PROBLEMS IN THE RELATIVISTIC THEORY OF GRAVITATIONAL COLLAPSE

PETER DAVID GOODSTEIN

R. H. C. LIBRARY
CLASS No. Ayz.GGE Goo
ACC. No. 111871
DATE ACQ Nov.72

ProQuest Number: 10107289

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10107289

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Gravitational collapse is analysed in terms of a simple model. Both Newtonian and Relativistic treatments are given, and the curious phenomenon of "overtaking" is discussed from a Newtonian viewpoint. It is shown that the Relativistic description of the motion is closely analogous to the classical treatment.

The predicted final stage of the motion is collapse to a point singularity. The asymptotic behaviour near this singularity is examined, and it is shown that the presence of small inhomogeneities in a collapsing dust-sphere will radically affect the motion.

The question of Boundary Conditions in General Relativity is considered, insofar as this affects the Relativistic description of the motion. It is shown that the Lichnerowicz conditions may be too restrictive, and conditions in respect of the first and second fundamental forms are proposed.

It is shown that the spectral shift of a collapsing body becomes unstable as the gravitational radius is approached.

The final stage in the collapse process is examined. In order to suggest possible models involving motion beyond the point singularity, it is necessary to consider General Relativity from a modern mathematical viewpoint; this treatment leads to a consideration of models involving multiply connected manifolds.

The concept of "time orientation" is developed. It is shown that if, in a particular model, a collapsing body passes through the point singularity and then expands into the same spatial region, the model necessarily involves causality violations.

ACKNOWLEDGEMENTS

The author wishes to acknowledge his deep indebtness to Professor W. H. McCrea, Sc.D., D.Sc., F.R.S., for invaluable criticism, assistance and advice at every stage of the preparation of this Thesis. A grateful acknowledgement is also due to Mrs. Buxton, for undertaking the formidable task of typing the Thesis.

The research underlying this Thesis was supported by a Research Studentship and Grant provided by the Science Research Council (and previously by the Department of Scientific and Industrial Research).

CONTENTS

CHAPTER I

INTRODUCTION

1. The problem of Gravitational Collapse, considered from the viewpoint of the General Theory of Relativity, has been the subject of renewed interest recently. Various aspects of the question have been discussed, notably by Hoyle & Fowler (1962), Hoyle & Narlikar (1963), Hoyle, Fowler, Burbidge & Burbidge (1964), Israel (1967) and Penrose & Hawking (1970).

Cosmological questions often feature in these discussions, since in some respects the behaviour of an ideal collapsing body is similar, in a mirror-image fashion, to certain cosmological models exhibiting expansion. In particular, mention must be made of papers by Penrose (1965) and Hawking (1966), in which general conditions are derived for the existence of Space-Time singularities. The subjects of Gravitational Collapse and Relativistic Cosmology are indeed linked in a recent investigation (Barnes (1970)).

Progress in the field is described in the recent book by R.U.Sexl (1970), as well as in the earlier comprehensive review "Quasi-Stellar Sources and Gravitational Collapse", (Robinson, Schild & Schucking, 1965)

2. Oppenheimer & Volkoff (1938) considered the possible terminal behaviour of stellar matter subsequent to the thermonuclear burning stages. Normally, a star is prevented from collapsing under the influence of its own gravitation by the effect of radiation pressure;

R.H.C. LIBRARY

however, when the thermonuclear "fuel" is exhausted the star may "degenerate" into a cold cloud of neutrons.

Such a "neutron star" may nevertheless resist its self-gravitation by means of a new type of pressure, known as "degeneracy pressure". However, Quantum Mechanical considerations predict an upper limit for the amount of matter which can be supported by degeneracy pressure.

The critical mass is, very approximately, of the order of magnitude of the mass of the sun. Larger accretions of matter must either fragment (if sufficient angular momentum is present) or collapse catastrophically. As such collapse proceeded, the effects of degeneracy pressure would become relatively insignificant, and the body might be regarded as consisting of pressure-free dust (apart from possible effects of radiation).

3. Hoyle & Fowler (1962) start from a rather different viewpoint, and consider various stages in stellar evolution. Quasi-equilibrium configurations exhibiting "relaxation oscillations" are investigated, and it is again shown that for a sufficiently large mass, catastrophic collapse may be predicted.

