Origin of nonunitarity in quantum gravity

P. Hajicek
Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland
(Received 24 March 1986)

We study the mathematical structure of finite-dimensional parametrized systems. We find that the systems must possess a so-called global time, if they are to be reducible. We list all conditions on the topology of the Hamiltonian vector field, which follow from the existence of a global time. The topological obstructions to the existence of a global time are analogous to the Gribov effect. We consider the canonical quantization and show that each known method of constructing a unitary quantum theory is based on the use of a global time. Studying cosmological models we show how choices of wrong candidates for time, as well as extensions of the configuration spaces, lead to violations of unitarity. We give simple examples of configuration spaces in which there is no global time. The interpretation of these results is that we are quantizing in wrong coordinates. Indeed, there is a class of global times for each parametrized system. These times cannot be functions of the internal and external three-geometry and of the instantaneous states of all fields only. We give an example of the transformation from geometrodynamical variables to variables containing a global time; its inverse is a sort of covering map and the topology of the Hamiltonian vector field is changed.

I. INTRODUCTION

Recent proposals[1,2] that unitarity is to be abandoned, if one is going to quantize gravity, as well as the subsequent discussion,[3,4] show that the problem of unitarity acquires some new aspects in quantum gravity.

In the present paper, we will study this phenomenon for finite-dimensional models. We will see that the construction of a unitary quantum theory can be very difficult even for such simple systems—if only the time is mixed under the other canonical variables (the so-called parametrized systems) and the corresponding constraint is quadratic in momenta.

We will use the canonical quantization methods. First, these methods are particularly suitable for the study of unitarity. Second, in all cases we shall consider every other used method is equivalent to at least some part of some canonical quantization procedure.

We will consider the canonical reduction, and the Dirac method of quantization. Within the first, one chooses a coordinate t, say, for time, and solves the constraint, $\mathscr{H}=0$ for the conjugated momentum p_t:

$$p_t=H(t,q^k,p_k) .$$

H plays the role of the Hamiltonian. Thus, unitarity depends on whether or not H is real and can be turned into a self-adjoint operator (see, e.g., Ref. 5).

In the Dirac method, the constraint is turned into the so-called Wheeler-DeWitt equation

$$\hat{\mathscr{H}}\Psi=0 ,$$

for functions Ψ on the configuration space $\overline{\mathscr{C}}$. This equation is of the hyperbolic type and there can be families of corresponding Cauchy hypersurfaces foliating $\overline{\mathscr{C}}$. To obtain a unitary quantum theory, one has to construct a conserved and positive-definite bilinear form of initial data along such Cauchy surfaces (see, e.g., Ref. 6). The Cauchy surfaces can be considered as levels of some function on the configuration space—the time coordinate. Thus, the choice of a time coordinate plays an important role in both methods.

We find that—independently of which quantization method is used—the quantum theory becomes nonunitary, if either the configuration space $\overline{\mathscr{C}}$ is chosen too large or the time function on $\overline{\mathscr{C}}$ is chosen in an improper way.

The configuration space $\overline{\mathscr{C}}$ is too large if it contains points through which no classical trajectory passes. The time coordinate t on $\overline{\mathscr{C}}$ is improperly chosen if the inequality

$$v^\mu\partial_\mu t>0$$

does not hold at any point $q\in\overline{\mathscr{C}}$ and for any tangent vector v^μ to some classical trajectory through q. These structures, defined by classical dynamical equations are relevant to the question of unitarity.

The problem will remain open as to how these results generalize to systems of infinite-dimension field theories, with their necessary renormalization procedure. However, an already existing analysis[6] of a simplified model (cylindrical waves) suggests strongly that we are on the right track.

In any case, the results of the present paper throw new light on the analysis of the Hawking effect as presented in Refs. 7 and 8. There, the Berger-Chitre-Moncrief-Nutku (BCMN) model of gravity has been studied, some theorems about its classical solutions have been shown, and claims about the relevance of these classical properties for the properties of the corresponding quantum theory have been put forward. The attempts of Refs. 7 and 8 to justify these claims contain errors and gaps; the fundamental ideas of the present paper provide a more reliable basis for such a justification.

II. GENERAL FORM OF REDUCIBLE SYSTEMS

For an adequate description of reducible systems, we shall need the following notions: global phase time, dynamical cone, and global time. They make sense only for *parametrized* Hamiltonian systems.

Consider a system with the action

$$\bar{I}=\int d\tau(p_\mu \dot{q}^\mu-\alpha\mathscr{H}) . \tag{1}$$

Here, $\mu=0,\ldots,N$, $(q^0,\ldots,q^N)$ span the configuration space $\overline{\mathscr{C}}$ of the system, (q^μ,p_μ) are coordinates on the cotangent bundle $T^*(\overline{\mathscr{C}})$, the phase space, α is a Lagrange multiplier, and

$$\mathscr{H}=\mathscr{H}(q^\mu,p_\mu)$$

is a smooth function on $T^*(\overline{\mathscr{C}})$, the constraint.

For the sake of simplicity, we limit ourselves to (1) topologically trivial configuration spaces, $\overline{\mathscr{C}}\simeq\mathbb{R}^{N+1}$, and (2) to systems with just one constraint. The topology which will be relevant for our problem is that of the Hamiltonian vector field. A generalization to systems with more constraints is in preparation.

We will use the abbreviation X^A for the coordinates (q^μ,p_μ) on the phase space; so capital latin indices will run through $2N+2$ values. Tangent vectors to $T^*(\overline{\mathscr{C}})$ will be denoted by, e.g., $H^A,L^A,\ldots,$ the gradient by ∂_A, etc.

