Information measures for inferring quantum mechanics

2005 J. Phys. A: Math. Gen. 38 6231

(http://iopscience.iop.org/0305-4470/38/27/009)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 6231–6237 doi:10.1088/0305-4470/38/27/009

Information measures for inferring quantum mechanics

Rajesh R Parwani

Department of Physics and University Scholars Programme, National University of Singapore, Kent Ridge, Singapore

E-mail: parwani@nus.edu.sg

Received 22 March 2005
Published 22 June 2005
Online at stacks.iop.org/JPhysA/38/6231

Abstract
Starting from the Hamilton–Jacobi equation describing a classical ensemble, one may infer a quantum dynamics using the principle of maximum uncertainty. That procedure requires an appropriate measure of uncertainty. Such a measure is constructed here from physically motivated constraints. It leads to a unique single parameter extension of the classical dynamics that is equivalent to the usual linear quantum mechanics.

PACS numbers: 03.65.Ca, 03.65.Ta, 89.70.+c

1. Deconstructing the Schrödinger equation

Despite its remarkable quantitative success, quantum mechanics continues to puzzle us with its seemingly counter-intuitive predictions. Even the mathematical formalism most widely used for its description appears very different from that used in classical mechanics: one sees in quantum mechanics the appearance of complex numbers, probability amplitudes and an apparently exact linear evolution equation.

In this paper the structure of Schrödinger's equation, in particular its linearity, will be derived within an information theoretic framework to be elaborated on below. The various assumptions involved in the derivation will also be discussed at length.

Let us begin with a review of the Schrödinger equation for N particles in $d+1$ dimensions,

$$i\hbar\dot{\psi} = \left[-\frac{\hbar^2}{2} g_{ij}\partial_i\partial_j + V\right]\psi \tag{1}$$

where $i, j = 1, 2, \ldots, \mathrm{d}N$ and the metric is defined as $g_{ij} = \delta_{ij}/m_{(i)}$ with the symbol (i) defined as the smallest integer $\geqslant i/d$. That is, $i = 1, \ldots, d$, refer to the coordinates of the first particle of mass m_1, $i = d+1, \ldots, 2d$, to those of the second particle of mass m_2 and so on. The overdot refers to a partial time derivative, and the summation convention is used unless

otherwise stated. Cartesian coordinates have been chosen as these allow an unambiguous correspondence between observables such as momenta and their operator representation [1].

The metric g_{ij} occurs naturally in the description of the system in configuration space [2, 3] and plays a crucial role in the discussion below. It is pertinent to note that the metric g_{ij} is diagonal and positive-definite. This is a consequence of the form of the kinetic term in the Schrödinger equation in Cartesian coordinates.

Since our intuition is mostly classical, it is useful to rewrite the Schrödinger equation in a form which allows comparison with Newtonian physics. The Madelung transformation [4] $\psi = \sqrt{p}\,\mathrm{e}^{\mathrm{i}S/\hbar}$ decomposes the Schrödinger equation into two real equations,

$$\dot{S} + \frac{g_{ij}}{2}\partial_i S \partial_j S + V - \frac{\hbar^2}{8} g_{ij}\left(\frac{2\partial_i\partial_j p}{p} - \frac{\partial_i p \partial_j p}{p^2}\right) = 0, \tag{2}$$

$$\dot{p} + g_{ij}\partial_i(p\partial_j S) = 0. \tag{3}$$

The first equation is a generalization of the usual Hamilton–Jacobi equation, the term with explicit $\hbar$ dependence (the 'quantum potential' [1]) summarizing the peculiar and nonlocal aspects of quantum theory. The second equation is a continuity equation expressing the conservation of probability, $\int p(x,t)\,\mathrm{d}x^{Nd}$.

Equations (2) and (3) may be obtained from a variational principle [3]; one minimizes the action

$$\Phi = \int p\left[\dot{S} + \frac{g_{ij}}{2}\partial_i S\partial_j S + V\right]\mathrm{d}x^{Nd}\,\mathrm{d}t + \frac{\hbar^2}{8} I_F \tag{4}$$

with respect to the variables p and S. Interestingly, the quantity

$$I_F \equiv \int \mathrm{d}x^{Nd}\,\mathrm{d}t\; g_{ij}\; p(\partial_i \log p)(\partial_j \log p) \tag{5}$$

resembles the 'Fisher information' [5], whose inverse sets a lower bound on the variance of the probability distribution $p(x)$ through the Cramer–Rao inequality [6, 3]. Since a broader probability distribution $p(x)$ represents a greater uncertainty in x, the term I_F is actually an inverse uncertainty measure.