This paper immediately preceded the discovery of the "Quasi-Stellar Objects", or "Quasars". These intense sources of radio emission were identified with optical sources and substantial spectral shifts were found. These red-shifts were presumably cosmological in origin, and implied that the sources were extra-galactic, and probably very distant.

If this was so, a problem arose: the intensity of the sources would then imply an exceptional intrinsic luminosity. Every though measurements indicated a mass several orders of magnitude greater than that of the sun, it appeared that a process of conversion of mass into energy more efficient than any known thermonuclear reaction might be involved.

These discoveries intensified interest in the study of theoretical models of Gravitational Collapse. Possibly catastrophic collapse itself implied an efficient mass-energy conversion mechanism; it was also possible that a portion of the observed quasar spectral shift was gravitational in origin, which further justified a General Relativistic model.

Whilst the precise nature of Quasars remains uncertain, our understanding of the collapse process has increased considerably. Gravitational Collapse is now recognised as a worth-while study in its own right: among interesting predictions of the Relatavistic theory is the probable existence of dense, invisible bodies ("black holes") observable only through the effect of their gravitational field. There is some prospect of experimental verification of this in the near future.

4. In this Thesis, we shall confine our attention to a very simple model, which may be called the "dust fluid". This ideal body consists of matter whose physical properties can be specified in terms of one parameter - the distribution of its mass in space. The body will further be supposed to possess spherical symmetry. The model seems

to possess one serious inadequacy only in describing a state of catastrophic collapse in a degenerate star: it does not possess rotation.

In practice, all bodies possess angular momentum, and as collapse progresses this feature may be expected to assume increasing importance. However, the General Relativistic treatment of collapse is substantially complicated if rotation is taken into account. A satisfactory treatment is still awaited, although significant progress has occurred within the last ten years (notably Kerr (1963)).

It seems probable that the effect of rotation will be to convert a process of collapse to a point of one of collapse to a disc. Nevertheless, it is unlikely that the presence of rotation will materially affect such parameters as the "gravitational radius" (see Chapter IV) or the time of collapse. Accordingly, the dust fluid model seems adequate for the general description of the motion, and it highlights the principal problems.

5. Milne and McCrea (1934) demonstrated that the Relativistic treatment of the motion of a homogeneous dust-sphere was analogous to the Newtonian treatment. In Chapter II we review the description of the motion of an inhomogeneous dust-sphere from the standpoint of Newtonian Mechanics. Certain interesting consequences of this treatment are noted: in particular the predicted "overtaking" phenomena that should occur in certain circumstances.

The Relativistic treatment is developed in Chapters III and IV, and it is shown that the analogy between the Newtonian and Relativistic results is carried over to the inhomogeneous case. Chapter III outlines some of the principles upon which the Relativistic analysis is based; it also contains a small section which endeavours to cast new light on the question of the "perfect fluid", although the Thesis as a whole is concerned for the most part with the Dust Fluid model.

Chapters IV and VI consider the various problems which form the crux of the present investigation. Chapter V is devoted to an exposition of the mathematical techniques and concepts which are a prerequisite to the understanding of these problems. A summary of the aims of the investigation and the results achieved appears at the end of Chapter VI.

CHAPTER II

NEWTONIAN COLLAPSE

We shall now consider the "Dust Fluid" model in more detail. It is instructive to consider first the properties of the associated Newtonian model: it will be shown that the two models possess many analogous characteristics.

We suppose that the body whose motion we seek to describe is a sphere of Perfect, Newtonian Fluid, of which the equation of state is simply

$$P = 0 \tag{1}$$

The body is supposed to be isolated, and only its self-gravitational force acts on it. We shall use the Lagrangian approach to the characterisation of the motion, supposing that the radius and density distribution in the fluid are known at some initial time, as is the distribution of velocity at this instant.

We shall firstly consider the homogeneous case, which is already familiar from discussions of Newtonian cosmology. (REF. 28).

Here we suppose

$$\rho = \rho(t) \tag{2}$$

throughout.

In discussing all these models, we shall ignore rotation, and therefore assume that the motion is purely radial. If a particle of the fluid is initially at the point of space

$$(r, \theta, \phi)$$

its position at time t will be

$$(R,\theta,\phi)$$

say, where

$$R = R(r,t).$$

Let r_b be the initial boundary of the sphere. Consider a concentric sphere of fluid of radius r, where

$$r < r_b.$$

At time t this will have become a sphere of radius R. Let M_R denote the mass of the sphere at time t. By Conservation of Matter

$$M_R = M_r \tag{3}$$

which gives

$$\frac{4}{3}\pi R^3\rho(t) = \frac{4}{3}\pi r^3\rho_0$$

where ρ_0 is the initial density.

