From probabilities to wave functions: A derivation of the geometric formulation of quantum theory from information geometry

Marcel Reginatto

Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

E-mail: `marcel.reginatto@ptb.de`

Abstract.
It is shown that the geometry of quantum theory can be derived from geometrical structure that may be considered more fundamental. The basic elements of this reconstruction of quantum theory are the natural metric on the space of probabilities (information geometry), the description of dynamics using a Hamiltonian formalism (symplectic geometry), and requirements of consistency (Kähler geometry). The theory that results is standard quantum mechanics, but in a geometrical formulation that includes also a particular case of a family of nonlinear gauge transformations introduced by Doebner and Goldin. The analysis is carried out for the case of discrete quantum mechanics. The work presented here relies heavily on, and extends, previous work done in collaboration with M. J. W. Hall.

1. Introduction
It is now known that quantum mechanics has a rich geometrical structure which allows for a geometric formulation of the theory. The geometric approach was introduced by Kibble [1] and later further developed by a number of other authors. One successful strategy has been to start from a known formulation of quantum mechanics and to identify geometrical features that can be used for the reformulation of the theory. This paper inverts this procedure: the aim is to *derive* the geometry of quantum theory from geometrical structure that may be considered more fundamental, and to examine the assumptions that are needed to do this. The work presented here relies heavily on, and extends, previous work of Reginatto and Hall [2, 3]. The basic elements of this geometrical reconstruction of quantum theory are the natural metric on the space of probabilities (information geometry), the description of dynamics using a Hamiltonian formalism (symplectic geometry), and requirements of consistency (Kähler geometry). One may associate a Hilbert space with the Kähler space, which leads to the standard version of quantum theory. The analysis is carried out for the case of discrete quantum mechanics; a similar approach has been carried out previously for continuous systems [2].

2. Information geometry
Consider a system with a discrete configuration space. If the configuration of the system is subject to uncertainty, the state of the system will be described by a probability $P = (P^1, ..., P^n)$ where n is the number of states. The probability that the system is in state i is P^i, where P^i

satisfies $P^i \geq 0$ and $\sum_i P^i = 1$. The space of probabilities can be visualized as points in the simplex $S_{n-1} = \{P \in \mathbf{R}_n^+ : \sum P^i = 1\}$.

There is a natural line element in this space, given by

$$ds^2 = G_{ij} \, dP^i \, dP^j = \frac{\alpha}{2P^i} \, \delta_{ij} \, dP^i \, dP^j, \tag{1}$$

where α is a constant. The value of α can not be determined *a priori*; it is usually set to $\frac{1}{2}$, but for the purposes of this paper it will be convenient not to follow this convention. Instead, α will be treated as a parameter. The line element of Eq. (1) defines a distance on the space of probabilities. This distance seems to have been introduced into statistics by Bhattacharyya [4, 5] as a way of providing a measure of divergence for multinomial probabilities [6]. Wootters has shown that this distance can be motivated using arguments based on the concept of the distinguishability of states and refers to this distance as the "statistical distance" [7].

The metric G_{ij} that appears in Eq. (1) is known as the *information metric*,

$$G_{ij} = \frac{\alpha}{2P^i} \, \delta_{ij}. \tag{2}$$

It has been shown by Čencov that the information metric is the only metric that is invariant under a family of probabilistically natural mappings known as *congruent embeddings by a Markov mapping* [8]. A simpler proof, which also makes use of these mappings, was given later by Campbell [9]. It will be useful to give a brief description of these mappings because their generalization (see below) play an important role in the derivation of the main result of this paper.

I follow the presentation of Ref. [9]. A Markov mapping is a particular type of linear transformation between a simplex S_{m-1} and a simplex S_{n-1} (with $m \leq n$) which preserves the probability; i.e., $\sum_{a=1}^m P^a = \sum_{b=1}^n \tilde{P}^b = 1$. For $m = n$, the mapping is just a permutation of the components P^i, but for $m < n$, the mapping relates spaces of different dimensions. A Markov mapping may be constructed in the following way. Let $A = \{A_1, ..., A_m\}$ be a partition of the set $\{1, 2, ..., n\}$ into disjoint sets. Associate a probability vector $Q_{(a)} = (q_{a1}, ..., q_{an})$ to each of the A_a, where the q_{ab} satisfy

$$q_{ab} = 0 \ \texttt{if} \ b \notin A_a, \qquad q_{ab} > 0 \ \texttt{if} \ b \in A_a, \qquad \sum_{b=1}^n q_{ab} = 1. \tag{3}$$

The probability vector $Q_{(a)}$ is therefore concentrated on A_a. Note that the $m \times n$ matrix $\mathbf{Q}$ with elements q_{ab} has the following properties: Each column has precisely one non-zero element and each row sums to one.

Define mappings between $f : S_{m-1} \to S_{n-1}$ and $g : S_{n-1} \to S_{m-1}$ by

$$\begin{aligned}
\tilde{P}^b &= \sum_{a=1}^m P^a q_{ab}, \\
P^a &= \sum_{b \in A_a} \tilde{P}^b, \qquad a \in \{1, 2, ..., m\}.
\end{aligned} \tag{4}$$

Following Čencov, the mapping f is known as a congruent embedding of S_{m-1} in S_{n-1} by a Markov mapping. The mapping g, which is also defined in terms of the partition A, has the property that the composition $g \circ f$ is the identity map on S_{m-1}.

