On the Linearity of the Schrödinger Equation

Iwo Bialynicki-Birula
Center for Theoretical Physics, Lotnikow 32/46, 02-668 Warszawa, Poland
Received on 12 December, 2004

The problem of the linearity of the Schrödinger equation is described from a historical perspective. It is argued that the Schrödinger picture on which this equation is based cannot be retained in relativistic quantum theory. A closer analysis of realistic experiments might offer a clue how to modify the evolution equation for the state vectors in quantum field theory.

I. INTRODUCTION

The Principle of Superposition of States was recognized by Dirac [1] as a fundamental law of quantum theory. The application of this principle not only to "the states at one particular time" but also to the states of motion ("the states throughout the whole of time after the preparation") leads necessarily to the linearity of the evolution equation — the Schrödinger equation. The more fundamental is the law, however, the more important it is to conduct precise, quantitative tests of this law. Thirty years ago [2–4] we have proposed a nonlinear version of the Schrödinger equation — the logarithmic Schrödinger equation — that is best suited to test the linearity of time evolution of the wave function.

The idea of introducing some kind of nonlinearity into the evolution of quantum systems is as old as wave mechanics itself. Already in 1927 Louis de Broglie [5] proposed his theory of double solution to provide a causal background for the probabilistic interpretation. After suffering a severe criticism at the 1927 Solvay Congress he abandoned the theory of the double solution but later he returned to it and in 1956 he has written the whole book entitled "Nonlinear Wave Mechanics" [6, 7]. His change of mind was caused by the appearance in early fifties of a series of papers by David Bohm [8, 9]. De Broglie and Bohm theories treat quantum mechanics as a statistical average of a nonlinear, classical motions of "a particle with well defined position and momentum" [10]. These theories, while offering a different interpretation of quantum theory, "lead to the same prediction for the experimental results" [10].

II. LOGARITHMIC SCHRÖDINGER EQUATION

In our work, we have introduced a nonlinear term directly in the Schrödinger equation, thus *modifying the predictions* of the theory. The logarithmic nonlinearity was chosen because it has several unique properties. First of all, it guarantees the separability of noninteracting subsystems. Other nonlinear modifications of the Schrödinger equation (for example, the commonly used quartic nonlinearity — Gross-Pitaevski equation) introduce the interaction between two subsystems, due to the nonlinear term even when there are no real forces acting between them. The electron on the Moon will feel the influence of a proton on the Earth. Secondly, the stationary states can always be normalized, according to the standard probabilistic interpretation. For other nonlinearities, stationary solutions have their norms fully determined and after the multiplication by a constant they cease to satisfy the equation. Finally, the Planck relation $\hbar\omega = E$ holds. For other nonlinearities the frequency of a stationary state is not, in general, equal to the energy (including the energy of the nonlinear interaction) divided by $\hbar$. An additional, pleasant feature of the logarithmic Schrödinger equation is that it possesses simple analytic solutions in any number of dimensions. These solutions, called by us Gaussons, because of the Gaussian shape, move in the absence of forces as nonspreading wave packets, similar to solitons. In view of all these similarities, the logarithmic Schrödinger equation

$$i\hbar\partial_t\psi(\mathbf{r},t) = \left(-\frac{\hbar^2}{2m}\Delta + V(\mathbf{r},t) - b\log(|\psi(\mathbf{r},t)|^2/a^3)\right)\psi(\mathbf{r},t) \quad (1)$$

can be used for testing the linearity of the time evolution of the wave function.

III. EXPERIMENTAL SEARCHES FOR A NONLINEARITY

The experimental upper limit on the constant b that measures the strength of the nonlinear term determines in a quantitative manner how small any departure from the linear regime must be. The constant b has the dimension of the energy and the most obvious test is to find corrections to the energy levels caused by the nonlinear term. For example, from the known agreement between the measured and calculated values of the Lamb shift in hydrogen we obtained [3] that $b < 4\times10^{-10}$eV. Energy measurements, however, are not the most sensitive tests of the linearity of the equation. Wave interference phenomena are much more sensitive test of the violation of the superposition principle due to the nonlinear term.