Hence

$$R = rS(t) \tag{4}$$

where

$$\rho(t) = \frac{\rho_0}{S^3(t)} \quad . \tag{5}$$

We shall choose our time coordinate such that initially

$$t = 0 .$$

Thus

$$S(0) = 1 . \tag{6}$$

According to Newtonian Gravitation, there is no force acting at any point within a spherical shell of matter. Thus at time t, the force acting on a particle of mass m distant R from the centre is that due only to the matter within the sphere of radius R, which is gravitationally equivalent to a point particle mass M_R, at the centre.

Thus the force acting towards the centre is of magnitude

$$G\frac{mM_R}{R^2}$$

and so the equation of motion is

$$\frac{\partial^2 R}{\partial t^2} = -G\frac{M_R}{R^2}$$

$$= -\frac{4}{3}\pi G R\rho(t).$$

Hence, using (4) and (5)

$$\ddot{S} = -\frac{4}{3}\pi G\rho_0 \cdot \frac{1}{S^2} .$$

Writing

$$\alpha = \frac{8}{3}\pi G\rho_0 \tag{7}$$

we find

$$\dot{S}^2 = \frac{\alpha}{S} + \text{const.}$$

There are essentially three solutions of this equation, according to our choice of constant.

Let

$$\xi(\tau), \quad \eta(\tau), \quad \zeta(\tau)$$

respectively satisfy

$$\left.\begin{aligned}\left(\frac{d\xi}{d\tau}\right)^2 &= \frac{1}{\xi} - 1\\ \left(\frac{d\eta}{d\tau}\right)^2 &= \frac{1}{\eta}\\ \left(\frac{d\zeta}{d\tau}\right)^2 &= \frac{1}{\zeta} + 1\end{aligned}\right\} \tag{8}$$

where we stipulate that ξ, η, ζ are each unity at

$$\tau = 0.$$

If the initial velocity of every point of the body is zero, then

$$S = \xi(\alpha^{\frac{1}{2}} t) \ . \tag{9}$$

Note that α has dimensions

$$[T^{-2}] \ .$$

This might be called the ELLIPTICAL SOLUTION, since the sphere has a maximum radius. It applies to gravitational collapse from rest.

Suppose instead that the boundary is initially moving with velocity

$$\left(\frac{\partial R_b}{\partial t}\right)_0 = V \ .$$

By (4) we must have

$$\begin{aligned}\frac{\partial R}{\partial t} &= r\ \dot{S}(t)\\ &= \frac{r}{r_b} r_b\ \dot{S}(t)\\ &= \frac{r}{r_b}\ \frac{\partial R_b}{\partial t}\end{aligned}$$

$$\therefore \quad \left(\frac{\partial R}{\partial t}\right)_0 = \frac{r}{r_b} V$$

i.e. a particle with coordinate r must have this velocity initially, for consistency with (2). Thus here

$$\dot{S}(0) = \frac{V}{r_b} \qquad (10)$$

It follows that the first integral of the equation of motion takes the form:

$$\dot{S}^2 = \alpha\left\{\frac{1}{S} - \left(1 - \frac{V^2}{\alpha r_b{}^2}\right)\right\}. \qquad (11)$$

The value

$$|V_0| = \alpha^{\frac{1}{2}} r_b$$

may be called the ESCAPE VELOCITY. If

$$|V| \geqslant |V_0|$$

then the boundary, and therefore every part of the body, other than the centre, reaches infinity during some part of the motion. If we take a negative V, then, as we shall see, the first part of the motion represents collapse from a dispersed state. If however

$$|V| < |V_0|$$

then the body is always of finite extent: again taking a negative value for V, we might regard this as collapse from an equilibrium state (e.g. if for some reason the pressure forces within a star suddenly become inadequate to support it against its own gravitation, and the pressure forces never "recover" the situation.