I now consider a simple example of a Markov mapping for the case where $n = m + 1$. Set

$$
\mathbf{Q} = \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 \\
0 & 0 & \cdots & 0 & k & (1-k)
\end{pmatrix},
\tag{5}
$$

with $0 < k < 1$. Then

$$
P = (P^1, ..., P^m) \rightarrow \tilde{P} = (\tilde{P}^1, ..., \tilde{P}^m, \tilde{P}^{m+1}) := (P^1, ..., kP^m, (1-k)P^m).
\tag{6}
$$

A vector in the tangent space of the simplex transforms in a similar way,

$$
V = (V^1, ..., V^m) \rightarrow \tilde{V} = (\tilde{V}^1, ..., \tilde{V}^m, \tilde{V}^{m+1}) := (V^1, ..., kV^m, (1-k)V^m).
\tag{7}
$$

To prove uniqueness of the information metric, Čencov [8] and Campbell [9] show that the only metric that preserves the inner product $< A, B >$ of two tangent vectors A, B under a Markov mapping is precisely the information metric. It is straightforward to show that the metric has this property. To see this for the simple example discussed above, simply compute

$$
\begin{aligned}
< \tilde{A}, \tilde{B} > &= \sum_{i=1}^{m+1} \left\{ \frac{\tilde{A}^i \tilde{B}^i}{\tilde{P}^i} \right\} = \sum_{i=1}^{m-1} \left\{ \frac{A^i B^i}{P^i} \right\} + \frac{kA^m\, kB^m}{kP^m} + \frac{(1-k)A^m\, (k-1)B^m}{(1-k)P^m} \\
&= \sum_{i=1}^{m} \left\{ \frac{A^i B^i}{P^i} \right\} = < A, B > .
\end{aligned}
\tag{8}
$$

A proof of uniqueness is, as one would expect, not straightforward. The monograph of Caticha [10], which includes a useful review of information geometry for both the discrete and continuous cases, provides a nice, accessible proof of uniqueness which follows the presentation of Campbell.

3. Dynamics and symplectic geometry

I now set the probabilities in motion. Assume that the time evolution of the P^i is generated by an action principle and write the equations of motion using a Hamiltonian formalism. To do this, introduce additional coordinates S^i which are canonically conjugate to the P^i and a corresponding Poisson bracket for any two functions $F(P, S)$ and $G(P, S)$,

$$
\{F, G\} = \sum_i \left(\frac{\partial F}{\partial P^i} \frac{\partial G}{\partial S^i} - \frac{\partial F}{\partial S^i} \frac{\partial G}{\partial P^i} \right).
\tag{9}
$$

As is well known, the Poisson bracket can be rewritten geometrically as

$$
\{F, G\} = (\partial F/\partial P, \ \partial F/\partial S) \, \Omega \begin{pmatrix} \partial G/\partial P \\ \partial G/\partial S \end{pmatrix},
\tag{10}
$$

where Ω is the corresponding symplectic form, given in this case by

$$
\Omega = \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix},
\tag{11}
$$

where $\mathbf{1}$ is the $n \times n$ unit matrix. We thus have a symplectic structure and a corresponding *symplectic geometry*. The equations of motion for P^i and S^i are given by $\dot{P}^i = \{P^i, H\}$, $\dot{S}^i = \{S^i, H\}$ where H is the Hamiltonian that generates time translations.

Notice that the $2n$-dimensional phase space with coordinates P^i and S^i has a richer structure than the n-dimensional space of probabilities P^i; in particular, one may introduce the notion of *observables*, which are functions $O(P, S)$ of the coordinates, together with an algebra of observables defined in terms of the Poisson brackets of these functions [11]. However, not every function $O(P, S)$ qualifies as an observable because observables are also generators of infinitesimal transformations and these transformations must satisfy certain requirements. For example, the infinitesimal canonical transformation generated by any observable O must preserve the normalization and positivity of P. This implies

$$O(P, S + \chi) = O(P, S), \qquad \partial O/\partial S^i = 0 \text{ if } P^i = 0. \tag{12}$$

Note that the first condition implies gauge invariance of the theory under $S^i \to S^i + \chi$, where χ is a constant [11].

4. Kähler geometry

The $2n$-dimensional phase space with coordinates P^i and S^i is an extension of the n-dimensional space of probabilities P^i. It is natural to ask the following question: Is it possible to extend the metric G_{ij} in Eq. (2), which is only defined on the n-dimensional space of probabilities P^i, to the full $2n$-dimensional phase space of the P^i and S^i? This can be done, but certain conditions which ensure the compatibility of the metric and symplectic structures have to be satisfied. These conditions are equivalent to requiring that the space have a Kähler structure (see the Appendix of Ref. [2] for a proof).

A Kähler structure brings together metric, symplectic and complex structures in a harmonious way. To define such a space, introduce a complex structure J^a_b and impose the following conditions [12],

$$\Omega_{ab} = g_{ac} J^c_b \,, \tag{13}$$
$$J^a_c g_{ab} J^b_d = g_{cd} \,, \tag{14}$$
$$J^a_b J^b_c = -\delta^a_c \,. \tag{15}$$

Eq. (13) is a compatibility equation between the symplectic structure Ω_{ab} and the metric g_{ab}, Eq. (14) is the condition that the metric should be Hermitian, and Eq. (15) is the condition that J^a_b should be a complex structure.

The metric over the n-dimensional space of probabilities is the information metric, Eq. (2). Then, metric over the full space will be of the form

$$g_{ab} = \begin{pmatrix} \mathbf{G} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{F} \end{pmatrix}, \tag{16}$$

where $\mathbf{G} = \mathtt{diag}(\frac{\alpha}{2P^i})$, and $\mathbf{E}$ and $\mathbf{F}$ are $n \times n$ matrices that need to be determined.

A simple matrix calculation using the Kähler conditions and the expression for Ω_{ab}, Eq. (11), leads to general forms for the metric g_{ab} and the complex structure J^a_b,

$$g_{ab} = \begin{pmatrix} \mathbf{G} & \mathbf{A}^T \\ \mathbf{A} & (\mathbf{1} + \mathbf{A}^2)\mathbf{G}^{-1} \end{pmatrix}, \qquad J^a_b = \begin{pmatrix} \mathbf{A} & (\mathbf{1} + \mathbf{A}^2)\mathbf{G}^{-1} \\ -\mathbf{G} & -\mathbf{GAG}^{-1} \end{pmatrix}. \tag{17}$$

where the $n \times n$ matrix $\mathbf{A}$ satisfies $\mathbf{GAG}^{-1} = \mathbf{A}^T$ but is otherwise arbitrary.