Equations (4) and (5) were used in [3, 7] to derive Schrödinger's equation through a procedure analogous to the principle of maximum entropy (uncertainty) [8, 9] used in statistical inference theory. The idea is that without the term I_F, variation of equation (4) gives rise to equations describing a classical ensemble. As the probability distribution $p(x)$ characterizing the ensemble is supposed to represent some fluctuations of unknown origin, we would like to be as unbiased as possible in its choice. This is achieved by choosing the broadest distribution possible, representing our maximum uncertainty. Technically, this is implemented in (4) by minimizing (5) when varying the classical action: $\hbar^2/8$ is the Lagrange multiplier.

2. Constructing the measure

However, the intriguing approach of [3] does not explain, *a priori*, the form of the information measure that should be used. That is, why must I_F be minimized rather than something else?

The goal of this section is to construct, from first principles, information measures that are permissible. To fix the notation, consider therefore the same classical ensemble as in section 1, but now constrained by a general information measure I whose form is to be determined. The relevant action is then

$$\mathcal{A} = \int p\left[\dot{S} + \frac{g_{ij}}{2}\partial_i S\partial_j S + V\right]\mathrm{d}x^{Nd}\,\mathrm{d}t + \lambda I, \tag{6}$$

with λ a Lagrange multiplier. Varying this action will give rise, in general, to a nonlinear Schrödinger equation after an inverse Madelung transformation,

$$i\hbar\dot{\psi} = \left[-\frac{\hbar^2}{2}g_{ij}\partial_i\partial_j + V\right]\psi + F(\psi, \psi^\dagger)\psi, \tag{7}$$

with F representing the nonlinearity.

In order to construct an explicit form for the information measure I, constraints need to be imposed. These constraints are of two types. The first type are those that are required for I to be sensibly interpreted as a measure of inverse uncertainty (information). Condition [$S1$] below belongs to this type.

The other constraints [$S2$]–[$S6$] to be discussed below are of a different type. One may adopt two alternative perspectives to motivate these constraints. The first, classical, perspective is to view action (6) as a generalized form of classical dynamics. In that case a minimal deformation of the usual classical dynamics is achieved if the additional constraints are the same as those already satisfied by the $I = 0$ part of (6): locality, homogeneity, separability, Galilean invariance, and absence of more than two derivatives in each product of terms in the action. Thus by using [$S2$]–[$S6$] one is not imposing any constraints on the action that do not already exist. In this classical perspective the physical motivations for the constraints are either the usual ones with an obvious interpretation (locality, separability, Galilean invariance), or else they are explained below.

The second, quantum, perspective is to view conditions [$S2$]–[$S6$] with respect to the interpretation of equation (7), which represents a generalized Schrödinger equation. As the usual linear quantum mechanics has been experimentally well tested, one could argue that it makes sense to only consider those potential deformations of the linear theory that permit as much of the usual interpretations of the wavefunction as possible. Remarkably, as discussed below, the same conditions [$S2$]–[$S6$] motivated by the classical perspective are needed also for this quantum viewpoint.

Here then are the axioms:

- [$S1$]
  - (i) Firstly, by definition, the measure I should be a real valued and positive-definite functional for *all* $p = \psi^\dagger\psi$. More specifically, we would like the measure to be universal in the sense of being independent of the external potential V.
  - (ii) Also, the interpretation of I as an information (inverse uncertainty) measure requires that it should approach a minimum when p is uniformly distributed. (A minimum exists because by (i) I is positive definite.)
- [$S2$] Locality: I should be of the form $I = \int \mathrm{d}x^{Nd}\,\mathrm{d}t\; pH(p)$ where H is a function of the probability $p(x, t)$ and its spatial derivatives. In the quantum perspective this local form ensures the validity of the weak superposition principle in the equations of motion (7): states with negligible overlap will not influence each other strongly. (In principle one should also allow I to depend on S, x and time derivatives. These generalizations are discussed in the next section.)
- [$S3$] Homogeneity: H should be invariant under scaling, $H(\lambda p) = H(p)$. The normalization of probability, $1 = \int \mathrm{d}x^{Nd}\; p(x, t)$, implies that the dimensions of $p(x, t)$ depend on the dimensions of the configuration space. Thus by demanding H to be scale invariant one ensures that the resulting equations of motion, whether in the classical or quantum perspective, have a universal form independent of the number of particles. Thus this condition may be viewed as restricting the search to universal dynamics.