The first such test was proposed by Shimony [11] and realized by Shull *et. al.* [12]. In the experiment, the neutron beam was split and an absorber was placed in the path of one part of the beam. The absorption decreases the value of $|\psi(\mathbf{r},t)|^2$ and this, in turn, if the nonlinear term is present, introduces a change of the phase of the wave function. Such a change of phase would produce interference fringes when the two parts of the beam are combined. "No measurable phase shift beyond experimental uncertainty was found and an upper limit of 3.4×10^{-13}eV for the energy constant b was established." [12] The second experiment performed by Gähler, Horne and Zeilinger [13] was even more directly testing the validity of

the linear superposition principle. It consisted of very precise measurements of the Fresnel diffraction of a slow neutron beam by an absorbing straight edge. No departure from the linear evolution was found and the upper limit on b was set at the incredibly small value of 3.3×10^{-15}eV. These two ground-breaking experiments practically ruled out the presence of a nonlinear term in the Schrödinger wave equation. The logarithmic Schrödinger equation outlived these defeats but not as a fundamental theory. Owing to its unique properties it has been used as an exactly soluble model of nonlinear phenomena in nonlinear optics [16, 17], in nuclear physics [14], in the study of dissipative systems [15], in geophysics [18], and even in computer science [19]. The logarithmic nonlinearity is also theoretically appealing due to its connection with stochastic dynamics [22–24].

Our attempt to produce an equation that might be used to test the linearity of the quantum theory was the first but not the only one. In 1988 Steven Weinberg [20] proposed a direct test of the linearity of the time evolution in the hyperfine transition observed in the $^9Be^+$ ion. By comparing the results of his calculation with the existing measurements [21], Weinberg estimated the upper limit of the nonlinearity in the evolution, of this essentially two-level system, to be of the order of 10^{-15}eV, the same as in the neutron Fresnel-diffraction experiments.

IV. LINEAR VS. NONLINEAR

Since the experiments rule out nonlinearities in the quantum evolution equation, one should focus the attention on finding an explanation of this fact. An almost obvious explanation is along the lines already envisioned by de Broglie and Bohm. Namely, one may hypothesize that the linearity is achieved at the cost of abandoning a causal microscopic description and replacing it by some kind of a statistical average. This line of reasoning may be illustrated with a very simple example — the logistic iteration.

$$x_{n+1} = r x_n (1 - x_n). \tag{2}$$

It is well known that the essential feature of this iteration, leading to chaotic behavior, is its nonlinear (quadratic) character. Of course, the nonlinearity is not sufficient to generate chaos. For example, the rational iteration

$$x_{n+1} = \frac{a x_n + b}{c x_n + d} \tag{3}$$

does not lead to chaos because it is exactly soluble. In this case one can write a closed formula for the k-th iteration. The nonlinear iteration rule of the logistic iteration, however, can be replaced by a a linear rule by changing the description "microscopic" to "macroscopic". Indeed, one may replace the transformation formula (2) that describes the trajectory of a "point" x_n by an equivalent linear formula that describes the evolution of the distribution function (i.e. the density of points for different initial conditions) $f(x,n) \to f(x,n+1)$, where n is the discrete "time variable" counting the iterations. This change of description is very similar to going from the Lagrangian description to the Eulerian description in fluid dynamics (cf., for example, [25]). The transformation from Lagrangian picture to the Eulerian picture is usually obtained with the help of the Frobenius-Perron equation (cf., for example, [26, 27]). When one does that in the present case, one obtains

$$f(x,n+1) = \frac{f(\frac{1+\sqrt{1-4x/r}}{2},n) + f(\frac{1+\sqrt{1-4x/r}}{2},n)}{\sqrt{r^2-4xr}}. \tag{4}$$

This linear transformation carries the complete information about the behavior of an ensemble of points subjected to the logistic iteration. Thus, the nonlinear evolution of the logistic iteration may be described by a linear operator. Clearly, we have paid the price: the description of the system became more complicated. The state of the system at time n is described now by a function $f(x,n)$ instead of a single number x_n. Obviously, the connection between the flow of probability, resulting from the Schrödinger equation, and some hypothetical nonlinear process underlying this flow cannot be as simple as in the case of the logistic map.