5. On the geometry of the space of probabilities in motion

At this point, it is useful to make some general remarks regarding the approach and the results obtained so far. The starting point is a system with a finite number of states (i.e., a discrete configuration space) and a probability $P = (P^1, ..., P^n)$ where n is the number of states. The first step of the procedure is to introduce the information metric, the natural metric on the space of probabilities. This leads to the most basic geometrical structure of the construction, *information geometry*. The second step is to consider dynamics and to introduce an action principle to derive equations of motion. This is done using a Hamiltonian formalism: introduce coordinates S^i canonically conjugate to the P^i, a Poisson bracket structure, and a Hamiltonian. This leads to additional geometrical structure, a *symplectic structure*. The third step is to extend the metric structure of information geometry, to define a geometry over the full space of the P^i and S^i. This can not be done in an arbitrary way. Consistency between the metric tensor and the symplectic form leads to a *Kähler geometry*.

Notice that the construction is very general: It applies to *any* system with a finite number of states that is described probabilistically and which admits equations of motion that can be derived from an action principle. In particular, the construction does not require any assumptions regarding classical or quantum physics.

The few assumptions that enter into the analysis lead to the beautiful result that *the natural geometry of the space of probabilities in motion is a Kähler geometry*.

6. Uniqueness of the Kähler metric via generalized Markov mappings

The Kähler conditions impose strong restrictions on the form of the metric g_{ab}, Eq. (17), but that they do not determine the metric uniquely because it depends on a matrix $\mathbf{A}$ which is to some extent arbitrary. Additional requirements are therefore needed to determine the form of g_{ab}.

As discussed before, Markov mappings play a crucial role in the proof of uniqueness of the information metric. In this section, it will be shown that the form of $\mathbf{A}$ can be determined by requiring invariance of the metric g_{ab} under a particular type of canonical transformation which extends the notion of a Markov mapping to the phase space with coordinates P^i and S^i. The simplest way to introduce a generalization of a Markov mapping is to treat it as a point transformation (i.e., a transformation of the P^i) and to use the well known fact that a point transformation can always be extended to a canonical transformation.

Consider first the simple case of Markov mappings between spaces of the same dimension; i.e., where $n = m$. In this case, the Markov mappings are just permutations of the components P^i and their extension to canonical transformations is trivial: carry out simultaneous permutations of the P^i and S^i. These transformations have two important properties: they are linear and they do not mix the P^i and the S^i.

The case that is non-trivial is the case where $n > m$. Here there are some subtle issues that need to be addressed, foremost that such canonical transformations will map spaces of different dimensions. This, however, is not a fundamental difficulty because there is a well developed theory of "nonclassical canonical transformations" which extends the concept of canonical transformations to allow for mappings of phase spaces of different dimensions [13, 14]. Although the canonical transformation that I derive in this section may be formulated within this formalism, a simpler approach is possible and therefore there will be no need to make use of the full theory of "nonclassical canonical transformations." The simpler approach followed here consists of extending the dimensionality of the space in a trivial way and imposing constraints.

For the purpose of this paper, it will be sufficient to work out the generalization for the case of the simple example of a Markov mapping that I discussed before, where the dimensionality of the space of probabilities is increased by one; i.e., $n = m + 1$. More generals cases can be derived by considering a series of successive transformations that are either permutations or

which increase the dimensionality of the space by one at each step.

Consider then a system with states described by the coordinates P^j, S^j, $j = 1, ... m$, and a Hamiltonian $H(P^j, S^j)$ which describes the dynamics.

As a first step, increase the dimensionality of the space in a trivial way by adding coordinates P^{m+1}, S^{m+1} to the phase space. This increase in the dimensionality does not change the Hamiltonian. Therefore, the time evolution of the system remains the same and the additional coordinates P^{m+1}, S^{m+1} are *constants of the motion*. Consider now a point transformation relating old and new coordinates $P^k, \tilde{P}^k$ which satisfies the following relations,

$$
\begin{aligned}
P^i &= \tilde{P}^i, \\
P^m &= \tilde{P}^m + \tilde{P}^{m+1}, \\
P^{m+1} &= (1-k)\tilde{P}^m - k\tilde{P}^{m+1} \approx 0.
\end{aligned}
\tag{18}
$$

where $i = 1, ..., m-1$, and the symbol "$\approx$" is used to indicate a weak equality (i.e., a constraint in the sense of an equality of numerical values, not of functions of the phase space coordinates). Using the last two relations of Eqs. (18), it is straightforward to show that $\tilde{P}^m$ and $\tilde{P}^{m+1}$ satisfy the constraints

$$
\begin{aligned}
\tilde{P}^m &\approx kP^m, \\
\tilde{P}^{m+1} &\approx (1-k)P^m,
\end{aligned}
\tag{19}
$$

which agree with Eq. (6). Therefore, the relations defined in Eqs. (18) are equivalent to a Markov mapping. Notice that these constraints are preserved because P^{m+1} is a constant of the motion.

