SINGULARITIES OF COSMOLOGICAL SOLUTIONS OF GRAVITATIONAL EQUATIONS. III

E. M. LIFSHITZ, V. V. SUDAKOV, and I. M. KHALATNIKOV

Institute of Physics Problems, Academy of Sciences, U.S.S.R.

Submitted to JETP editor January 25, 1961

A general geometric analysis is given of the situation that leads to the appearance of a time singularity in solutions of gravitational equations in a synchronous system of reference [a system satisfying the conditions (1)]. This analysis, together with the previous results, leads to the conclusion that such a singularity is absent in the general case of an arbitrary distribution of matter and gravitation field in space.

In our previous papers (cited below as I and II) the problem was posed of the investigation of the form of the cosmological solutions of the gravitational equations close to a time singularity; different types of such solutions were found. These results, together with the considerations set forth below, allow us to draw definite conclusions on the fundamental question—whether the existence of a time singularity is inevitable in cosmological models of the general theory of relativity.

1. GENERAL SOLUTION WITH FICTITIOUS SINGULARITY

As before, we shall use a reference frame that obeys the conditions

\[ g_{00} = -1, \quad g_{0a} = 0. \]  \hspace{1cm} (1)

The vanishing of the components \( g_{00} \) of the metric tensor is, as is well known, the necessary condition that guarantees the possibility of synchronization of clocks throughout all space (see, for example, reference 3, Sec. 84); therefore, the reference frame under consideration can be called synchronous. By virtue of the condition \( g_{00} = -1 \), the coordinate \( t \) is the world time in this case.

We have already mentioned in II the fact that the metric determinant \( g \) must necessarily vanish over a finite time duration in this reference frame because of one of the gravitational equations. To be precise, from the equation*

\[ T_{\alpha}^{\mu} = \frac{1}{2} \frac{\partial}{\partial t} x_{\alpha}^2 + \frac{1}{2} x_{\alpha}^2 x_{\beta}^2 = T_{00} - \frac{1}{2} T_{i i} \leq 0 \]  \hspace{1cm} (2)

(all the notation is the same as in I and II), and with the aid of the algebraic inequality

\[ x_{\alpha}^2 x_{\beta}^2 \geq \frac{1}{2} (x_{\beta}^2)^2 \]

we have

\[ \frac{\partial}{\partial t} x_{\alpha}^2 + \frac{1}{6} (x_{\alpha}^2)^2 \leq 0 \quad \text{or} \quad \frac{\partial}{\partial t} \frac{1}{x_{\alpha}^2} \geq \frac{1}{6}. \]  \hspace{1cm} (3)

For example, let \( \kappa_{\alpha}^\alpha > 0 \) at some instant of time. Then, upon decrease in \( t \), the quantity \( 1/\kappa_{\alpha}^\alpha \) falls off, having always a finite (non-vanishing) derivative; it therefore must vanish (from the positive side) over a finite time interval. In other words, \( \kappa_{\alpha}^\alpha \) goes to \( +\infty \) and, inasmuch as \( \kappa_{\alpha}^\alpha = \delta \ln (-g)/\delta t \) [see I, (1.8)], then this means that the determinant \( -g \) vanishes [but not more rapidly than \( t^6 \), from the inequality (3)]. If now \( \kappa_{\alpha}^\alpha < 0 \) at the initial moment, then this same result is obtained for increasing time.

We note that the presence of matter is not essential in this derivation: a zero on the right hand side of Eq. (2) (in the case of empty space) is sufficient to obtain the inequality (3).

However, this result still in no way proves the necessity of the existence in the metric of a real (physical) singularity that is not eliminated by a transformation of the reference frame. The singularity can be shown to be nonphysical, fictitious, associated simply with the character of the reference frame selected. The geometrical considerations given below show that this singularity, which is inescapable in the synchronous system, is actually fictitious in the general case.

