Solution of Dirac equation for a step potential and the Klein paradox

2007 J. Phys. A: Math. Theor. 40 8991

(http://iopscience.iop.org/1751-8121/40/30/021)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 8991–9001 doi:10.1088/1751-8113/40/30/021

Solution of Dirac equation for a step potential and the Klein paradox

S Danko Bosanac

R Boskovic Institute, Bijenicka 54, Zagreb, Croatia

Received 15 March 2007, in final form 9 May 2007
Published 12 July 2007
Online at stacks.iop.org/JPhysA/40/8991

Abstract
The Klein paradox is one of the cornerstones in the development of quantum mechanics, and its consequences were used in various branches of physics, ranging from elementary particles to solid state. Yet its mathematical derivation is questionable in a number of steps, resulting in the wrong solution of Dirac equation. In this paper, the paradox is analysed in more detail, and its mathematical content is emphasized in order to show that in the study of extreme conditions of matter simple arguments may result in erroneous predictions.

PACS numbers: 03.65.Sq, 11.80.−m

1. Introduction

The effect that is known as the Klein paradox [1, 2][1] is one of the cornerstones in the development of (relativistic) quantum mechanics. It resulted from the analysis of Dirac equation for a particle that is subject to a one-dimensional impulsive repulsive force, and the epithet 'paradox' was given because the solution was in discord with anything that the intuition would have expected. Namely, the essence of the paradox is that if the force is of sufficient strength, or equivalently the potential step that represents it is sufficiently high, a particle is not reflected from its source, as one would expect. In fact in the limit when the force is infinitely repulsive the probability for no reflection is not zero. Not only that, but the probability current that gets transmitted is negative whilst that of the reflected is larger than the probability current of the incoming particle. It appears as if more current is reflected than its source provides (the incoming probability current). The intuition would have required to reexamine the analysis because of this 'non-physical' result, but this did not happen, and in order to learn why it is very instructive to learn the historic circumstances. The history of that period was nicely depicted by Pais [3], where all dilemmas of that period are put into a condensed form. It all started with the Dirac equation in 1928, which could not be solved without introducing the 'positive and negative' energy states. The 'negative energy' states could not be dismissed because they

[1] For a more educational presentation of the Klein paradox see the video titled: the Klein Paradox 1, 2 at the web address http://mediasite.oddl.fsu.edu/mediasite/Catalog/Front.aspx?cid=3bd4c40c-e410-4ba4-8594-5b9891cfeefd.

were needed for the correct solution of relativistic dynamics. The symmetry arguments were then used to show that the two states in fact represent solutions for the 'negative and positive' charges. In the history of physics this is considered as the birth of idea of the positron[2]. Klein published his analysis in 1929, when discussion about the physical meaning of the two states was getting momentum, and so it was natural to put his results into that context. Coupled with already prevailing philosophy that when considering dynamics of atoms one should forsake classical intuition, the paradox of Klein was accepted as the evidence of mixing the two states. The fact that the reflected current has larger value than the incoming, the argument went, is the evidence of particles being created whilst the negative current is the evidence of creation of particles with the opposite charge. Today this explanation is the bases for the argument that merging relativity and quantum dynamics is not possible without quantizing fields [4]. Namely, the currents for 'charges' and 'anti-charges' have continuous values, thus implying that they could take any value, which contradicts the evidence that they have discrete values. Discovery of positron in 1931 was therefore the proof that Klein result was not a paradox but evidence of a real effect, although needed quantum field theory. Pais (p 319) summarized it nicely as '. . . the prediction and subsequent discovery of the positron had eliminated many of the problems and paradoxes of the late twenties concerning the Dirac equation including the Klein paradox'.

By accepting the Klein paradox and its interpretation one must be contented with its consequences. One of them is that in the presence of a very strong force (large potential) there is spontaneous creation of particle–antiparticle pairs, and the other is that a particle cannot be confined within a potential well with very high walls. The first was essentially the basis for predicting that in the presence of strong gravitational field, near the event horizon of a black hole, there is spontaneous emission of particle–antiparticle pairs thus leading to its instability [5, 6]. The other consequence causes difficulties in attempting to formulate a potential-like model, of the MIT bag model kind, for the confinement of quarks [7]. Experimental evidence is that individual quarks are not observed, therefore the potential well that contains them should have (nearly) infinite walls, if the non-relativistic arguments are used. According to the paradox, however, such a potential well cannot be modelled because it does not support bound states. The problem was circumvented by postulating the so called Lorentz scalar potential [8, 9], whose essential feature is to modify the mass term in the Dirac equation and not the energy operator (time derivative). The things get even more complicated when more advanced potential models are used, such as the one with a barrier. One expects resonances in this case; however, the analysis produces rather disturbing result that the outgoing current increases without limits [10, 11]. The predictions of the paradox are not only limited to the relativistic processes, recently there are suggestions that it could play important role in solid state [12, 13] although the systems involved are not relativistic.