The second step is to extend this point transformation to a canonical transformation. In analogy to the case discussed above in which $n = m$, I will look for a linear canonical transformation which does not mix the P^i and the S^i. Notice that these two conditions lead to a *unique* canonical transformation (up to additive constants which are unimportant). To define the canonical transformation, introduce the generating function

$$
K = \sum_{i=1}^{m-1} \left\{ \tilde{P}^i S^i \right\} + (\tilde{P}^m + \tilde{P}^{m+1}) S^m + [(1-k)\tilde{P}^m - k\tilde{P}^{m+1}] S^{m+1}.
\tag{20}
$$

Derive the canonical transformation from the generating function in the standard way; i.e. $P^k = \partial K / \partial S^k$ and $\tilde{S}^k = \partial K / \partial \tilde{P}^k$. This leads to the following equations,

$$
\begin{aligned}
P^i &= \tilde{P}^i, & P^m &= \tilde{P}^m + \tilde{P}^{m+1}, & P^{m+1} &= (1-k)\tilde{P}^m - k\tilde{P}^{m+1}, \\
S^i &= \tilde{S}^i, & S^m &= k\tilde{S}^m + (1-k)\tilde{S}^{m+1}, & S^{m+1} &= \tilde{S}^m - \tilde{S}^{m+1},
\end{aligned}
\tag{21}
$$

and

$$
\begin{aligned}
\tilde{P}^i &= P^i, & \tilde{P}^m &= kP^m + P^{m+1}, & \tilde{P}^{m+1} &= (1-k)P^m - P^{m+1}, \\
\tilde{S}^i &= S^i, & \tilde{S}^m &= S^m + (1-k)S^{m+1}, & \tilde{S}^{m+1} &= S^m - kS^{m+1}.
\end{aligned}
\tag{22}
$$

where $i = 1, ..., m-1$.

I have increased the dimensionality of the phase space by two dimensions and I have two constants of the motion, P^{m+1} and S^{m+1}. As shown in Appendix A, consistency requires both $P^{m+1} \approx 0$ and $S^{m+1} \approx 0$, which leads to

$$
\begin{aligned}
\tilde{P}^m &\approx kP^m, & \tilde{P}^{m+1} &\approx (1-k)P^m, \\
\tilde{S}^m &\approx S^m, & \tilde{S}^{m+1} &\approx S^m.
\end{aligned}
\tag{23}
$$

It is also shown in Appendix A that the time evolution is preserved; i.e., the dynamics in the $2(n+1)$-dimensional phase space (with coordinates with tildes) reproduces precisely the dynamics in the $2n$-dimensional phase space (with coordinates without tildes).

Given the canonical transformation, Eq. (21), and its inverse, Eq. (22), one can examine the restrictions imposed on the matrix $\mathbf{A}$ by the requirement of invariance of the metric g_{ab} under this generalization of a Markov mapping. The calculation is summarized in Appendix B. The result is that the matrix $\mathbf{A}$ must be proportional to the $n \times n$ unit matrix, $\mathbf{A} = A\mathbf{1}$, where A is a constant. The line element depends on two parameteres only, α and A, and it takes the remarkably simple form

$$dl^2 = \sum_k \left[\frac{\alpha}{2P^k}(dP^k)^2 + 2A dP^k dS^k + \frac{2P^k}{\alpha}(1+A^2)(dS^k)^2 \right]. \tag{24}$$

7. Complex coordinates and wave functions

Up to now, I have made use of real coordinates P^i, S^i. Kähler geometry, however, is best expressed in terms of complex coordinates. I carry out a complex transformation that shows that the metric of Eq. (24) describes in fact a flat Kähler space.

Set $\mathbf{A} = A\mathbf{1}$ in Eqs. (17) and consider first the particular case $A = 0$. The tensors that define the Kähler structure take the form

$$\Omega_{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad g_{ab} = \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{G}^{-1} \end{pmatrix}, \quad J^a_b = \begin{pmatrix} 0 & \mathbf{G}^{-1} \\ -\mathbf{G} & 0 \end{pmatrix}. \tag{25}$$

Define now the *Madelung transformation*,

$$\psi^k = \sqrt{P^k}\exp(iS^k/\alpha), \qquad \bar{\psi}^k = \sqrt{P^k}\exp(-iS^k/\alpha). \tag{26}$$

A simple calculation shows that the tensors that define the Kähler geometry, expressed in terms of ψ^k, $\bar{\psi}^k$, take the standard form which is characteristic of flat-space [12],

$$\Omega_{ab} = \begin{pmatrix} 0 & i\alpha\mathbf{1} \\ -i\alpha\mathbf{1} & 0 \end{pmatrix}, \quad g_{ab} = \begin{pmatrix} 0 & \alpha\mathbf{1} \\ \alpha\mathbf{1} & 0 \end{pmatrix}, \quad J^a_b = \begin{pmatrix} -i\mathbf{1} & 0 \\ 0 & i\mathbf{1} \end{pmatrix}. \tag{27}$$

One may conclude that in this case ($A = 0$) there is a natural set of fundamental variables given by ψ^i and $\bar{\psi}^i$. In terms of these variables, the tensors that define the Kähler geometry take their simplest form. If the constant α is set equal to $\hbar$, these fundamental variables are precisely the *wave functions* of quantum mechanics. This is a remarkable result because it is based on geometrical arguments only. The derivation does not use any assumptions from quantum theory.

Consider now the more general case $A \neq 0$. The tensors that define the Kähler structure take the form

$$\Omega_{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad g_{ab} = \begin{pmatrix} \mathbf{G} & A\mathbf{1} \\ A\mathbf{1} & (1+A^2)\mathbf{G}^{-1} \end{pmatrix}, \quad J^a_b = \begin{pmatrix} A\mathbf{1} & (1+A^2)\mathbf{G}^{-1} \\ -\mathbf{G} & -A\mathbf{1} \end{pmatrix}. \tag{28}$$

In this case, define the *modified Madelung transformation*

$$\phi^k = \sqrt{P^k}\exp\left[i\left(\Lambda S^k/\alpha - \gamma\ln\sqrt{P^k}\right)\right], \quad \bar{\phi}^k = \sqrt{P^k}\exp\left[-i\left(\Lambda S^k/\alpha - \gamma\ln\sqrt{P^k}\right)\right], \tag{29}$$

where $\Lambda = 1/(1+A^2)$ and $\gamma = -A/(1+A^2)$. Once more, the tensors that define the Kähler geometry, expressed now in terms of ϕ^k, $\bar{\phi}^k$, take the standard form which is characteristic of flat-space,

$$\Omega_{ab} = \begin{pmatrix} 0 & i\alpha\Lambda^{-1}\mathbf{1} \\ -i\alpha\Lambda^{-1}\mathbf{1} & 0 \end{pmatrix}, \quad g_{ab} = \begin{pmatrix} 0 & \alpha\Lambda^{-1}\mathbf{1} \\ \alpha\Lambda^{-1}\mathbf{1} & 0 \end{pmatrix}, \quad J^a_b = \begin{pmatrix} -i\mathbf{1} & 0 \\ 0 & i\mathbf{1} \end{pmatrix}. \tag{30}$$