In the synchronous reference frame, the time lines are geodesics in four-space. Actually, the four-vector \( u^4 = dx^4/ds \) tangent to the world line \( x^1, x^2, x^3, x^4 = \text{const} \) has the components \( u^0 = 0 \) and
u^0 = 1, and automatically satisfies the geodesic equations
\[ du^i/ds + \Gamma^i_{jk} u^j u^k = \Gamma^i_{00} = 0, \]
inasmuch as the Christoffel symbols \( \Gamma^i_{00}, \Gamma^i_{0j} \) in the conditions (1) are identically equal to zero. It is also easy to see that these lines are normal to the hypersurfaces \( t = \text{const} \). Actually, the four-vector normal to such a hypersurface \( n_1 = -\partial t/\partial x^1 \) has the covariant components \( n_\alpha = 0 \) and \( n_0 = -1 \). The corresponding contravariant components for the conditions (1) are \( n^\alpha = 0 \) and \( n^0 = 1 \), i.e., they are identical to the components of the four-vector \( u^1 \) tangent to the time lines.

Conversely, these properties can be used for the geometrical construction of a synchronous reference frame in an arbitrary space-time. For this purpose, we choose as a reference any spatial hyper-surface, i.e., a hypersurface the normal to which at each point has a time-like direction (lying within the light cone with vertex at this same point); all the interval elements on such a hypersurface are space-like. We then construct a family of geodesic lines normal to this hypersurface. If we now choose these lines as the coordinate time lines, and define the time coordinate \( t \) as the length of the geodesic line measured from the original hypersurface, we obtain a synchronous reference frame.

It is clear that such a construction, and thereby the choice of the synchronous reference frame, is always possible in principle.

But the geodesic lines of an arbitrary family generally intersect one another on certain enveloping hypersurfaces—the four-dimensional analogs of the caustic surfaces of geometrical optics. In other words, there is a geometrical reason for the appearance of the singularity, connected simply with the specific properties of the synchronous reference frame; obviously, therefore, it has no physical character.

An arbitrary metric of four-space also generally allows the existence of non-intersecting families of geodesics. On the other hand, the property of the curvature of the real space-time, which is expressed by the inequality \( R^i_{jk} \leq 0 \), shows that the metric allowed by the gravitational equations generally excludes the existence of such families, so that the time lines in any synchronous system of reference must necessarily intersect.*

*We naturally avoid the trivial exception—that of pencils of parallel lines in flat four-space. We note, however, that in the arbitrary selection of an initial hypersurface, the family of geodesic lines normal to it intersects also in the flat four-space. This circumstance reveals with particular clarity the fictitious character of the resultant singularity.

From the analytic point of view, this means that in the synchronous reference frame the equations of gravitation have a general solution with a fictitious time singularity. Such a solution (for empty space) must contain eight arbitrary functions of the three space coordinates: 1) four "physically different" functions that are necessary to specify the gravitational field at some initial moment (see II, Sec. 1), 2) a single function which determines the initial hypersurface in the geometric construction described above, 3) three functions associated with the fact that the conditions (1) still permit arbitrary transformations of the space coordinates not involving the time.

The character of the singularity in the metric is already clear from geometric considerations. Above all, the caustic hypersurface must be time-like, inasmuch as it in any event includes time-like intervals in itself—elements of length of the geodesics at points of their tangency with the caustic.*

Furthermore, one of the principal values of the metric tensor vanishes on the caustic, corresponding to the fact that the distance between two neighboring geodesics (which intersect each other at their point of tangency to the caustic) vanishes (the corresponding principal direction obviously lies along the normal to the caustic). This distance vanishes as the first power of the distance to the point of intersection. Therefore, the principal value of the metric tensor, and with it the entire determinant — \( g \), vanishes as the square of the distance just mentioned.

The method of analytic determination of the form of the general solution under study, close to a singularity, is described in the Appendix. One can make use of the arbitrariness in the selection of the spatial coordinates to write the first terms of the expansion of the metric in the neighborhood of the singularity in a form for which the spatial element of length \( d \) is given by the formula
\[ dl^2 = g_{ab} dx^a dx^b = a_{ab} dx^a dx^b + (t - \varphi)^2 a_{ss} dx^s + 2(t - \varphi)^3 a_{ss} dx^s dx^a. \]