In the former case, the "velocity at infinity" is

$$r\left(\frac{V^2}{V_c^2} - 1\right)^{\frac{1}{2}} .$$

Note also that

$$V_c^2 = \frac{8}{3}\pi G\rho_0 r_b^2$$

(using (7)), and so, writing M for the total mass of the body, we have:

$$V_c^2 = \frac{2GM}{r_b} .$$

This is, of course, the escape velocity from any sphere of matter, radius r_b and mass M, no matter what is happening within the sphere. Note that this is the speed at which a particle initially at rest at infinity arrives at the surface of a static sphere of this radius and mass. In particular if the particle arrives with the velocity of light, c, then

$$r_b = \frac{2GM}{c^2}$$

which provides a Newtonian interpretation of the General Relativistic notion of "Gravitational Radius", which we shall consider later in this chapter.

We now give the explicit solutions of (11) in terms of the functions

$$\xi, \eta, \zeta$$

given in (8).

If

$$|V| < |V_c|$$

we shall write

$$p = 1 - \frac{V^2}{V_c^2} \tag{12}$$

whereas, if

$$|V| > |V_c|$$

we write

$$q = \frac{V^2}{V_c^2} - 1 \quad . \tag{13}$$

Note that we must have

$$\left.\begin{array}{l} 0 < p \leqslant 1 \\ \text{and} \\ 0 < q \end{array}\right\} \tag{14}$$

<u>Case I</u>: $|V| < |V_c|$

Write

$$\tilde{S} = pS$$

Then

$$\dot{\tilde{S}}^2 = \alpha p^2 \left[\frac{p}{\tilde{S}} - p\right]$$

$$= \alpha p^3 \left[\frac{1}{\tilde{S}} - 1\right]$$

Now write

$$\xi(a) = p \tag{15}$$

where by a we mean the least, positive root of

$$\xi(\tau) = p.$$

Then

$$\tilde{S} = \xi(\alpha^{1/2}p^{3/2}t + a)$$

since we have postulated (see (6))

$$S(0) = 1.$$

Thus

$$S = p^{-1}\xi(\alpha^{1/2}p^{3/2}t + a) \quad . \tag{16}$$

<u>Case II</u>: $|V| = |V_c|$

Here we have simply

$$S = \eta(\alpha^{1/2}t) \tag{17}$$

<u>Case III</u>: $|V| > |V_c|$

Writing

$$\zeta(b) = q \tag{15a}$$

and proceeding as in case I, we find

$$S = q^{-1}\zeta(\alpha^{1/2}q^{3/2}t + b) \ . \tag{18}$$

Equations (16), (17), (18) fully describe the motion of the dust-fluid, no matter what initial conditions, consistent with (2), are applied.

Equations (8) may be solved to give

$$\left.\begin{aligned} &(a)\quad \cos^{-1}\pm\sqrt{\xi}\pm\sqrt{\xi(1-\xi)} &&= \tau \\ &(b)\quad \pm\tfrac{2}{3}(1-\eta^{3/2}) &&= \tau \\ &(c)\quad \sinh^{-1}\zeta-\sqrt{\zeta(1+\zeta)} &&= \pm\tau-\sqrt{2}+\log(1+\sqrt{2}) \end{aligned}\right\} \tag{19}$$

where we take the principal value of

$$\cos^{-1}$$

and take the positive square root during the contraction phase, and the negative root during expansion.

Cases II and III might be called the PARABOLIC and HYPERBOLIC solutions, respectively. The three cases are precisely the (pressure-free) Newtonian Cosmologies, since (2) is the condition for spatial homogeneity, and for a dust fluid, isotropy is a consequence of equation (4).

Cases I and III are respectively typified by

(i) $V = C$

giving equation (9), and

(ii) $V = \sqrt{2}\, V_0$

giving

$$S = \zeta(\alpha^{1/2} t). \tag{20}$$

Thus the cosmological solutions apply to a dust sphere of arbitrary radius, and are the only homogeneous solutions.

We now consider general initial conditions, resulting in inhomogeneous solutions. However the three functions (8) still figure in the description of the motion.

We retain (1) but instead of (2) assume that

$$\rho(r,0) = \rho_0(r) . \tag{21}$$

Equation (3) still applies, but we now deduce

$$4\pi \int_0^R R^2 \rho(r,t)\, dR = 4\pi \int_0^r r^2 \rho_0(r)\, dr \tag{22}$$

where the integration on the L.H.S. is to be performed for fixed t, so that ρ may be regarded as a function of R.

