THE LAGRANGIAN IN QUANTUM MECHANICS.

By P. A. M. Dirac.

(Received November 19, 1932).

Quantum mechanics was built up on a foundation of analogy with the Hamiltonian theory of classical mechanics. This is because the classical notion of canonical coordinates and momenta was found to be one with a very simple quantum analogue, as a result of which the whole of the classical Hamiltonian theory, which is just a structure built up on this notion, could be taken over in all its details into quantum mechanics.

Now there is an alternative formulation for classical dynamics, provided by the Lagrangian. This requires one to work in terms of coordinates and velocities instead of coordinates and momenta. The two formulations are, of course, closely related, but there are reasons for believing that the Lagrangian one is the more fundamental.

In the first place the Lagrangian method allows one to collect together all the equations of motion and express them as the stationary property of a certain action function. (This action function is just the time-integral of the Lagrangian). There is no corresponding action principle in terms of the coordinates and momenta of the Hamiltonian theory. Secondly the Lagrangian method can easily be expressed relativistically, on account of the action function being a relativistic invariant; while the Hamiltonian method is essentially non-relativistic in form, since it marks out a particular time variable as the canonical conjugate of the Hamiltonian function.

For these reasons it would seem desirable to take up the question of what corresponds in the quantum theory to the Lagrangian method of the classical theory. A little consideration shows, however, that one cannot expect to be able

to take over the classical Lagrangian equations in any very direct way. These equations involve partial derivatives of the Lagrangian with respect to the coordinates and velocities and no meaning can be given to such derivatives in quantum mechanics. The only differentiation process that can be carried out with respect to the dynamical variables of quantum mechanics is that of forming Poisson brackets and this process leads to the Hamiltonian theory.[1]

We must therefore seek our quantum Lagrangian theory in an indirect way. We must try to take over the ideas of the classical Lagrangian theory, not the equations of the classical Lagrangian theory.

Contact Transformations.

Lagrangian theory is closely connected with the theory of contact transformations. We shall therefore begin with a discussion of the analogy between classical and quantum contact transformations. Let the two sets of variables be p_r, q_r and P_r, Q_r, $(r = 1, 2 \ldots n)$ and suppose the q's and Q's to be all independent, so that any function of the dynamical variables can be expressed in terms of them. It is well known that in the classical theory the transformation equations for this case can be put in the form

$$p_r = \frac{\partial S}{\partial q_r}, \quad P_r = -\frac{\partial S}{\partial Q_r}, \tag{1}$$

where S is some function of the q's and Q's.

[1] Processes for partial differentiation with respect to matrices have been given by Born, Heisenberg and Jordan (ZS. f. Physik **35**, 561, 1926) but these processes do not give us means of differentiation with respect to dynamical variables, since they are not independent of the representation chosen. As an example of the difficulties involved in differentiation with respect to quantum dynamical variables, consider the three components of an angular momentum, satisfying

$$m_x m_y - m_y m_x = ihm_z.$$

We have here m_z expressed explicitly as a function of m_x and m_y, but we can give no meaning to its partial derivative with respect to m_x or m_y.

In the quantum theory we may take a representation in which the q's are diagonal, and a second representation in which the Q's are diagonal. There will be a transformation function $(q'|Q')$ connecting the two representations. We shall now show that this transformation function is the quantum analogue of $e^{\ell S/h}$.

If α is any function of the dynamical variables in the quantum theory, it will have a „mixed" representative $(q'|\alpha|Q')$, which may be defined in terms of either of the usual representatives $(q'|\alpha|q'')$, $(Q'|\alpha|Q'')$ by

$$(q'|\alpha|Q') = \int (q'|\alpha|q'')\,dq''\,(q''|Q') = \int (q'|Q'')\,dQ''\,(Q''|\alpha|Q').$$

From the first of these definitions we obtain

$$(q'|q_r|Q') = q_r'(q'|Q') \tag{2}$$

$$(q'|p_r|Q') = -ih\frac{\partial}{\partial q_r'}(q'|Q') \tag{3}$$

and from the second

$$(q'|Q_r|Q') = Q_r'(q'|Q') \tag{4}$$

$$(q'|P_r|Q') = ih\frac{\partial}{\partial Q_r'}(q'|Q'). \tag{5}$$

Note the difference in sign in (3) and (5).