The fundamental ideas, of de Broglie, Bohm and others, about the origin of probabilistic nature of quantum mechanics have not so far been widely accepted. However, the gist of their programs, somewhat surprisingly, found applications in the computational methods known as "the quantum trajectory" method (used in the study of wave packet dynamics in atomic and molecular physics) and "the quantum jumps" method (used in quantum optics). In both cases a linear evolution equation is replaced by a nonlinear one to reduce the computational effort. In the first case, the solutions of Schrödinger equation for a complex system are modeled as arising from a motion of particles of the "probability fluid" subjected to the quantum potential arising in the Bohmian interpretation of quantum mechanics (see, for example, [28]. This procedure is computationally more effective than solving the multidimensional Schrödinger equation on a grid. In the second case (sometimes also called the method of quantum trajectories), the aim is to solve the master equation of the Lindblad type (i.e. taking into account the interaction with the reservoir). This linear evolution problem is solved by "unraveling" the linear evolution equation for the density matrix into a sum of trajectories. The evolution of each trajectory is decomposed into two alternating steps: unitary evolution of the wave function and the reduction of the state (quantum jump) to one of its components due to the interaction with the environment. The quantum jump method has its historical roots in the old hypothesis of Niels Bohr concerning the nature of atomic transitions and to the von Neumann's hypothesis of the (nonlinear) reduction of the state in the measurement process but it is also related to direct observations [29, 30] of jumps in dynamics of trapped ions driven by laser beams. The quantum jump method reduces the computational effort because one replaces the evolution of density matrix (linear, but size N^2 problem) by the evolution of the wave function (nonlinear, but only size N problem). This approach was developed and applied to various problems in quantum optics by Howard Carmichael [31].

V. IS TIME SHARP OR DIFFUSE?

Despite all the successes in many branches of physics, the conclusion that there are no weak points in the linear Schrödinger equation is not justified. The problem arises when this equation is applied to relativistic theory. In this case, one must include the processes of particle creation and annihilation and the usual infinities that affect all relativistic theories appear with full strength. This problem for the Schrödinger equation is even more acute than in the S-matrix theory since one cannot renormalize away the infinities. The standard renormalization methods developed over last half century do not work. This was discovered in 1950 by Ernst Stueckelberg who wrote in the abstract of his paper [32]: "If transition probabilities are evaluated for transitions occurring during a finite time interval, additional divergencies occur different from those commonly encountered for infinite time intervals. The expressions obtained can however be made convergent, if an indeterminacy of time is attributed to each epoch of observation." Later Paul Dirac tried to attribute these divergencies to the non-equivalence of the Schrödinger and the Heisenberg pictures [1]. Dirac correctly identified the problem as arising from the fact that "The interaction is so violent that if we start with a particular state vector in Hilbert space to represent the initial state, it gets knocked right out of Hilbert space... It moves about in some more general space, in which it cannot be represented by coordinates, and thus one cannot construct a Schrödinger wave function". However, the statement that he made in the abstract "One can set up quantum electrodynamics entirely in the Heisenberg picture and thereby avoid the worst difficulties encountered in the Schrödinger picture" was not substantiated by a convincing analysis. In my opinion, the problem still exists and cannot be solved simply by changing the picture because it is deeply rooted in our inability to formulate a theory of physical interacting particles without starting from the abstract, bare and noninteracting particles (free fields). The same divergencies appear in both pictures and the Heisenberg picture seems to be even more complicated because one is dealing with operators instead of vectors. So far, no one has been able to write down a sensible set of field equations in quantum electrodynamics (not to speak about non-Abelian gauge theories) that would be free of infinities. The problem one is facing is basically the same in both pictures: What is the basis in Hilbert space that can be used to identify the states or the operators? The only such basis known to us is the Fock basis of noninteracting particles used in the description of scattering. This is why the S-matrix method works. This is why the S-matrix elements can be renormalized (expressed in terms of physical parameters). The existence of the free-particle basis, however, is not sufficient to give meaning to the Schrödinger equation. The reason is quite simple. In this basis the time evolution is that of a free system of particles and, therefore, the Hamiltonian is equal to the sum of the energies of *noninteracting* particles. The conclusion must, therefore, be that the Schrödinger equation cannot hold in relativistic quantum mechanics in the standard, general form

$$i\hbar\partial_t|\psi(t)\rangle = H|\psi(t)\rangle, \tag{5}$$

because one cannot define the action of the Hamiltonian on the state vectors for interacting particles. This does not mean, however, that one cannot salvage the idea of the linear evolution. All one has to do is to employ the concept of a diffuse time introduced by Stueckelberg and elaborated further by Bogolubov and Shirkov [34]. The evolution equation of the state vector can then still be described by a linear Schrödinger-like equation (5). The essential difference is that the label t of the state vectors would refer now to some average value of time associated with the state vector. Such a diffuse time is much more complicated mathematically. However, as I have argued before [36], it is much more realistic. After all, every preparation of a quantum state and its detection does not happen instantaneously but takes finite time. The main technical problem is how to handle the overcompleteness of state vectors prepared over a finite time interval.