More extensive criticism of the Klein paradox will be made in this paper, but few brief comments here are appropriate. First, there is the question of initial conditions which are used in the derivation of the Klein paradox. It should be assumed for a particle to be initially away from the source of the impulsive force, whilst in describing the Klein paradox the stationary states are used, which are delocalized over the whole space. This is a standard approach to the analysis of scattering problems, taken over from the non-relativistic theory, that in many circumstances could be very misleading. In particular this is evident in the next step, time propagation of the solution, which is in a form of an integral over the energy (momentum) variable. Solution must obey a fundamental condition, the causality requirement, which states

[2] Originally Dirac thought that solutions for the negative charge is that for the electron and for the positive it is for the proton. It was later in 1931 that instead of the proton the positron was postulated, and in the same year it was discovered.

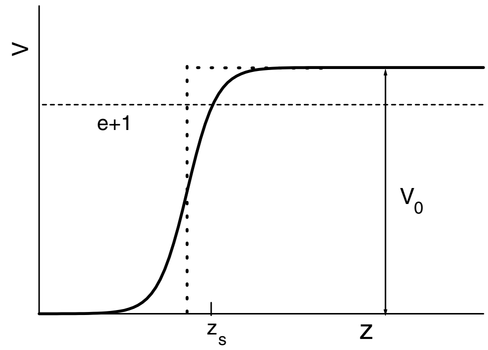

Figure 1. Step potential.

that no part of the probability density could move faster than the speed of light. This means that if the probability density is initially strictly localized within a certain interval then it is also so at any later time, but the end points move at the speed of light. This requirement could only be fulfilled if the integrand is an analytic function, which requires careful definition of the relationship between the energy and momentum variables. However, in the analysis with the stationary solutions this observation is not self-evident, and in particular it is not clear how to implement the causality requirement. As it will be shown, the implementation of this requirement drastically changes the conclusions of the Klein analysis.

In this paper the derivation of the Klein paradox is critically reviewed, and it is shown that it results from the oversimplified approach to solving relativistic quantum problems, in particular the Dirac equation. In a proper analysis one derives a solution that is in accord with what one would expect from the intuition, with slight modifications that are appropriate for relativistic dynamics. This in particular means that very high potential barriers indeed form an insurmountable obstacle for a particle.

2. Dirac equation

2.1. One-dimensional solution

Dirac equation in the convenient units and in one dimension is

$$\mathrm{i}\partial_t F = -\mathrm{i}\partial_z G + VF + F, \qquad \mathrm{i}\partial_t G = -\mathrm{i}\partial_z F + VG - G \tag{1}$$

and by writing the solution as

$$F(z,t) = \int_{-\infty}^{\infty} \mathrm{d}e\, A(e) f(z,e)\,\mathrm{e}^{-\mathrm{i}te}, \qquad G(z,t) = \int_{-\infty}^{\infty} \mathrm{d}e\, A(e) g(z,e)\,\mathrm{e}^{-\mathrm{i}te} \tag{2}$$

the equation for the stationary solution f is

$$\partial_z^2 f + \frac{1}{e - V + 1}\,(\partial_z V)\,(\partial_z f) = -[(e-V)^2 - 1]f \tag{3}$$

and

$$g = -\mathrm{i}\frac{\partial_z f}{e - V + 1}. \tag{4}$$

The simplest example is a step potential, which is shown in figure 1 by the solid line, and its limiting form that is commonly used in the analysis leading to the Klein paradox is shown by the broken line. Initial conditions for the functions $F(z,t)$ and $G(z,t)$ are defined in the space $z < 0$, where the stationary solution f is given in terms of the momentum variable $p = \sqrt{e^2 - 1}$. This implies that in the integrals in equations (2) the energy variable should not take values $|e| < 1$ because the momentum variable is then imaginary. Therefore to exclude this interval from analysis the integrals should be reformulated in terms of the momentum variable, and because of the arbitrariness in the sign between p and e the amplitude F is

