AMERICAN JOURNAL of PHYSICS

Path integral solution to the infinite square well

Mark Goodman

Citation: American Journal of Physics **49**, 843 (1981); doi: 10.1119/1.12720
View online: http://dx.doi.org/10.1119/1.12720
View Table of Contents: http://scitation.aip.org/content/aapt/journal/ajp/49/9?ver=pdfcov
Published by the American Association of Physics Teachers

[3]A. Boivin, J. Dow, and E. Wolf, J. Opt. Soc. Am. **57**, 1171 (1967).
[4]C. O. Hines, Can. J. Phys. **30**, 123 (1952).
[5]G. L. Wallace and R. F. O'Connell, Can. J. Phys. **58**, 744 (1980).
[6]D. Villarroel, Ann. Phys.(NY) **89**, 241 (1974).

Path integral solution to the infinite square well

Mark Goodman[a)]
Instituto de Fisica Teorica dell'Università di Trieste, Trieste, Italy
(Received 21 July 1980; accepted 20 November 1980)

The infinite square-well potential in one dimension is solved using Feynman path integration. This solution uses an image point method equivalent to a sum over classical paths.

INTRODUCTION

The infinite square-well potential is one of the simplest bound-state problems in wave mechanics; it is usually one of the first examples given in any introductory quantum mechanics course. The problem is to solve for the motion in one dimension of a particle under the influence of the potential

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < L \\ \infty & \text{for } x \leq 0 \text{ or } x \geq L. \end{cases} \quad (1)$$

The normalized solutions to the time-independent Schrödinger equation for this potential are

$$\psi_n(x) = \begin{cases} \sqrt{2/L}\,\sin(\pi n/Lx) & \text{for } 0 < x < L \\ 0 & \text{for } x \leq 0 \text{ or } x \geq L, \end{cases} \quad (2)$$

with the energy eigenvalues

$$E_n = (1/2m)(\pi n\hbar/L)^2. \quad (3)$$

For $t_b > t_a$ and $0 < x_a, x_b < L$ the propagator is given by

$$\langle x_b,t_b|x_a,t_a\rangle = \sum_{n=1}^{\infty} \exp[(i/\hbar)E_n(t_b-t_a)]\psi_n^*(x_b)\psi_n(x_a); \quad (4)$$

otherwise it is zero.

This is a very familiar result from wave mechanics. It has the standard interpretation in terms of the normal modes of oscillation of a wave in a resonant cavity. This interpretation, however, deals only with the wave part of the wave–particle duality of matter. A standard tool for exhibiting the particle aspects of this duality is the Feynman integral over paths. Applying directly the formalism for this integral[1] yields the relation

$$\begin{aligned}\langle x_b,t_b|x_a,t_a\rangle &= \int_{x_a}^{x_b} \exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b}\frac{m}{2}\dot{x}^2(t)\right)\mathcal{D}x(t) \\ &= \lim_{n\to\infty}\int_0^L\cdots\int_0^L \delta(x_b-x_n)\delta(x_0-x_a) \\ &\times \exp\left(\sum_{k=1}^{n}\frac{m}{2\epsilon}(x_k-x_{k-1})^2\right)\left(\frac{m}{2\pi\hbar\epsilon}\right)^{n/2}\prod_{k=0}^{n}dx_k. \quad (5)\end{aligned}$$

The limits for each integration are 0 and L, since for the infinite square well, the particle is confined even quantum mechanically to this region. There is no apparent way to demonstrate this from within the path integral formulation, and it can be argued that this is because the problem is stated improperly (see below). This fact can be seen, however, by following the development of the path integral from the Heisenberg formalism[2]

$$\begin{aligned}\langle x_b,t_b|x_a,t_a\rangle &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\delta(x_b-x_n)\delta(x_0-x_a) \\ &\times \prod_{k=1}^{n}\langle x_k,t_k|x_{k-1},t_{k-1}\rangle\prod_{k=0}^{n}dx_k. \quad (6)\end{aligned}$$

For the square well, each factor in the integrand vanishes for $x_k \leq 0$ or $x_k \geq L$. In effect, then, each integral is limited to the domain $0 < x_k < L$.

The difficulties in evaluating the integral (5) are now obvious; a Gaussian integral cannot be evaluated in closed form if it is restricted to a bounded domain. The prospects for repeating this integration, and then proceeding to the limit, seem bleak indeed. This direct approach to the path integral does not readily yield a solution. It will be seen that a less direct but more intuitive attack will give a solution that can be justified rigorously using path integrals.