This shows that the geometry of the Kähler space is the same whether $A = 0$ or $A \neq 0$. In fact, it is possible to map one case to the other using an A-dependent canonical transformation. It is clear then that both cases lead to the same theory (provided one sets $\alpha = \hbar$ when $A = 0$ or $\alpha \Lambda^{-1} = \hbar$ when $A \neq 0$), and in the following sections I will set $A = 0$ and use the complex coordinates (wave functions) ψ^i and $\bar{\psi}^i$.

The transformation that takes you from the coordinates of Eq. (26) to the coordinates of Eq. (29) is a particular case of a family of nonlinear gauge transformations introduced by Doebner and Goldin [15] (compare to their Eq. (2.2)). As pointed out by Doebner and Goldin, the theory that results from this particular family of nonlinear gauge transformations is physically equivalent to standard quantum mechanics. Here we arrive at the same conclusion, but now on the basis of the equivalence of the two cases $A = 0$ and $A \neq 0$ via a canonical transformation. One may therefore view the present derivation of the geometric formulation of quantum mechanics as providing a new route to this family of Doebner-Goldin nonlinear gauge transformations.

8. Group of unitary transformations and Hilbert space formulation

In the Hamiltonian formalism, any (smooth) function of the canonically conjugate variables generates an infinitesimal transformation. The group associated with these generators is the group of canonical transformations. The group of transformations of the theory developed here, however, can not be the full group of canonical transformations: only a subset of such generators is permissible because the infinitesimal transformations are subject to certain restrictions. In this section, I show that the group of transformations of the theory is the group of unitary transformations and that one may introduce a Hilbert space formulation of the theory. Both of these results are needed to establish the equivalence of the geometric formulation derived here to standard quantum mechanics.

To understand the role of the unitary group, one can argue as follows. Since the Kähler structure includes a symplectic structure, the group of symplectic transformations, Sp(2n,R), will play an important role in the theory. But the group of transformations of the theory can not be the symplectic group due to the requirements that the transformations have to satisfy. The first requirement is that they preserve the normalization of the probability, $\sum_i P^i = \sum_i \psi^i \bar{\psi}^i = 1$. The second requirement is that the metric be form invariant under the transformations; i.e., that the line element $dl^2 = 2\alpha \sum_j d\bar{\psi}^j d\psi^j$ of the Kähler space be preserved by the transformations. Requiring normalization of the probability and metric invariance leads to the group of rotations on the $2n$-dimensional sphere, O(2n, R). Unitary transformations are the only symplectic transformations which are also rotations; i.e., Sp(2n,R) $\cap$ O(2n, R)= U(n) [16]. Therefore, we are led to the group of unitary transformations. One can show that time evolution is described by a one-parameter group of unitary transformations (see Appendix C for a proof).

One can now introduce a Hilbert space formulation. There is a standard construction that associates a complex Hilbert space with any Kähler space. Given two complex vectors ψ^i and φ^i, define the Dirac product by [1]

$$
\begin{aligned}
\langle \psi | \varphi \rangle &= \frac{1}{2\alpha} \sum_i \left\{ \left(\psi^i, \bar{\psi}^i \right) \cdot [g + i\Omega] \cdot \begin{pmatrix} \varphi^i \\ \bar{\varphi}^i \end{pmatrix} \right\} \\
&= \frac{1}{2} \sum_i \left\{ \left(\psi^i, \bar{\psi}^i \right) \left[\begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} + i \begin{pmatrix} 0 & i\mathbf{1} \\ -i\mathbf{1} & 0 \end{pmatrix} \right] \begin{pmatrix} \varphi^i \\ \bar{\varphi}^i \end{pmatrix} \right\} \\
&= \sum_i \bar{\psi}^i \varphi^i
\end{aligned}
\tag{31}
$$

In this way one arrives at the Hilbert space formulation of quantum mechanics.

9. Concluding remarks

The geometry of quantum theory can be derived from information geometry, the natural geometry on the space of probabilities, using only a few assumptions. The derivation has a number of interesting features:

- Doubling of the dimensionality of the space (i.e., $\{P^i\} \rightarrow \{P^i, S^i\}$) from dynamical considerations,

- Complex structure from consistency between metric and symplectic structures,

- Wave functions as the natural complex coordinates of the Kähler space,

- Representation in terms of canonical transformations of a particular case of a family of Doebner-Goldin nonlinear gauge transformations,

- Time evolution described by a one-parameter group of unitary transformations,

- Hilbert space formulation expressed in terms of geometrical quantites associated with the Kähler space.

The derivation presented here relies heavily on, and extends, a geometrical reconstruction of quantum theory by Reginatto and Hall which takes information geometry as its starting point [2, 3]. Mehrafarin [17] and Goyal [18, 19] have also developed reconstructions of quantum theory using information-geometrical approaches. A detailed comparison to their approaches has not been carried out yet; however, one of the main differences is in the handling of *dynamics*, which plays a crucial role here. In particular, the use of an action principle to describe the dynamics of probabilities leads in a natural way to geometrical structure that goes beyond information geometry. A generalization of Markov mappings has proven to be very useful for deriving a unique Kähler geometry; these types of transformations may be of interest for other classes of problems where probabilities play an important role.