Here the indices \( a, b \) run over the values 1, 2; the quantities \( a_{ab}, a_{ss} \) are functions of all three coordinates. This form still permits an arbitrary transformation \( x^3 = x'^3(x^1, x^2) \), which reduces to the transformation of the quantities \( a_{ab}, a_{ss} \), and the first terms of the expansion

*A hypersurface is called time-like when the normal to it lies outside the light cones. The elements of the interval in such a hypersurface can lie both inside and outside the light cones, i.e., they can be both time- and space-like.
of the components $g_{ab}$. For example, one can use them to cause the function $\varphi$, which gives the shape of the caustic hypersurface, to be $\varphi = x^3$. After all this, only the transformations of the two coordinates $x^1$ and $x^2$ into one another remain. Therefore, only five arbitrary functions should remain in the metric (three coordinates), so that the six functions $a_{ab}$, $a_{ab}$, $a_{23}$ must be related, as the consequence of the equations of gravitation, by a single equation.

The singularity in the metric (5) is not simultaneous—different space points achieve it at different moments of time. However, it is easy to see that one can always construct such a synchronous reference frame in which the singularity (fictitious) will be reached simultaneously throughout the entire space. It is clear that such a singularity cannot be located on the hypersurface tangent to the time lines at the points of their intersection, since the existence of time-like intervals at these points excludes beforehand the simultaneity of the singularity. Therefore, the time lines must intersect at a "manifold of points," which has a smaller number of dimensions than the hypersurface, i.e., which is a certain two-dimensional surface in four-space; it can be called the focal surface of the corresponding family of geodesics. By selecting an arbitrary focal surface, constructing all possible directions of normals to it from each of its points (all directions in the two-dimensional plane normal to the focal surface), and drawing geodesics in these directions, we at the same time construct a synchronous reference frame possessing the required property.

Thus the general solution of the gravitational equations can also be represented (by corresponding choice of a synchronous reference frame) in a form in which the singularity is simultaneous for the entire space. Of course, in such a form it contains the same four physically different arbitrary functions (three space coordinates) which suffice to specify the arbitrary initial distribution of the gravitational field. In comparison with the solution in the form (5), it contains one less arbitrary function: if we construct a synchronous reference frame, beginning with some initial hypersurface, then nowhere can an arbitrary hypersurface lead to the focusing of the geodesic lines constructed along the normals to it.*

As has already been mentioned, the fictitious nature of the singularity in the solutions considered is already obvious from its mode of construction; the singularity can be removed by transformation of the coordinates, but only at the price of foregoing synchronism of the reference frame. One can also establish directly the fact that, for example, the scalar component of the curvature tensor $R_{iklm}R^{iklm}$ has no singularity.

The introduction of matter does not change the qualitative character of the considered solutions, and the density of matter remains finite. This becomes evident, in particular, if we note that matter moves (in the synchronous reference frame) along world lines that do not coincide with the time lines, and are, generally speaking, not even geodesics.

Only the case of "dust-like" matter is an exception (the equation of state is $p = 0$). Such matter moves along geodesic lines. Therefore, the condition of synchronism of the reference frame in this case does not contradict the condition of its "accompaniment" of matter (which means that the matter moves along time lines), so that the reference frame can be chosen to be not only synchronous but at the same time accompanying (in the general case of an arbitrary equation of state, this is of course impossible). The density of matter then becomes infinite on the caustic simply as the result of the intersection of the trajectories of the particles. However, it is clear that this singularity of the density has no physical character, and is removed by the introduction of a conveniently small, but non-zero, value of the pressure of the matter.

2. GENERAL CONCLUSIONS

The geometric considerations set forth in the previous section essentially solve the problem of obtaining for the gravitational equations a general solution that has a singularity in the time—a solution the existence of which in the synchronous reference frame follows, as we have seen, from the inequality (3). However, the singularity in this solution is shown to be not physical but connected only with the specific properties of the reference frame employed.

