tude $K(b,a)$ to go from a to b. This amplitude is the sum of contribution $\phi[x(t)]$ from each path.

$$K(b,a) = \sum_{\substack{\text{over all paths} \\ \text{from } a \text{ to } b}} \phi[x(t)] \tag{2-14}$$

The contribution of a path has a phase proportional to the action S:

$$\phi[x(t)] = \text{const } e^{(i/\hbar)S[x(t)]} \tag{2-15}$$

The action is that for the corresponding classical system [see Eq. (2-1)]. The constant will be chosen to normalize K conveniently, and it will be taken up later when we discuss more mathematically just what we mean in Eq. (2-14) by a sum over paths.

2-3 THE CLASSICAL LIMIT

Before we go on to making the mathematics more complete, we shall compare this quantum law with the classical rule. At first sight, from Eq. (2-15) all paths contribute equally, although their phases vary, so it is not clear how, in the classical limit, some particular path becomes most important. The classical approximation, however, corresponds to the case that the dimensions, masses, times, etc., are so large that S is enormous in relation to $\hbar$ ($= 1.05 \times 10^{-27}$ erg-sec). Then the phase of the contribution $S/\hbar$ is some very, very large angle. The real (or imaginary) part of ϕ is the cosine (or sine) of this angle. This is as likely to be plus as minus. Now if we move the path as shown in Fig. 2-1 by a small amount δx, *small on the classical scale*, the change in S is likewise small on the classical scale, but not when measured in the tiny unit $\hbar$. These small changes in path will, generally, make enormous changes in phase, and our cosine or sine will oscillate exceedingly rapidly between plus and minus values. The total contribution will then add to zero; for if one path makes a positive contribution, another infinitesimally close (on a classical scale) makes an equal negative contribution, so that no net contribution arises.

Therefore, no path really needs to be considered if the neighboring path has a different action; for the paths in the neighborhood cancel out the contribution. But for the special path $\bar{x}$, for which S is an extremum, a small change in path produces, in the first order at least, no change in S. All the contributions from the paths in this region are nearly in phase, at phase S_{cl}, and do not cancel out. Therefore, only for paths in the vicinity of $\bar{x}$ can we get important contributions, and in the classical limit we need only consider this particular trajec-

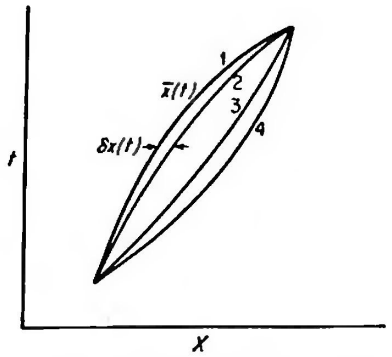

Fig. 2-1 The classical path 1, $\bar{x}(t)$, is that for which a certain integral, the action S, is minimum. If the path is varied by $\delta x(t)$, to path 2, the integral suffers no first-order change. This determines the equation of motion.

In quantum mechanics, the amplitude to go from a to b is the sum of amplitudes for each interfering alternative path. The amplitude for a given path, $e^{iS/\hbar}$, has a phase proportional to the action.

If the action is very large compared to $\hbar$, neighboring paths such as 3 and 4 have slightly different actions. Such paths will (because of the smallness of $\hbar$) have very different phases. Their contributions will cancel out. Only in the vicinity of the classical path $\bar{x}(t)$, where the action changes little when the path varies, will neighboring paths, such as 1 and 2, contribute in the same phase and constructively interfere. That is why the approximation of classical physics—that only the path $\bar{x}(t)$ need be considered—is valid when the action is very large compared to $\hbar$.

tory as being of importance. In this way the classical laws of motion arise from the quantum laws.

We may note that trajectories which differ from $\bar{x}$ contribute as long as the action is still within about $\hbar$ of S_{cl}. The classical trajectory is indefinite to this slight extent, and this rule serves as a measure of the limitations of the precision of the classically defined trajectory.