Equation (22) may be rewritten as:

$$\int_0^r R^2 R' \rho(r,t) dr = \int_0^r r^2 \rho_0(r) dr$$

where "dash" signifies partial differentiation with respect to r. Since this holds for all r, we have

$$R^2 R' \rho(r,t) = r^2 \rho_0$$

and so

$$\frac{\rho_0(r)}{\rho(r,t)} = \frac{\partial R^3}{\partial r^3} \qquad (23)$$

which is the Lagrangian form of the equation of continuity for a spherically symmetric fluid.

Here

$$M_r = 4\pi \int_0^r r^2 \rho_0(r) dr \quad .$$

A valuable aid to discussing the motion is a function which we may call the "mean initial density up to radius r", defined as

$$\bar{\rho}_0(r) = 4\pi \int_0^r r^2 \rho_0(r) dr \,/\, 4\pi \int_0^r r^2 dr \qquad (24)$$

whereby we may discuss the general solution on an analogous basis to the homogeneous one.

Note

$$M_r = \frac{4}{3}\pi r^3 \bar{\rho}_0(r)$$

and

$$\rho_0(r) = \frac{d(r^3 \bar{\rho}_0(r))}{dr^3}$$

$$= \frac{1}{3} r \frac{d\bar{\rho}_0}{dr} + \bar{\rho}_0(r) \quad . \qquad (25)$$

We shall write

$$R = rS(r,t) \tag{26}$$

(cf.(4)).

The equation of motion is

$$\frac{\partial^2 R}{\partial t^2} = -G\frac{M_R}{R^2}$$

$$= -G\frac{M_r}{R^2}$$

$$= -G \cdot \frac{4\pi r^3 \bar{\rho}_0(r)}{R^2}$$

using (24).

Thus, by (26), the equation of motion reduces to

$$\ddot{S} = -\frac{4}{3}\pi G \bar{\rho}_0(r) \cdot \frac{1}{S^2} .$$

Put

$$\alpha(r) = \frac{8}{3}\pi G \bar{\rho}_0(r) \tag{27}$$

(cf.(7)).

Then the first integral of the equation of motion gives

$$\dot{S}^2 = \frac{\alpha(r)}{S} + F(r) .$$

Firstly, let us assume that the body is started from rest with a given density distribution. Using (26), we see that

$$S(r,0) = 1 \tag{28}$$

and since

$$\dot{S}(r,0) = 0$$

here, we have

$$F(r) = -\alpha(r)$$

and so

$$S = \xi(\alpha^{1/2}t) \tag{29}$$

which is identical to (9), except that in the definition of α, the constant ρ_0 is replaced by the function

$$\bar{\rho}_0(r) .$$

Suppose instead that there is an initial velocity distribution

$$\left(\frac{\partial R}{\partial t}\right)_0 = V(r) .$$

Then, using (26),

$$\dot{S}(r,0) = \frac{1}{r}V(r)$$

and so the first integral of the equation of motion is

$$\dot{S}^2 = \alpha\left(\frac{1}{S} - \left(1 - \frac{V^2}{\alpha r^2}\right)\right) \tag{30}$$

(cf.(11)), where α,V are now functions of r. Here the motion may be essentially different for differing values of r. There is an "escape velocity" associated with each value of r:

$$|V_0(r)| = \alpha^{1/2}r .$$

(Note that whereas in the homogeneous solution "V" means the velocity at the boundary, here "V(r)" means the velocity at any point r.)

If, for some range of r,

$$|V(r)| < V_0(r)|$$

then (30) can be solved in the form (16), where now

$$p(r) = 1 - \frac{V^2(r)}{\alpha r^2} \tag{31}$$

and

$$a(r)$$

is

is such that

$$\xi(a(r)) = p(r) \tag{32}$$

where we take the least non-negative root of this latter equation.

Similarly for other values of $V(r)$, giving solutions of the form (17) or (18), except that the constants

$$\alpha, q, b$$

(see (13) and (15a)) are replaced by appropriate functions of r.

If we put

$$V(r) = \frac{r}{r_b} V$$

$$\rho_0 = \text{const}$$

we find

$$\alpha = \text{const}$$

and

$$p, q = \text{const}$$

so that we recover our homogeneous solutions. Indeed the general solution may be regarded as a sort of "ensemble" of homogeneous ones.

Non-homogeneous solutions may exhibit curious behaviour. The initial density and velocity distributions may be such that "cavitation", resulting in fragmentation, occurs. This could only happen if the initial velocity distribution is discontinuous at a point. For simplicity we shall suppose that

$$V(r)$$

is twice-differentiable, and that

$$\rho_0(r)$$

is, at any rate, piecewise twice-differentiable.