Finally, a few remarks concerning the limit $\alpha \rightarrow 0$ (i.e., the limit $\hbar \rightarrow 0$ in quantum mechanics). In this case, neither the Madelung transformation, Eq. (26), nor the modified Madelung transformation, Eq. (26), are well defined so I consider the description in terms of the real coordinates P^i, S^i. Of the three tensors that describe the Kähler structure, Eq. (25), only the symplectic structure Ω_{ab} remains well defined: both the metric g_{ab} and the complex structure J^a_b become singular when $\alpha = 0$. Therefore, in the limit $\alpha \rightarrow 0$, the metric g_{ab} and complex structure J^a_b are no longer available and the Kähler structure is lost. The only mathematical structure that remains is the symplectic structure, which can be used to define equations of motion for the canonically conjugate variables P^i, S^i given an appropriate Hamiltonian. One interesting consequence of taking the limit $\alpha \rightarrow 0$ is that the argument in Appendix C is no longer valid (because without a metric one can no longer require metric invariance). A more careful analysis of the limit $\hbar \rightarrow 0$ will be presented in a future publication.

10. Acknowledgments

Appendix A. Generalized Markov mappings: constants of the motion and dynamics

To derive the canonical transformation that generalizes the Markov mapping of Eq. (6), the dimensionality of the original phase space was increased by two in a trivial way. This led to

two constants of the motion, P^{m+1} and S^{m+1}. P^{m+1} was set to $P^{m+1} \approx 0$, with corresponding constraints for the $\tilde{P}^k$ of the form

$$
\begin{aligned}
\tilde{P}^m &\approx kP^m, \\
\tilde{P}^{m+1} &\approx (1-k)P^m.
\end{aligned}
\tag{A.1}
$$

These are precisely the conditions that are needed to get a generalization of the Markov mapping of Eq. (6). When $k = 1/2$, $\tilde{P}^m \approx \tilde{P}^{m+1}$, which is expected because in this case there should be invariance under the re-labeling $m \leftrightarrow m+1$. To fix the value of S^{m+1}, notice that $S^{m+1} \approx c$ leads to constraints for the $\tilde{S}^k$ of the form

$$
\begin{aligned}
\tilde{S}^m &\approx S^m + (1-k)c, \\
\tilde{S}^{m+1} &\approx S^m - kc.
\end{aligned}
\tag{A.2}
$$

Argue once more that there should be invariance under the re-labeling $m \leftrightarrow m+1$ in the case when $k = 1/2$. But this can only be satisfied if $c = 0$. On can conclude therefore that the constants of the motion must satisfy

$$
\begin{aligned}
P^{m+1} &= (1-k)\tilde{P}^m - k\tilde{P}^{m+1} \approx 0, \\
S^{m+1} &= \tilde{S}^m - \tilde{S}^{m+1} \approx 0.
\end{aligned}
\tag{A.3}
$$

The corresponding constraints for the unprimed coordinates are of the form

$$
\begin{aligned}
\tilde{P}^m &\approx kP^m, \qquad \tilde{P}^{m+1} \approx (1-k)P^m, \\
\tilde{S}^m &\approx S^m, \qquad \tilde{S}^{m+1} \approx S^m.
\end{aligned}
\tag{A.4}
$$

I check now that the dynamics in the $2(n+1)$-dimensional phase space (with coordinates with tildes) reproduces precisely the dynamics in the original $2n$-dimensional phase space (with coordinates without tildes). Using Eqs. (21-22), one can show that

$$
\begin{aligned}
\dot{\tilde{P}}^i &= \frac{\partial H}{\partial S^i}, \qquad \dot{\tilde{P}}^m = \frac{\partial H}{\partial S^m}\,k, \qquad \dot{\tilde{P}}^{m+1} = \frac{\partial H}{\partial S^m}\,(1-k), \\
\dot{\tilde{S}}^i &= -\frac{\partial H}{\partial P^i}, \qquad \dot{\tilde{S}}^m = -\frac{\partial H}{\partial P^m}, \qquad \dot{\tilde{S}}^{m+1} = -\frac{\partial H}{\partial P^m}.
\end{aligned}
\tag{A.5}
$$

These equations lead to

$$
\begin{aligned}
\dot{P}^i &= \dot{\tilde{P}}^i = \frac{\partial H}{\partial S^i}, \qquad \dot{P}^m = \dot{\tilde{P}}^m + \dot{\tilde{P}}^{m+1} = \frac{\partial H}{\partial S^m} \\
\dot{S}^i &= \dot{\tilde{S}}^i = -\frac{\partial H}{\partial P^i}, \qquad \dot{S}^m = k\dot{\tilde{S}}^m + (1-k)\dot{\tilde{S}}^{m+1} = -\frac{\partial H}{\partial P^m},
\end{aligned}
\tag{A.6}
$$

which are the correct equations of motion for the original space.

Appendix B. Invariance of the Kähler metric under generalized Markov mappings

The metric of the Kähler space is given by

$$
g_{ab} = \begin{pmatrix} \mathbf{G} & \mathbf{A}^T \\ \mathbf{A} & (\mathbf{1} + \mathbf{A}^2)\mathbf{G}^{-1} \end{pmatrix},
\tag{B.1}
$$

where $\mathbf{G} = \mathtt{diag}(\frac{\alpha}{2P^i})$ and the $n \times n$ matrix $\mathbf{A}$ satisfies $\mathbf{GAG}^{-1} = \mathbf{A}^T$. For the calculations in this Appendix it is convenient to introduce the matrix $\mathbf{B}$ with matrix elements given by

$$
B_{jk} = \sqrt{P_j/P_k}\, A_{jk}.
\tag{B.2}
$$

It is straightforward to show that **B** is a symmetric matrix, $B_{jk} = B_{kj}$.

To restrict the form of **B**, it will be sufficient to consider the invariance of the metric under the particular generalized Markov mapping which corresponds to the inverse canonical transformation of Eq. (22). After taking into consideration the constraints, Eq. (23), the generalized Markov mapping can be written in the form

$$
\begin{aligned}
P &= (P^i, P^m) \rightarrow \tilde{P} = (\tilde{P}^i, \tilde{P}^m, \tilde{P}^{m+1}) := (P^i, kP^m, (1-k)P^m), \\
S &= (S^i, S^m) \rightarrow \tilde{S} = (\tilde{S}^i, \tilde{S}^m, \tilde{S}^{m+1}) := (S^i, S^m, S^m),
\end{aligned}
\tag{B.3}
$$

where $i = 1, ..., m-1$.