The constraint hypersurface $\bar{\Gamma}$ in $T^*(\overline{\mathscr{C}})$ is defined by

$$\mathscr{H}=0 ,$$

and the Hamiltonian vector field on $T^*(\overline{\mathscr{C}})$ is given by

$$H^A=\alpha(\Omega^{-1})^{AB}\partial_B\mathscr{H} ,$$

where Ω_{AB} is the symplectic form,

$$\Omega=dq^\mu\wedge dp_\mu .$$

The integral curves of the vector field H^A on $\bar{\Gamma}$ will be called classical trajectories. The field $(\Omega^{-1})^{AB}\partial_B\mathscr{H}$ gives an orientation to each classical trajectory; this orientation is important for the construction of the quantum theory and we will, therefore, require its invariance with respect to reparametrization. This can be achieved by the condition

$$\alpha>0 . \tag{2}$$

Definition 1. Let

$$T:\bar{\Gamma}\to\mathbb{R}^1$$

be a smooth real function on $\bar{\Gamma}$ with the property

$$\{T,\mathscr{H}\}\equiv(\Omega^{-1})^{AB}\partial_A T\partial_B\mathscr{H}>0 \tag{3}$$

everywhere on $\bar{\Gamma}$. Then, T is called a *global phase time.*

Clearly, the values of the function T along any classical trajectory can be used as an allowed (properly oriented) parametrization of the trajectory.

The property of being a global phase time is invariant with respect to canonical transformations of the form

$$Q^\mu=Q^\mu(q^\mu,p_\mu) ,$$

$$P_\mu=P_\mu(q^\mu,p_\mu) ,$$

because they do not change the function $\mathscr{H}$ on $T^*(\overline{\mathscr{C}})$ Ω_{AB} is an invariant tensor.

Definition 2. Let a global phase time T have the following property: if $\pi(X)=\pi(Y)$, then

$$T(X)=T(Y) .$$

In this case, there is a smooth function

$$t:\overline{\mathscr{C}}\to\mathbb{R}^1 ,$$

defined by

$$t(q)=T(X)$$

for any $X\in\pi^{-1}(q)\cap\bar{\Gamma}$ and any $q\in\overline{\mathscr{C}}$. t is called a *global time* on $\overline{\mathscr{C}}$. [π is the projection map $\pi:T^*(\overline{\mathscr{C}})\to\overline{\mathscr{C}}$.]

Definition 3. Let $q\in\overline{\mathscr{C}}$ and let

$$K(q)=\{v^\mu\in T_q(\overline{\mathscr{C}})\mid v^\mu=\pi_*(H^A(X)) \text{ for some } X\in\pi^{-1}(q)\cap\bar{\Gamma} \text{ and for some } \alpha>0\} . \tag{4}$$

$K(q)$ is called a *dynamical cone* at q.

Each element of $K(q)$ is a tangent vector to some classical trajectory through q as projected down to $\overline{\mathscr{C}}$. As $T(X)=t(\pi(X))$ must be a global phase time, we have immediately the following theorem.

Theorem 1. A smooth function t on $\overline{\mathscr{C}}$ is a global time, if and and only if

$$v^\mu\partial_\mu t>0 \tag{5}$$

for any $v^\mu\in K(q)$ and any $q\in\overline{\mathscr{C}}$.

Example 1. A free relativistic particle of mass m on the Minkowski spacetime with coordinates x^μ, $\mu=0,1,2,3$, and the metric $\eta^{\mu\nu}=\text{diag}(-1,1,1,1)$ has the following action:[6]

$$I=\int d\tau(p_\mu\dot{x}^\mu-\alpha\mathscr{H}) ,$$

where

$$\mathscr{H}=\eta^{\mu\nu}p_\mu p_\nu+m^2 .$$

$\overline{\mathscr{C}}$ is $\mathbb{R}^4$ spanned by x^μ. The constraint hypersurface $\bar{\Gamma}$ has two sheets $\bar{\Gamma}_\epsilon$, $\epsilon=\pm1$, where on $\bar{\Gamma}_\epsilon$,

$$p_0=\epsilon(m^2+p_kp_k)^{1/2} ,$$

$k=1,2,3$.

In the coordinates (x^μ,p_μ), the Hamiltonian vector field H^A has the components

$$H^A=2\alpha(-\epsilon(m^2+p_kp_k)^{1/2},+p_1,+p_2,+p_3 \ , \ 0,0,0,0) .$$

An example of a global phase time is

$$T=\epsilon x^0 .$$

The dynamical cone $K(x)$ at any point x is the interior of the light cone at x. $K(x)$ has the following property: if $v^\mu\in K(x)$, then $-v^\mu\in K(x)$. This, and theorem 1, imply together that there is no global time.

However, all physically interesting classical trajectories lie on $\bar{\Gamma}_-$. We can, therefore, supplement the constraint

$\mathscr{H}=0$ by the condition $p_0<0$. With $\overline{\Gamma}_-$ as constraint hypersurface, we obtain, for any $v^\mu \in K(x)$,

$$v^0=+2\alpha(m^2+p_kp_k)^{1/2} ,$$

$$v^k=2\alpha p_k ,$$

where (p_1,p_2,p_3) is some point of $\mathbb{R}^3$. The dynamical cone is, now, just the future half of the interior of the light cone, and an example of the global time is $t=x^0$ (in fact, any boost of this function will also do).

We observe that the dynamical cones define a causal structure on $\overline{\mathscr{C}}$: it is just that of the time-oriented Minkowski spacetime.

This analysis leads to the following proposal: let us call the configuration space and the phase space of any parametrized system *configuration spacetime* and *phase spacetime*.

We say that the parametrized system (1) is reducible, if its classical trajectories coincide with those of the following action:

$$I=\int dt(P_k\dot{Q}^k-H) , \tag{6}$$

where

$$H=H(Q^k,P_k,t)$$

and $k=1,\ldots,N$.

The action $\overline{I}$ results from "parametrization" of I; the parametrization consists of the following steps.