*In a certain sense, this solution corresponds to a zero value of the function $\varphi$ in the solution of (5); in this case the square of the interval $ds^2 = dt^2 - dl^2$ reduces at the singularity ($t = 0$) to the quadratic form $ds^2 = -a_{ab}dx^adx^b$ of only two differentials. We emphasize, however, that it is not at all impossible to obtain the expansion of the metric in the neighborhood of such a singularity simply by setting $\varphi = 0$ in the formulas which refer to the solution of the form (5). We also point out that such a reference frame does not include all of space-time. This is clear from the fact that all the points of each hypersurface $t = \text{const}$ lie in it at the same time-distance from the spatial focal surface, i.e., these hypersurfaces as a whole are located in regions of absolute future or absolute past relative to the focal surface.
By the same token, any basis vanishes for the existence of a solution in addition to this one which would possess a real singularity and would also be general. In fact, an investigation, carried out by two of us, of the possible forms of such singularities has shown that the broadest solution with a real singularity is the solution found in II, with the metric

\[ g_{ab} \approx t^{2m_1 \lambda_1} + t^{2m_2 \lambda_2}m_5 + t^{2p_1 \mu_1} \,

p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1. \tag{6} \]

However, it contains one less arbitrary function of the coordinates then would be required for the general solution. Therefore, this solution also is only a special case, in spite of its broadness.*

Such a solution is unstable; there exists a type of small perturbations whose action destroys the state described by this solution. Inasmuch as the singularity in the synchronous system cannot vanish in general, this means that it goes over into a fictitious singularity under the action of the perturbation. If in the transformation process one sees to it that the singularity is simultaneous (which is always possible), then the process must terminate in a transition to the solution described in Sec. 1, with a simultaneous singularity for all space.

It is interesting to note that a solution of the form (6) exists even in the absence of matter, i.e., for empty space (and it then contains three physically different arbitrary functions—see II). The geometric constructions given in Sec. 1 also do not depend on the presence or absence of matter. All this supports the idea that the most general properties of cosmological solutions relative to the singularities in time appear already in the case of empty space, and matter does not change these properties in any qualitative fashion. This result is natural if we note that the gravitational properties of a set of short-wave gravitational waves chosen in appropriate fashion can imitate the gravitational properties of matter (with an equation of state \( p = \epsilon/3 \)). The isotropic (Friedmann) solution occupies an exceptional position in this sense, as does its generalization, considered in I; these solutions exist only for a space filled with matter. However, this exception is connected with the high symmetry (homogeneity)

---

*The systematic construction of all possible types of solutions with real singularities will be given in a subsequent paper. We take this opportunity to correct two errors which appear in II, in the formulas referring to the solution of (6). In Eq. (2.16) for \( P'_{1}, P'_{8} \) and \( P'_{n} \) the factor \( t^{-2p} \) is omitted. On the right hand side of Eq. (3.6) of II, \( u_{0}^{(5)} \) should appear instead of \( u_{n}^{(5)} \).

of the distribution of matter peculiar to this solution, a distribution which cannot be realized in the imitation shown.

All the foregoing leads to the fundamental conclusion that the presence of a time singularity is not an essential property of the cosmological model of the general theory of relativity, and the general case of an arbitrary distribution of matter and gravitational field does not lead to the appearance of such a singularity.

We have referred constantly (here and in I, II) to the direction of approach to the singularity as being one wherein the time is decreased. In actuality, in view of the symmetry of the gravitational equations relative to time reversal, one could speak with equal success of the approach to the singularity in the direction of increasing time.

Physically, however, in view of the physical non-equivalence of future and past, there is an essential difference between the two cases relative to the very statement of the problem. A singularity in the future can have physical meaning only if it is attainable under completely arbitrary conditions given at some previous instant of time; it is clear that there is no basis for maintaining that the distribution of matter and field, attained at some given moment in the process of evolution of the universe, corresponds to the specific conditions required for the realization of some particular solution of the equations of gravitation which possess a real singularity. Furthermore, if such a distribution is attained for any reason at some instant of time it will be inevitably destroyed subsequently, at least by the inevitable thermodynamic (or quantum) fluctuations. Therefore, the results that have been set forth exclude the possibility of the existence of a singularity in the future, and mean that the contraction of the world (if it should occur in general) should in final analysis alternate with its expansion. As to the past, a consideration based only on gravitational equations can only superimpose definite limitations on the attainable forms of the initial conditions, the full explanation of the character of which is impossible on the basis of the existing theory.