POTENTIAL BARRIER

Before proceeding with this method, we will first solve the simpler problem of the infinite potential barrier, whose potential function is

$$V(x) = \begin{cases} 0 & \text{for } x > 0 \\ \infty & \text{for } x \leq 0. \end{cases} \quad (7)$$

From this problem, we will develop the tools, both intuitive and mathematical, for dealing with infinite potential discontinuities, which we will then apply to the square-well problem.

The momentum eigenstates for this potential are not those of a free particle, but rather

$$\psi_p(x) = \begin{cases} (1/\sqrt{\pi\hbar})\sin[(p/\hbar)x] & \text{for } x > 0 \\ 0 & \text{for } x \leq 0. \end{cases} \quad (8)$$

This can be thought of as coming from the reflection coefficient of -1 and the transmission coefficient of 0, or equivalently from the application of the boundary conditions at $x = 0$. The propagator is now

 0002-9505/81/090843-05$00.50

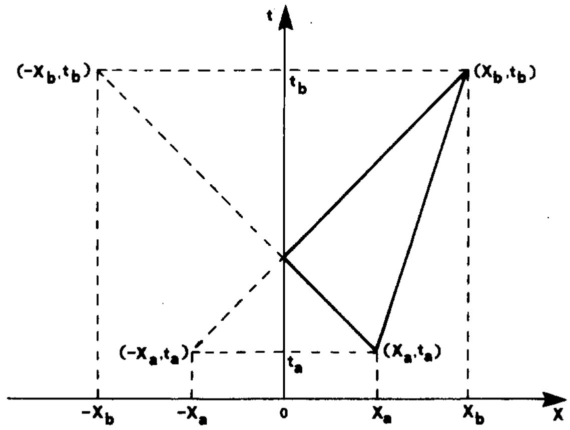

Fig. 1. Two classical paths connecting (x_a,t_a) and (x_b,t_b) and their corresponding image points.

$$\langle x_b,t_b|x_a,t_a\rangle = \left(\frac{m}{2\pi i\hbar(t_b-t_a)}\right)^{1/2} \times \left[\exp\left(\frac{im}{2\hbar(t_b-t_a)}(x_b-x_a)^2\right) - \exp\left(\frac{m}{2\hbar(t_b-t_a)}(-x_b-x_n)^2\right)\right] = \langle x_b,t_b|x_a,t_a\rangle_F - \langle -x_b,t_b|x_a,t_a\rangle_F, \quad (9)$$

where the subscript F denotes the propagator for a free particle.

This result has two explanations using path integral concepts. The first of these is not mathematically rigorous, relying more on physical intuition. Since the particle is essentially free for $x > 0$, it would seem that its behavior would be closely linked to that of a free particle. It is well known that for a free particle, or for any other particle whose Lagrangian is quadratic in position and velocity, the propagator has the form $Ne^{i/\hbar S_{\text{Cl}}}$, where N is a normalizing factor and S_{Cl} is the classical action, the integral of the Lagrangian along the classical path. For the potential barrier the key difference is that there are two classical paths (see Fig. 1). The first is that of a free particle, while the second is that of a particle that bounces off the wall on its way from (x_a,t_a) to (x_b,t_b). Geometrically, this second path can be constructed by first reflecting (x_b,t_b) about the line $x = 0$, then constructing the free-particle path $x(t)$ from (x_a,t_a) to this image point $(-x_b,t_b)$, and finally reflecting the path back to (x_b,t_b) so that $x(t) \geq 0$ everywhere.

Alternatively, we may perform these reflections with (x_a,t_a) rather than (x_b,t_b), and obtain exactly the same results. This corresponds to the fact that the second term of Eq. (4) may also be written $\langle x_b,t_b| -x_a,t_a\rangle_F$. For simplicity, we will consider from now on only reflections of the final point.

Applying the superposition principle, we would expect the propagator to be the sum of contributions from these two classical paths, which should be the free-particle propagators from (x_a,t_a) to (x_b,t_b) and $(-x_b,t_b)$, respectively, except perhaps for some effect of the barrier on the reflected path. Equation (9) confirms these expectations, and indicates that the effect of the barrier on the reflected path to multiply its propagator by the phase factor -1. This phase factor can be thought of as arising from the bound end reflection of the wave function at the barrier.