Alternatively, "overtaking" of one portion of the fluid by another may occur: the conditions under which this may happen will be investigated shortly.

We now proceed to a detailed discussion of our solutions. Consider first the basic solutions (16), (17), (18). All solutions of type (16) are characterised by the function

$$\xi(\tau)$$

satisfying (8). This function, by (19), is periodic with period Π. The function

$$\sqrt{\xi}$$

takes all values in the range [-1,1], and so ξ takes all values in the range [0,1].

At

$$\tau = \pi/2$$

we have

$$\xi = 0, \quad \dot{\xi} = \pm\infty ;$$

in fact

$$\lim_{\tau \uparrow \pi/2} \dot{\xi} = -\infty$$

$$\lim_{\tau \downarrow \pi/2} \dot{\xi} = +\infty$$

The special solution (9) represents collapse from rest, the collapse taking a time (half-period)

$$t_o = \frac{\pi}{2} \alpha^{-1/2} \qquad (33)$$

at the end of which the whole body occupies the point

$$r = 0.$$

At this moment, ρ, by (5), becomes infinite, and the velocity of every particle of the body switches instantaneously from infinite inward radial motion to a similar outward motion. It may be objected that such behaviour is highly unphysical, especially the velocity "reflection". Collapse into such a "singularity" is nevertheless the predicted behaviour of a dust fluid on the basis of Newtonian Gravitation; as we shall see, it is also the prediction of Relativity Theory. In fact, the solution, for a range of τ which includes the value $\frac{\pi}{2}$, and/or any odd multiple of $\frac{\pi}{2}$, is not unique. As will be explained in Chapter V, the point

$$R = 0$$

is a singularity of the polar coordinate system itself, at which our criterion of "Spherical Symmetry" becomes devoid of meaning. It is quite admissible to join the solution for

$$-t_o < t < t_o$$

with fixed

$$\theta = \theta_0, \quad \phi = \phi_0$$

to the solution for

$$t > t_o$$

with

$$\theta = \theta_1, \quad \phi = \phi_1$$

say. If this is done in such a way that

$$\theta_1 = \frac{\pi}{2} - \theta_0 \ ; \quad \phi_1 = \pi - \phi_0$$

i.e. we take diametrically opposite points on the surface of any sphere, radius the origin, then we might expect that the fluid velocity would not jump, sharply. This interpretation of the "follow through" of collapse is, of course, quite arbitrary. It would mean that a collapsing sphere "turned itself inside out". Further investigation of this question will be carried out in Chapter VI.

Consider now the more general solution (16). This is essentially the prototype solution (9). The effect of the initial velocity is threefold. Firstly, in the equation

$$R = r\xi\,(\alpha^{1/2}t)$$

we replace the coordinate r with the diminished value:

$$r' = pr.$$

This operation increases the effective velocity of any particle.

Next the time coordinate t is replaced by

$$t' = p^{-3/2}t$$

which decreases the effective velocity of a particle. Finally the whole solution is advanced by an amount

$$t'_0 = \alpha^{-1/2}p^{-3/2}a$$

which, during the contraction phase, increases the effective velocity at a given time.

Note

$$\frac{\partial R}{\partial t'} = p^{-1} r' \cdot \alpha^{1/2} p^{3/2} \dot{\xi} \, (\alpha^{1/2} p^{3/2} t' + a)$$

which shows that the second change dominates the first. However, the third operation must do more than restore the balance.

The time of collapse is therefore given by

$$t_c' = p^{-3/2} t_c - t_0'$$

$$= p^{-3/2} \alpha^{-1/2} \left(\frac{\pi}{2} - a\right) \qquad (34)$$

Of course, we are inserting the "dashed" coordinates merely in order to compare (16) and (9): we shall not use this notation again.

Using equation (a) of (19), we see that

$$a = \cos^{-1}\sqrt{p} + \sqrt{p(1-p)} \qquad (35)$$

since we take the positive square root during contraction.

