A NEW EXPLANATION OF THE RECESSION OF THE NEBULAE

E. R. R. Holmberg

(Received 1956 October 5)*

Summary

The basis of the explanation is the proposal that the space appropriate to cosmology has three, not four, dimensions. The coordinates $x, y, z, c t$ of the so-called light cone are taken to be homogeneous coordinates of a projective space which is then specified to have a uniform elliptic metric. This involves replacing the time coordinate by

$$W = \sqrt{R^2 - c^2 t^2}$$

in a manner analogous to the replacement of classical kinetic energy by the relativistic quantity

$$mc^2/\sqrt{1 - v^2/c^2}.$$ 

The radius of the universe, $R$, is thus introduced explicitly. The components of the 4-velocity vector are shown to be interpretable as the plane coordinates of this space.

Clifford’s work on the parallels of elliptic space is recalled and used to show that the physical identifications above of a point and plane manifold in elliptic space would lead one to expect the illusion of a recession velocity proportional to distance.

1. Introduction.—Although cosmology provides the simplest field of application for the theory of relativity—the metric is uniform—no completely satisfactory explanation of the observed facts in terms of it has yet been put forward.

Three stages in the development of ideas can be traced. In the first stage the metric is supposed to be isotropic with respect to all four dimensions of space and time. No solution of this type can be found unless an arbitrary negative pressure is introduced to stop the universe expanding. In the second stage the isotropy condition with respect to the time coordinate is dropped and the universe is allowed to expand. Support for this solution is found in the Hubble red shift of the spectral lines in the light from distant stars. In order to avoid the implications of a starting point in the expansion of the universe the third stage is to assume a steady state in which the loss of matter from any one neighbourhood caused by the expansion is made good by continual creation.

Since the discovery of the Hubble red shift and the analogous effect in the $21$ cm line of radio astronomy, both indicating proportionality of recession velocity and distance, the static solution of the cosmological equations has had to be abandoned. Unfortunately neither the simple expanding universe hypothesis nor its continuous creation variant are without their difficulties. The latter in particular has recently been threatened by the radio-astronomical observations of Ryle (1) which seem to indicate that at very great distances there is a crowding together of matter more in accord with the simple hypothesis of expansion: this crowding together can be taken as evidence for a change in size

* Received in original form 1956 July 12.
of the universe with time, because the distant stars are seen at an epoch in the past, whereas the continuous creation hypothesis apparently implies complete isotropy to the limit of the assessable universe. On the other hand the simple expansion hypothesis fails to account for the fact that in every neighbourhood of space stars can be found at all stages of development: if the universe is not in a steady state one would expect to find the younger stars farther away, that is, back in the past.

Besides their disagreement with some of the observational evidence neither of the expanding universe hypotheses finds easy acceptance on intuitive and philosophical grounds. Both imply creation of one kind or another and although continuous creation can be represented as a process ultimately capable of mathematical description it is difficult at the moment to see how it will be reconciled with the established laws of physics. The need for such drastic proposals would be avoided however if it could be shown that the effects like the expansion of the universe and the crowding together of distant matter were only illusions inherent in the methods of observation allowed us. For if these effects do not correspond to actually occurring physical processes we must not make the usual physical deductions from them. We could then revert to a static or stable cosmological solution in which the universe has no beginning and no end but, being closed and isolated and therefore losing no energy, is in a state of perpetual change with stationary characteristics.

The possibility of finding such a solution, and of showing the illusory nature of some of the astronomical evidence, is offered by the circumstance that all the cosmological theories mentioned have in them a certain arbitrary feature. The negative pressure of the original theory, the ratio of the curvature of space to the extent of the accessible neighbourhood set by the velocity of light and the rate of expansion of the universe in the later ones. The existence of these arbitrary constants indicates that the basic theory on which cosmology is at present founded may allow for more degrees of freedom than are physically necessary or possible. Accordingly the desired result may be achieved only by a revision of the underlying theory and not by finding yet another model based on that theory as it stands at present.