As a first step, look at the contribution to the line element dl^2 from the mixed terms $dP^k dS^k$. In terms of the coordinates without tildes,

$$
dl^2 = \sum_{i=1}^{m-1} \left\{ B_{ii} dP^i dS^i + \sqrt{\frac{P^i}{P^m}} B_{im} dP^i dS^m + \sqrt{\frac{P^m}{P^i}} B_{mi} dP^m dS^i \right\} + B_{mm} dP^m dS^m.
\tag{B.4}
$$

There is a corresponding expression for the coordinates with tildes, and with the help of Eq. (B.3) it can be rewritten in terms of coordinates without tildes. This leads to

$$
\begin{aligned}
dl^2 =\ & \sum_{i=1}^{m-1} \left\{ \tilde{B}_{ii} dP^i dS^i + \left[\sqrt{\frac{P^i}{kP^m}} \tilde{B}_{im} + \sqrt{\frac{P^i}{(1-k)P^m}} \tilde{B}_{i(m+1)} \right] dP^i dS^m \right\} \\
& + \sum_{i=1}^{m-1} \left\{ \left[\sqrt{\frac{k^3 P^m}{P^i}} \tilde{B}_{mi} + \sqrt{\frac{(1-k)^3 P^m}{P^i}} \tilde{B}_{(m+1)i} \right] dP^m dS^i \right\} \\
& + \left[k\tilde{B}_{mm} + \sqrt{\frac{k^3}{1-k}} \tilde{B}_{m(m+1)} + \sqrt{\frac{(1-k)^3}{k}} \tilde{B}_{(m+1)m} + (1-k)\tilde{B}_{(m+1)(m+1)} \right] dP^m dS^m.
\end{aligned}
\tag{B.5}
$$

Equate terms in Eqs. (B.4) and (B.5) proportional to the same $dP^a dS^b$, where $a, b = 1, ..., m$. This leads to the four relations

$$
\begin{aligned}
B_{ii} &= \tilde{B}_{ii}, \\
B_{im} &= \sqrt{\frac{1}{k}} \tilde{B}_{im} + \sqrt{\frac{1}{(1-k)}} \tilde{B}_{i(m+1)}, \\
B_{mi} &= \sqrt{k^3} \tilde{B}_{mi} + \sqrt{(1-k)^3} \tilde{B}_{(m+1)i}, \\
B_{mm} &= k\tilde{B}_{mm} + \sqrt{\frac{k^3}{1-k}} \tilde{B}_{m(m+1)} + \sqrt{\frac{(1-k)^3}{k}} \tilde{B}_{(m+1)m} + (1-k)\tilde{B}_{(m+1)(m+1)}.
\end{aligned}
\tag{B.6}
$$

Since the matrix **B** is symmetric, $B_{im} = B_{mi}$, which leads to

$$
\sqrt{\frac{1}{k}} \tilde{B}_{im} + \sqrt{\frac{1}{(1-k)}} \tilde{B}_{i(m+1)} = \sqrt{k^3} \tilde{B}_{mi} + \sqrt{(1-k)^3} \tilde{B}_{(m+1)i}.
\tag{B.7}
$$

By symmetry, $\tilde{B}_{im} = \tilde{B}_{i(m+1)}$ at $k = 1/2$, but this relation can only be satisfied if $\tilde{B}_{im} = \tilde{B}_{i(m+1)} = 0$, which in turn implies $B_{im} = B_{mi} = 0$. Since B_{im} and B_{mi} are independent of k, it follows that they must always be zero. This shows that the off-diagonal elements of the matrix **B** are zero.

Now look at the contribution to the line element dl^2 from terms proportional to $dP^a dP^a$ and $dS^a dS^a$. The terms proportional to $dP^a dP^a$ give the two relations

$$\begin{aligned} B_{ii} &= \tilde{B}_{ii}, \\ B_{mm} &= k\tilde{B}_{mm} + (1-k)\tilde{B}_{(m+1)(m+1)}, \end{aligned} \quad (B.8)$$

while the terms proportional to $dS^a dS^a$ give the two relations

$$\begin{aligned} 1 + B_{ii}^2 &= 1 + \tilde{B}_{ii}^2, \\ 1 + B_{mm}^2 &= k(1 + \tilde{B}_{mm}^2) + (1-k)\left(1 + \tilde{B}_{(m+1)(m+1)}^2\right). \end{aligned} \quad (B.9)$$

Combining Eqs. (B.8) and (B.9) leads to

$$\begin{aligned} B_{ii} &= \tilde{B}_{ii}, \\ B_{mm} &= \tilde{B}_{mm} = \tilde{B}_{(m+1)(m+1)}. \end{aligned} \quad (B.10)$$

Notice that this result is valid for arbitrary values of k. Since there is nothing special about the particular labels m and $(m+1)$, *all* the diagonal elements of the matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ must be equal. Then

$$\begin{aligned} \mathbf{B} &= B\mathbf{1}_{m\times m}, \\ \tilde{\mathbf{B}} &= B\mathbf{1}_{(m+1)\times(m+1)}, \end{aligned} \quad (B.11)$$

where $\mathbf{1}_{n\times n}$ is the $n \times n$ unit matrix and B still has to be determined.