(a) Transformation of t to an arbitrary parameter τ along each classical trajectory preserving the orientation

$$\frac{dt}{d\tau}>0 .$$

Then,

$$I=\int d\tau\left[P_k\frac{dQ^k}{d\tau}-H\frac{dt}{d\tau}\right] .$$

(b) Setting $Q^0=t$, $P_0=-H$, and $\mu=0,1,\ldots,N$, the action can be written in the form

$$I=\int d\tau P_\mu \dot{Q}^\mu$$

subjected to the constraint

$$\mathscr{H}_1\equiv P_0+H(Q^0,\ldots,Q^N,P_1,\ldots,P_N)=0 . \tag{7}$$

Thus, a new action,

$$\overline{I}=\int d\tau(P_\mu\dot{Q}^\mu-\alpha\mathscr{H}) ,$$

has the same trajectories as I, if α is a Lagrange multiplier and

$$\mathscr{H}=\mathscr{F}\times\mathscr{H}_1 , \tag{8}$$

where $\mathscr{F}$ is an arbitrary smooth function on $T^*(\overline{\mathscr{C}})$ satisfying

$$\mathscr{F}\,|_{\mathscr{H}_1=0}>0 . \tag{9}$$

(One can allow for more general $\mathscr{F}$, but this one will be sufficient for our purposes.) Now, the constraint hypersurface $\overline{\Gamma}$ defined by $\mathscr{H}=0$ will, in general, contain $\overline{\Gamma}_1$ defined by

$$\mathscr{H}_1=0$$

as a proper subset; only $\overline{\Gamma}_1$ is physical.

(c) Performing a canonical transformation

$$Q^\mu=Q^\mu(q^\mu,p_\mu) ,$$
$$P_\mu=P_\mu(q^\mu,p_\mu) , \tag{10}$$

the action takes the form

$$\overline{I}=\int d\tau(p_\mu\dot{q}^\mu-\alpha\mathscr{H}) ,$$

where

$$\mathscr{H}(q^\mu,p_\mu)=\mathscr{H}(Q^\nu(q^\mu,p_\mu),P_\nu(q^\mu,p_\mu)) .$$

Lemma 1. Q^0 is a global phase time.

Proof. The Poisson brackets $\{Q^0,\mathscr{H}\}$ is invariant with respect to any canonical transformation, so we can calculate it in the coordinates Q^μ,P_μ using (7) and (8):

$$\{Q^0,\mathscr{H}\}=\frac{\partial\mathscr{H}}{\partial P_0}=\mathscr{F}+\frac{\partial\mathscr{F}}{\partial P_0}\mathscr{H}_1 .$$

Thus, on $\overline{\Gamma}_1$,

$$\{Q^0,\mathscr{H}\}=\mathscr{F}>0 .$$

Q.E.D.

This lemma shows that only a global phase time can be used as a time coordinate for reduction.

Theorem 2. Let the system (1) be reducible. Then, there is a global phase time $T(q^\mu,p_\mu)$. Moreover, there is a canonical transformation of the form (10) such that there is a global time on the new configuration spacetime spanned by Q^μ.

This theorem follows from the above construction.

We observe that the reducibility of the system (1) is a topological property of the Hamiltonian vector field H^A on $\overline{\Gamma}$. We have the following.

Theorem 3. The existence of a global phase time on $\overline{\Gamma}$ is equivalent to the existence of a manifold Γ and a diffeomorphism

$$h:\overline{\Gamma}\to\Gamma\times\mathbb{R}^1 ,$$

with the following property: every classical trajectory C on $\overline{\Gamma}$ determines a unique point $X\in\Gamma$ such that

$$h(\{C\})=X\times\mathbb{R}^1 .$$

Proof. Suppose that there is a global phase time T on $\overline{\Gamma}$. Then, H^A has no critical points on $\overline{\Gamma}$, and the gradient $\partial_A\mathscr{H}$ is nonzero on $\overline{\Gamma}$. This implies that $\overline{\Gamma}$ is a smooth hypersurface in $T^*(\overline{\mathscr{C}})$. The set of points in $\overline{\Gamma}$ determined by the equation

$$T=0$$

forms a hypersurface Γ in $\overline{\Gamma}$. Γ intersects each classical trajectory exactly once. Thus, given an arbitrary point

$X\in\bar{\Gamma}$, the unique classical trajectory C through X intersects Γ at a unique point, $Y(X)$. We define

$$h(X)=(Y(X),T(X)) .$$

From the construction, it follows that h is a differentiable map from $\bar{\Gamma}$ on $\Gamma\times\mathbb{R}^1$, and that it is one to one. Thus, we have the desired diffeomorphism.

Suppose that h with the above properties exists. We denote by π the natural projection:

$$\pi:\Gamma\times\mathbb{R}^1\to\mathbb{R}^1 .$$

Then,

$$T=\pi\circ h .$$

Q.E.D.

Theorem 3 implies that the critical points and cycles of the Hamiltonian vector field form obstructions to the existence of a global phase time. The absence of critical points and cycles is, however, not sufficient for the existence. Consider $\bar{\Gamma}$ of the form $\mathbb{R}^2-\{0\}$, with the Hamiltonian vector field

$$H^1=x^2 ,\quad H^2=x^1 ,$$

where x^1,x^2 are coordinates on $\mathbb{R}^2$. There are no cycles and no critical points. The classical trajectories on $\bar{\Gamma}$, however, do not even form a fibering (if $\bar{\Gamma}$ with the classical trajectories were a fiber bundle with a typical fiber $\mathbb{R}^1$, then the existence of a global phase time is guaranteed, see, e.g., Ref. 9).

For the parametrized systems, the existence of a global phase time is an analogue of the existence of a global gauge for gauge fields. The possible obstructions are of topological nature in both cases; for the gauge fields, the phenomenon is called the "Gribov effect."

III. RELATION TO UNITARITY

In the previous section, we studied some aspects of the behavior of classical trajectories. In the present one, we want to understand the implications of this behavior for the canonical quantization of the classical theory.

For this purpose, we will study a simple model which was dealt with many times in the literature: the so-called "quantum cosmology." We will use this system just as an abstract model of parametrized systems (so we are not going to make any claims about the creation, the Universe, etc.).

