{out}
\]

\[
S_{I(\alpha+\pi)}a_{1\perp\alpha}^{in}S_{I(\alpha+\pi)}^\dagger = a_{1+}^{out}.
\] (8.78)

A comparison of (8.78) with (8.75) shows that a change of the phase shift from \(\alpha\) to \(\alpha + \pi\) is analogous to a rotation of a polarizer (say, the Wollaston prism) by \(\pi/2\), or a rotation of the Stern-Gerlach device by \(\pi\).
Bibliography


[38] P. A. M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Science, Yeshiva University (1964).


[56] U. Fano,


