GENERAL COSMOLOGICAL SOLUTION OF THE GRAVITATIONAL EQUATIONS WITH A SINGULARITY IN TIME

I. M. Khalatnikov and E. M. Lifshitz
Landau Institute for Theoretical Physics, Institute for Physical Problems,
Academy of Sciences of the U.S.S.R., Moscow, U.S.S.R.
(Received 6 October 1969)

A way is indicated to construct a general solution of the Einstein equations with a singularity, starting from a previously known solution of a lesser degree of generality. A qualitative description is given of the evolution of the metric in this general solution towards the singularity, which is of a complex, oscillatory nature.

In this paper we return to a question which we discussed extensively before,[1] the question of whether there exists a general solution of the Einstein equations with a singularity. Recall that the statement of the problem consists of the search for the limiting form (in the neighborhood of the singularity) of the broadest class of solutions containing a physical singularity; the degree of generality is defined by the number of physically different arbitrary functions of the space coordinates contained in the solution. A general solution is one composed of a sufficient number of arbitrary functions to specify arbitrary initial conditions at a given moment of time (four for an empty universe, eight for the universe with matter). Note that what we have in mind is a physical singularity—infinite density of matter or (in an empty space) infinite curvature invariants.

Our previous investigations[1] have led us to the conclusion that the most general properties of the cosmological solutions with respect to their singularities manifest themselves already in the case of empty space, and that matter does not change these properties in a qualitative way. These investigations also provided a class of solutions which contain only one arbitrary function less than is necessary for the general case. These solutions represent a generalization of the homogeneous Kasner solution and have the form

$$ds = dt^2 - (a^2 l_\alpha l_\beta + b^2 m_\alpha m_\beta + c^2 n_\alpha n_\beta) dx^\alpha dx^\beta, \quad (1)$$

$$a = t^{p_1}, \quad b = t^{p_2}, \quad c = t^{p_3}, \quad (2)$$

with

$$p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1. \quad (3)$$

(One of the numbers p_1, p_2, p_3 is negative; let it be p_1.) The vectors $\vec{1}, \vec{m}, \vec{n}$ and the numbers p_1, p_2, p_3 are functions of the space coordinates. We introduce also the notation

$$\lambda = (\vec{1}\cdot\mathrm{curl}\,\vec{1})/v, \quad \mu = (\vec{m}\cdot\mathrm{curl}\,\vec{m})/v,$$
$$\nu = (\vec{n}\cdot\mathrm{curl}\,\vec{n})/v, \quad v = \vec{1}\cdot\vec{m}\times\vec{n}, \quad (4)$$

where each mathematical expression is constructed as though x^1, x^2, x^3 were Cartesian coordinates.

Apart from the "natural" conditions, imposed on the coordinate functions in (1) by the equations $R_0^{\ \alpha} = 0$, it turned out to be necessary to impose also an additional condition,

$$\lambda = 0, \quad (5)$$

for that one of the vectors $\vec{1}, \vec{m}, \vec{n}$ which is associated with the negative power of t. It is just this condition which leads to the "loss" of one arbitrary function from the solution.

A general solution is, by definition, completely stable: It does not change its character under application of any perturbation or, equivalently, under any change of initial conditions. But for the solution (1), (2), the presence of the restriction (5) causes an instability with respect to perturbations which violate this condition. Under the influence of such a perturbation the model must evolve into another regime which ipso facto will already be quite general. Of course in this process the perturbation ought not to be considered as small: The transition to the new regime lies outside the domain of infinitesimal perturbations.

The equations which determine the functions a, b, and c in the metric (1) are (cf. Ref. 1)

$$R_0^{\ 0} = (\ddot{a}/a) + (\ddot{b}/b) + (\ddot{c}/c) = 0; \quad (6)$$

$$R_l^{\ l} = [(\dot{a}bc)^\cdot/abc] + (\lambda^2 a^2/2b^2c^2) = 0,$$
$$R_m^{\ m} = [(a\dot{b}c)^\cdot/abc] - (\lambda^2 a^2/2b^2c^2) = 0,$$
$$R_n^{\ n} = [(ab\dot{c})^\cdot/abc] - (\lambda^2 a^2/2b^2c^2) = 0. \quad (7)$$

(The dots denote differentiation with respect to t.) Apart from the principal terms $\sim t^{-2}$ which lead to (2) with the conditions (3), terms of a still higher power in $1/t$ $[\sim t^{-2(1-2p_1)}]$ also are retained here. It is just to eliminate these latter terms that the additional constraint (5) had to be introduced. And it is just these terms, when switched on, which represent that perturbation whose influence is to be followed. [With $p_2 < 0$ or $p_3 < 0$, instead of the terms in λ^2, the analogous terms in μ^2 or ν^2 ought to be included in (7).] Of course in the process of evolution of the perturbation, the quantity λ becomes dependent on time, and hence additional terms appear in the Eqs. (6) and (7). But the evaluation of $\dot{\lambda}$ with the help of the equation $R_l^m = 0$ shows that these terms can be neglected.