(There is another motivation sometimes given for this homogeniety condition: it allows solutions of (7) to be (re)normalized and thus allow for the usual interpretation of states after a measurement process, as discussed in [10].)

- [$S4$] Separability: H should be separable for the case of two independent sub-systems described by probability distributions p_1 and p_2: $H(p = p_1 p_2) = H(p_1) + H(p_2)$.
- [$S5$] H should be Galilean invariant[1].
- [$S6$] H should not contain more than two derivatives in any product of terms that appears in it. As each derivative involves an inverse length, this condition obviously restricts the number of new dimensional parameters that can appear in (6) and (7). This condition will be referred to in brief as 'absence of higher number of derivatives' or 'AHD'.
  As will be apparent later, the implementation of this condition means that the lagrange multiplier λ in (6), and hence Planck's constant, is the only new parameter that is required in making the transition from classical to quantum mechanics. Conversely, relaxing the AHD condition would imply the appearance of other parameters, with the dimensions of length, and a generalized form of quantum dynamics.

I would like to reiterate that conditions [$S2$]–[$S6$] above are all already satisfied by the *classical* part of action (6), so demanding them of the additional piece I is actually quite natural and minimalist. The additional motivations provided by viewing the equations from the quantum perspective are simply a bonus.

Having enumerated and motivated the axioms, one may begin constructing the measure. Clearly the homogeneity requirement [$S3$] cannot be satisfied if H depends only on p: it must also contain derivatives of p. The AHD condition and rotational invariance imply that the building blocks of H must be

$$g_{ij} U_1 \partial_i U_2 \partial_j U_3 \qquad \text{and} \qquad g_{ij} V_1 \partial_i \partial_j V_2, \tag{8}$$

where the U_i, V_i are the functions of p. Separability can now be used to deduce that H must be linear in g_{ij}:

$$H = g_{ij}(U_1 \partial_i U_2 \partial_j U_3 + V_1 \partial_i \partial_j V_2). \tag{9}$$

One may consider sums of such structures and so place an additional index n on the U_i, V_i but it is easy to check that the final result below remains unchanged. One could also use the separability condition to restrict the explicit forms that U, V may take but I will use the homogeneity condition for that purpose below. (Note that if the form of $H(p)$ were restricted so that derivatives of p occurred only in polynomial form, as is commonly done in physics, then separability would not be required in obtaining (9).)

Using the chain rule one can rewrite (9) as

$$H = g_{ij}\left(\frac{\partial_i p \partial_j p}{p^2}[U_1 U_2' U_3' p^2 + V_1 V_2'' p^2] + \frac{\partial_i \partial_j p}{p}[V_1 V_2' p]\right), \tag{10}$$

where the prime symbol denotes a derivative with respect to p. Consider now the scaling $p \to \lambda p$ under which H is required to be invariant. The terms inside the square brackets become dependent only on the product λp. Since the forms $\frac{\partial_i p \partial_j p}{p^2}$ and $\frac{\partial_i \partial_j p}{p}$ are distinct and independently scale invariant, this means that the terms in square brackets must also be independent of λ: but since those terms depend only on λp, this implies that the terms in square brackets are simply constants.

Thus one obtains

$$I = \int \mathrm{d}t\, \mathrm{d}x^{Nd}\, p g_{ij}\left(A\frac{\partial_i p \partial_j p}{p^2} + B\frac{\partial_i \partial_j p}{p}\right). \tag{11}$$

[1] In this paper, when discussing the symmetry of the Schrödinger equation one will always refer to the case of vanishing potential, $V = 0$.

The 'B' term gives a surface contribution which might not vanish for some wavefunctions and so its contribution to I is of indefinite sign. The positivity and universality of I therefore requires us to choose $B = 0$.

Hence one concludes that the unique solution of conditions [$S1$]–[$S6$] is the information measure I_F given in (5). The Lagrange multiplier λ in (6) must then have the dimension of (action)^2 thereby introducing the Planck constant into the picture; the equation of motion is then the *linear* Schrödinger equation.

Interestingly, neither the second part of [$S1$] nor the full Galilean invariance was used explicitly in the above construction even though the final result, the measure I_F, does satisfy all the conditions. However these additional constraints will be useful in the next section.

It should be noted that, as shown in [11], action (4) increases for variations of p that keep S fixed, the increase being due to an increase in I_F, so that the resulting solutions are not just an extremum but do indeed minimize the information measure I_F.