Equations (2) and (4) may be generalised as follows. Let $f(q)$ be any function of the q's and $g(Q)$ any function of the Q's. Then

$$\begin{aligned}(q'|f(q)g(Q)|Q') &= \iint (q'|f(q)|q'')\,dq''\,(q''|Q'')\,dQ''\,(Q''|g(Q)|Q') \\ &= f(q')g(Q')(q'|Q').\end{aligned}$$

Further, if $f_k(q)$ and $g_k(Q)$, $(k=1, 2\ldots, m)$ denote two sets of functions of the q's and Q's respectively,

$$(q'|\Sigma_k f_k(q)g_k(Q)|Q') = \Sigma_k f_k(q')g_k(Q')\cdot(q'|Q').$$

Thus if α is any function of the dynamical variables and we suppose it to be expressed as a function $\alpha(qQ)$ of the q's and Q's in a „well-ordered" way, that is, so that it consists of a sum of terms of the form $f(q)g(Q)$, we shall have

$$(q'|\alpha(qQ)|Q') = \alpha(q'Q')(q'|Q'). \tag{6}$$

This is a rather remarkable equation, giving us a connection between $\alpha(qQ)$, which is a function of operators, and $\alpha(q'Q')$, which is a function of numerical variables.

Let us apply this result for $\alpha = p_r$. Putting

$$(q' \mid Q') = e^{iU/\hbar}, \tag{7}$$

where U is a new function of the q''s and Q''s we get from (3)

$$(q' \mid p_r \mid Q') = \frac{\partial U(q'Q')}{\partial q'_r}(q' \mid Q').$$

By comparing this with (6) we obtain

$$p_r = \frac{\partial U(qQ)}{\partial q_r}$$

as an equation between operators or dynamical variables, which holds provided $\partial U/\partial q_r$ is well-ordered. Similarly, by applying the result (6) for $\alpha = P_r$ and using (5), we get

$$P_r = -\frac{\partial U(qQ)}{\partial Q_r},$$

provided $\partial U/\partial Q_r$ is well-ordered. These equations are of the same form as (1) and show that the U defined by (7) is the analogue of the classical function S, which is what we had to prove.

Incidentally, we have obtained another theorem at the same time, namely that equations (1) hold also in the quantum theory provided the right-hand sides are suitably interpreted, the variables being treated classically for the purpose of the differentiations and the derivatives being then well-ordered. This theorem has been previously proved by Jordan by a different method.[1]

The Lagrangian and the Action Principle.

The equations of motion of the classical theory cause the dynamical variables to vary in such a way that their values q_t, p_t at any time t are connected with their values q_T, p_T at any other time T by a contact transformation, which may be put into the form (1) with $q,\ p = q_t,\ p_t$; $Q,\ P = q_T,\ p_T$ and S equal to the time integral of the Lagrangian over the range

[1] Jordan, ZS. f. Phys. **38**, 513, 1926.

T to t. In the quantum theory the q_t, p_t will still be connected with the q_T, p_T by a contact transformation and there will be a transformation function $(q_t \mid q_T)$ connecting the two representations in which the q_t and the q_T are diagonal respectively. The work of the preceding section now shows that

$$(q_t \mid q_T) \text{ corresponds to } \exp\left[i \int_T^t L dt/h\right], \tag{8}$$

where L is the Lagrangian. If we take T to differ only infinitely little from t, we get the result

$$(q_{t+dt} \mid q_t) \text{ corresponds to } \exp\left[iL\, dt/h\right]. \tag{9}$$

The transformation functions in (8) and (9) are very fundamental things in the quantum theory and it is satisfactory to find that they have their classical analogues, expressible simply in terms of the Lagrangian. We have here the natural extension of the well-known result that the phase of the wave function corresponds to Hamilton's principle function in classical theory. The analogy (9) suggests that we ought to consider the classical Lagrangian, not as a function of the coordinates and velocities, but rather as a function of the coordinates at time t and the coordinates at time $t + dt$.