2. Metrical geometry.—The mathematical basis for all the previous cosmologies is the metrical geometry of 4-dimensions. There are a number of ways of developing this geometry which have formerly been regarded as equivalent. From what follows in this paper, however, it seems possible that one of these ways, namely, through projective geometry and the invariant theory of Cayley (2), may be superior, at least insofar as it enables one to keep the status of the various axioms and definitions clearly in mind. After following this development one is led to a somewhat novel representation of the geometry of the universe which appears to remove the arbitrary elements of the earlier cosmologies described above and yet leave room for a natural explanation of most of the observational facts.

For an account of the principles of projective geometry the reader is referred to the standard textbooks,* where it will be found that the coordinates required correspond only to labels of the points, planes and other manifolds of space and that differences between pairs of coordinate sets have no invariant significance and can therefore not be considered as representing distances. The Cayley

* e.g. Baker, H. F., Principles of Geometry, Vol. I.
A new explanation of the recession of the nebulae

No. 6, 1956

theory (2) of distance between two elements of a manifold defines it as the logarithm of the cross ratio of the coordinates of two elements and the coordinates of the two intersections of their join with an absolute quadric surface. This expression is invariant under the group of transformations which takes the absolute quadric into itself, and Lie (3) has shown that there are three such groups only corresponding to different forms of quadric. If the quadric is real and closed one obtains the hyperbolic geometry of Lobatchewsky; if it is imaginary, the elliptic geometry of Riemann; if it degenerates into a single plane the geometry is affine, and if in this plane a closed conic is latent the geometry is Euclidian.

We give now the principal result of the Cayley theory which is the basis of the proposal in this paper. It is that the distance, $D$, between two points $x_i$ and $x_j (i = 1 \ldots n + 1)$ in an $n$ dimensional manifold can be put in the form:

$$\cos(D/k) = \frac{a_{11}x_1x_1' + a_{12}(x_1x_2 + x_2'x_1') + a_{22}x_2x_2' + \ldots}{\sqrt{(a_{11}x_1'^2 + 2a_{12}x_1'x_2 + \ldots)}\sqrt{(a_{11}x_1^2 + 2a_{12}x_1x_2 + \ldots)}}$$

(2.1)

where

$$\phi = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + \ldots = 0$$

is the equation of the absolute quadric in homogeneous coordinates and $k$ is an arbitrary scale factor which connects the numbers used for the coordinates and those which measure distance.

In Euclidian geometry the limiting procedures which define it and which correspond to the degeneration of the metric quadric result in the cancellation of the expressions in the denominator of (2.1) and the reduction of the numerator to a form which involves coordinate differences only. But in the non-Euclidian cases it appears impossible to reduce the R.H.S. of 2.1 to a form involving only differences of coordinate values, even when these differences are small, without explicitly setting the expressions under the root signs in the denominator equal to a constant, that is imposing the condition:

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + \ldots = k^2$$

(2.2)

on all the coordinate sets. (Incidentally, it is necessary to make this new constant equal to the scale factor on the left of 2.1 in order to preserve the usual definition of distance.) This condition can then be used to eliminate one of the coordinates, giving an expression for distance involving the same number of variables as dimensions. This reduction to inhomogeneous coordinates occurs automatically in Euclidian geometry described relative to orthogonal axes, and only in this case are the coefficients in the distance expression constants. In a sense, therefore, the passage from Euclidean to non-Euclidian geometry can be regarded as a generalization: the replacement of constant by non-constant coefficients. The cosmological theories which have developed from Einstein’s relativity have all taken as their starting point a distance formula in which the reduced number of variables, that is, the number of inhomogeneous coordinates, was four, and in which the coefficients were not all constant. All these theories are thus based on the premise that the universe to be described is four-dimensional.

This paper adopts another premise, that the universe, as accessible to observation, is only three dimensional. This is not an arbitrary procedure, as will be shown. It takes its justification from the fact that the equations of special relativity can be interpreted as referring to a three-dimensional quite
as well as to a four-dimensional world. Other more cogent arguments in favour of the new premise can be adduced but these would be outside the scope of this paper. Here therefore after demonstrating that the premise need not contradict established facts we rest inductively on its logical consequences for its acceptance.