To carry out this last step, use the relations

$$\begin{aligned} B(P,S) &= B(P^i, P^m, S^i, S^m), \\ \tilde{B}(\tilde{P}, \tilde{S}) &= B(P(\tilde{P}), S(\tilde{S})) = B(\tilde{P}^i, \tilde{P}^m + \tilde{P}^{m+1}, \tilde{S}^i, k\tilde{S}^m + (1-k)\tilde{S}^{m+1}). \end{aligned} \quad (B.12)$$

The functional form of $B(P,S)$ must be the same as the functional form of $\tilde{B}(\tilde{P}, \tilde{S})$, and these expressions must be both invariant under permutations and independent of k. The only functional form that seems to satisfy all these conditions appears to be $B(P,S) = B(\sum_i P^i)$. But $\sum_i P^i = 1$, therefore one can conclude that B is a *constant* and the matrix

$$\mathbf{B} = B\mathbf{1} \quad (B.13)$$

is a constant matrix proportional to the unit matrix. This in turn implies that

$$\mathbf{A} = A\mathbf{1} \quad (B.14)$$

where A is a constant.

Appendix C. Group of unitary transformations

The generators of infinitesimal transformations (for example, Hamiltonians that generate the time evolution of the system) have to satisfy certain requirements; therefore, not every function of the coordinates qualifies as a Hamiltonian. To determine the allowed functional form of Hamiltonians, it is convenient to set $A = 0$ and work with the complex coordinates ψ^k, $\bar{\psi}^k$ introduced in Section 7. In this Appendix, I follow the derivation presented in Ref. [3] and show that Hamiltonians must be of the form $H = \sum_{j,k} M_{jk} \bar{\psi}^j \psi^k$ (up to an additive constant), where M_{jk} is a Hermitian matrix. Thus, time evolution is described by a one-parameter group of unitary transformations.

To show this, one needs to consider three requirements. The first requirement is that the transformations preserve the normalization of the probability, $\sum_i P^i = \sum_i \psi^i \bar{\psi}^i = 1$. The second requirement is that the metric be form invariant under those transformations; i.e., that the transformations preserve the line element $d\sigma^2 = 2\alpha \sum_j d\bar{\psi}^j d\psi^j$ of the Kähler space. Requiring normalization of the probability and metric invariance leads to the group of rotations on the $2n$-dimensional sphere. Such rotations are linear with respect to ψ^j and $\bar{\psi}^j$. For an infinitesimal transformation, it follows that

$$\dot{\psi}^j = -i\frac{\partial H}{\partial \bar{\psi}^j}, \qquad \dot{\bar{\psi}}^j = i\frac{\partial H}{\partial \psi^j}, \tag{C.1}$$

are linear in ψ and $\bar{\psi}$, where H is the Hamiltonian that generates the motion. Then H must be of the form

$$H = E(t) + \sum_{j,k} \left[M_{jk}\bar{\psi}^j\psi^k + N_{jk}\psi^j\psi^k + \bar{N}_{jk}\bar{\psi}^j\bar{\psi}^k \right] \tag{C.2}$$

where $E(t)$ is a arbitrary function of time, M is Hermitian, and N is symmetric.

The third and final requirement is to consider rotations on the $2n$-dimensional sphere that are compatible with the equations of motion. Conservation of probability requires that the Hamiltonian be invariant (up to an additive constant) under $S^j \to S^j + \chi$, since to first-order [11]

$$0 = \epsilon \sum_j \dot{P}^j = \epsilon \sum_j \frac{\partial H}{\partial S^j} = H(P, S+\epsilon) - H(P, S). \tag{C.3}$$

This condition, when written in terms of complex coordinates, is equivalent to invariance of the Hamiltonian under $\psi \to \psi e^{i\chi}$. Using the notation $Q := \sum N_{jk}\psi^j\psi^k$, Eq. (C.2) leads to the equality

$$\left[Qe^{2i\chi} + \bar{Q}e^{-2i\chi} \right] = 0 \tag{C.4}$$

which must be valid for all χ. Differentiating with respect to χ gives the additional equality

$$2i\left[Qe^{2i\chi} - \bar{Q}e^{-2i\chi} \right] = 0. \tag{C.5}$$

Combining these two expressions leads to $Q = 0$. Since this must hold for all ψ, it follows that $N_{jk} \equiv 0$, i.e., the Hamiltonian has the Hermitian form

$$H = E(t) + \sum_{j,k} M_{jk}\bar{\psi}^j\psi^k \tag{C.6}$$

as desired.

References
[1] Kibble T W B 1979 *Commun. math. Phys.* **65** 189
[2] Reginatto M and Hall M J W 2012 *AIP Conf. Proc.* **1443** 96
[3] Reginatto M and Hall M J W 2013 *AIP Conf. Proc.* **1553** 246
[4] Bhattacharyya A 1943 *Bull. Calcutta Math. Soc.* **35** 99
[5] Bhattacharyya A 1946 *Sanhkyā* **7** 401
[6] Good I J 1990 *J. Stat. Comput. Simulation* **36** 179
[7] Wootters W K 1981 *Phys. Rev. D.* **23** 357
[8] Čencov N N 1981 *Statistical decision rules and optimal inference* (Transl. Math. Monographs **53**) (Providence, R. I.: Amer. Math. Soc.)
[9] Campbell L L 1986 *Proc. Amer. Math. Soc.* **98** 135
[10] Caticha A 2012 *Entropic Inference and the Foundations of Physics* (Sao Paulo, Brazil: Brazilian Chapter of the International Society for Bayesian Analysis – ISBrA)

[11] Hall M J W 2004 *J. Phys. A* **37** 7799
[12] Goldberg S I 1982 *Curvature and Homology* (New York: Dover Publications)
[13] Scheifele G 1970 *Celest. Mech.* **2** 296
[14] Kurcheeva I V 1977 *Celest. Mech.* **15** 353
[15] Doebner H-D and Goldin G A 1996 *Phys. Rev. A* **54**, 3764
[16] Arnold V I 1978 *Mathematical methods of classical mechanics* (Berlin: Springer)
[17] Mehrafarin M 2005 *Int. J. Theor. Phys.* **44** 429
[18] Goyal P 2008 *Phys. Rev. A* **78** 052120
[19] Goyal P 2010 *New Journal of Physics* **12** 023012