$$F(z,t) = \int_{-\infty}^{\infty} \mathrm{d}p\, A^+(p) f^+(z,p)\, \mathrm{e}^{-\mathrm{i}te} + \int_{-\infty}^{\infty} \mathrm{d}p\, A^-(p) f^-(z,p)\, \mathrm{e}^{\mathrm{i}te}; \qquad e > 0 \tag{5}$$

and similarly the amplitude G is

$$G(z,t) = -\mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}p\, A^+(p) \frac{\partial_z f^+(z,p)}{e - V + 1}\, \mathrm{e}^{-\mathrm{i}te} - \mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}p\, A^-(p) \frac{\partial_z f^-(z,p)}{-e - V + 1}\, \mathrm{e}^{\mathrm{i}te}, \tag{6}$$

where the superscript indicates the sign of e, e.g. $f^-(z,e) = f(z,-e)$.

An important feature of the step potential is that if V_0 is sufficiently large then there is a point $z = z_s$ (shown in figure 1) at which (for more details see [14])

$$e - V(z_s) + 1 = 0 \tag{7}$$

and in its vicinity equation (3) approximates as

$$\mathrm{d}_z^2 f = \frac{1}{z - z_s} \mathrm{d}_z f$$

with the solution

$$f = a_1 + a_2 (z - z_s)^2 . \tag{8}$$

It could be shown that a_1 is arbitrary (the proof is relatively straightforward and it is not elaborated here), which means that it is always possible to find a solution of equation (3) that is zero at the barrier by requiring

$$a_1 = 0.$$

As a result of this choice there is no flow of probability across the point $z = z_s$. This finding also applies in the extreme case of a step potential (impulsive force), shown by the broken line in figure 1, when z_s coincides with the position of the step.

Formal solution f for the (impulsive) step potential is

$$f^\pm(z,p) = \begin{cases} \mathrm{e}^{\mathrm{i}pz} + a^\pm\, \mathrm{e}^{-\mathrm{i}pz}; & z < 0 \\ b^\pm\, \mathrm{e}^{\mathrm{i}P^\pm z}; & z > 0 \end{cases} \tag{9}$$

and g is

$$g^\pm(z,p) = \begin{cases} \dfrac{p}{1 \pm e}(\mathrm{e}^{\mathrm{i}pz} - a^\pm\, \mathrm{e}^{-\mathrm{i}pz}); & z < 0 \\ \dfrac{P^\pm}{1 \pm e + V_0} b^\pm\, \mathrm{e}^{\mathrm{i}P^\pm z}; & z > 0 \end{cases}$$

$$g = -\mathrm{i}\frac{\partial_z f}{e - V_0 + 1},$$

where

$$P^\pm = [(\pm e - V_0)^2 - 1]^{1/2},$$

which anticipates that initially the particle is localized in the space $z < 0$ and moves towards the step. In the space $z > 0$ there is only the transmitted amplitude, where by analogy with the non-relativistic scattering it is assumed that e^{iPz} represents the wave that propagates in the direction $z \to \infty$. The solution does not take into account that at $z = z_s$ it has the form (8), which, strictly speaking, should be included. However, at this stage the derivation follows the standard treatment, in which case the coefficients are

$$a^{\pm} = \frac{p\,(1 - V_0 \pm e) - P^{\pm}(1 \pm e)}{p(1 - V_0 \pm e) + P^{\pm}(1 \pm e)} \qquad b^{\pm} = \frac{2p\,(1 - V_0 \pm e)}{p\,(1 - V_0 \pm e) + P^{\pm}\,(1 \pm e)}, \tag{10}$$

which are normalized as

$$|a^{\pm}|^2 + \frac{P^{\pm}(1 \pm e)}{(1 \pm e - V_0)\,p}|b^{\pm}|^2 = 1. \tag{11}$$

The energy variable is defined as

$$e(p) = (p^2 + 1)^{1/2}.$$

The use of stationary solutions (9) in analysis of dynamics of a particle is very dangerous without knowing their precise analytic properties as the functions of the complex momentum variable p. These are not simple because $f^{\pm}(z, p)$ are explicitly dependent on two important variables, $e(p)$ and $P^{\pm}(p)$, which define three pairs of square-root branch points in the p plane. One pair is $p_{1,2} = \pm i$ and the other two are the solutions of the equation

$$P^{+}(p) = 0,$$

which are given by (throughout the analysis it is assumed that $V_0 > 2$)

$$p_3 = \sqrt{V_0(V_0 - 2)}, \qquad p_4 = \sqrt{V_0(V_0 + 2)}$$
$$p_5 = -\sqrt{V_0(V_0 - 2)}, \qquad p_6 = -\sqrt{V_0(V_0 + 2)}.$$

The momentum function $P^{-}(p)$ does not have roots, and hence it is not a source of the branch points. In order to make functions $f^{\pm}(z, p)$ analytic in the plane that contains these branch points one must define Riemann cuts that connect them, however, the procedure is not unique. For example, either one cut joins the points $p_{1,2}$ or two cuts are defined along the intervals $(i, i\infty)$ and $(-i, -i\infty)$. Physical arguments determine their choice.