By the substitution $a = e^\alpha$, $b = e^\beta$, $c = e^\gamma$, $dt = abcd\tau$, Eqs. (7) are reduced to

$$\alpha_{\tau\tau} = -(\tfrac{1}{2}\lambda^2)e^{4\alpha}, \quad \beta_{\tau\tau} = \gamma_{\tau\tau} = (\tfrac{1}{2}\lambda^2)e^{4\alpha}. \tag{8}$$

These simple equations must be solved with the initial conditions at $\tau \to \infty$: $\alpha_\tau = p_1$, $\beta_\tau = p_2$, $\gamma_\tau = p_3$ [the initial metric (1), (2)]. The solution shows that after a certain period of strong influence the perturbation is damped and we return (at $\tau \to -\infty$) to a metric of the form (1), (2), but with the new power labels

$$(p_1', p_2', p_3') = [1/(1+2p_1)] \times (-p_1, p_2 + 2p_1, p_3 + 2p_1). \tag{9}$$

If, initially, $p_1 < p_2 < p_3$, $p_1 < 0$, then $p_2' < p_1' < p_3'$, $p_2' < 0$. The negative power of t is transferred from the $\vec{l}$ to the $\vec{m}$ direction, the previously decreasing function b begins to increase, and the increasing function a begins to decrease.

It is convenient to present the substitution law (9) with the aid of the parametrization

$$p_{\mathrm{I}}(u) = -u/(1+u+u^2),$$
$$p_{\mathrm{II}}(u) = (1+u)/(1+u+u^2),$$
$$p_{\mathrm{III}}(u) = u(1+u)/(1+u+u^2), \tag{10}$$

where the parameter u assumes values in the region $u \geq 1$. If $u < 1$, then it can be reset into the region $u > 1$ by means of the relations

$$p_{\mathrm{I}}(1/u) = p_{\mathrm{I}}(u), \quad p_{\mathrm{II}}(1/u) = p_{\mathrm{III}}(u),$$
$$p_{\mathrm{III}}(1/u) = p_{\mathrm{II}}(u). \tag{11}$$

Now the rule (9) is formulated as follows: If $p_1 = p_{\mathrm{I}}(u)$, $p_2 = p_{\mathrm{II}}(u)$, and $p_3 = p_{\mathrm{III}}(u)$, then

$$p_1' = p_{\mathrm{II}}(u-1), \quad p_2' = p_{\mathrm{I}}(u-1),$$
$$p_3' = p_{\mathrm{III}}(u-1). \tag{12}$$

This process of the replacement of "Kasner-like epochs" with a bouncing of the negative power of time from one direction to another (mentioned already in Ref. 2) is the key to an understanding of the character of the evolution of the metric toward the singularity.

The successive replacements (12), with the negative power bouncing from a to b and back, continue until u becomes less than one. The value $u < 1$ is reset into $u > 1$, according to (11). The next series of replacements will bounce the negative power between c and a (or c and b), and so on. With an arbitrary (irrational) initial value of u, this process of replacements will continue indefinitely, and acquires a stochastic character. Let $w_n(x)$ be the probability for the nth series of u values to end with a value $u = x < 1$. It can be shown that when $n \to \infty$ it tends to a stationary distribution $w(x) = 1/\sqrt{2}\,(1+x)$, which does not depend on the initial conditions.[3] The length of each successive series is determined by the integer part of $1/x$, and the probability $W(k)$ for a series to be of a length k for large k is $W(k) \approx 1/\sqrt{2}\,k^2$. Hence the mean value of k diverges logarithmically, i.e., in successive series of replacements a large portion of the u values will be very large, which means (p_1, p_2, p_3) values near to $(0, 0, 1)$.

The qualitative meaning of these regularities in the replacements is the following: The evolution of the metric proceeds through successive periods (call them eras) which condense towards $t = 0$. During every era the spatial distances in two directions oscillate, and in the third direction decrease monotonically. In the transition from one era to the next the direction of monotonic decrease is bounced over to another axis. This monotonic decrease in successive eras proceeds, for long periods of time, according to a law which approximates closely to $\sim t$ [i.e., the metric labels are close to $(0, 0, 1)$]; but still, at the end of each era, the metric swerves away from this limit, without ever reaching it.