The positivity condition in [$S1$] plays an important physical role beyond ensuring the existence of a minimum for I (state of maximum uncertainty). It also guarantees that the following energy functional is bounded from below for potentials V that are likewise bounded:

$$\mathcal{E}[S, p] = \int \mathrm{d}^{Nd}x p \left(\frac{g_{ij}}{2} \partial_i S \partial_j S + V + \lambda H(p) \right) \tag{12}$$

where the function H is defined in [$S2$]. This functional is conserved for stationary states and it also reduces to the average of the usual quantum mechanical Hamiltonian for the case of the linear theory. These properties of the energy functional qualify it as the most natural to use for defining the energy of the system in a potential generalization (nonlinear) of quantum theory.

The meaning of the AHD condition can be elucidated with an explicit example. Consider

$$H_1(p) = g_{ij} \partial_i (\log p + \eta f(p)) \partial_j (\log p + \eta f(p)) \tag{13}$$

with $f(p) = g_{kl} (\partial_k \log p)(\partial_l \log p)$ and η a constant. This H_1 satisfies all the constraints except AHD. The price to pay is the appearance of an additional parameter, η, required to balance the dimensions of the higher order derivatives.

Thus the AHD condition ensures that, within the information theoretic approach, the Schrödinger equation is the unique *single* parameter extension of the classical statistical Hamilton–Jacobi equations. Since the information theoretic approach attempts to provide an unbiased description of data, one may say that the AHD condition further restricts us to the simplest unbiased description.

3. Relaxing some conditions

It has been implicitly assumed in the previous section that the metric g_{ij} that appears in I is the same as the metric in the classical part of the action. If one allows in the measure a metric $\bar{g}_{ij}$ which is still diagonal but different from the classical metric g_{ij} then a nonlinear Schrödinger equation apparently ensues. However the nonlinearity can be removed by a change of variables (a nonlinear gauge transformation) [12, 13] with the result that for a range of values of the Lagrange multiplier λ one actually recovers the usual linear Schrödinger equation. (This example highlights the point that a nonlinear Schrödinger equation cannot immediately, and in all generality, be declared pathological.) However for the remaining values of λ one obtains after the change of variables [13] a linear *diffusion* equation. It will be assumed for the rest of this paper that the classical metric is used also in the information measure when symmetries are preserved.

Consider now allowing I and hence H to depend also on the phase factor $S(x, t)$ in addition to the probability density $p(x, t)$. By definition, the phase factors S_1 and S_2 for two

independent systems are additive in the composite system, $S = S_1 + S_2$. One can proceed as before and consider the most general structures restricted by rotational invariance, AHD, homogeneity, separability and positivity. The result is a generalized measure of the form

$$I_Q = \int \mathrm{d}t\, \mathrm{d}x^{Nd}\, p g_{ij} \partial_i(\alpha_1 \log p + \alpha_2 S)\partial_j(\alpha_1 \log p + \alpha_2 S), \tag{14}$$

where the α_k are constants. In arriving at this structure, homogeneity has only been used to imply that derivatives of p occur, while separability is the stronger constraint as it acts also on the S variable. However the second part of condition $[S1]$ requires that $\alpha_2 = 0$, thus eventually one is again led to the Fisher form. (Again, more generally one may place another index n on the α_k and sum over such terms but the conclusion remains unchanged.)

Nevertheless, the special case of (14) with $\alpha_1 = 0$ but $\alpha_2 \neq 0$ is sufficiently interesting to deserve further study because like the classical measure I_F, it is positive definite by itself, but unlike I_F it also contains some information about the phase of the wavefunction. Now, if used in the variational action, this S-dependent term can be absorbed, after a scaling of the metric, by a similar term already existing in the classical part of action (6). The net result is therefore still a linear Schrödinger equation but with a mass, $\bar{m}_{(i)}$ which is renormalized with respect to the original mass parameter $m_{(i)}$ in the classical theory. Empirically this renormalization will have no consequence if all calculations, as usual, refer to the mass parameter appearing in the quantum theory.

One may also consider allowing time derivatives of p and S in H. However the demands of positivity conflict with those of separability unless one relaxes the AHD condition: then it is possible to have structures such as

$$H_2(p) = g_{ij}\partial_i \left(\log p + \eta \frac{\dot{p}}{p}\right) \partial_j \left(\log p + \eta \frac{\dot{p}}{p}\right) \tag{15}$$

that contain dimensionful parameters.