Next consider the dependence of the phase on the position of the end point (x_b,t_b). If we change the end point a little, this phase changes a great deal, and $K(b,a)$ changes very rapidly. If by a "smooth function" we mean one like S_{cl} which changes only when changes in argument which are appreciable on a classical scale are made, we note that $K(b,a)$ is far from smooth, but in this classical approximation our arguments show that it is of the form

$$K(b,a) = \text{"smooth function"} \cdot e^{(i/\hbar)S_{cl}} \tag{2-16}$$

All these approximate considerations apply to a situation on a scale for which classical physics might be expected to work ($S \gg \hbar$). But

at an atomic level, S may be comparable with $\hbar$, and then all trajectories must be added in Eq. (2-14) in detail. No particular trajectory is of overwhelming importance, and of course Eq. (2-16) is not necessarily a good approximation. To deal with such cases, we shall have to find out how to carry out such sums as are implied by Eq. (2-14).

2-4 THE SUM OVER PATHS

Analogy with the Riemann Integral. Although the qualitative idea of a sum of a contribution for each of the paths is clear, a more precise mathematical definition of such a sum must be given. The number of paths is a high order of infinity, and it is not evident what measure is to be given to the space of paths. It is our purpose in this section to give such a mathematical definition. This definition will be found rather cumbersome for actual calculation. In the succeeding chapters we shall describe other and more efficient methods of computing the sum over all paths. As for this section, it is hoped that the mathematical difficulty, or rather inelegance, will not distract the reader from the physical content of the ideas.

We can begin our understanding with a consideration of the ordinary Riemann integral. We could say, very roughly, that the area A, under a curve, is the sum of all its ordinates. Better, we could say that it is proportional to that sum. But to make the idea precise, we do this: take a subset of all ordinates (e.g., those spaced at equal intervals h). Adding these ordinates, we obtain

$$A \sim \sum_i f(x_i) \tag{2-17}$$

where the summation is carried out over the finite set of points x_i, as shown in Fig. 2-2.

The next step is to define A as the limit of this sum as the subset of points, and thus the subset of ordinates, becomes more complete or—because a finite set is never any measurable part of the infinite continuum—we may better say as the subset becomes more representative of the complete set. We can pass to the limit in an orderly manner by taking continually smaller and smaller values of h. In so doing, we would obtain a different sum for each value of h. No limit exists. In order to obtain a limit to this process, we must specify some normalizing factor which should depend on h. Of course, for the Riemann integral, this factor is just h itself. Now the limit exists

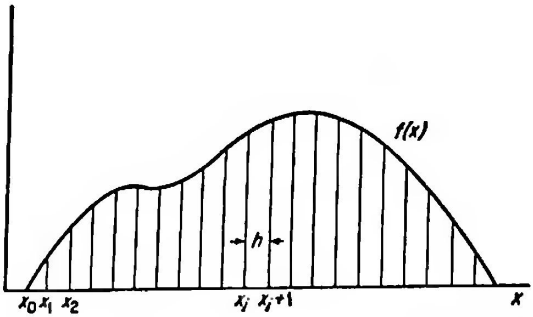

Fig. 2-2 In the definition of the ordinary Riemann integral, a set of ordinates is drawn from the abscissa to the curve. The ordinates are spaced a distance h apart. The integral (area between the curve and the abscissa) is approximated by h times the sum of the ordinates. This approximation approaches the correct value as h approaches zero.

An analogous definition can be used for path integrals. The measure which goes to zero in the limit process is the time interval ϵ between discrete points on the paths.

and we may write the expression

$$A = \lim \left[h \sum_i f(x_i) \right] \tag{2-18}$$

Constructing the Sum. We can follow through an analogous procedure in defining the sum over all paths. First, we choose a subset of all paths. To do this, we divide the independent variable time into steps of width ϵ. This gives us a set of values t_i spaced a distance ϵ apart between the values t_a and t_b. At each time t_i we select some special point x_i. We construct a path by connecting all the points so selected with straight lines. It is possible to define a sum over all paths constructed in this manner by taking a multiple integral over all values of x_i for i between 1 and $N - 1$, where

$$\begin{aligned} N\epsilon &= t_b - t_a \\ \epsilon &= t_{i+1} - t_i \\ t_0 &= t_a \qquad t_N = t_b \\ x_0 &= x_a \qquad x_N = x_b \end{aligned} \tag{2-19}$$

The resulting equation is

$$K(b,a) \sim \int\int \cdots \int \phi[x(t)]\, dx_1\, dx_2 \cdots dx_{N-1} \tag{2-20}$$

We do not integrate over x_0 or x_N because these are the fixed end points x_a and x_b. This equation corresponds formally to Eq. (2-17).

In the present case we can obtain a more representative sample of the complete set of all possible paths between a and b by making ϵ smaller. However, just as in the case of the Riemann integral, we cannot proceed to the limit of this process because the limit does not exist. Once again we must provide some normalizing factor which we expect will depend upon ϵ.