A. The classical model

Consider general relativity with a cosmological constant, and a massive scalar field coupled minimally to gravity. The action is

$$ {}^4I=\frac{1}{16\pi G}\int d^4x\sqrt{|g|}(\mathscr{R}-2\tilde{\Lambda}) + \frac{1}{4\pi}\int d^4x\sqrt{|g|}(g^{\mu\nu}\partial_\mu\tilde{\phi}\partial_\nu\tilde{\phi}+\tilde{m}^2\tilde{\phi}^2) .$$

We limit ourselves to spatially isotropic metrics and fields (the Robertson-Walker cosmology):

$$dt^2=\tilde{\alpha}^2d\tau^2-R^2(\tau)d\Sigma_k{}^2 ,$$

$$\tilde{\phi}=\tilde{\phi}(\tau) ,$$

where

$$d\Sigma_k{}^2=\frac{dr^2}{1-kr^2}+r^2d\theta^2+r^2\sin^2\theta\, d\varphi^2$$

and $k=\pm1$ or 0.

The dynamical equations for the remaining degrees of freedom can be obtained from the new action (see, e.g., Ref. 10)

$$I=\int d\tau(\dot{\phi}\pi_\phi+\dot{\Omega}\pi_\Omega-\alpha\mathscr{H}) , \tag{11}$$

where

$$\mathscr{H}=\tfrac{1}{4}e^{-3\Omega}(\pi_\phi{}^2-\pi_\Omega{}^2)+m^2e^{3\Omega}\phi^2-ke^{\Omega}+\Lambda e^{3\Omega} , \tag{12}$$

and where

$$\Omega=\ln\left[\left[\frac{3}{4\pi G}\right]^{1/2}R\right] ,$$

$$\alpha=\left[\frac{3}{4\pi G}\right]^{1/2}\tilde{\alpha} ,$$

$$\Lambda=\frac{4\pi G}{3}\tilde{\Lambda} ,$$

$$\phi=\left[\frac{4\pi G}{3}\right]^{1/2}\tilde{\phi} ,\quad m=\left[\frac{4\pi G}{3}\right]^{1/2}\tilde{m} .$$

The action (11) represents a typical parametrized system with a constraint $\mathscr{H}$ and a Lagrange multiplier α. The range of the coordinates ϕ and Ω is

$$-\infty<\Omega<\infty ,$$

$$-\infty<\phi<\infty ,$$

so our configuration spacetime $\bar{\mathscr{C}}$ is $\mathbb{R}^2$.

The constraint hypersurface $\bar{\Gamma}$ is given in $T^*(\bar{\mathscr{C}})$ by

$$\pi_\phi{}^2-\pi_\Omega{}^2=V(\Omega,\phi) , \tag{13}$$

where

$$V(\Omega,\phi)=4ke^{4\Omega}-4m^2e^{6\Omega}\phi^2-4\Lambda e^{6\Omega} . \tag{14}$$

$\bar{\Gamma}$ is (for general values of k, m, and Λ) a topologically nontrivial, three-dimensional algebraic surface with cusps and bifurcations.

By variation of the action (11), and using the constraint (13), we obtain the Hamiltonian vector field H^A on $\bar{\Gamma}$:

$$H^\Omega=\alpha\left[-\frac{e^{-3\Omega}}{2}\pi_\Omega\right] ,$$

$$H^\phi=\alpha\left[\frac{e^{-3\Omega}}{2}\pi_\phi\right] ,$$

$$H^{\pi_\Omega}=\alpha(4ke^{\Omega}-6m^2e^{3\Omega}\phi^2-6\Lambda e^{3\Omega}) ,$$

$$H^{\pi_\phi}=\alpha(-2m^2e^{3\Omega}\phi) .$$

The dynamical cone $K(\Omega,\phi)$ depends on the sign of $V(\Omega,\phi)$. $K(\Omega,\phi)$ consists of all vectors v^μ whose components, v^Ω and v^ϕ, satisfy the following relations: if $V(\Omega,\phi)<0$, then $(v^\Omega)^2-(v^\phi)^2>0$, if $V(\Omega,\phi)>0$, then $(v^\Omega)^2-(v^\phi)^2<0$, if $V(\Omega,\phi)=0$, then $v^\Omega=\pm v^\phi$.

B. The problem with noninvertibility of time

We come to our first problem, which, as far as I know, has not yet been noticed in the literature.

$K(\Omega,\phi)$ has the property that if $v^\mu \in K(\Omega,\phi)$, then $-v^\mu \in K(\Omega,\phi)$. This seems to be analogous to the situation encountered in example 1, but it is not. We illustrate this claim by another example.

Example 2. Let us set, in (12),

$$m=0\ ,\quad k=-1\ ,\quad \Lambda=0\ .$$

The constraint takes the form

$$\pi_\phi{}^2-\pi_\Omega{}^2+4e^{4\Omega}=0\ ;$$

the hypersurface $\overline{\Gamma}$ has two sheets, $\overline{\Gamma}_\epsilon$, $\epsilon=\pm 1$, and, on $\overline{\Gamma}_\epsilon$,

$$\pi_\Omega=\epsilon(\pi_\phi{}^2+4e^{4\Omega})^{1/2}\ . \tag{15}$$

If we want to reduce the theory in the configuration space $\overline{\mathscr{C}}$ spanned by Ω and ϕ, we must select one of the sheets $\overline{\Gamma}_+$ and $\overline{\Gamma}_-$ as "physical," or else there is no global time on $\overline{\mathscr{C}}$ (theorem 1). If we select $\overline{\Gamma}_\epsilon$, then $-\epsilon\Omega$ can be chosen as time. However, such a step means that we allow only those classical trajectories along which $-\epsilon\Omega$ is increasing. This, clearly, is only a half of all allowed classical trajectories, unlike the situation in example 1. The origin of the difficulty is that, by choosing some variable as a time coordinate, one definitely makes this variable noninvertible. Indeed, the time-inversion transformation in classical mechanics does not mean inversion of the time flow, but rather the inversion of the dependence of all other coordinates on time. It is, therefore, the following transformation:

$$q^k(t)\to q^k(-t)\ ,\quad t\to t\ .$$

(This is what represents "letting the film go backwards.") In our case,

$$\phi(\Omega)\to\phi(-\Omega)\ ,\quad \Omega\to\Omega\ .$$

The construction of a unitary quantum theory confirms this judgment.