The above arguments provide a rather nice intuitive interpretation of the solution (4), but are open to serious objections from the point of view of the path integral formalism on which they rely. The fundamental objection is that in principle every allowed path between (x_a,t_a) and (x_b,t_b) must be counted in determining the propagator $\langle x_b,t_b|x_a,t_a\rangle$. There is no *a priori* reason for preferring the classical paths. Furthermore, the reflected path is not even allowed physically, since it intersects at one point the forbidden region $x \leq 0$.

There is a second, more rigorous path integral interpretation of Eq. (9) that takes care of these objections. Consider the first term on the left in this equation as an integral over paths. Clearly, there are many paths in this integral that enter the forbidden region $x \leq 0$. It will be shown that the effect of the second term is to cancel exactly the contribution from these forbidden paths. More explicitly, it will be shown that

$$\int_{x_a\ (F)}^{x_b} \exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b}\frac{m}{2}\dot{x}^2(t)\right)\mathcal{D}x(t) = \int_{x_a}^{x_b} \exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b}\frac{m}{2}\dot{x}^2(t)dt\right)\mathcal{D}x(t), \quad (10)$$

where the (F) on the first integral indicates that only the forbidden paths are to be counted. Equation (10) is equivalent to Eq. (4).

To derive Eq. (10), we need only demonstrate a one-to-one correspondence between forbidden paths from (x_a,t_a) to (x_b,t_b) and paths of equal free-particle action from (x_a,t_a) to $(-x_b,t_b)$. Consider an arbitrary forbidden path $x(t)$ from (x_a,t_a) to (x_b,t_b), and let $t_{\max}$ be the maximum time for which $x(t) < 0$. Clearly, $t_{\max}$ exists since $x(t)$ is continuous. Define a new path $\tilde{x}(t)$ (see Fig. 2)

$$\tilde{x}(t) = \begin{cases} x(t) & \text{for } t \leq t_{\max} \\ -x(t) & \text{for } t > t_{\max}. \end{cases} \quad (11)$$

This establishes a correspondence between the two classes of paths. Conversely, we could have started with $\tilde{x}(t)$ from (x_a,t_a) to $(-x_b,t_b)$ and defined $x(t)$. This means that the correspondence is one-to-one and onto. In addition, the free-particle actions along these two paths are equal, so the correspondence has the desired features. Thus from path integral considerations alone, we have derived the solution to the infinite potential barrier problem.

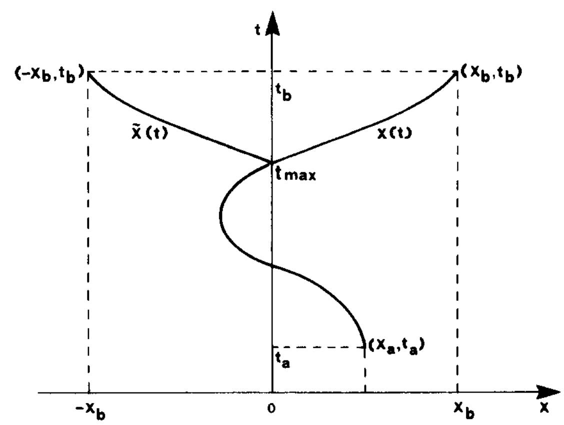

Fig. 2. Typical pair of cancelling paths $x(t)$, a forbidden path, and $\tilde{x}(t)$, its corresponding image path.

INFINITE SQUARE WELL

We now understand, more or less, the effect of an infinite potential discontinuity on the motion of a particle, and can apply this understanding to a problem with two such discontinuities, the infinite square well. The intuitive interpretation involving classical paths will serve as a guide for constructing an Ansatz solution to the square-well problem. This Ansatz is equivalent to the solution given by the Schrödinger equation, and can be justified by path cancellation techniques based on those used above for the potential barrier.

The key to the intuitive interpretation of the infinite potential barrier problem was the observation that there are two classical paths connecting the points (x_a,t_a) and (x_b,t_b). This is because the classical particle may go directly between the two points, or it may bounce off the wall once on the way. With the infinite square well, the particle may bounce back and forth an arbitrary number of times on its way from (x_a,t_a) to (x_b,t_b). This leads to an infinite number of classical paths connecting the two points. These paths correspond to an infinite sequence of images of the final point (see Fig. 3) having coordinates

$$x_r = \begin{cases} rL + x_b & \text{for } r \text{ even} \\ (r+1)L - x_b & \text{for } r \text{ odd,} \end{cases} \qquad (11')$$

where the subscript r is used to denote the number of reflections in the classical path corresponding to the image point, and to distinguish from the subscript n used for the energy eigenstates.