3. A modification of special relativity.—It will be observed immediately that the relation (2.2) for a 3-dimensional space is identical with the condition that a $4$-vector in Euclidian space should have a constant length. The fact that in such a condition the principal coefficients are all unity and the cross products vanish is merely a consequence of choosing the coordinate system so that the equation of the metric quadric has the canonical form:

$$x_1^2 + x_2^2 + \ldots = 0.$$ 

There is therefore no formal difference between the properties of constant vectors in 4-dimensional Euclidian space and general vectors in 3-dimensional non-Euclidian space of constant curvature, and the question poses itself: whether we have been physically correct in relativity in supposing that constant 4-vectors do in fact refer to a 4-dimensional space. As remarked above, this paper is hardly the place to discuss the geometrical implications of the problem, so we shall proceed by taking the newly exposed alternative as a working hypothesis and show what its consequences are.

The specific 4-vector which concerns us is the 4-velocity, which we can write for our present purpose in its usual form:

$$v_1 = \{i\xi v/c / \sqrt{1 - \sigma^2/c^2}, \eta v/c / \sqrt{1 - \sigma^2/c^2}, \xi v/c / \sqrt{1 - \sigma^2/c^2}, 1/\sqrt{1 - \sigma^2/c^2}\}$$

where $\xi^2 + \eta^2 + \zeta^2 = 1$. It therefore satisfies the condition

$$v_1^2 = 1.$$ 

Our proposal is that the $v_1$ should be regarded as the homogeneous plane coordinates of a uniformly curved 3-space. There is no contradiction of the accepted principles of special relativity involved in doing this for the rules of composition of velocities in relativity correspond exactly to the Cayley formula (2.1). Taking as an example two velocities in the same direction $(\xi, \eta, \zeta)$, (2.1) becomes:

$$\cos(D/c) = \frac{1 - \nu \nu'/c^2}{\sqrt{(1 - \sigma^2/c^2)} \sqrt{(1 - \sigma^2/c^2)}}$$

or

$$\tan(D/c) = \frac{i(\nu' - \nu)/c}{1 - \nu \nu'/c^2}.$$ 

The novel consequences of the proposal appear only because of the fact that if the velocity manifold is taken as a plane manifold it must have a point manifold associated with it, which also has non-Euclidian metrical characteristics. This is the nub of the matter and seems easily to be overlooked if any of the alternative approaches to relativity are employed. This point manifold must, like the plane manifold, be isotropic and so its sets of four homogeneous coordinates must be connected by a relation of the type (2.2). It follows therefore that we cannot be dealing here with the ordinary space-time of relativity because the coordinates of this are in principle independent.

The solution offered by the author is that the point manifold corresponding to the velocity manifold is the so-called light cone. The coordinates, $x, y, z, ct$ of this space are not independent and can with modification be regarded as
referring to a non-Euclidean space of 3-dimensions. The initial difficulty in the way of making this identification is occasioned by the fact that as they stand \( x, y, z, ct \) satisfy

\[
x^2 + y^2 + z^2 - c^2t^2 = 0 \tag{3.1}
\]

in which the characteristic constant is zero. This is clearly not compatible with the Cayley theory for it would force the distance between two points to be always infinite. To overcome this one can adopt the device used by Einstein in constructing the 4-velocity, that is, to add to the time term a constant so large that it is normally neglected. By putting:

\[
c^2t^2 = R^2 - w^2 \tag{3.2}
\]

(3.1) becomes:

\[
x^2 + y^2 + z^2 + w^2 = R^2. \tag{3.3}
\]