The necessity to introduce diffuse time may also be linked with the time-energy uncertainty principle. By forcing the state to be localized in time (in the sense that all its characteristics refer to one value of time), one introduces infinite fluctuations in energy. In nonrelativistic physics this was not leading to any abnormalities but in field theory creation and annihilation of particles the state vector, in Dirac's words, "gets knocked right out of Hilbert space".

VI. A MATTER OF LABELS

The problem with the Schrödinger equation in relativistic quantum theory is in essence a matter of *labels* that one has to put on vectors in the Hilbert space to identify the states in terms of their physical properties. In order to resolve this problem one has to move as close to the experiments as possible and ask the questions: how are the quantum states prepared and how they are measured? A typical characteristic of a state carries a high level of abstraction. One may say, for example, "a state in question describes an electron with momentum $\mathbf{p}$ and with some spin value s". Then, one attaches the labels $\mathbf{p}$ and s on the state vector. In reality, a state that has such a label is always prepared in a process that has finite time duration. First, one has to extract the electron from some neutral entity (hydrogen atom, metallic wire, etc.). Next the electron must be accelerated (most often by an electric field) to acquire the desired momentum. Finally acting on the electron's magnetic moment one may filter out the undesired spin component. When all that is taken into account, the instantaneous characteristic of the state by $\mathbf{p}$ and s gets replaced by a configuration of electromagnetic fields that supposedly produce the required result. Obviously, however, the new characteristic involves finite time intervals. The state is characterized by a configuration of electromagnetic fields in a finite region of spacetime. The right label to be associated with this state is just this particular configuration of fields. One may write down the state vector in the form $|\Psi[\mathcal{A}]\rangle$ indicating a

functional dependence on the electromagnetic field representing the experimental setup. This state vector also depends on some fiducial state vector taken as the starting point in the experiment. All experimental arrangements in quantum electrodynamics can, in principle, be characterized in such a manner. The same description can be used for the final state of the system. This mode of description is similar to the Schwinger's *source theory* since this theory is also meant to stay close to reality in order to overcome "physical remoteness of operator field theory" [37]. The state vectors labeled in this realistic manner are free from ultraviolet infinities (after the standard mass and charge renormalization). Also the infrared infinities do not show up because there are no charged particles moving all the way to infinity [36]. However, the price to be paid for this departure from an abstract description ("physical remoteness") is very high. Many extremely useful tools and tricks of trade are gone. The set of vectors labeled by the characteristics of an experimental setup is highly overcomplete. This precludes the absolute normalization of the probabilities because one does not know what is the set of all possibilities (fortunately, the relative probabilities can still be defined). In this realistic setting, it is not easy, or may be it is not even possible, to formulate the Schrödinger equation. In the absence of the evolution equation for the state vector, the problem of its linearity cannot even be stated.

The conclusion of this analysis is that while in the low energy domain (in nonrelativistic quantum mechanics) the linearity of the Schrödinger equation seems to be perfect, at high energies (in quantum field theory) one does not even know how to write such an equation. In my opinion, the resolution of these difficulties might come from some mode of description, intermediate between the standard one — very abstract and unrealistic — and the one proposed in [36] and outlined here — fully realistic but useless when it comes to practical calculations.

[1] P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, Oxford, 1957), p. 12.

[2] I. Bialynicki-Birula and J. Mycielski, Wave equations with logarithmic nonlinearities, Bull. Acad. Polon. Sci. Cl. III **23**, 461 (1974).

[3] I. Bialynicki-Birula and J. Mycielski, Nonlinear wave mechanics, **100**, 62 (1976).

[4] I. Bialynicki-Birula and J. Mycielski, Gaussons: Solitons of the logarithmic Schrödinger equation, Physica Scripta **20**, 539 (1979).

[5] L. de Broglie, La méchanique ondulatoire et la structure atomique de la matère et du rayonnement, J. de Physique et du Radium **8**, 225 (1927).

[6] L. de Broglie, *Une Tentative Interprétation Causale et Nonlinéare de la Méchanique Ondulatoire* (Gauthier-Villars, Paris, 1956); English translation *Non-linear Wave Mechanics* (Elsevier, Amsterdam, 1960).

[7] L. de Broglie, Interpretation of quantum mechanics by double solution theory, Ann. de la Fond. L. de Broglie **12**, 1 (1987).

[8] D. Bohm, A suggested interpretation of the quantum theory in terms of "hidden" variables. I, Phys. Rev. **85**, 166 (1952); II, ibid., 180 (1952).

[9] D. Bohm, Proof that probability density approaches $|\psi|^2$ in causal interpretation of the quantum theory, Phys. Rev. **89**, 458 (1953).

