The double-slit experiment: A paradox-free kinematic description

A. Kryukov
Department of Mathematics, University of Wisconsin Colleges, 780 Regent Street, Madison, WI 53708
(Dated: June 27, 2007)

The paradoxes of the double-slit experiment with an electron are shown to originate in the implicit assumption that the electron is always located in the classical space. It is demonstrated that there exists a natural substitute for this assumption that provides a method of resolving the paradoxes.

PACS numbers: 03.65.-w

Two physical experiments have captured the paradoxical nature of quantum mechanics in an elementary yet essentially complete way: the double-slit experiment and the EPR experiment. Few experiments in the history of science have generated so many ferocious debates, prompted so many controversial interpretations and at the end left us with such a deep feeling of discomfort with the current state of affairs. Recall the mysterious notions of the "wave-particle duality" and of "quantum non-locality" originating in the experiments.

The double-slit and the EPR experiments are essentially similar. In fact, in both of them one deals with superpositions of the classically meaningful states and those superpositions are the source of all controversies in the theory. Accordingly, there is essentially only one mystery in quantum mechanics: the existence of superpositions of classically meaningful states. So, to understand quantum mechanics is to understand superpositions of states.

In particular, the electron in the double-slit experiment is in a superposition of states that describe electron passing through one of the slits. There are three logical possibilities for such an electron:

(A) the electron passes through both slits

(B) the electron passes through only one of the slits

(C) the electron does not pass through the slits at all

Which of these possibilities is realized?

The most common answer within the quantum community is (A). Namely, one says that the electron in the given state behaves like a wave, rather than a particle. The wave passes through both slits at once causing an interference pattern behind the slits. One goes on to say that the wave function gives a complete quantum description of the electron's state. Namely, it yields the probability of finding the electron at an arbitrary spatial point (Born's rule). It is normally assumed that the electron itself is real (physical), while the wave function is not. The sketched position is due primarily to Bohr and it serves a basis for the famous Copenhagen interpretation.

In this approach the electron in the given state has no definite position. In particular, it cannot be near a single slit as otherwise the wave function would be concentrated at that slit. Paradoxically, the fact that the wave function vanishes away from the slits indicates that the electron must be near the pair of the slits. It follows that the electron splits somehow into two parts. The density of the resulting "electron cloud" coincides with the square of the modulus of the wave function. This relationship of the wave function with the physically meaningful density contradicts its earlier mentioned non-physicality.

However, whenever necessary, the Copenhagen interpretation distances itself from such problematic conclusions and logical contradictions. Instead, it retreats to the view that one should only be concerned with measurements, in which no electron parts can ever be observed and no need for a physical wave function ever arises. This runaway argument may indeed eliminate the problem, but it leaves one with a feeling of guilt for the ostrich-like behavior.

Einstein on the other hand maintained that the possibility (B) is realized. That is, the electron in the experiment goes through only one of the slits, but the standard quantum mechanics does not tell us the whole story. His famous question"Do you really think the moon isn't there if you aren't looking at it?" pushed for the development of the more detailed, "hidden variables" theories. However, as well known after Bell, those theories can only be reconciled with experiment if they admit some form of "action at a distance" i.e., nonlocality. Ironically, one of the main motivations for Einstein to promote that type of approach was to eliminate nonlocality from the theory.

Following Bohr and Einstein, the possibilities (A) and (B) were extensively explored and various more advanced interpretations of these possibilities were considered. On the other hand, the possibility (C) has not been seriously investigated. If realized, this possibility would mean that the electron disappears somehow between the source and the screen with the slits and then reappears on the other side of the screen, when absorbed by the particle detector. The goal of the Letter is to explore this radical scenario in detail. At the end we will see that (C) offers a possible way out of the major conceptual difficulties of quantum mechanics. The paper analyzes the possibility (C) within the context of the double-slit experiment. The upcoming, more technical publication [2] will do the same in the context of the EPR experiment.