For simplicity in the further discussion in this section we shall take the case of a single degree of freedom, although the argument applies also to the general case. We shall use the notation

$$\exp\left[i \int_T^t L\, dt/h\right] = A\,(tT),$$

so that $A(tT)$ is the classical analogue of $(q_t \mid q_T)$.

Suppose we divide up the time interval $T \to t$ into a large number of small sections $T \to t_1$, $t_1 \to t_2$, ..., $t_{m-1} \to t_m$, $t_m \to t$ by the introduction of a sequence of intermediate times t_1, t_2, ... t_m. Then

$$A\,(tT) = A\,(tt_m)\, A\,(t_m t_{m-1}) \ldots A\,(t_2 t_1)\, A\,(t_1 T). \tag{10}$$

Now in the quantum theory we have

$$(q_t \mid q_T) = \int (q_t \mid q_m)\, dq_m (q_m \mid q_{m-1})\, dq_{m-1} \ldots (q_2 \mid q_1) dq_1 (q_1 \mid q_T), \tag{11}$$

where q_k denotes q at the intermediate time t_k, $(k = 1, 2 \ldots m)$. Equation (11) at first sight does not seem to correspond properly to equation (10), since on the right-hand side of (11) we must integrate after doing the multiplication while on the right-hand side of (10) there is no integration.

Let us examine this discrepancy by seeing what becomes of (11) when we regard t as extremely small. From the results (8) and (9) we see that the integrand in (11) must be of the form $e^{iF/h}$ where F is a function of q_T, q_1, $q_2 \ldots q_m$, q_t which remains finite as h tends to zero. Let us now picture one of the intermediate q's, say q_k, as varying continuously while the others are fixed. Owing to the smallness of h, we shall then in general have F/h varying extremely rapidly. This means that $e^{iF/h}$ will vary periodically with a very high frequency about the value zero, as a result of which its integral will be practically zero. The only important part in the domain of integration of q_k is thus that for which a comparatively large variation in q_k produces only a very small variation in F. This part is the neighbourhood of a point for which F is stationary with respect to small variations in q_k.

We can apply this argument to each of the variables of integration in the right-hand side of (11) and obtain the result that the only important part in the domain of integration is that for which F is stationary for small variations in all the intermediate q's. But, by applying (8) to each of the small time sections, we see that F has for its classical analogue

$$\int_{t_m}^{t} L\,dt + \int_{t_{m-1}}^{t_m} L\,dt + \ldots + \int_{t_1}^{t_2} L\,dt + \int_{T}^{t_1} L\,dt = \int_{T}^{t} L\,dt,$$

which is just the action function which classical mechanics requires to be stationary for small variations in all the intermediate q's. This shows the way in which equation (11) goes over into classical results when h becomes extremely small.

We now return to the general case when h is not small. We see that, for comparison with the quantum theory, equa-

tion (10) must be interpreted in the following way. Each of the quantities A must be considered as a function of the q's at the two times to which it refers. The right-hand side is then a function, not only of q_T and q_t, but also of q_1, q_2, ... q_m, and in order to get from it a function of q_T and q_t only, which we can equate to the left-hand side, we must substitute for q_1, q_2 ... q_m their values given by the action principle. This process of substitution for the intermediate q's then corresponds to the process of integration over all values of these q's in (11).

Equation (11) contains the quantum analogue of the action principle, as may be seen more explicitly from the following argument. From equation (11) we can extract the statement (a rather trivial one) that, if we take specified values for q_T and q_t, then the importance of our considering any set of values for the intermediate q's is determined by the importance of this set of values in the integration on the right-hand side of (11). If we now make h tend to zero, this statement goes over into the classical statement that, if we take specified values for q_T and q_t, then the importance of our considering any set of values for the intermediate q's is zero unless these values make the action function stationary. This statement is one way of formulating the classical action principle.