We give two approximate forms of (35):

1. If $V << V_c$, we have

$$p \approx 1$$

and so

$$a \approx \cos^{-1}\left(1 - \frac{1}{2}\frac{V^2}{V_c^2}\right) + \frac{V}{V_c}\left(1 - \frac{1}{2}\frac{V^2}{V_c^2}\right)$$

$$\approx 2\frac{V}{V_c}\left(1 - \frac{1}{6}\frac{V^2}{V_c^2}\right)$$

2. If $V \approx V_c$,

$$\cos^{-1}\sqrt{p} \approx \frac{\pi}{2} - \sqrt{p}\,\left(1 + \frac{1}{6}p\right)$$

and

$$\sqrt{p(1-p)} \approx \sqrt{p}\,(1 - \tfrac{1}{2}p)$$

whence

$$a \approx \frac{\pi}{2} - \frac{2}{3}\,p^{3/2} \quad .$$

In case 1.,

$$t_c(v) \approx \left(1 + \frac{3}{2}\frac{V^2}{V_c^2}\right)\left(\frac{\pi}{2} - 2\frac{V}{V_c}\left[1 - \frac{1}{6}\frac{V^2}{V_c^2}\right]\right)\alpha^{-1/2}$$

which shows the dominance of the term due to "a".

To the first order

$$t_c(V) \approx \left(\frac{\pi}{2} - 2\frac{V}{V_c}\right)\alpha^{-1/2}$$

which shows that for small V, the time of collapse is reduced by the presence of V.

In case 2.,

$$t_c(V) \approx \frac{2}{3}\,\alpha^{-1/2}(1 + O(p))$$

so that

$$t_c(V) \to \frac{2}{3}\,\alpha^{-1/2}$$

$$V \to V_c \ .$$

There is only one solution of type (17), for given α.

The function

$$\eta(\tau)$$

is not periodic. From (b) of (19), we take

$$S(t) = (1 - \tfrac{3}{2}\,\alpha^{1/2}t)^{2/3} \tag{36}$$

when

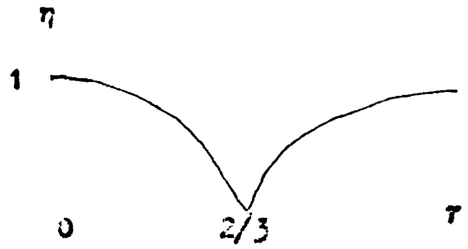
$$-\infty < t < \frac{2}{3}\alpha^{-1/2}$$

and

$$\eta(t) = \left(\frac{3}{2}\alpha^{1/2} t - 1\right)^{2/3} \qquad (36a)$$

when

$$t > \frac{2}{3}\alpha^{-1/2} .$$



The graph of η is symmetrical about the line

$$\tau = 2/3 .$$

Note that, using (36), collapse to a point singularity occurs after a time

$$t_c = 2/3\,\alpha^{-1/2} \quad . \qquad (37)$$

Thus if $V < V_0$

$$t_c(V) \to t_c(V_c)$$

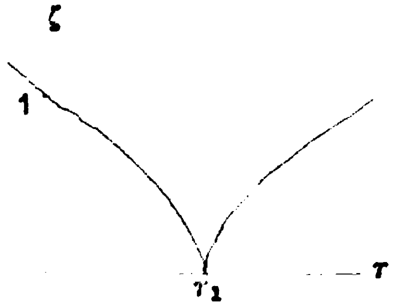
as might be expected.

This solution is by far the simplest analytically. Clearly it would be convenient to use it rather than the others.

The third class of solutions, (13), derives from the function

$$\zeta(\tau)$$

given by (8) or (19).



Here

$$\tau_1 = \sqrt{2} - \log_e(1 + \sqrt{2}) \ .$$

In general, we see that collapse to a point singularity occurs after a time

$$t_c = q^{-3/2}\alpha^{-1/2}(\sqrt{2} - \log(1 + \sqrt{2}) - b) \qquad (34a)$$

where

$$b = \sqrt{2} - \log(1 + \sqrt{2}) + \sinh^{-1}\sqrt{q} - \sqrt{q(1 + q)} \qquad (35a)$$

Note that when

$$V/V_0 \approx 1$$

we have

$$b \approx \sqrt{2} - \log(1 + \sqrt{2}) - 2/3\ q^{3/2}$$

so that

$$t_c \to \frac{2}{3}\alpha^{-1/2}$$

as

$$V \to V_0 \ .$$

When

$$\frac{V}{V_o} >> 1$$

we have

$$q >> \log_e q >> 0$$

giving

$$b \approx q$$

and

$$t_o \simeq (\alpha q)^{-1/2} .$$

We now come to the following question; does

$$t_o(V)$$

decrease as V increases? This appears obvious on physical grounds. However, reference to (34) or (34a) indicates that this is by no means obvious.