A generalization of the above arguments for the potential barrier would suggest that the propagator $\langle x_b,t_b|x_a,t_a\rangle$ should be the sum of contributions from each of these classical paths. The contribution from each classical path should be the free-particle propagator for its corresponding image point, multiplied by -1 for each time it is reflected. This leads to the Ansatz

$$\langle x_b,t_b|x_a,t_a\rangle = \sum_{r=-\infty}^{\infty} (-1)^r \langle x_r,t_b|x_a,t_a\rangle_F. \qquad (12)$$

Although different in form, this is exactly the solution given by the Schrödinger equation. To see this, we rewrite it

$$\begin{aligned}\langle x_b,t_b|x_a,t_a\rangle &= \sum_{r=-\infty}^{\infty} \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \\ &\times\left\{\exp\left[\frac{-i}{\hbar}\left(\frac{p^2}{2m}(t_b-t_a) - p(2rL + x_b - x_a)\right)\right]\right. \\ &\left. - \exp\left[\frac{-i}{\hbar}\left(\frac{p^2}{2m}(t_b-t_a) - p(2rL - x_b - x_a)\right)\right]\right\} \qquad (13) \\ &= \frac{-i}{2\pi\hbar}\int_{-\infty}^{\infty}\exp\left(\frac{-i}{\hbar}\frac{p^2}{2m}(t_b-t_a)\right)\exp\left(\frac{i}{\hbar}px_a\right) \\ &\times \sin\left(\frac{p}{\hbar}x_b\right)\left[\sum_{r=-\infty}^{\infty}\exp\left(\frac{-2irLp}{\hbar}\right)\right]dp.\end{aligned}$$

The sum is given by the Poisson summation rule

$$\sum_{r=-\infty}^{\infty}\exp\left(\frac{-2irLp}{\hbar}\right) = \frac{\pi\hbar}{L}\sum_{n=-\infty}^{\infty}\delta\left(p - \frac{\pi n\hbar}{L}\right). \qquad (14)$$

Substituting this into Eq. (13) simply picks out the values $p = \pi n\hbar/L$ and converts the integral to a sum

$$\begin{aligned}\langle x_b,t_b|x_a,t_a\rangle &= \frac{1}{iL}\sum_{n=-\infty}^{\infty}\exp\left(\frac{-i}{\hbar}E_n(t_b-t_a)\right) \\ &\times \exp(ik_nx_a)\sin(k_nx_b), \qquad (15)\end{aligned}$$

where $E_n = (1/2m)(\pi n\hbar/L)^2$ and $k_n = \pi n/L$. Noting that the term with $n = 0$ vanishes, we can combine the positive and negative terms to get

$$\begin{aligned}\langle x_b,t_b|x_a,t_a\rangle &= \frac{2}{L}\sum_{n=1}^{\infty}\exp\left(\frac{-i}{\hbar}E_n(t_b-t_a)\right) \\ &\times \sin(k_nx_a)\sin(k_nx_b). \qquad (16)\end{aligned}$$

This is the same as Eq. (4), which was derived from the Schrödinger equation.

PATH CANCELLATION

We have shown that the Ansatz solution to the square-well problem (12) agrees with the solution to Schrödinger equation (4). Of course, this is no mere coincidence, and the Ansatz can be justified by path cancellation arguments similar to these used for the potential barrier. Because there is an infinite number of image points, the arguments in this case will necessarily be more complicated, but their underlying content is the same. The effect of the image points is simply to cancel the contribution to the free-particle propagator $\langle x_b,t_b|x_a,t_n\rangle_F$ from paths that enter the forbidden regions $x \le 0$ and $x \ge L$. We write this symbolically

$$\sum_{r=-\infty}^{\infty}\int_{x_a(F)}^{x_r}\exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b}\frac{m}{2}\dot{x}^2(t)dt\right)Dx(t) = 0, \qquad (17)$$

where the (F) indicates that the integration is to be performed only over forbidden paths, and is included in each term to simplify notation. Note that this is equivalent to the Ansatz solution (12).