In the exact solution of the equations, the labels (p_1, p_2, p_3) lose their literal meaning, of course, and only the described qualitative properties persist. But there is a need for a special elucidation of the question of the approach of the solution to the metric with the $(0, 0, 1)$ labels, since the singularity in this metric in itself is not physical (the geometrical meaning of such a metric was explained in Ref. 1, and its analytical construction was given by Belinsky and Khalatnikov[4]). This question disturbed us for a long time in connection with the fear that some small effect in the equations might divert the evolution of the metric precisely into $(0, 0, 1)$ with a consequent disappearance of the singularity. It is to be added that with two of the numbers $(p_1, p_2,$

p_3) being small and thus close to each other, one may question the validity of those qualitative properties of the model at which we arrive by consideration of the perturbations, characterized by only one parameter.

However these doubts are removed by an analytical investigation of the simplest case of $\lambda = \mu = \nu = \text{const}$ (the homogeneous model); this case was considered by one of us jointly with Belinsky,[2] and also by Misner.[5] The quantity (say, c) which decreases monotonically during a long era becomes small in comparison with a and b. The analytical consideration of the above case confirms that such evolution cannot last indefinitely; the decreasing function c begins finally to increase, and the transition to the next era begins. It should be added that this property persists also in the general representation of the metric during an era (not subject to the assumption of $\lambda = \mu = \nu = \text{const}$), which has been constructed in an analytical form by Belinsky and Khalatnikov[6] and contains a complete required set of arbitrary functions of the space coordinates.

It should be mentioned that this latter representation contains functions which are periodic with respect to one coordinate. This suggests that there might be some sort of general connection between the existence of a singularity and closure of the universe (note that the particular model with $\lambda = \mu = \nu = \text{const}$ is closed). This question requires, of course, a special investigation.

The character of the singularity in this general solution opens new vistas for cosmological applications of the theory. Most important appears to be a property of the model pointed out by Misner[5]: the opening of the light horizons (which motivated the appropriate nickname of "Mixmaster universe").

Also, new light is cast on the problem of the gravitational collapse of nonspherical bodies. The final stage of such a collapse may well be a singularity of this same type.

It remains to make some remarks on the connection of the present results with our earlier work, which led us previously to infer that a singularity is absent from the general solutions.[1] We do this also with the aim of bringing to an end a prolonged discussion (see, e.g., Ref. 7) which over the last few years has become increasingly pointless.

Since there exists no systematic method for examining the singularities of the solutions of the Einstein equations, our search for increasingly more general solutions of this kind proceeded essentially by trial and error. A negative result from such a procedure could of course never be completely conclusive in itself; construction of a new solution with the required generality reverses the conclusion without affecting the results pertaining to the concrete solutions considered previously. A heuristic notion, which served as a guiding principle in our search, was the conviction that if the existence of a singularity is a general property of solutions, then there must exist indications of this based only on the most general properties of the Einstein equations themselves (although these indications may be insufficient by themselves to reveal the nature of the singularity). The only indication of this kind known at the time of our earlier work was the nullification of the metric determinant as a consequence of the equation $R_0^{\ 0} = 0$ in a synchronous[1] reference system. But this indication disappeared after it became clear that it was due purely to the geometrical intersection of the time-coordinate lines in the synchronous reference system. It was just this circumstance which we had in mind when we wrote that, by that very fact, the grounds for the existence of a singularity in a general solution essentially disappear. However the situation has changed since the discovery, by Penrose[8] (and later by Hawking[9] and Geroch[10]), of new theorems which reveal a connection between the existence of a singularity (of an unknown type) and some very general properties of the equations, which bear no relation to the choice of reference system.[11]

The new developments finally clarify the problem of singularities in general solutions and remove all previous contradictions.

---

[1] I. M. Khalatnikov, E. M. Lifshitz, and V. V. Sudakov, Phys. Rev. Letters 6, 311 (1961); E. M. Lifshitz and I. M. Khalatnikov, Advan. Phys. 12, 185 (1963) (we use here the notation of this paper).

[2] V. A. Belinsky and I. M. Khalatnikov, Zh. Eksperim. i Teor. Fiz 56, 1700 (1969).

[3] We are indebted to I. M. Lifshitz for examining this question.

[4] V. A. Belinsky and I. M. Kahalatnikov, Zh. Eksperim. i Teor. Fiz. 49, 1000 (1965) [translation: Soviet Phys.-JETP 22, 694 (1966)].