To begin with, one needs a positive statement consistent with (C). Indeed, as stated, the possibility (C) can-

not be used in a constructive way. In such a form (C) is in fact consistent with a stronger form of the Copenhagen interpretation that denies existence of the electron before it is absorbed by a particle detector. The exact opposite position will be taken here. Namely, it will be assumed that *the electron in the double-slit experiment exists (in some physical form) throughout the entire experiment.*

From this assumption, (C) and the topology of the space R^3 divided by the screen with the slits, one concludes that between emission and absorption the electron lives outside R^3. So the only way to reconcile the above statements is by giving up the common perception that the electron is "attached" to the classical space. Instead, this space itself must be a "part" of a physical space of more dimensions, into which the electron can escape.

To develop this thought, recall that the electron's state in quantum mechanics is captured by the wave function. In the case when one is interested only in the electron's position, this is a complex-valued function of spatial coordinates. The evolution of the electron in the double-slit experiment is given by a path φ_t with values in a Hilbert space H of such functions. This path originates at a point φ_{t_1}, given by the wave function of the electron that is about to be emitted by the source. If the source is located at $\mathbf{a} \in R^3$, the corresponding point φ_{t_1} in H is ideally the Dirac delta function $\delta^3(\mathbf{x}-\mathbf{a})$ [7]. The endpoint φ_{t_2} of the path is given by the wave function of the electron absorbed by the particle detector. If the electron was detected at $\mathbf{b} \in R^3$, then $\varphi_{t_2}(\mathbf{x}) = \delta^3(\mathbf{x}-\mathbf{b})$.

Assuming H contains the set M_3 of all delta functions $\delta^3(\mathbf{x}-\mathbf{u})$ with $\mathbf{u} \in R^3$, one can identify the classical Euclidean space R^3 with the set M_3. Indeed, there is an obvious one-to-one correspondence between R^3 and M_3 via the map $\omega : \mathbf{u} \longrightarrow \delta^3(\mathbf{x}-\mathbf{u})$. This correspondence is physically meaningful: if the electron is located at a point $\mathbf{u} \in R^3$, then the electron's wave function is the eigenstate $\delta^3(\mathbf{x}-\mathbf{u})$ of the position operator $\widehat{\mathbf{x}}$ and vice versa [8]. Moreover, for an appropriately chosen Hilbert space H the map ω is an *isometric embedding* which means that R^3 and M_3 are identical manifolds with a metric. As a result, the classical Euclidean space R^3 can be identified in a physically meaningful way with the submanifold M_3 of a Hilbert space H of wave functions [9]. Accordingly, the following statement will be accepted:

(S) The classical space arena in physics is a part of a larger, Hilbert space arena. Physical processes with an electron on the classical space R^3 are particular cases of physical processes on the Hilbert space H of the electron's states. More precisely, the classical space R^3 can be physically identified with the manifold M_3 of the wave functions $\delta^3(\mathbf{x}-\mathbf{u})$, $\mathbf{u} \in R^3$ of the electron. The evolution of the electron is a path φ_t in the space H. Whenever the electron is detected at a point $\mathbf{a}$ in R^3, the electron's path φ_t passes through the point $\delta^3(\mathbf{x}-\mathbf{a})$ in M_3.

How does the statement (S) help interpret the double-slit experiment? It answers the question of what happens to the electron between its emission and absorption. Recall that the initial and the terminal points of the electron's path φ_t are in the classical space R^3 (identified with the manifold M_3). However, between these points the electron is in a superposition of the delta-like states. Such a superposition is not given by a delta function and therefore is *not* a point in the classical space. So the electron's path begins at a point in the classical space, then leaves this submanifold while staying in the space of states, passes *over* the screen with the slits (located in classical space) and then returns to the classical space as it is absorbed by the detector.