Finally, one may consider an explicit dependence on the coordinates, x_i, in H. However such terms are ruled out by translational invariance.

4. Conclusion

If one adopts the philosophy that the laws of physics should be constructed so as to provide the most economical and unbiased representation of empirical facts, then the principle of maximum uncertainty [8, 9] is the natural avenue by which to investigate the foundations of quantum theory [7, 3].

The investigation here has extended the initiative of [3] in several ways. Firstly, the constraints that a relevant information measure should satisfy have been made explicit and motivated from two alternative perspectives. Secondly, the measure has been constructed from the constraints rather than postulated, thus motivating the structure of the linear Schrödinger equation. Indeed, it has been shown here that within the information theoretic approach, the linear Schrödinger equation is the unique one-parameter extension of the classical dynamics.

One should compare the approach here with an alternative, axiomatic, but not information theoretic based, construction of the Schrödinger equation from classical mechanics discussed in [14]. Starting from the classical Hamilton–Jacobi equations, the authors add constrained fluctuations to the kinetic term. The result is an equation similar to equation (6) above with an explicitly positive definite $H(p)$ and with the symmetries of the classical action. However, in addition to the major differences in motivation and epistemology, there is a significant technical contrast between this paper and [14]: there the authors postulate and use an 'exact uncertainty relation' instead of the homogeneity condition $[S3]$ adopted here. As discussed above, the

homogeneity condition is already satisfied by the classical action and has the interpretation of making the form of the equations of motion independent of the dimension of configuration space, and thus it is a simple restriction to universal dynamics.

An open and interesting problem is to extend the constructive approach adopted in section 2 to include spin [15] and relativistic effects [3]. This might involve further refinement of the conditions in section 2 and might result in the use of more general information measures such as I_Q of equation (14).

Finally, one may enquire into the possible generalizations of quantum mechanics (typically nonlinear) that result from omitting one or more of the conditions [$S1$]–[$S6$] and the ensuing phenomenological consequences. These issues are discussed elsewhere [16, 17]. It is of interest to note that one of the earliest suggestions for a nonlinear Schrödinger equation [18] had a log p nonlinearity, which is allowed if the homogeneity condition is abandoned.

Acknowledgment

I am grateful to Michael Hall and Marcel Reginatto for critical readings of earlier drafts of this paper and for many stimulating discussions.

References

[1] Ballentine L 1998 *Quantum Mechanics: A Modern Development* 2nd edn (River Edge, NJ: World Scientific)
[2] Synge J L 1960 Classical dynamics *Encyclopaedia of Physics* vol III/I ed S Flugge (Berlin: Springer)
[3] Reginatto M 1998 *Phys. Rev.* A **58** 1775
Reginatto M 1999 *Phys. Rev.* A **60** 1730
[4] Madelung E 1926 *Z. Phys.* **40** 322
[5] Fisher R A 1925 *Proc. Cambridge Phil. Soc.* **22** 700
[6] Kullback S 1959 *Information Theory and Statistics* (New York: Wiley)
[7] Frieden B R 1988 *J. Mod. Opt.* **35** 1297
Frieden B R 1989 *Am. J. Phys.* **57** 1004
[8] Jaynes E T 1957 *Phys. Rev.* **106** 620
Jaynes E T 1957 *Phys. Rev.* **108** 171
[9] Jaynes E T 2004 *Probability Theory, The Logic of Science* (Cambridge: Cambridge University Press)
[10] Kibble T W B 1978 *Commun. Math. Phys.* **64** 73
Weinberg S 1989 *Ann. Phys., NY* **194** 336
[11] Reginatto M 1999 *Preprint* quant-ph/9909065
[12] Doebner H D, Goldin G A and Nattermann P 1999 *J. Math. Phys.* **40** 49
[13] Auberson G and Sabatier P C 1994 *J. Math. Phys.* **35** 4028
[14] Hall M and Reginatto M 2002 *J. Phys. A: Math. Gen.* **35** 3289
[15] Reginatto M 1998 *Phys. Lett.* A **249** 355
[16] Parwani R 2005 *Ann. Phys.* **315** 419–52
[17] Parwani R 2004 Why is Schrödinger's equation linear? *Proc. DICE 2004 Workshop (Piombino, Italy)* (*Preprint* quant-ph/0412192, at press)
[18] Bialynicki-Birula I and Mycielski J 1976 *Ann. Phys.* **100** 62