The coefficients $A^{\pm}$ in solutions (5) and (6) are determined from their initial values F_0 and G_0, and under assumption that at $t = 0$ they are localized entirely in the space $z < 0$ one gets

$$A^{\pm}(p) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dz \left[\pm \frac{1 \pm e}{e} F_0(z) \pm \frac{p}{e} G_0(z) \right] e^{-ipz}. \tag{12}$$

2.2. Analytic structure of the solution for the Klein paradox

In the standard treatment of scattering of a particle from the step potential in figure 1 the momentum variable p is implicitly assumed to have positive values, but its range of definition should also include the negative values. The physical argument for this is simple. If any physically relevant information about the dynamics of a particle is required then the initial condition must assume that the particle is strictly localized within a certain interval outside the reach of the force. However, by the rules of quantum dynamics this implies that the momentum variable must extend over the whole available space, which also includes the negative values. Extension to the negative values is not an easily implemented task, because it directly reflects on the properties of the solutions $f^{\pm}(z, p)$. The problem lies in the fact that these functions are not explicitly dependent on p but through the momentum functions $P^{\pm}(p)$

and $e(p)$. Superficially all three functions are symmetric with respect to reflection $p \to -p$, but on closer examination this is not the case.

One starts by analysing the analytic properties of $P^+(p)$, which in the standard treatment gets the values for real p

$$P^+(p) = \begin{cases} |P^+|; & p > p_4 \\ \mathrm{i}|P^+|; & p_3 < p < p_4 \\ |P^+|; & p_5 < p < p_3 \\ \mathrm{i}|P^+|; & p_6 < p < p_5 \\ -|P^+|; & p < p_6, \end{cases} \tag{13}$$

where its definition for negative p is anticipated from physical arguments: the finiteness of the probability amplitude and the direction of motion for large $|p|$. Already on this simple level there is inconsistency with the definition of this momentum function. If it changes the sign when $p > p_4$ goes to $p < p_6$, why it should not when $0 < p < p_3$ goes into $p_5 < p < 0$? Enforcing the change of sign in the latter case is not simple because that could only happen if $P^+(p)$ is zero for $p = 0$. All other possibilities to change the sign encounter obstacles that result in the non-physical solutions (their detailed investigation is not of relevance here). However, this is not the biggest problem that one encounters with the definition of the momentum function (13). Its sign for $p > p_4$ is determined from the physical argument that in the space $z > 0$ the solution $f^+(p)$ represents the wave that propagates away from the potential. By the same argument the sign of $P^+(p)$ in the interval $0 < p < p_3$ is also chosen positive, but this is not correct. Indeed if $p \gg V_0$ the limit of $P^+(p)$ is p and could be chosen positive to represent the physical solution. On the other hand, in the interval in question $V_0 > p$ and in the limit of very large V_0 the function $P^+(p)$ has the estimate

$$P^+(p) \approx V_0 - e(p) \approx V_0 - p \tag{14}$$

and in the space $z > 0$ the solution $f^+(p)$ represents motion of the particle towards the potential. This is not the only problem, but if the positive sign of $P^+(p)$ is retained then one encounters additional problem in the definition of the initial conditions. It could be shown, but not elaborated in details, that if initially the particle has the average momentum $p_0 \ll V_0$ and the amplitudes $A^\pm(p)$ are calculated from (12) then besides reproducing correctly their values F_0 and G_0 in the space $z < 0$ one gets additional, 'ghost', functions in the space $z > 0$. These 'ghost' functions move towards the potential step, resulting in an additional contribution in the probability current in the space $z < 0$ once they reach the point $z = 0$. Combining the two observations one puts into doubt the results of standard analysis and therefore the Klein paradox, yet it is too early to say that it could be dismissed as non-physical. In order to do that one needs to find mathematically correct alternative to the previous analysis.