The path cancellation arguments are conceptually quite simple, though rather tedious when written out in detail. For this reason I present them first in outline. The method is to create a partition of all the paths considered in Eq. (17), such that the contribution at each equivalence class is zero. This will make the integral vanish, as desired. In order that the contribution of each equivalence class be zero, we impose two requirements. First, all paths in any one equivalence class must have the same free-particle action. Second, an equal number of these paths must go to even-numbered image points as to odd-numbered ones, so that an equal number will be counted positively as negatively.

Now we construct the equivalence class $[x(t)]$ of a path $x(t)$ from (x_a,t_a) to one of the image points (x_r,t_b). It is necessary to make a number of definitions, whose meaning

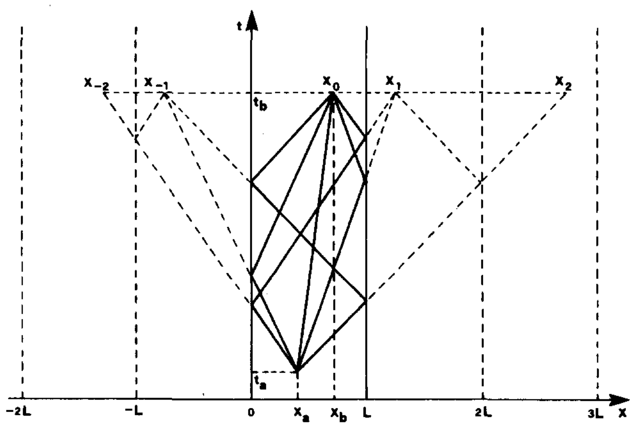

Fig. 3. Several of the classical paths connecting (x_a,t_a) and (x_b,t_b) and their corresponding image points.

is summed up in Fig. 4. First, let n_0 be the number of the first line $x = n_0 L$ intersected by $x(t)$. Let t_1 be the largest time when $x(t) = n_0 L$, before $x(t)$ terminates or intersects any other of these lines. Now we produced by induction, assuming that $n_0, \ldots, n_{i-1}$ and $t_1, \ldots, t_i$ have already been defined. Let n_i be the number of the next line $x = n_i L$ intersected by $x(t)$, if $x(t)$ intersects any more lines before terminating. Let t_{i+1} be the largest time when $x = n_i L$, before $x(t)$ terminates or intersects another line. When we have come to the last intersection of $x(t)$ with any of the lines $x = nL$, labeled n_{k-1} for definiteness, define $n_k = n_{k-1} + 1$ if $x(t_b) > n_{k-1}L$ and $n_k = n_{k-1} - 1$ if $x(t_b) < n_{k-1}L$. We are tacitly assuming k is finite, an assumption that will be discussed later.

Now we use these sequences $t_1, \ldots, t_k$ and $n_0, \ldots, n_k$ to define the 2^k elements of the equivalence class $[x(t)]$. First, we label $x_1(t) = x(t)$. Now we define $x_2(t)$, by reflecting $x_1(t)$ about the line $x = n_{k-1}L$ for $t > t_k$,

$$x_2(t) = \begin{cases} 2n_{k-1}L - x_1(t) & \text{for } t > t_k \\ x_1(t) & \text{for } t \leq 5k. \end{cases} \tag{18}$$

Here again we proceed by induction, this time assuming that the paths $x_1(t), \ldots, x_{2^i}(t)$ have already been defined. Define the paths $x_{2^i+1}(t), \ldots, x_{2^{i+1}}(t)$ by reflecting these paths about the line $x = n_{k-i-1}L$ for $t > t_{k-i}$

$$x_{2^i+j}(t) = \begin{cases} 2n_{k-i-1}L - x_j(t) & \text{for } t > t_{k-i} \\ x_j(t) & \text{for } t \leq t_{k-i}. \end{cases} \tag{19}$$

Repeating this process, we end up with 2^k paths $x_1(t), \ldots, x_{2^k}(t)$.

We now make several observations that will make it clear that these 2^k paths form an equivalence class with the desired properties. First of all, they do form an equivalence class; the property that two paths can be transformed into one another by this reflection process is an equivalence relation. Reflexivity, symmetry, and transitivity are all trivial consequences of the definitions.

Second, all these paths have the same free-particle action. This is because the free-particle Lagrangian depends only the square of the velocity, which is left unchanged except at a finite number of points, a set of measure zero.