If \( R \) is very large compared with \( x, y, z, ct \) on the Earth we find ourselves dealing with the coordinates

\[
\frac{x}{R}, \quad \frac{y}{R}, \quad \frac{z}{R}, \quad \frac{1}{R}
\]

in most situations, where \( R \) can be absorbed into the scale factor and space appears to be straight. The law of compounding the coordinates also reduces to the classical form in the limit. For example (3.3) is satisfied if the coordinates are put in the form

\[
\xi ct, \quad \eta ct, \quad \zeta ct, \quad \sqrt{R^2 - c^2t^2}
\]

where \( \xi^2 + \eta^2 + \zeta^2 = 1 \). The distance between two points in the same direction is then given by

\[
\cos \left( \frac{D}{R} \right) = \frac{I}{R^2} \left( c^2tt' + \sqrt{(R^2 - c^2t^2)(R^2 - c^2t'^2)} \right).
\]

When \( ct \) and \( ct' \) are small the R.H.S. reduces to:

\[
1 - \frac{1}{2} \frac{c^2}{R^2} (t' - t)^2
\]

so that \( D = c(t' - t) \). If the sign of the spatial components of one of the sets of coordinates is made negative the same procedure yields the sum of two times instead of the difference, as it should. The sign of the time-like component must not be changed since this is in the nature of a cosine.

It is clear from the work just done that \( R \) plays an analogous role in the point manifold to the velocity of light in the plane manifold. With this difference: the plane manifold, on account of the spatial components being counted imaginary, is hyperbolic, whereas the point manifold is elliptic. In the plane manifold, therefore, the metric quadric is real and constitutes a boundary which is the velocity of light. The point manifold is, on the other hand, unbounded although finite. Its extent is a matter of choice and is settled by the scale factor \( R \). This may be regarded as "the radius of the universe". This is not a trivial identification for it can be seen to correspond closely with what we know of the physical universe. Our experience of events outside the solar system and particularly outside our own galaxy is, because of the great distances and finite velocity of light, effectively confined to one observation at one instant of time. The time taken for light to reach us with information about these events is so great compared even with the whole of written human history that these events can be considered
as lying on a light cone. The point manifold whose geometry we have just
analysed is therefore entirely appropriate for the representation of our knowledge
of the physical universe and the radius of this manifold can be identified with
the radius of the actual universe.

4. A new interpretation of nebular recession.—The properties of 3-dimensional
non-Euclidian space were examined by W. K. Clifford (4) who discovered a
number of significant differences between them and those of Euclidian space.
At the time Clifford wrote it was customary to regard elliptic and hyperbolic
spaces as distinct alternatives and, in speaking of point and plane manifolds,
therefore, he regarded it possible to have both in either an elliptic or hyperbolic
space. From what has been said in the last section however it is clear that the
difference between elliptic and hyperbolic spaces is a matter of whether one
counts all the coordinates as real or whether one (or three) of them is taken as
imaginary. The fact that Clifford's results referred to an elliptic space therefore
need not prevent us from applying them to the association of an elliptic point
manifold and hyperbolic plane manifold we have discussed above. In fact
the point of view implied by this appears the more natural one. In Newtonian
space the point manifold is Euclidian and open and the plane manifold is closed.
It accords with our intuitive view of things to have point and plane manifolds of
a space complementary in this respect.

To return to the main argument, Clifford pointed out that in elliptic space
points and planes stand in a definite relation to each other which has no parallel
in Euclidian space. Thus each point defines a unique plane, its polar with respect
to the metric quadric, and each plane defines a point, its pole with respect
to the same quadric. It follows from this that the points of a line define a pencil
of polar planes, so that each line has a polar line with respect to the metric quadric.
Accordingly, translation along a line is equivalent to rotation about the polar line
as axis, and rotation about the same line is equivalent to translation along the polar.
Clifford also showed that two self-polar lines lie in the surface generated by
translating one of them parallel to itself. There are in general two lines parallel
to a given line and through a point not on it, but these coincide when the point
lies on the polar line; this line is the polar itself.