[10] D. Bohm and B.J. Hiley, *The Undivided Universe* (Routledge, London, 1993), Ch. 1.

[11] A. Shimony, Proposed neutron interferometer test of some nonlinear variants of wave mechanics, Phys. Rev. A **20**, 394 (1979).

[12] C.G. Shull, D.K. Atwood, J. Arthur, and M.A. Horne, Search for a nonlinear variant of the Schrödinger equation by neutron interferometry, Phys. Rev. Lett. **44**, 765 (1980).

[13] R. Gähler, A.G. Klein, and A. Zeilinger, Neutron optical tests of nonlinear wave mechanics, Phys. Rev. A **23**, 1611 (1981).

[14] E.F. Hefter, Application of the nonlinear Schrödinger equation with a logarithmic inhomogeneus term to nuclear physics, Phys. Rev. A **32**, 1201 (1985).

[15] E.S. Hernandez and B. Remaud, General properties of Gausson-conserving descriptions of quantal damped motion, Physica **105A**, 130 (1980).

[16] W. Krolikowski, D. Edmundson, and O. Bang, Unified model for partially coherent solitons in logaritmically nonlinear media, Phys. Rev. E **61**, 3122 (2000).

[17] H. Buljan, A. Šiber, M. Soljačic, T. Schwartz, M. Segev, and D.N. Christodoulides, Incoherent white light solitons in logarithmically saturable noninstantaneous nonlinear media, Phys. Rev. E **68**, 036607 (2003).

[18] S. De Martino, M. Falanga, C. Godano, and G. Lauro, Logarithmic Schrödinger-like equation as a model for magma transport, Europhys. Lett. **63**, 472 (2003).

[19] M.H. Jakubowski, K. Steiglitz, and R. Squier, When can solitons compute?, Complex Systems, **10**, 1 (1996).

[20] S. Weinberg, Precision tests of quantum mechanics, Phys. Rev. Lett. **62**, 485 (1989).

[21] J.J. Bolinger, J.D. Prestage, Wayne M. Itano, and D.J. Wineland, Laser-cooled-atomic frequency standard, Phys. Rev. Lett. **54**, 1000 (1985).

[22] N.A. Lemos, Stochastic derivation of the Birula-Mycielski nonlinear wave equation, Phys. Lett. **94 A**, 20 (1983).

[23] A.B. Nassar, New method for the solution of the logarithmc Schrödinger equation via stochastic dynamics, Phys. Rev. A **33**, 3502 (1986).

[24] C. Castro, J. Mahecha, and B. Rodríguez, Nonlinear QM as a fractal Brownian motion with complex diffusion constant, arXiv: quant-ph/0202026.

[25] A.L. Fetter and J.D. Walecka *Theoretical Dynamics of Particles and Continua* (McGraw Hill, New York, 1980), p.294.

[26] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1994), Ch. 2.

[27] J.R. Dorfman, *An Introduction to Chaos in Nonequiibrium Statistical Mechanics* (Cambridge University Press, Cambridge, 1999), Ch. 10.

[28] C.L. Lopreore and R.E. Wyatt, Quantum wave packet dynamics with trajectories, Phys. Rev. Lett. **82**, 5190 (1999).

[29] W. Nagourney, J. Sandberg, and H. Dehmelt, Shelved optical electron amplifier: Observation of quantum jumps, Phys. Rev. Lett. **56**, 1696 (1986).

[30] Th. Sauter, W. Neuhauser, R. Blatt, and P. E. Toschek, Observation of quantum jumps, Phys. Rev. Lett. **57**, 2797 (1986).

[31] H.J. Carmichael, *An Open System Approach to Quantum Optics*

(Springer-Verlag, Berlin, 1993).
[32] E.C.G. Stueckelberg, Relativistic quantum theory for finite time intervals, Phys. Rev. **81**, 130 (1951).
[33] P.A.M. Dirac, Quantum electrodynamics without dead wood, Phys. Rev. **139**, B684 (1965).
[34] N.N. Bogolubov and D.V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley, New York, 1980), Ch. VI.
[35] I. Bialynicki-Birula, Is time sharp or diffused?, in: *Physical Origins of Time Asymmetry*, Eds. J.J. Halliwell, J. Peres-Mercador and W.H. Zurek (Cambridge University Press, Cambridge, 1994).
[36] I. Bialynicki-Birula, Triumphs and failures of quantum electrodynamics, Acta Phys. Polon. B **27**, 2403 (1996).
[37] J. Schwinger, *Particles, Sources and Fields* (Addison-Wesley, Reading, 1970).