Unfortunately, to define such a normalizing factor seems to be a very difficult problem and we do not know how to do it in general terms. But we do know how to give the definition for all situations which so far seem to have practical value. For example, take the case where the lagrangian is given by Eq. (2-2). The normalizing factor turns out to be A^{-N}, where

$$A = \left(\frac{2\pi i\hbar\epsilon}{m}\right)^{1/2} \tag{2-21}$$

We shall see later (e.g., Sec. 4-1) how this result is obtained. With this factor the limit exists and we may write

$$K(b,a) = \lim_{\epsilon\to 0} \frac{1}{A} \int\int \cdots \int e^{(i/\hbar)S[b,a]} \frac{dx_1}{A} \frac{dx_2}{A} \cdots \frac{dx_{N-1}}{A} \tag{2-22}$$

where

$$S[b,a] = \int_{t_a}^{t_b} L(\dot{x},x,t)\, dt \tag{2-23}$$

is a line integral taken over the trajectory passing through the points x_i with straight sections in between, as in Fig. 2-3.

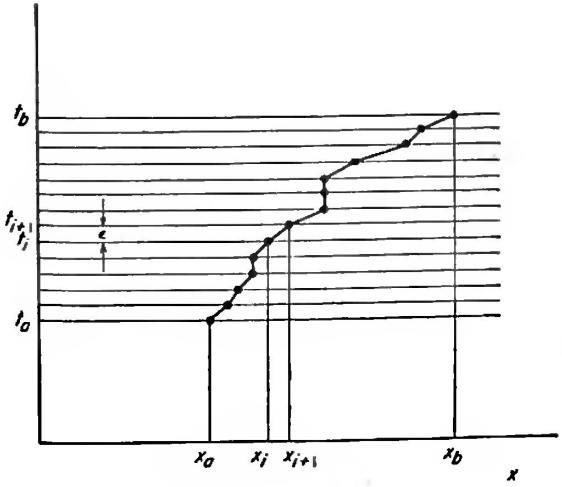

Fig. 2-3 The sum over paths is defined as a limit, in which at first the path is specified by giving only its coordinate x at a large number of specified times separated by very small intervals ϵ. The path sum is then an integral over all these specific coordinates. Then to achieve the correct measure, the limit is taken as ϵ approaches 0.

It is possible to define the path in a somewhat more elegant manner. Instead of straight lines between the points i and $i + 1$, we could use sections of the classical orbit. Then we could say that S is the minimum value of the integral of the lagrangian over all the paths which go through the specified points (x_i,t_i). With this definition no recourse is made to arbitrary straight lines.

The Path Integral. There are many ways to define a subset of all the paths between a and b. The particular definition we have used here may not be the best for some mathematical purposes. For example, suppose the lagrangian depends upon the acceleration of x. In the way we have constructed the path, the velocity is discontinuous at the various points (x_i,t_i); that is, the acceleration is infinite at these points. It is possible that this situation would lead to trouble. However, in the few such examples with which we have had experience the substitution

$$\ddot{x} = \frac{1}{\epsilon^2}(x_{i+1} - 2x_i + x_{i-1}) \tag{2-24}$$

has been adequate. There may be other cases where no such substitution is available or adequate, and the present definition of a sum over all paths is just too awkward to use. Such a situation arises in ordinary integration in which the Riemann definition, as in Eq. (2-18), is not adequate and recourse must be had to some other definition, such as that of the Lebesgue.

The necessity to redefine the method of integration does not destroy the concept of integration. So we feel that the possible awkwardness of the special definition of the sum over all paths [as given in Eq. (2-22)] may eventually require new definitions to be formulated. Nevertheless, the concept of the sum over all paths, like the concept of an ordinary integral, is independent of a special definition and valid in spite of the failure of such definitions. Thus we shall write the sum over all paths in a less restrictive notation as

$$K(b,a) = \int_a^b e^{(i/\hbar)S[b,a]}\,\mathfrak{D}x(t) \tag{2-25}$$

which we shall call a *path integral.* The identifying notation in this expression is the script $\mathfrak{D}$. Only rarely shall we return to the form given in Eq. (2-22).