A specific form of the path φ_t depends on details of interaction of the electron with the source, the detector and the screen with the slits. However, any such path consists of the same basic segments: propagation from the source toward the screen with the slits, passing "through" the slits, propagation behind the screen toward the detector, and "collapse" on the detector. The first three of these segments are shown in Fig. 1. In the figure,

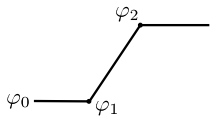

FIG. 1: Passing through the slits as a refraction of φ_t

the horizontal segments represent the propagation of the electron toward and away from the screen with the slits. The middle segment represents the motion of the electron "through" the slits when the initial electron's wave packet "splits" into a superposition of two wave packets.

Whatever the actual form of the path in Fig. 1 may be, it is clearly *single-valued* and *continuous*, i.e., it is a path in the mathematical sense. In particular, for each value of the parameter t there is only one point φ_t in H. The screen with the slits simply causes a *refraction* of the electron's path. Notice the stunning difference between Fig. 1 and the standard picturing of the double-slit experiment shown in Fig. 2. In the figure, ξ, χ are the wave

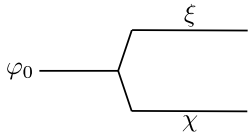

FIG. 2: The standard picturing of the double-slit experiment

functions of the electron passing through one of the slits with the second slit closed. They are the (normalized) components of the superposition $\psi = c_1\xi + c_2\chi$ that represents the electron behind the slits. The splitting of the electron's path in Fig. 2 is due to attaching the path to

the classical space and is responsible for the paradox associated with the experiment. Namely, by insisting that the electron is in the classical space R^3 one is forced to accept that the electron goes along *two* different paths in R^3. That is, both components ξ, χ must be real. This by itself is contradictory. In fact, if the same wave function is written as a superposition of eigenstates of a different observable, then, by the same logic, the new components must be real as well. Since there are many observables, the notion of reality becomes ill-defined. This is known as a "preferred basis problem" in quantum mechanics. A very similar situation arises in classical physics. Namely, when a physical vector (say, a velocity vector) is written in terms of its components in a certain basis, should we count the components as real? The answer is obvious: the physical vector itself is real because it is basis independent. However, relative to the vector, the components are just "shadows" of the real thing as they change with the change of basis, very much like shadows change when the sources of light are moved around. In quantum mechanics too, one should say that the state itself is a "real thing", while the components are not.

For instance, consider the event of passing through the slits in one dimension with the X-axis along the screen with the slits and orthogonal to the slits. If $\delta_{x_1}(x) \equiv \delta(x - x_1)$, $\delta_{x_2}(x) \equiv \delta(x - x_2)$ are the (idealized) wave functions of the electron passing through the slits at $x = x_1$ and $x = x_2$ respectively, then the wave function of the electron "passing through both slits" is a superposition $c_1\delta_{x_1} + c_2\delta_{x_2}$ with $c_1, c_2 \neq 0$. This superposition is the actual state of the electron in one dimension right behind the screen with the slits, so it is a "real thing". On the other hand, the components δ_{x_1}, δ_{x_2} themselves, no matter how familiar and real they seem to us, are only secondary and "representation dependent". So instead of having two "real" components one has now a single superposition. Instead of having two "real" electron's paths one now has a *single* path in the space of states. Instead of passing through both slits at once, the electron in the experiment does *not* pass through either of them.

The physical meaning of the superposition $\psi = c_1\delta_{x_1} + c_2\delta_{x_2}$ in the double-slit experiment is now transparent and consistent with its mathematical meaning. Namely, the superposition is the decomposition of the actual state in a basis. Once again, it is wrong to think that the components δ_{x_1}, δ_{x_2} of this decomposition are real, while the actual state ψ is not. Rather, the exact opposite is true in the experiment. So, instead of *superposing* the "real" states δ_{x_1}, δ_{x_2} to obtain a state that is not real, one *decomposes* the actual state into the components that do not enjoy an independent existence in the experiment.