In the complex p plane the table of values (13) is reproduced by a path that is shown in figure 2(a) by the broken line. The thick lines represent the Riemann cuts between two parallel complex p planes, therefore the path for $\mathrm{Re}(p) < p_3$ is in the lower one. Going into the complex p plane is not only of superficial interest, but required if one is to get some general properties of the solution for dynamics, which is summarized in the integrals (5) and (6). In particular, a very important property to show is that by integrating along such a path the causality principle is not violated. This principle requires that if one chooses initial probability amplitude that is strictly zero outside certain interval in space $z < 0$ then at any later time it is also zero outside certain interval, where the boundaries of the probability amplitude do not move faster than the speed of light. In general, this is shown if for the parameters that violate this principle the integration path could be distorted without acquiring new values into far upper or lower half complex p plane. Along such a path the integrand should be zero. For

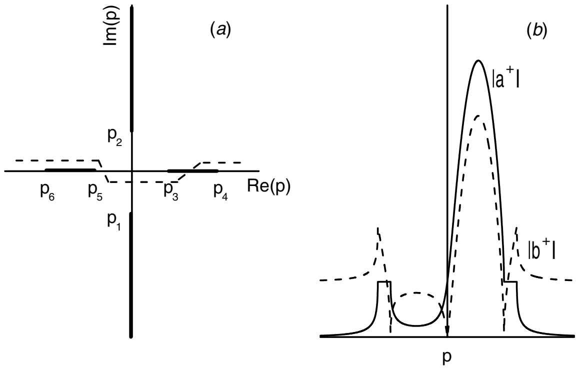

Figure 2. Riemann surface in the standard analysis of the Klein paradox (*a*) together with the integration path (broken line) for calculating probability amplitude. Corresponding modules of the reflection (solid line) and transmission (broken line) amplitudes are shown (*b*).

the path in figure 2(*a*) this is not possible because it cannot be distorted in neither halves of the plane without returning to the real axes where it crosses from one Riemann plane into the other.

The appropriate modules of the coefficients $a^+(p)$ and $b^+(p)$ are shown in figure 2(*b*), where the intervals in figure 2(*a*) are not indicated but noted as the flat line in $|a^+(p)|$. In the interval $0 < p < p_3$ these modules increase without bounds when $V_0 \to \infty$, which is interpreted as creation of the particle–antiparticle pairs.

3. Klein paradox reexamined

There are two aspects of the Klein paradox that are crucial for its understanding: (a) proper analysis of the analytic structure of the integrand in solutions (5) and (6), and (b) inspecting whether it is possible to prevent the flow of probability across the step when the potential step is very high. Each of these aspects is important in its own way, but only together they give the answer to the question of whether there is a paradox or not.

3.1. Analytic structure of integrands

As it was shown in the previous section one of the greatest obstacles in the standard treatment of the Klein paradox comes from the wrong sign of $P^+(p)$ in the interval $0 < p < p_3$, as given in equation (14), with all the consequences that are deduced from that. One could also make criticism of the 'non-physical' values of the coefficients $a^+(p)$ and $b^+(p)$ in the same interval, but because there is already an 'interpretation' of this feature further discussion of it is omitted at this stage. The possible remedy is to assume that $P^+(p)$ has negative sign in the interval $p_5 < p < p_3$, but that does not solve the problems that were indicated earlier on, except that the 'ghost' solution is no longer present. A radically new step is required, and this is to use different Riemann cuts in the complex p plane. For those that connect p_3 with p_4 and p_5 with p_6 there is no alternative and the only option is to define the cut that connects the points p_1 with p_2. In this case, the new complex p plane is shown in figure 3(*a*), where the path along which $P(p)$ is defined goes above the branch point p_2. The path now scans all the values of $e(p)$ as required by the transformation from equations (2) to equations (5) and (6), which was dismissed in the previous analysis with the argument that along the interval from $p = 0$ to $p = i = p_2$ the solution $f(z, p)$ is not physical. As it turns out that although this portion of the path is formally included in the integrals its contribution in the solutions is zero. In fact, the inclusion of the path has nothing to do with its contribution in the solution, instead it is there for entirely different reason. By going around the branch point p_2 the sign of $e(p)$ changes, which means that $P^+(p)$ becomes $P^-(p)$ (shown in figure 3(a)) and hence in the space $\mathrm{Re}(p) < 0$ there are no cuts. The table of values for $P(p)$ in this case is

$$P^+(p) = \begin{cases} |P^+|; & p > p_4 \\ \mathrm{i}|P^+|; & p_3 < p < p_4 \\ -|P^+|; & 0 < p < p_3 \\ -|P^-|; & p < 0. \end{cases}$$