Third, these 2^k paths are distinct. This can be seen by comparing their corresponding sequences $n_0, \ldots, n_k$, keeping track of how they transform under Eqs. (18) and (19). For our purposes, however, it suffices to observe that each of the last 2^{k-1} paths $x_{2^{k-1}+j}(t)$ is distinct from the path $x_j(t)$ from which it is derived, but this is obvious, since their endpoints do not coincide.

Fourth, and last, an equal number of these paths terminate at even-numbered as at odd-numbered image points. If $x_j(t)$ terminates at an even-numbered image point, then $x_{2^{k-1}+j}(t)$ terminates at an odd-numbered one, and vice versa. This is guaranteed by the minus sign in the top of Eq. (19). This, coupled with the above facts, is enough to demonstrate that the contribution of the equivalence class $[x(t)]$ to the integral (17) is zero, since even if some of the first 2^{k-1} paths were not distinct, they would in any case be counted equally positively and negatively.

These considerations make it clear that Eq. (17) is correct, provided that we can justify the assumption that the chosen path has finite action. This depends on the paths being somewhat well behaved, a sufficient condition being that they be piecewise continuously differentiable. In general, however, the paths are not at all well behaved. It can be shown for instance, that the contribution of paths with finite action to the integral (17) is zero (see Ref. 3, though the proof there is not quite correct). However, if we return to the definition of the integral (17) as a limit of integrals over piecewise linear paths,[4] then the above considerations are valid. Specifically, they are valid at each step in the limiting process, then they are valid in the limit.

REMARKS

This image-point solution to the infinite square well bears a striking resemblence to a result[5] for the free-particle propagator on certain types of manifolds. These manifolds are ones that are covered by flat Euclidean or Minkowski space-time. The similarity is that in both cases an image-point method used to deal with a situation in which there are a number of classical paths connecting any pair of points.

The square well is not a special case of this result, however. Rather, it appears to be prototype for the general problem of quantum mechanics on manifolds with boundary. To see this, we first observe that the infinite square well can be restated as problem of describing the motion of a free particle on the manifold $I = [0,1]$ with the usual topological structures inherited from R. This also appears to be a preferable way of stating the square-well problem. The boundary conditions in the Schrödinger picture, that the wave function be zero at the endpoints, are then a simple consequence of probability conservation, that the particle cannot pass through the boundary of its manifold.

The relationship between the square-well problem and the problem of manifolds covered by flat space-time is this: if we form the double[6] of the manifold I on which the particle moves, by connecting this manifold to a copy of itself and identifying the corresponding boundary points, then the resulting space is isomorphic to the circle S_1. The solution (12) can then be written in terms of the solutions for S_1 as

$$\langle x_b, t_b | x_a, t_a \rangle_I = \langle x_b, t_b | x_a, t_a \rangle S_1 - \langle x_b', t_b | x_a, t_a \rangle S_1, \tag{20}$$

where the x_b' in the second term indicates the image of x_b on the other half of the double manifold. Each of those terms is itself an infinite sequence of terms, the first term accounting for all of the positive terms of (12) and the second term taking care of all the negative terms.

This result appears to be a special case of a more general

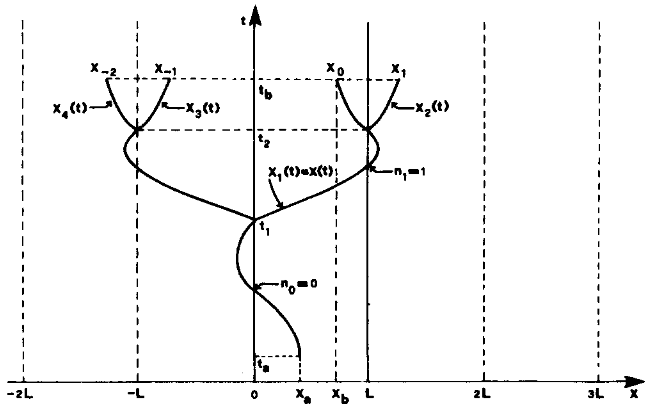

Fig. 4. Typical class of cancelling paths $x(t)$, a forbidden path, and its images $x_1(t)$, $x_2(t)$, $x_3(t)$, $x_4(t)$.

relation for a manifold with boundary M and its double $2M$:

$$\langle x_b,t_b|x_a,t_a\rangle_M = \langle x_b,t_b|x_a,t_a\rangle_{2M} - \langle x'_b,t_b|x_a,t_a\rangle_{2M}. \quad (21)$$