(a) Reduction method (see Ref. 5). Within the choice $\overline{\Gamma}_\epsilon$, one obtains the Schrödinger equation

$$i\frac{\partial\psi}{\partial\Omega}=-\epsilon\left[-\frac{\partial^2}{\partial\phi^2}+4e^{4\Omega}\right]^{1/2}\psi$$

(the square root is defined with the help of the spectral theorem). The time inversion

$$\psi(\Omega,\phi)\to\psi^*(-\Omega,\phi)$$

does not change the sign of the Hamiltonian: $-\epsilon\Omega$ is always increasing.

A provisional mending of this defect is enabled by the possibility to construct both quantum mechanics: one for $\overline{\Gamma}_-$ and another for $\overline{\Gamma}_+$. The resulting pair of Hilbert spaces, each of which is equipped with its own Schrödinger equation, will reproduce the whole classical theory. A possible state of the system will be a pair of wave functions, one of the first Hilbert space, the second of the other one. The Hamiltonian operator will be a matrix

$$\begin{bmatrix}\left[-\dfrac{\partial^2}{\partial\phi^2}+4e^{4\Omega}\right]^{1/2} & 0\\ 0 & -\left[-\dfrac{\partial^2}{\partial\phi^2}+4e^{4\Omega}\right]^{1/2}\end{bmatrix}.$$

(b) Dirac method. The Wheeler-DeWitt equation reads

$$-\frac{\partial^2\Psi}{\partial\phi^2}+\frac{\partial^2\Psi}{\partial\Omega^2}+4e^{4\Omega}\Psi=0\ .$$

We can choose either only positive frequencies (with respect to Ω) or only negative frequencies to span the "physical space"

$$-i\frac{\partial}{\partial\Omega}<0\ ,\quad -i\frac{\partial}{\partial\Omega}>0\ .$$

The conserved scalar product, defined by

$$(\Psi_1,\Psi_2)=\frac{i}{2}\int_{\Omega=\text{const}}d\phi\left[\Psi_1^*\frac{\partial\Psi_2}{\partial\Omega}-\Psi_2\frac{\partial\Psi_1^*}{\partial\Omega}\right],$$

is positive definite on the first, and negative definite on the second subspace. The quantum-mechanical time inversion

$$\Psi(\Omega,\phi)\to\Psi^*(-\Omega,\phi)$$

does not change the sign of the probability current (Ψ,Ψ) through $\Omega=\text{const}$ hypersurfaces. Again, one can take both subspaces together.

C. The reduction method

In example 2, we were able to construct a unitary quantum theory. The reason was that (1) there was a global time (Ω or $-\Omega$, after restricting the constraint hypersurface to $\overline{\Gamma}_-$ or $\overline{\Gamma}_+$) and (2) we chose the global time as the time coordinate.

What happens if we choose a variable which is not a global time, as, for example, ϕ? The solution of the constraint with respect to the conjugated momentum, π_ϕ, reads

$$\pi_\phi=\pm(\pi_\Omega{}^2-4e^{4\Omega})^{1/2}\ .$$

The expression on the right-hand side differs, now, from that of (15) in the following property: it is *not* real for all allowed values of π_Ω and Ω. Hence, the Hamiltonian operator will not be self-adjoint (it can be made normal, so that the spectral theorem is, again, applicable). This behavior is typical for all known examples (see Ref. 5 for various choices of time and the resulting self-adjoint or not self-adjoint Hamiltonians).

There is one more "possibility" of how one can make the Hamiltonian imaginary. This is shown in the next example.

Example 3. $k=1$, $\Lambda>0$, $\phi\equiv 0$. This corresponds to the "minisuperspace model" of Ref. 2. The constraint reads

$$\pi_\Omega{}^2+4e^{4\Omega}-\Lambda e^{6\Omega}=0\ ,$$

and the "Hamiltonian" is given by

$$\pi_\Omega=\pm e^{3\Omega}(\Lambda-4e^{-2\Omega})^{1/2}\ .$$

The configuration spacetime $\mathscr{C}$ is spanned by Ω in the present case. For

$$\ln\frac{2}{\sqrt{\Lambda}}\le\Omega<\infty\ , \tag{16}$$

π_Ω is real. If we extend $\mathscr{C}$ beyond this boundary, then the Hamiltonian becomes imaginary. Relation (16) gives the so-called *natural size* of the configuration spacetime in the sense that there is at least one classical trajectory through each point.

An example of an analogous effect is given in Ref. 11. The system in Ref. 11 is not parametrized—it is just a gauge system with a quadratic constraint. The momentum which becomes imaginary is, therefore, not the Hamiltonian, but just some momentum (p_θ).

We can conclude that the classical dynamical equations of a given parametrized system determine the natural size of its configuration spacetime $\mathscr{C}$. Every point of $\mathscr{C}$ within this size defines a nonempty dynamical cone. The system of dynamical cones determines what is a global time on $\mathscr{C}$. If we choose $\mathscr{C}$ larger than its natural size, or if we choose, as a time coordinate, a function on $\mathscr{C}$ which is not a global time, we will obtain a nonunitary quantum theory.

If this is true, then the following question naturally arises: does a global time always exist? The next example gives a negative answer.

Example 4. We choose $k=1$, $m>0$, $\Lambda=0$. The constraint has the form

$$\pi_\phi{}^2-\pi_\Omega{}^2+4m^2\phi^2e^{6\Omega}-4e^{4\Omega}=0\ .$$

The properties of the constraint hypersurface can be summarized as follows.

(a) $\overline{\Gamma}$ is a smooth three-dimensional hypersurface in $T^*(\mathscr{C})\simeq\mathbb{R}^4$. This is true because H^A has no critical points on $\overline{\Gamma}$.