[5] C. W. Misner, Phys. Rev. Letters 22, 1071 (1969).

[6] V. A. Belinsky and I. M. Khalatnikov, to be published.

[7] C. W. Misner, Phys. Rev. 186, 1328 (1969).

[8] R. Penrose, Phys. Rev. Letters 14, 57 (1965).

[9] S. W. Hawking, Phys. Rev. Letters 15, 689 (1965).

See also S. W. Hawking and G. F. R. Ellis, Astrophys. J. 152, 25 (1968) and references therein.

[10]R. P. Geroch, Phys. Rev. Letters 17, 445 (1966).

[11]It was just the realization of this change of the situation that prompted one of us to omit the relevant section [L. O. Landau and E. M. Lifshitz, The Classical Theory of Fields (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962), 2nd ed., Sec. 110] from the latest Russian edition of the book, which appeared in 1967.

CONSEQUENCE OF HIGH-ENERGY NEUTRINO EXPERIMENT ON THE LEPTON NUMBER CONSERVATION LAW*

C. Y. Chang
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 17 November 1969)

Using the quoted μ^+-to-μ^- ratio obtained from the Berne-CERN-Fribourg neutrino experiment, the upper limit of C_M was estimated to be of the order of $5C_V$, where C_V is the universal vector-coupling constant of β decay, and C_M characterizes the coupling strength of a four-fermion weak interaction vertex only allowed by a multiplicative lepton-number conservation law.

As a result of the recent CERN high-energy neutrino experiment,[1] additional remarks can be made on the form of the lepton-number conservation law using the quoted μ^+-to-μ^- ratio. All present experimental evidence[2] is consistent with the existence of either (A) an additive conservation law of muonic and electronic lepton numbers, in which the sums $\sum L_\mu$ and $\sum L_e$ are separately conserved, or (B) a multiplicative conservation law[3,4] in which only the sum $\sum(L_\mu+L_e)$ and the sign $(-1)^{\Sigma L_\mu}$ are separately conserved. If the μ-e conservation law should take the multiplicative form, one would expect an overall production of two positive muons in the CERN experiment[1] via the double process

$$\pi^+ \text{ or } K^+ \to \mu^+ + \nu_\mu, \quad \nu_\mu + {}_ZA \to {}_ZA + \mu^+ + e^- + \nu_e, \tag{1}$$

where coherent scattering of the virtual charged leptons from the Coulomb field of a target nucleus, ${}_ZA$, dominates Reaction (1).

Process (1) can be described in the lowest approximation of the weak and electromagnetic interactions by the sum of the two diagrams (a) and (b) of Fig. 1. For a zero-spin target, the total cross section is given by[5]

$$d\sigma(\mu^+e^-) = \frac{F(q^2)}{q^4} Z^2e^2 \frac{T_{\mu\nu}(p+p')_\mu(p+p')_\nu}{[(kp)^2-k^2p^2]^{1/2}} \cdot \frac{d^3p'}{(2\pi)^3 2p_0'}, \tag{2}$$

where $F(q^2) \simeq F(0)$ is the nuclear form factor of the nucleus (with $q=p-p'$), and $T_{\mu\nu}$ is a tensor describing the upper vertices of Fig. 1. From the gauge-invariance requirements we obtain[5]

$$T_{\mu\nu} = a[(kq)\delta_{\mu\nu} + q^2 k_\mu k_\nu/(kq) - k_\mu q_\nu - k_\nu q_\mu] + b[q^2\delta_{\mu\nu} - q_\mu q_\nu], \tag{3}$$

where a and b are scalar functions.

It can be readily verified that as $q^2 \to 0$ and $(kq) \ll (kp)$ the quantity a reduces to $\sigma_{\rm ph}/(kq)$ and the b de-

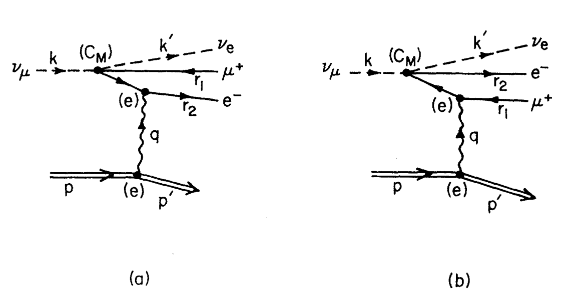

FIG. 1. The relevant Feynman diagrams of process (1). k, k', r_1, etc. are the four-momenta of the corresponding particles. (C_M) and (e) indicate the coupling strength of the corresponding vertex.