Likewise $P^-(p)$ has the values on the real axes

$$P^-(p) = \begin{cases} |P^-|; & p > 0 \\ |P^+|; & p_5 < p < 0 \\ \mathrm{i}|P^+|; & p_6 < p < p_5 \\ -\mathrm{i}|P^+|; & p < p_6. \end{cases}$$

An important feature of these integration paths is that along them the modules of the coefficients $a^\pm$ and $b^\pm$ has 'physically acceptable' values, as shown in figure 3(b) for $P^+(p)$. This means that their modules are limited within the bounds of unity; however, they still have a 'non-acceptable' feature that for $V_0 \to \infty$ the reflection coefficient is zero and the transmission is unity.

Another equally important feature of this definition of the Riemann plane is that the integration path could be distorted into its far upper part thus enabling to satisfy the causality principle. As it was argued in the previous section this is not possible in the standard treatment, because the integration path must always return to the vicinity of the real p axes.

3.2. *Implementation of boundary condition*

Most of the problems are solved in the plane with the cuts that are shown in figure 3(a); the path is on a single plane of the Riemann surface and it is easily distorted into the upper half plane, the propagation of the solution in the space $z > 0$ reverses direction when $p > 0$ goes to $p < 0$ and the modules of the coefficients $a(p)$ and $b(p)$ are within the bounds that are acceptable, as shown in figure 3(b). The only remaining obstacle to finding the correct solution is the wrong limit of the coefficients $a^+(p)$ and $b^+(p)$ when $V_0 \to \infty$, the same problem as in figure 3(b). In other words, for the infinite barrier the particle goes through it as if it is not present. One consequence of this would be that a particle could not be contained in a potential well with infinitely high walls, say by a harmonic force, because it would leak out as if it is not present. Therefore, it is essential to make the final step to resolve this obstacle, and the basis is the Riemann plane in figure 3(a). In fact, there are two Riemann planes, the mentioned one and the other that starts with the function $P^-(p)$ in the space $p > 0$. The latter is a mirror image with respect to the imaginary p axes of that shown in figure 3(a). The existence of two Riemann planes means that solution (5) is again a sum of two integrals, each one in a separate plane. The solution is therefore

$$F(z,t) = \int_1 \mathrm{d}p\, A_1(p) f_1(z,p)\, \mathrm{e}^{-\mathrm{i}te} + \int_2 \mathrm{d}p\, A_2(p) f_2(z,p)\, \mathrm{e}^{\mathrm{i}te},$$

where the indices 1 and 2 indicate the path in the two Riemann planes. Similarly, one defines the solution for the amplitude G.

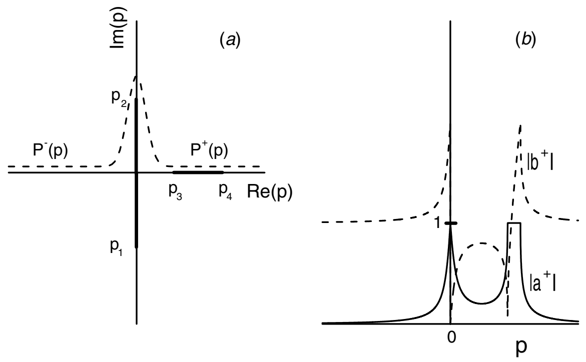

Figure 3. Riemann surface that correctly reproduces the essential requirements on the probability amplitude (*a*) together with the integration path (broken line) for calculating it. Corresponding modules of the reflection (solid line) and transmission (broken line) amplitudes are shown (*b*).

The stationary functions $f_{1,2}(z, p)$ are defined in terms of the stationary function $f(z, p)$ by taking into account the value of $e(p)$ and $P(p)$ along the path in the complex p plane, the one that is shown in figure 3(*a*). However, as it was shown in section 2 when condition (7) is satisfied then solution $f(z, p)$ has the functional form (8) around $z = 0$, for example, the function $f_1(z, p)$ in the intervals $p > 0$ and $p \in (0, i)^{\pm}$. The important feature is that in the interval of p where the condition is satisfied also coincides with the interval of the 'non-physical' character of the function $f(z, p)$, of the sort that was discussed in figure 3(*b*). Therefore, by imposing the boundary condition $f(0, p) = 0$ in these intervals this character of the solution disappears, and $f(z, p)$ becomes perfectly acceptable 'physical' solution. Thus in the space $z < 0$ the two solutions on the real and imaginary (in the interval between 0 and p_2) axes are