More generally, the superposition principle is essentially similar to writing equations of the classical particle mechanics in components. However, instead of dealing with the motion of a classical particle along a path $\mathbf{x}_t$ in the Euclidean space R^3 one deals now with the motion of the electron along a path φ_t in the space of states H. Instead of representing $\mathbf{x}_t$ by its components in an appropriate basis, one now does the same for the path φ_t. In this sense quantum mechanics becomes an extension of the classical particle mechanics onto the space of states.

One may wonder how this "mechanical" motion may account for the wave-like properties of the electron. The answer is simple: these properties follow from the functional nature of the points on the path φ_t. For instance, when two wave packets are superposed, the square of the modulus of the resulting state ψ contains the interference term. So, when φ_t passes through the point ψ, the electron behaves like a wave. One concludes that the "wave-particle duality" is completely captured by the electron's motion in the space of states. Whenever the path crosses the classical space M_3, we see it as a particle. Whenever it leaves M_3 and passes through the regions represented by less localized states in H, it behaves like a wave.

The final part of the electron's evolution in the double-slit experiment is the collapse on the detector behind the slits. The collapse is mysterious because of its apparent discontinuity and non-locality. In particular, how could the electron's wave function, which is in general non-vanishing over large distances in R^3, instantaneously "shrink" to a point supported state? Also, how could finding the electron in one place instantaneously affect results of measurements at a distant place? This is especially paradoxical if one thinks of the electron in the state $\varphi(\mathbf{x})$ as a "cloud" of the density $|\varphi(\mathbf{x})|^2$. Once again, the Copenhagen interpretation discards this problem by saying that no such cloud can be observed and no superluminal signaling based on the collapse can be achieved in experiments. In other words, the Copenhagen school denies collapse a physical status. However, here the electron is assumed to exist throughout the entire experiment. Consequently, the collapse in the experiment must be described both mathematically and physically.

Consider once again the wave function $\psi = c_1\delta_{x_1} + c_2\delta_{x_2}$ of the electron right behind the slits in one dimension. Suppose that under a measurement this function collapses into δ_{x_1}. In this case the electron is assumed to pass through the first slit. The paradox of collapse resides once again in thinking that both terms of ψ represent a reality. That is, that the electron is in both places at once. Because of that the process of collapse seems to require an instantaneous transfer of the electron from x_2 to x_1. However, the electron in the state ψ is at *neither* of these two places. Rather, it is at the point ψ in H which is not on the classical space submanifold of H. The collapse is *not* a process on the classical space. It *does not* collect the electron's pieces into a particle. Indeed, there are no pieces to collect! Rather, the electron is represented by a *single point* in the space of states and collapse is a motion φ_t that connects that point to a point in the classical space. The fact that the electron was found at x_1 does *not* mean that it has passed through

the first slit. Instead, provided both slits were open, the electron did not pass through either of them!

This resolves the paradox of the delayed-choice experiment, when the decision to determine "which slit the electron went through" is made *after* the electron interacted with the slits. Such a delayed measurement is known to destroy the interference pattern on the photographic plate. The paradox is: how could the pattern disappear if the measurement occurred *after* the electron has already "made up its mind" and "passed through both slits"? The answer is now obvious: whether or not the measurement occurred, the electron in the experiment did *not* pass through the slits and the segments of the electron's path in Fig. 1 did not change. If the measurement (delayed or not) occurred, the collapse segment is added to the path. This transforms the superposition into a single concentrated packet and destroys the interference.

One can see that collapse is in a way opposite to the process of passing "through" the slits. Namely, whereas the screen with the slits "splits" a wave packet into a superposition of two packets, the collapse reduces the superposition into a single packet. Whereas the slits "push" the electron away from the classical space, the collapse returns it back to that space. In this respect collapse on the detector is another refraction of the electron's path in the space of states. The full path of the electron in the double-slit experiment is shown in Fig. 3.