(b) The topology of $\overline{\Gamma}$ is $\mathbb{R}^2\times S^1$, i.e., $\overline{\Gamma}$ is connected, but not simply connected. This can be seen as follows. From the constraint, we have

$$\pi_\phi{}^2+4m^2e^{6\Omega}\phi^2=\pi_\Omega{}^2+4e^{4\Omega}\ .$$

If we fix the values of Ω and π_Ω in the range

$$-\infty<\Omega<\infty\ ,\quad -\infty<\pi_\Omega<\infty\ ,$$

then the possible values of ϕ and π_ϕ are given by

$$\phi=\frac{(\pi_\Omega{}^2+4e^{4\Omega})^{1/2}}{2me^{3\Omega}}\sin\omega\ ,$$

$$\pi_\phi=(\pi_\Omega{}^2+4e^{4\Omega})^{1/2}\cos\omega\ ,$$

for any $\omega\in[0,2\pi]$. Thus, we have $\overline{\Gamma}=\mathbb{R}^2\times S^1$.

(c) H^A has cycles on $\overline{\Gamma}$; each of the well-known cycles is incontractible on $\overline{\Gamma}$. This is well known (the results on classical trajectories for the present case are summarized in Ref. 12).

Conclusion: There is an obstruction to the existence of a global phase time (the Hamiltonian vector field has cycles). The system is not reducible; no continuous canonical transformation can help.

D. The Dirac method

The Dirac method for parametrized systems is based on the Wheeler-DeWitt equation,[6] which reads, for our model, as

$$-\frac{\partial^2\Psi}{\partial\Omega^2}+\frac{\partial^2\Psi}{\partial\phi^2}-4e^{6\Omega}(m^2\phi^2-ke^{-2\Omega}+\Lambda)\Psi=0\ . \tag{17}$$

It has the form of a two-dimensional Klein-Gordon equation in an external potential. It is a hyperbolic equation. The Cauchy hypersurfaces in $\mathscr{C}$ are either $\Omega=$const or $\phi=$const (this is a curiosity of two dimensions), or deformations of each of these two which preserve the signature.

Construction of a unitary quantum theory makes use of the conserved-current-valued bilinear form

$$j^\mu=\frac{i}{2}g^{\mu\nu}\sqrt{|g|}(\Psi_1^*\partial_\nu\Psi_2-\Psi_2\partial_\nu\Psi_1^*)\ ,$$

where $g_{\mu\nu}$ is the DeWitt metric:

$$g_{\Omega\Omega}=-1\ ,\quad g_{\phi\phi}=1\ ,\quad g_{\Omega\phi}=0\ .$$

We must integrate the current j^μ over a Cauchy surface and require (i) $j^\mu\to 0$ at the boundary of the Cauchy hypersurface (otherwise the integral need not be conserved), and (ii) the integral is a positive-definite bilinear form on some subspace of all solutions to Eq. (17). It must be a proper subspace, because the equation is real.

In general, we will have two difficulties: (1) the boundary properties of the semiclassical wave packets (which follow roughly the classical trajectories) can be in conflict with (i); (2) a division of the solutions to Eq. (17) into some positive- and negative-frequency functions required by (ii) can lead to wrong properties of the corresponding semiclassical wave packets.

Example 5. We choose $m=\Lambda=0$, $k=1$. This case has been studied in Ref. 13; we will use those results.

The Wheeler-DeWitt equation

$$-\frac{\partial^2\Psi}{\partial\Omega^2}+\frac{\partial^2\Psi}{\partial\phi^2}-4e^{4\Omega}\Psi=0$$

is separable, and we have

$$\Psi=A(\phi)B(\Omega)\ ,$$

$$A_\nu(\phi)=ae^{-i|\nu|\phi}+be^{i|\nu|\phi}\ ,$$

$$B_\nu(\Omega)=cI_{i|\nu/2|}(\tfrac{1}{2}e^{2\Omega})+dI_{-i|\nu/2|}(\tfrac{1}{2}e^{2\Omega})\ .$$

Here, ν is the separation constant,

$$-\infty<\nu<\infty\ ,$$

and $I_{i\kappa}(x)$ is the modified Bessel function of imaginary order.

The boundary conditions mentioned in (i) require that

$$B_\nu\to 0\quad\text{for }\Omega\to\infty\ .$$

This leads to $c=-d$. It follows immediately that $\Omega=\text{const}$ is not a good Cauchy surface. In Ref. 13, it is shown that

$$\frac{i}{2}\int_{\Omega=\text{const}} d\phi \left[\Psi^* \frac{\partial \Psi}{\partial \Omega} - \Psi \frac{\partial \Psi^*}{\partial \Omega}\right] = 0$$

for all Ψ of the form $\Psi = A_\nu(\phi) B_\nu(\Omega)$. However, the integral

$$\frac{i}{2}\int_{\phi=\text{const}} d\Omega \left[\Psi \frac{\partial \Psi}{\partial \phi} - \Psi \frac{\partial \Psi^*}{\partial \phi}\right]$$

is positive if $b=0$, and negative if $a=0$. This leads to the division of all solutions into the following subspaces: H_0^+, defined by the basis $\Psi_\nu^+ = e^{-i\nu\phi} B_\nu(\Omega)$, H_0^-, defined by the basis $\Psi_\nu^- = e^{i\nu\phi} B_\nu(\Omega)$, where $\nu > 0$.

Conclusion: With Ω as a time coordinate, no unitary quantum theory can result. With ϕ as a time coordinate, we can construct a unitary theory, if we restrict ourselves either just to positive or just to negative frequencies with respect to ϕ. In these cases, ϕ or $-\phi$ are global times (Ω can never be a global time).

We have studied several examples with a separable Wheeler-DeWitt equation, and we have always found similar results. If there is a global time t, then the choice of $t=\text{const}$ as Cauchy hypersurface does not violate any boundary conditions, and a unitary quantum theory can be constructed. If we choose a time coordinate which is not a global time, or if there is no global time, the construction is in trouble.