$$f_1(z, p) = \begin{cases} e^{ipz} + a(p)\,e^{-ipz}; & p > 0, & e(p) > 0, & P(p) = P^+(p) \\ e^{ipz} - e^{-ipz}; & p_3 > p > 0, & e(p) > 0, & P(p) = P^+(p) \\ e^{ipz} - e^{-ipz}; & p \in (0, i)^+, & e(p) > 0, & P(p) = P^+(p^+) \\ e^{ipz} - e^{-ipz}; & p \in (0, i)^-, & e(p) < 0, & P(p) = -P^-(p^+) \\ e^{ipz} + a(p)\,e^{-ipz}; & p < 0, & e(p) < 0, & P(p) = -P^-(-p) \end{cases}$$

and

$$f_2(z, p) = \begin{cases} e^{ipz} + a(p)\,e^{-ipz}; & p > 0, & e(p) > 0, & P(p) = P^-(p) \\ e^{ipz} - e^{-ipz}; & p \in (0, i)^+, & e(p) > 0, & P(p) = P^-(p^+) \\ e^{ipz} - e^{-ipz}; & p \in (0, i)^-, & e(p) < 0, & P(p) = -P^+(p^+) \\ e^{ipz} - e^{-ipz}; & p_5 < p < 0, & e(p) < 0, & P(p) = -P^+(-p) \\ e^{ipz} + a(p)\,e^{-ipz}; & p < 0, & e(p) < 0, & P(p) = -P^{+*}(-p), \end{cases}$$

where the superscripts $\pm$ of p indicate infinitesimal positive or negative real contribution. In the space $z > 0$ these solutions are

$$f_1(z, p) = \begin{cases} b(p)\,e^{iPz}; & p > 0, & e(p) > 0, & P(p) = P^+(p) \\ 0; & p_3 > p > 0, & e(p) > 0, & P(p) = P^+(p) \\ 0; & p \in (0, i)^+, & e(p) > 0, & P(p) = P^+(p^+) \\ 0; & p \in (0, i)^-, & e(p) < 0, & P(p) = -P^-(p^+) \\ b(p)\,e^{iPz}; & p < 0, & e(p) < 0, & P(p) = -P^-(-p) \end{cases} \qquad (15)$$

and

$$f_2(z,p)=\begin{cases} b(p)\,\mathrm{e}^{\mathrm{i}Pz}; & p>0, \quad e(p)>0, \quad P(p)=P^-(p) \\ 0; & p\in(0,i)^+, \quad e(p)>0, \quad P(p)=P^-(p^+) \\ 0; & p\in(0,i)^-, \quad e(p)<0, \quad P(p)=-P^+(p^+) \\ 0; & p_5<p<0, \quad e(p)<0, \quad P(p)=-P^+(-p) \\ b(p)\,\mathrm{e}^{\mathrm{i}Pz}; & p<0, \quad e(p)<0, \quad P(p)=-P^{+*}(-p), \end{cases} \tag{16}$$

where the zero value of the functions was chosen because across the point $z=0$ there is no flow of the probability. What the solution now says is that in the intervals where V_0 is much larger than $e(p)$, and not necessarily infinite, the particle is totally reflected from the potential, but otherwise it behaves as expected from the standard non-relativistic treatment.

Typical interval of p, which is assumed in the following analysis, in which the Klein paradox is described is easily identified: this is where p is positive and for which $e(p)$ is much smaller than V_0. If the amplitudes $A_{1,2}(p)$ are negligible outside this interval then the selected stationary solutions $f_{1,2}(z,p)$ produce probability density that never crosses into the space $z>0$. Strictly speaking, however, solutions for the functions $f_{1,2}(z,p)$ are not zero in the space $z>0$, contrary to the suggested values in equations (15) and (16). Furthermore, the amplitude $F(z,t)$ is only one component in the probability density, the final object of the calculation. The other component is the $G(z,t)$ amplitude, given by equation (6) but appropriately formulated in terms of the functions $g_{1,2}(z,p)$ that are derived from the functions $f_{1,2}(z,p)$ through the relationship (4). At the turning point $z=z_s=0$ the functions $f_{1,2}(z,p)$ are zero, by the choice of the boundary condition, but $g_{1,2}(z,p)$ are not zero, which implies that the object of the previous exercises is nullified. Namely, the object was to select the solution that gives zero for the probability density

$$P(z,t)=|F(z,t)|^2+|G(z,t)|^2$$

in the space $z=z_s>0$ (when the barrier is very high). First impression is that in order to obtain this result one requires that $g_{1,2}(z,p)$ are zero at $z=z_s$; however, this is not the correct criterion. Sufficient condition is that the probability current