We observe immediately that the infinite potential barrier is also a special case of this relation with $M = [0,\infty)$, and $2M = R$ both with the usual structures. The topological problems involved in attempting to prove Eq. (21) are significant, and would seem to depend on finding a relationship between the fundamental group of the double manifold $2M$ and the topological structure of M as a manifold with boundary. This is a very difficult mathematical problem that is, to my knowledge, unsolved.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Namik Pak, with whom I had many useful discussions that cleared up several points of detail in this paper, and also Rai Choudury and David Storey. In addition, I wish to thank the Theoretical Physics Institute of the University of Trieste, especially R. Iengo, G. Furlan, and the secretary Mrs. L. Doria Vici, for making their facilities available to me.

[a]On leave of absence from Brown Univ., Providence, RI 02912.

[1]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1950).
[2]E. S. Abers and B. W. Lee, Phys. Rep. **9C**, 1 (1973).
[3]S. Coleman, The Uses of Instantons, 1977 Erice Lectures HUTP-78/A004 (unpublished).
[4]Reference 1, Chap. 2.
[5]J. S. Dowker, J. Phys. A **5**, 936 (1972).
[6]W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Academic, New York, 1975), p. 252; J. R. Munkres, *Elementary Differential Topology*, Ann. Math. Stud. 54, 2nd rev. ed. (Princeton University, Princeton, NJ, 1968), Chap. 1, Sec. 6.

Diamagnetism, gauge transformations, and sum rules

J. L. Friar
Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545

S. Fallieros
Physics Department, Brown University, Providence, Rhode Island 02912
(Received 7 July 1980; accepted 15 December 1980)

The dependence of conventional definitions of the atomic diamagnetic susceptibility on the gauge adopted for the description of the magnetic field is illustrated. The demonstration of the gauge invariance of the complete magnetic susceptibility leads to a discussion of the connection between gauge transformations and sum rules.

The conventional treatment[1] of atomic diamagnetism starts with the standard nonrelativistic Hamiltonian representing the interaction of a system with an external, static, uniform magnetic field $\mathbf{B}$, adopts the gauge $\mathbf{A}(\mathbf{r}) = (\mathbf{B} \times \mathbf{r})/2$ for the vector potential, and treating the interaction term proportional to $\mathbf{A}^2$ in first-order perturbation theory obtains for the diamagnetic susceptibility the result

$$\beta_d = -q^2\langle \mathbf{r}^2\rangle/6mc^2 = \chi_d/N_A,$$

where q is the charge of the particle, $\langle \mathbf{r}^2\rangle$ is the mean-square radius of the system, assumed to be spherically symmetric, N_A is Avogadro's number, and β_d and χ_d represent the atomic and molar susceptibilities, respectively. In other problems involving the interaction of particles with magnetic fields, e.g., in discussions of the Landau levels, it is often useful to consider the vector potential in a different gauge. If $\mathbf{B}$ is along the z axis we may, e.g., choose[2] $\mathbf{A}'(\mathbf{r}) = -By\hat{u}_x$ (where $\hat{u}_x$ is a unit vector along the x axis) without changing the value of the magnetic field $\mathbf{B} = \nabla \times \mathbf{A} = \nabla \times \mathbf{A}'$. Predictably, if we adopt this gauge and repeat the steps that led to the result for β_d in terms of the vector potential $\mathbf{A}(\mathbf{r})$, we find a different result! Since physical quantities must be gauge independent, it is clear that (a) the quantity β_d is not a physical observable and (b) the discussion of the magnetic susceptibility outlined above, is incomplete.

In this article we present a somewhat more complete discussion of the magnetic susceptibility and demonstrate its gauge invariance. This is quite straightforward and the conclusion is certainly not surprising. What is probably more instructive, is the method used in obtaining the results. It involves the application of a sum rule, similar but not identical to the one originally written down by Thomas, Reiche, and Kuhn.[3] The discussion presented below may, therefore, be considered also as an illustration of the intimate connection between gauge invariance and quantum-mechanical sum rules.

We shall consider for simplicity only a single spinless particle in a spherically symmetric potential $V(r)$. The Hamiltonian representing this system in interaction with a uniform static magnetic field is

$$H = H_0 + H_1 + H_2$$

with

$$H_0 = \mathbf{p}^2/2m + V(r),$$

 0002-9505/81/090847-03$00.50