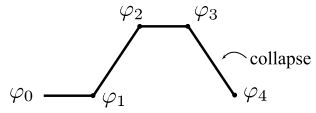

FIG. 3: Double-slit experiment with collapse as a path in H

The collapse segment of the path in Fig. 3 is shown to be continuous. This seems to contradict the known discontinuous, nonlocal character of collapse. For instance, how could a continuous process account for the instantaneous effect that finding the electron at x_1 has on the wave function and measurements at a possibly distant point x_2? Recall however that collapse is happening on the space of states. So, instead of worrying about the distance between x_1 and x_2 in the classical space, one should worry about the distance between the points $c_1\delta_{x_1}+c_2\delta_{x_2}$ and δ_{x_1} *in the space of states*. Instead of asking about the speed of collapse in the classical space, one should ask about the speed of the evolution φ_t *in the space of states*. This shift allows one to model the discontinuous, non-local process of collapse on the classical space by a continuous, local process on the space of states.

In fact, the distance between $c_1\delta_{x_1}+c_2\delta_{x_2}$ and δ_{x_1} can be small even if the distance between x_1 and x_2 is large [2]. Namely, the correspondence ω allows one to identify the classical space with a spiral-like submanifold M_3 of an arbitrarily small sphere S^H in the space of states H (see Fig. 4). Under the embedding, the infinite "size" of the Euclidean space R^3 has its counterpart in the infinite dimensionality of S^H rather than its radius [3].

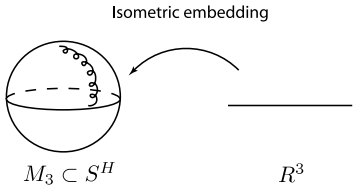

FIG. 4: R^3 as a submanifold of the sphere S^H

Take the radius of the sphere S^H to be, say, one Planck unit of length ($\approx 1.6\cdot 10^{-35}$m). Then the distance between any two states on S^H is at most π Planck units. Assume that collapse is the motion along a geodesic φ_t between the initial and the terminal states on S^H [6]. Since geodesics are continuous curves, the path φ_t is *continuous*. Also, because the equation of geodesics is a differential equation, the metric on a small neighborhood of a point is sufficient to find the path φ_t near that point. In other words, collapse is modeled by a *continuous local process* on the sphere of states. Suppose now that the speed of collapse on the sphere of states is equal to the speed of light. Then the collapse from $c_1\delta_{x_1}+c_2\delta_{x_2}$ onto δ_{x_1} happens in less than 10^{-43}s for *all* values of x_1 and x_2! The same process identified with a propagation from x_1 to x_2 in the classical space would require an infinite speed and would be a discontinuous action at a distance.

So, does this resolve the paradox of the double-slit experiment? Not quite. The presented analysis of the experiment was kinematical. The collapse process was assumed to be a geodesic motion on the sphere of states, but that assumption was not developed (see [6] for results in this direction). Despite of this, the provided analysis addresses the questions that normally appear in any discussion of the double-slit experiment (see for example the famous treatment by Feynman [1]). The analysis demonstrates that the paradigm shift from the classical space to the space of states resolves the paradox *in principle*.

[1] R. Feynman, *The character of physical law* (The M.I.T. Press, 1967)
[2] A. Kryukov, to be submitted
[3] A. Kryukov, *Int. J. Math. & Math. Sci.* **14**, 2241 (2005)
[4] A. Kryukov, *Found. Phys.* **34**, 1225 (2004)
[5] A. Kryukov, *Found. Phys.* **36**, 175 (2006)
[6] A. Kryukov, *Found. Phys.* **37**, 3 (2007)
[7] See [3] for examples of Hilbert spaces containing delta functions and related mathematical considerations.
[8] The phase of the delta functions in the identification is taken to be zero.
[9] The embedding ω can be used to *derive* the classical space from a Hilbert space of states [4].