$$J(z,t)=G^*(z,t)F(z,t)+F^*(z,t)G(z,t)$$

is zero at $z=z_s$, which is indeed the case. This means that if the particle is initially localized entirely in the space $z<0$, i.e. $P(z,0)=0$ for $z>0$, then because there is no flow of the probability density across the point $z=z_s$ the probability density stays zero in the space $z>0$ for all times. Formally this is achieved by setting $f_{1,2}(z,p)$ and $g_{1,2}(z,p)$ to zero for $z>0$, and when p is from the appropriate interval. This is the argument behind the choice of the values in equations (15) and (16).

There still remains a problem, it is connected with the function $f_2(z,p)$, and hence with $g_2(z,p)$ as well. In the space $z>0$ and for $p>0$ it is determined by the momentum function $P^-(p)$, as shown in equation (16). This means that in this space the potential barrier is in fact treated as the potential well, because the change in the sign of e is equivalent as the replacement $V_0\to -V_0$, but keeping positive e. It would appear that for this component the probability density gets through the barrier, and there is probability of finding particle where it should not be. In short, one obtains the Klein paradox again. However, it should be recalled that in the integral that propagates the 'negative energy' component of the amplitude $F(z,t)$ there is the exponential $\exp[\mathrm{i}e(p)t]$, which causes this component to move in the opposite direction then anticipated by the sign of p. Therefore, when the initial state is set up then in the space $z<0$ the component with $\exp(\mathrm{i}pz)$ propagates towards $z\to-\infty$, away from the step. The component with $\exp(-\mathrm{i}pz)$ has no contribution in the space $z<0$, which is the same as with the component $\exp[\mathrm{i}P^-(p)z]$ in the space $z>0$. This means that although formally the 'negative' energy component treats the barrier as a well, implying that part of the probability amplitude might cross into the space $z > 0$, this never happens because this component moves away from the barrier. The matter would have been different if the initial probability density is partly in the space $z < 0$ and partly in $z > 0$.

4. Conclusion

Complete, mathematically correct, solution of the Dirac equation for a one-dimensional step potential was obtained that conforms to the 'intuitive feel' of what it should be. It could be therefore argued that it invalidates the Klein paradox, which is only an artefact of the incomplete mathematical analysis of the solution of the Dirac equation. However, one could argue that the solution in this paper was gauged towards that goal, but the fact remains that in its derivation all the objections to the standard derivation were resolved. As it was shown, if the standard derivation is accepted then many mathematical problems are encountered. These problems could not be ignored if some general requirements are imposed, but historically they were because the solution represented particle–antiparticle pair creation. Despite several wave packet analysis of the Klein paradox in the past [15] their goal was not to invalidate it, which could have been done, instead the aim was to reconfirm it and show how the particle–antiparticle pairs are created. The 'intuitive feel' that this is the correct interpretation predetermined the solution.

References

[1] Bjorken J D and Drell S D 1964 *Relativistic Quantum Mechanics* (New York: McGraw-Hill)
[2] Dombey N and Calogeracos A 1999 *Phys. Rep.* **315** 41
[3] Pais A 1986 *Inward Bound* (New York: Oxford University Press)
[4] Greiner W, Muller B and Rafelski J 1985 *Quantum Electrodynamics of Strong Fields* (Berlin: Springer)
[5] Hawking S 1974 *Nature* **284** 30
[6] Page D N 2005 *New J. Phys.* **7** 203
[7] For a detailed review of the model see Squires E J 1979 *Rep. Prog. Phys.* **42** 1187
[8] Bogoliubov N P 1968 *Ann. Inst. Henri Poincare* **8** 163
[9] Ho C-H 2006 *Ann. Phys., NY* **321** 2170
[10] Katsnelson M I, Novoselov K S and Katsnelson M I 2006 *Nat. Phys.* **2** 620
[11] McKellar B H J and Stephenson G J 1987 *Phys. Rev.* A **36** 2566
[12] De Leo S and Rotelli P P 2006 *Phys. Rev.* A **73** 042107
[13] De Leo S and Rotelli P P 2006 *Eur. Phys. J.* C **46** 551
[14] Bosanac S D 2005 *Dynamics of Particles and the Electromagnetic Field* (Singapore: World Scientific)
[15] Krekor P, Su Q and Grobe R 2004 *Phys. Rev. Lett.* **92** 040406