Finally, let us make one remark. All the previous study has been based on Einstein's general equations of gravitation in the form in which they follow logically from the general theory of relativity. At the present time there exist no astronomical or theoretical grounds for introducing an additional "cosmological term" in these equations. This applies even more strongly to the completely arbitrary and groundless changes introduced into the gravitational equation by F. Hoyle.
In conclusion, we express our sincere gratitude to L. D. Landau for numerous stimulating discussions. We also thank L. P. Pitaevskii for discussion of a number of questions.

APPENDIX

ANALYTIC CONSTRUCTION OF A GENERAL SOLUTION WITH FICTITIOUS SINGULARITY

By choosing the space coordinates in the fashion shown in the text (we designate them as $x^1 = x$, $x^2 = y$, $x^3 = z$), we seek the first terms of the expansion of the components of the metric tensor in the form

$$g_{ab} = a_{ab} + \tau b_{ab}, \quad g_{aa} = \tau^2 (a_{aa} + \tau b_{aa}), \quad g_{aa} = \tau^4 (a_{aa} + \tau b_{aa}).$$

where $\tau = t - z$, and the indices, which are denoted by the Latin letters (a, b, c, ...), take on the values 1 and 2 everywhere. All the coefficients of the expansion are functions of $x$, $y$, and $t$ (for reasons which are made clear below, such an expansion is more convenient than one in $x$, $y$, and $t$). The components of the corresponding contravariant tensor are:

$$g^{ab} = a^{ab} - \tau b^{ab}, \quad g^{aa} = - a^{aa}, \quad g^{aa} = \tau^2 (a^{aa} - \tau b^{aa}).$$

Here $a^{ab}$ is a two-dimensional tensor inverse to $a_{ab}$ and $a^{33} = \frac{1}{a_{33}}$. All the operations of raising and lowering of indices for the other quantities are carried out with the help of $a_{ab}$ and $a_{33}$; thus,

$$b^{ab} = a^{ac} a^{bd} b_{cd}, \quad a^{aa} = a^{ab} a^{33} a_{33} \quad \text{etc.}$$

The components of the tensor $\kappa_{\alpha\beta}$ are computed, by definition, as

$$\kappa_{\alpha\beta} = \dot{g}_{\alpha\beta}, \quad \kappa^{\alpha\beta} = \ddot{g}^{\alpha\beta} \kappa_{\gamma\gamma}, \quad \kappa^{33} = - \dddot{g}^{33}$$

(the dot denotes everywhere differentiation with respect to $t$).

We carry out all the calculations below for the case of an empty space, and accordingly we use Eqs. (2.1) - (2.3) of II. First we find

$$R_0^{\alpha\beta} \approx \frac{1}{2} x^{33} + \frac{1}{4} (x^{0})^2 \approx \frac{1}{2} \left( b_{33} + \dot{a}_{33} \right) = 0,$$

whence $b_{33} = - \dot{a}_{33}$. Further, simple calculation yields

$$R_0 = P_{\alpha\beta} = - b_{ab} 2 \tau a_{33} = 0,$$

whence $b_{ab} = 0$, i.e., there are no terms linear in $\tau$ in the expansion of $g_{ab}$ (this circumstance naturally simplifies the subsequent calculations, including the advantage of an expansion for given $x$, $y$, and $t$). The remaining equations do not give anything new in the approximations used.

Taking these results into account, we now write

$$g_{ab} = a_{ab} + \tau^2 c_{ab} + \tau^4 d_{ab}, \quad g_{aa} = \tau (a_{aa} + \tau b_{aa}), \quad g_{aa} = \tau^3 (a_{aa} - \tau b_{aa} + \tau^3 c_{aa}).$$

(A.1)

Here we have written out those of the subsequent terms of the expansion which are jointly determined in the transition to the next approximation in the gravitational equations.