Problem 2-6 The class of functionals for which path integrals can be defined is surprisingly varied. So far we have considered func-

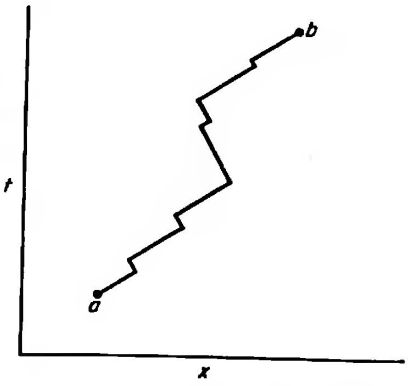

Fig. 2-4 The path of a relativistic particle traveling in two dimensions is a zigzag of straight segments. The slope of the segments is constant in magnitude and differs only in sign from zig to zag. The amplitude for a particular path, as well as the kernel to go from a to b, depends on the number of corners R along a path, as shown by Eqs. (2-26) and (2-27).

tionals such as that given in Eq. (2-15). Here we shall consider quite a different type. This latter type of functional arises in a one-dimensional relativistic problem. Suppose a particle moving in one dimension can go only forward or backward at the velocity of light. For convenience, we shall define the units such that the velocity of light, the mass of the particle, and Planck's constant are all unity. Then in the xt plane all trajectories shuttle back and forth with slopes of $\pm 45°$, as in Fig. 2-4. The amplitude for such a path can be defined as follows: Suppose time is divided into small equal steps of length ϵ. Suppose reversals of path direction can occur only at the boundaries of these steps, i.e., at $t = t_a + n\epsilon$, where n is an integer. For this relativistic problem the amplitude to go along such a path is different from the amplitude defined in Eq. (2-15). The correct definition for the present case is

$$\phi = (i\epsilon)^R \tag{2-26}$$

where R is the number of reversals, or corners, along the path.

As a problem, the reader may use this definition to calculate the kernel $K(b,a)$ by adding together the contribution for the paths of one corner, two corners, etc. Thus determine

$$K(b,a) = \sum_R N(R)(i\epsilon)^R \tag{2-27}$$

where $N(R)$ is the number of paths possible with R corners. It is best to calculate four separate K's, namely, the amplitude $K_{++}(b,a)$ of starting at the point a with a positive velocity and coming into the point b with a positive velocity, the amplitude $K_{+-}(b,a)$ of starting at the point a with a negative velocity and coming into the point b

with a positive velocity, and the amplitudes K_{-+} and K_{--} defined in a similar fashion.

Next suppose the unit of time is defined as $\hbar/mc^2$. If the time interval is very long ($t_b - t_a \gg \hbar/mc^2$) and the average velocity is small [$x_b - x_a \ll c(t_b - t_a)$], show that the resulting kernel is approximately the same as that for a free particle [given in Eq. (3-3)], except for a factor $\exp[(imc^2/\hbar)(t_a - t_b)]$. The definition given here for the amplitude, and the resulting kernel, is correct for a relativistic theory of a free particle moving in one dimension. The result is equivalent to the Dirac equation for that case.

2-5 EVENTS OCCURRING IN SUCCESSION

The Rule for Two Events. In this section we shall derive an important law for the composition of amplitudes for events which occur successively in time. Suppose t_c is some time between t_a and t_b. Then the action along any path between a and b can be written as

$$S[b,a] = S[b,c] + S[c,a] \tag{2-28}$$

This follows from the definition of the action as an integral in time and also from the fact that L does not depend on derivatives higher than the velocity. (Otherwise, we would have to specify values of velocity and perhaps higher derivatives at point c.) Using Eq. (2-25) to define the kernel, we can write

$$K(b,a) = \int e^{(i/\hbar)S[b,c]+(i/\hbar)S[c,a]}\, \mathfrak{D}x(t) \tag{2-29}$$

It is possible to split any path into two parts. The first part would have the end points x_a and $x_c = x(t_c)$, and the second part would have the end points x_c and x_b, as shown in Fig. 2-5. It is possible to integrate over all paths from a to c, then over all paths from c to b, and finally integrate the result over all possible values of x_c. In performing the first step of this integration $S[b,c]$ is constant. Thus the result can be written as

$$K(b,a) = \int_{x_c}\int_c^b e^{(i/\hbar)S[b,c]}K(c,a)\, \mathfrak{D}x(t)\, dx_c \tag{2-30}$$

where integrations must now be carried out not only over paths between c and b but also over the variable end point x_c. In the next step we carry out the integration over all paths between some point with an arbitrary x_c and the point b. All that is left is an integral