From these results of Clifford's certain peculiarities in the process of inter-
preting events occurring at great distances from an observer seem to be indicated.
For example, when translation is being observed at a distance this has to be
described, as far as magnitude and direction are concerned, by referring to a set
of parallel axes close at hand, so the peculiarities of parallels in elliptic space
are significant. The matter is simplest to visualise in the extreme case when
these parallel axes are the polars of the ones along which the actual translation
takes place. We must then take into account that translation along these is
equivalent to rotation about the latter. It seems plausible to suppose that this
equivalence means that translation taking place at the polar distance will appear
and must be represented as rotation about the local parallel axis and vice-versa.
Also, at intermediate distances a translation or a rotation would need to be
represented as a twist about a screw. These remarks have to be interpreted
somewhat before being applied to cosmology. Clifford had in mind a simple
generalization of Newtonian space as his model of the point manifold, whereas
we have shown that the light "cone" is more appropriate for our purposes. In
our case, therefore, instead of rotation we must understand velocity.
At first sight it seems implausible that position should be capable of being confused with velocity in this way and perhaps it will always remain so. However such confusion of qualitatively different observables is already familiar in the theory of relativity where the strict dichotomy of time from space has disappeared. Moreover the interchange of position and velocity goes back to a period before relativity in the shape of the constancy of the sum of potential and kinetic energy. Notwithstanding the implausibility or otherwise of the idea, however, it would seem that it puts in our hands a simple explanation of the recession of the distant nebulae.

The simplest way of understanding this is to take again the extreme case of observers situated at opposite poles of the universe and to appeal to the principle of relativity, that is, that what these observers see of each other should be a function only of their relative situations. The new feature introduced by the knowledge that space is elliptic is that what is locally displacement for one is interpreted locally by the other as velocity and vice-versa. Since they both see the other as at a great distance it would seem that they both see the other also in rapid motion, the motion of the one being the displacement of the other.

The case of observers at intermediate distances is not so straightforward because distance and velocity are for them not completely reversed. This brings to light certain subtleties in the concepts of distance and velocity when these are large, which could be neglected in the extreme case, and if the simple relativity argument were used confusion would arise. The matter is however made clear by a return to the Clifford concept of parallel translation of the axes of reference. We find there that during a translation the parallel is rotated through an angle proportional to the amount of the translation; in fact, a parallel translation in elliptic space consists of a twist about a screw with pitch unity. Now the rotation part of the twist represents acceleration of the reference frame in our geometry so the proportionality of recession velocity and distance observed in nature follows directly. In particular, of course, the translation which carries an axis into its polar is a twist through a right angle, that is, an acceleration to the maximum velocity, so this argument agrees with the one in the previous paragraph, in the case of polar observers.

5. Discussion.—The last three sections contain the main points of the proposed modification of the theoretical basis of relativity which removes the arbitrary features of the older cosmologies and makes the recession of the distant nebulae a purely observational illusion. Although these changes are drastic it has been shown that they do not conflict with accepted principles where these principles can be subjected to experimental test and, furthermore, detailed consideration shows them to be very much in accord with the spirit of special relativity in its earlier form. The special theory emphasized that observation depended on the propagation of light rays and that concepts such as absolute spatial length which could not be measured by sending and receiving light signals should not be represented. In constructing cosmologies which rest on the supposition that space and time are independent variables for distant events even though it is certain that we cannot separate them observationally we were thus abandoning the relativity principle. The present hypothesis of founding the metric in the space of the light "cone" is thus a return to it.

The present paper has deliberately not dealt with more than the very first principles of the new approach. It remains to show that these principles are
A new explanation of the recession of the nebulae

not so restrictive as to exclude explanations of other observable phenomena, such as the crowding together of distant matter reported by Ryle. Although this work has already been largely completed an exposition of it requires, as a preliminary, detailed discussions of the fundamentals of geometry and algebra which are probably outside the scope of an astronomical paper and which therefore may have to be published elsewhere. It is hoped, nevertheless, that the further astronomical consequences of the new proposals can be published at an early date.

Army Operational Research Group,
Broadoaks, Parvis Road,
West Byfleet, Surrey:
1956 October.

References

(1) Ryle, M., Observatory, 75, 137 (1955).
(2) Cayley, A., "Sixth Memoir upon Quantics", Phil. Trans., 149, 61 (1859).