We should now compute the first nonvanishing terms in Eqs. (2.1) - (2.3) of II, which appear upon substitution in them of Eqs. (A.1). These rather cumbersome computations are materially simplified if $g_{33}$ is represented temporarily in the form of an expansion in powers of $\tau$ for given $x$, $y$, and $z$ (and not $x$, $y$, and $t$). Carrying out the corresponding expansion of the function $a_{33} (x, y, z + \tau)$ [in the function $C_{33} (x, y, t)$ it suffices simply to replace $t$ by $z$] we get

$$g_{33} = \tau^3 [a_{33} (x, y, z) + \tau C_{33} (x, y, z)],$$

$$C_{33} = C_{33} (x, y, z) - \frac{1}{2} a_{33} (x, y, z)$$

(*) denotes differentiation with respect to $z$), i.e., the term $\sim \tau^3$ drops out of the expansion. Of course, application of different expansions for different components $g_{\alpha\beta}$ requires a subsequent re-reduction to like variables in the final result of the calculation; upon calculation of the first nonvanishing terms in the equations, however, this reduction is seen to amount simply to a replacement of $z$ by $t$.

The contravariant components of the metric tensor will now be*

$$g^{ab} = a^{ab} + \tau^2 \left( - c^{ab} + c_{ab}^{\alpha\beta} \right) + \tau^3 \left( - d^{ab} + a_{ab}^{\alpha\beta} + a_{ab}^{\alpha\beta} \right),$$

$$g^{aa} = - a^{aa} - \tau b^{aa}, \quad g^{aa} = \tau^2 a^{aa} - C_{33} - a_{33} a^{33}.$$ (A.2)

The determinant is

$$-g = \tau^2 a_{33} |a_{ab}| [1 + \tau^2 (C_{33} + C_{33} - a_{33} a_{33})].$$

The components of the tensors $\kappa_{\alpha\beta}$ and $P_{\alpha\beta}$ are computed according to the given metric. The left sides of Eqs. (2.1) - (2.3) of II are then computed. We write down only the final result here of the rather involved calculations [after which we again return to the expansion of $g_{33}$ in the form (A.1)]:

*If $g_{ab} = g_{ab}^{(0)} + h_{ab}$ (where $h_{ab}$ are small), then $g_{ab}^{\alpha\beta} = g_{ab}^{(0)} + h_{ab}^{\alpha\beta} + h_{ab}^{\alpha\beta}$. In the given case, we mean by $g_{ab}^{(0)}$ the tensor with components $g_{a0}^{(0)} = a_{ab}^{(0)}$, $g_{a3}^{(0)} = 0$, $g_{33}^{(0)} = \tau^4 a_{33}^{(0)}$.}
where K is the two-dimensional scalar curvature derived from the metric \( g_{ab} = a_{ab} \) (the two-dimensional analog \( K_{ab} \) of the tensor \( T^{ik} \) reduces, as is well known, to a scalar: \( K_{ab} = \frac{1}{2} K g_{ab} \)).

Equations (A.4) determine the coefficients \( b_{a3} \) (for the given functions \( a_{ab} \), \( a_{a3} \), and \( a_{33} \)). Then, eliminating \( d_{a}^{2} \) from Eqs. (A.5) and (A.6), we obtain the single equation

\[
\dot{a}_{ab} c^{ab} \dot{c}^{a} k_{c} + 2 \dot{a}^{a} c^{ab} \dot{a}^{b} a_{ab} c^{ab} \dot{a}^{c} \dot{a}_{c} \dot{a}_{d} = 0,
\]

which, together with the two equations of (A.7),

\[
2 \dot{c}^{a} c_{a} + \dot{c}^{a} k_{a} - \dot{c}^{b} k_{b} - \dot{c}^{c} k_{c} = - 2 \dot{a}^{a} c_{a} - \dot{a}^{a} c^{a} \dot{a}_{a} + \dot{a}_{a} \dot{a}_{a} + \dot{a}_{a} \dot{a}_{a},
\]

determines the three quantities \( c_{ab} \). The quantity \( c_{33} \) is determined from Eq. (A.3) and \( d_{a}^{2} \) from Eq. (A.6). Equation (A.8) can then be considered as the relation connecting the quantities \( a_{ab} \) with one another.

Thus, all the coefficients of the expansion (A.1) are determined from the six coefficients \( a_{ab} \), \( a_{a3} \), and \( a_{33} \), which are connected with one another by a single relation. In other words, the metric (A.1) contains five arbitrary functions.


Translated by R. T. Beyer 311