Application to Field Dynamics.

We may treat the problem of a vibrating medium in the classical theory by Lagrangian methods which form a natural generalisation of those for particles. We choose as our coordinates suitable field quantities or potentials. Each coordinate is then a function of the four space-time variables x, y, z, t, corresponding to the fact that in particle theory it is a function of just the one variable t. Thus the one independent variable t of particle theory is to be generalised to four independent variables x, y, z, t.[1]

[1] It is customary in field dynamics to regard the values of a field quantity for two different values of (x, y, z) but the same value of t as two different coordinates, instead of as two values of the same coordi-

We introduce at each point of space-time a Lagrangian density, which must be a function of the coordinates and their first derivatives with respect to x, y, z and t, corresponding to the Lagrangian in particle theory being a function of coordinates and velocities. The integral of the Lagrangian density over any (four-dimensional) region of space-time must then be stationary for all small variations of the coordinates inside the region, provided the coordinates on the boundary remain invariant.

It is now easy to see what the quantum analogue of all this must be. If S denotes the integral of the classical Lagrangian density over a particular region of space-time, we should expect there to be a quantum analogue of $e^{iS/h}$ corresponding to the $(q_t \mid q_T)$ of particle theory. This $(q_t \mid q_T)$ is a function of the values of the coordinates at the ends of the time interval to which it refers and so we should expect the quantum analogue of $e^{iS/h}$ to be a function (really a functional) of the values of the coordinates on the boundary of the space-time region. This quantum analogue will be a sort of „generalized transformation function". It cannot in general be interpreted, like $(q_t \mid q_T)$, as giving a transformation between one set of dynamical variables and another, but it is a four-dimensional generalization of $(q_t \mid q_T)$ in the following sense.

Corresponding to the composition law for $(q_t \mid q_T)$

$$(q_t \mid q_T) = \int (q_t \mid q_1)\, dq_1 (q_1 \mid q_T), \tag{12}$$

the generalized transformation function (g.t.f.) will have the following composition law. Take a given region of space-time and divide it up into two parts. Then the g.t.f. for the whole region will equal the product of the g.t.f.'s for the two parts, integrated over all values for the coordinates on the common boundary of the two parts.

Repeated application of (12) gives us (11) and repeated application of the corresponding law for g.t.f.'s will enable

nate for two different points in the domain of independent variables, and in this way to keep to the idea of a single independent variable t. This point of view is necessary for the Hamiltonian treatment, but for the Lagrangian treatment the point of view adopted in the text seems preferable on account of its greater space-time symmetry.

us in a similar way to connect the g.t.f. for any region with the g.t.f.'s for the very small sub-regions into which that region may be divided. This connection will contain the quantum analogue of the action principle applied to fields.

The square of the modulus of the transformation function $(q_t|q_T)$ can be interpreted as the probability of an observation of the coordinates at the later time t giving the result q_t for a state for which an observation of the coordinates at the earlier time T is certain to give the result q_T. A corresponding meaning for the square of the modulus of the g.t.f. will exist only when the g.t.f. refers to a region of space-time bounded by two separate (three-dimensional) surfaces, each extending to infinity in the space directions and lying entirely outside any light-cone having its vertex on the surface. The square of the modulus of the g. t. f. then gives the probability of the coordinates having specified values at all points on the later surface for a state for which they are given to have definite values at all points on the earlier surface. The g.t.f. may in this case be considered as a transformation function connecting the values of the coordinates and momenta on one of the surfaces with their values on the other.

We can alternatively consider $|(q_t|q_T)|^2$ as giving the relative a priori probability of any state yielding the results q_T and q_t when observations of the q's are made at time T and at time t (account being taken of the fact that the earlier observation will alter the state and affect the later observation). Correspondingly we can consider the square of the modulus of the g.t.f. for any space-time region as giving the relative a priori probability of specified results being obtained when observations are made of the coordinates at all points on the boundary. This interpretation is more general than the preceding one, since it does not require a restriction on the shape of the space-time region.

St John's College, Cambridge.