As far as I am aware, there is no general theory of Klein-Gordon equations of the above form. In the review paper by Seiler,[14] only test-function-like potentials are studied. In Ref. 15, potentials with Coulomb-type singularity and falloff are considered.

IV. A UNIVERSAL CLASS OF TIME COORDINATES

The results of the previous section can be interpreted in different ways. The most "revolutionary" view is that we are to abandon choices of time and unitarity altogether. The corresponding philosophy is described in Ref. 2.

Another possibility is that there are other methods leading to a unitary quantum theory which work without global time. Such methods are not known. Indeed, the covariant path-integral method (Lorentzian or Euclidean) does not yield a unitary theory. This has been shown in Ref. 16 for simple (finite-dimensional) models. The idea is as follows. The path integral formulation is based on integration over all four-geometries which interpolate between two fixed three-geometries, ${}^3g_1$ and ${}^3g_2$, say. The path integral yields a functional $J({}^3g_1, {}^3g_2)$ of the two three-geometries.

The result of Ref. 16 is that J is not a propagator of any unitary dynamics; rather, J is the kernel of the projector from the space of all functionals, $\Psi({}^3g)$, to the subspace of all solutions to Wheeler-DeWitt equation. Thus, from the knowledge of J, there is a long way to a Hilbert space of states and to a unitary evolution.

The third interpretation simply is that we are quantizing in wrong coordinates. Indeed, our coordinates Ω and ϕ are some invariants of the three-geometry (Ω), and state of the scalar field (ϕ). If there is a classical solution which oscillates periodically, Ω and ϕ will both be periodical functions of time, and the same holds for the corresponding momenta. Thus, even if the spacetime itself is not a causal loop, the classical trajectory in the coordinates $(\Omega, \phi, \pi_\Omega, \pi_\phi)$ is a cycle on $\overline{\Gamma}$. Similarly, if there is a static classical solution there will be a critical point on $\overline{\Gamma}$. Finally, the difficulty with the invertibility is clearly also due to the choice of some unworthy candidate for time: a true time cannot be invertible.

This interpretation sounds quite natural, but it has one striking consequence: a global time cannot be a function of internal or external three-geometry and of the instantaneous state of all fields only. Hence, the theories which use exclusively these kinds of quantities to describe parametrized systems (for example, the "superspace" geometrodynamics) need not lead, at least by well-known methods, to a unitary quantum theory.

If the variables used so far are not suitable for quantization, which variables, then, are to be chosen? Do they exist at all? How can a transformation of variables change the topology of the Hamiltonian vector field? We can answer all these questions, at least for the simple model considered in this paper.

First, only one variable is to be changed: the time coordinate. There is a very general class of times which will work and which is, moreover, quite obvious and a self-suggesting choice. The construction of such times works as follows.

Consider an arbitrary parametrized system. Any classical trajectory of it determines uniquely a spacetime in which the system is moving. This spacetime can be the same for all trajectories (as in example 1) or different from trajectory to trajectory (as in cosmology). Let us call this spacetime *trajectory spacetime.*

Each trajectory spacetime is globally hyperbolic by choice or by being a maximal solution corresponding to some initial data (see, e.g., Ref. 17). Each globally hyperbolic spacetime allows a foliation by a family of Cauchy hypersurfaces, and there is a smooth function t with a nonzero gradient which is constant along each hypersurface of the family.[18] There are many such families for each trajectory spacetime and many such functions for each family. Each such function will be called *trajectory time.*

Any trajectory time defines a parameter along the corresponding classical trajectory. Parametrize each classical trajectory on $\overline{\Gamma}$ by its trajectory time. If this can be done in a smooth way from trajectory to trajectory, we obtain a function t on $\overline{\Gamma}$ which is a global time.

To clarify these ideas, and to answer the rest of the questions, we describe a method giving an explicit transformation from the geometrodynamical variables to a global time. This method works only if the Hamilton-Jacobi equation of the considered system is separable.

Let the action of our system be (1). We define an auxiliary action for a system on the whole of $T^*(\overline{\mathscr{C}})$:

$$\overline{I}_1 = \int dt (p_\mu \dot{q}^\mu - \mathscr{H}) \ .$$

The corresponding Hamilton-Jacobi equation reads

$$\frac{\partial S}{\partial t}+\mathscr{H}\left[q^\mu,\frac{\partial S}{\partial q^\mu}\right]=0 .$$

Let a complete integral of this equation (depend on $N+1$ constants P_μ) be of the form

$$S(t,q^\mu)=S_0(q^\mu,P_\mu)-P_0 t .$$

Then,

$$\mathscr{H}\left[q^\mu,\frac{\partial S_0}{\partial q^\mu}\right]=P_0 . \tag{18}$$

A complete solution of the classical dynamical equations is given by

$$\frac{\partial S}{\partial P_\mu}=Q^\mu ,$$

where (Q^μ,P_μ) are canonically conjugated pairs of integrals of motion which span the whole $T^*(\overline{\mathscr{C}})$. The motion is physical only if the constraint is satisfied, or $P_0=0$, according to (18). The above equations take the form

$$\frac{\partial S_0}{\partial P_0}(q^\mu,P_\mu)=t+Q^0 , \quad \frac{\partial S_0}{\partial P_k}(q^\mu,P_\mu)=Q^k . \tag{19}$$

In the transformation equations, which consist of (19) and

$$\frac{\partial S_0}{\partial q^\mu}=p_\mu ,$$

the variable Q^0 appears only in the combination $t+Q^0$, so t and Q^0 are dependent: Q^0 is the integral of motion which specifies the origin of time t. We can write Q^0 for Q^0+t (or set $t=0$).

Thus, the new canonical variables, Q^μ,P_μ, contain a global time, namely, Q^0. The generating function of the transformation from (q^μ,p_μ) to $(P_\mu,-Q^\mu)$ is $S_0(q^\mu,P_\mu)$ (Ref. 19). Transforming further to (Q^μ,P_μ) and using (18), we obtain, for the action,

$$\bar{I}=\int dt(P_\mu\dot{Q}^\mu-\alpha P_0) .$$

If the Hamilton-Jacobi equation separates, then we have a solution of the form

$$S(t,q^\mu)=S_1(q^k,P_k)+S_0(q^0,P_0)-[E(P_k)+P_0]t ;$$

we can calculate S_0 and use it to transform just the pair (q^0,p_0) to (Q^0,P_0) as follows. The equations

$$\frac{\partial S_0}{\partial P_0}=t+Q^0 ,$$

$$\frac{\partial S_0}{\partial q_0}=p_0$$

imply

$$p_0dq^0=\frac{\partial S_0}{\partial q_0}dq^0+\frac{\partial S_0}{\partial P_0}dP_0-\frac{\partial S_0}{\partial P_0}dP_0$$
$$=dS_0-(t+Q^0)dP_0$$
$$=d(S_0-Q^0P_0-tP_0)+P_0d(t+Q^0) .$$

Setting $t=0$, we obtain

$$p_0dq^0=d(S_0-Q^0P_0)+P_0dQ^0 .$$

A canonical transformation of this class was performed in Ref. 13 for the cosmological model with $k=1$, $m=0$, and $\Lambda=0$. The procedure was slightly modified (instead of P_0, $P_0{}^2$ was used), it was performed for a different reason, and the authors of Ref. 13 did not explicitly mention that the defect of irreversibility of Ω and ϕ was cured by the transformation without any additional construction. No critical points or cycles were to be removed, however, so we would like to give still another example, in which the toplogy of the Hamiltonian vector field will be changed by the transformation.

Example 6. The simplest constraint which leads to a separable Hamilton-Jacobi equation on one hand, and to critical points and cycles on the other, is given by

$$\mathscr{H}=\tfrac{1}{2}(\pi_\phi{}^2-\pi_\Omega{}^2+\phi^2-\Omega^2)$$

(this is not a cosmology).

A complete solution to the classical dynamical equations is

$$\Omega=A\sin(t-t_0) ,$$

$$\phi=A\cos(t-t_1) ,$$

where $A\geq 0$, and t_0 and t_1 are arbitrary constants. All classical trajectories are cycles with the exception of $A=0$, which is a critical point with coordinates $\Omega=\phi=\pi_\Omega=\pi_\phi=0$.

The topology of $\overline{\Gamma}$ with the critical point removed is $(\mathbb{R}^1-\{0\})\times S^1$, and all cycles are incontractible within this part of $\overline{\Gamma}$.

The Hamilton-Jacobi equation separates; we set

$$S=S_1(\phi,P_1)+S_0(\Omega,P_0)-E(P_0,P_1)t$$

and obtain

$$S_0=-P_0\arcsin\frac{\Omega}{(-2P_0)^{1/2}}+\frac{\Omega}{2}(-2P_0-\Omega^2)^{1/2} .$$

This gives the following transformation of the pair Ω,π_Ω:

$$\Omega=(-2P_0)^{1/2}\sin Q^0 , \quad \pi_\Omega=(-2P_0)^{1/2}\cos Q^0 . \tag{20}$$

Thus, Q^0 is not a continuous function on the submanifold $\phi=$const, $\pi_\phi=$const of $\overline{\Gamma}$: each of these submanifolds is either a circle, for $\phi\neq 0$ or $\pi_\phi\neq 0$, or just a point, for $\phi=\pi_\phi=0$. Rather, (20) is a differentiable map from $\mathbb{R}^2$ spanned by (Q^0,P_0) to $\mathbb{R}^2$ spanned by (Ω,π_Ω), which maps the "straight lines" $P_0=$const to circles or points $\Omega^2+\pi_\Omega{}^2=$const, and which has no inverse.

The new action is

$$\bar{I}=\int dr\left[\pi_\phi\dot{\phi}+P_0\dot{Q}^0-\alpha\left[P_0+\frac{\pi_\phi{}^2+\phi^2}{2}\right]\right] ,$$

and the global time is Q^0; observe that the constraint becomes linear in P_0.

ACKNOWLEDGMENTS

Discussions with H. Leutwyler, S. Randjbar-Daemi, W. Simon, and P. Thomi are gratefully acknowledged. This work was supported in part by Schweizerischer Nationalfonds.

[1]S. W. Hawking, Commun. Math. Phys. **87**, 395 (1982).
[2]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
[3]D. J. Gross, Nucl. Phys. **B236**, 349 (1984).
[4]P. Hajicek, Report No. BUTP-85/22 (unpublished).
[5]W. F. Blyth and C. J. Isham, Phys. Rev. D **11**, 768 (1975).
[6]K. Kuchar, in *Relativity, Astrophysics, and Cosmology*, edited by W. Israel (Reidel, Dordrecht, 1973).
[7]P. Hajicek, Phys. Rev. D **31**, 785 (1985).
[8]P. Hajicek, Phys. Rev. D **31**, 2452 (1985).
[9]N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, NJ, 1951), p. 54ff, Theorem 12.2.
[10]M. A. H. MacCallum, in *Quantum Gravity, An Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. Sciama (Clarendon, Oxford, 1975).
[11]A. Ashtekar and G. T. Horowitz, Phys. Rev. D **26**, 3342 (1982).
[12]D. N. Page, Class. Quantum Gravit. **1**, 417 (1984).
[13]D. J. Kaup and A. P. Vitello, Phys. Rev. D **9**, 1648 (1974).
[14]R. Seiler, in *Invariant Wave Equations*, edited by G. Velo and A. S. Wightman (Springer, Berlin, 1978).
[15]R. Weder, Helv. Phys. Acta **50**, 105 (1977).
[16]P. Hajicek, Report No. BUTP-85/2 (unpublished).
[17]Y. Choquet-Bruhat and R. P. Geroch, Commun. Math. Phys. **14**, 329 (1969).
[18]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).
[19]L. D. Landau and E. M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960).