In (34), as V increases, p decreases, and so

$$p^{-3/2}$$

increases, whence it is not immediately clear that

$$t_c(V)$$

decreases. Also, in (34a), q increases with V: thus

$$q^{-3/2}$$

certainly decreases; however, we have (see (15a))

$$\zeta(b) = q$$

and so

$$\dot{\zeta}(b) \frac{db}{dq} = 1$$

and using the third equation (3):

$$\frac{db}{dq} = - \frac{1}{\sqrt{\frac{1}{\zeta(b)} + 1}}$$

$$= - \frac{\sqrt{q}}{\sqrt{1+q}}$$

since we are considering the expansion phase. Thus

$$\frac{d}{dq}(-b) > 0$$

whence the expression

$$\sqrt{2} - \log(1 + \sqrt{2}) - b$$

in (34a) increases with V. This latter fact could also be obtained from differentiation of (35a).

It therefore seems worthwhile to check this point. We shall only consider the case

$$V < V_c,$$

since the other case can be treated in similar fashion.

By (11)

$$\dot{S}^2 = \alpha\left(\frac{1}{S} - p\right)$$

and taking the negative square root, we find

$$\alpha^{1/2} t_c = \int_0^1 \frac{\sqrt{S}\, dS}{\sqrt{1-pS}} \quad . \qquad (38)$$

Thus

$$\alpha^{1/2} \frac{dt_c}{dp} = \int_0^1 \frac{S^{3/2}\, dS}{2(1-pS)^{3/2}}$$

and on integrating by parts,

$$= \frac{1}{p}\left[\left(\frac{S^{3/2}}{\sqrt{1-pS}}\right)_0^1 - \frac{3}{2}\int_0^1 \frac{\sqrt{S}\,dS}{\sqrt{1-pS}}\right]$$

$$= \frac{1}{p}\left[\frac{1}{\sqrt{1-p}} - \frac{3}{2}\alpha^{1/2}t_c\right] \tag{39}$$

(using (38)).

Thus

$$\alpha^{1/2}\left[\frac{dt_c}{dp} + \frac{3}{2p}t_c\right] = \frac{1}{p}\cdot\frac{1}{\sqrt{1-p}}$$

i.e

$$\frac{d}{dp}\left(\alpha^{1/2}p^{3/2}t_c\right) = \frac{\sqrt{p}}{\sqrt{1-p}}. \tag{40}$$

This could also have been obtained by using (34) and (35).

Write

$$Z = \frac{p^{3/2}}{\sqrt{1-p}} - \frac{3}{2}\alpha^{1/2}p^{3/2}t_c .$$

Then

$$\frac{dZ}{dp} = \frac{3}{2}\frac{\sqrt{p}}{\sqrt{1-p}} + \frac{1}{2}\frac{p^{3/2}}{(1-p)^{3/2}} - \frac{3}{2}\frac{\sqrt{p}}{\sqrt{1-p}} ,$$

the last term having been obtained from (40).

Hence

$$\frac{dZ}{dp} = \frac{1}{2}\frac{p^{3/2}}{(1-p)^{3/2}}$$

$$> 0 \text{ for } 0 < p \leqslant 1.$$

Thus Z is monotonically increasing for this range of p.

At

$$p = 0$$

we have

$$Z = 0$$

whence

$$Z > 0 \text{ in } 0 < p \leqslant 1$$

and so, using (39),

$$p^{3/2} . \alpha^{1/2} \frac{dt_c}{dp} > 0$$

in this range: since

$$p > 0$$

we deduce

$$\frac{dt_c}{dp} > 0 \tag{41}$$

and therefore

$$\frac{dt_c}{dV} < 0 \tag{41a}$$

which we required to prove.

Having discussed each homogeneous solution in detail, we now consider their behaviour near the singularity.

The elliptical solution may be written:

$$\cos^{-1}\sqrt{pS} + \sqrt{pS(1-pS)} = \alpha^{1/2} p^{3/2} t + \cos^{-1}\sqrt{p} + \sqrt{p(1-p)}$$

(see (16),(19)).

Put

$$t = t_0 - \alpha^{-1/2} \epsilon \tag{42}$$

Now, using (34), (35), we have

$$t_0 = p^{-3/2}\alpha^{-1/2}\left(\frac{\pi}{2} - \cos^{-1}\sqrt{p} - \sqrt{p(1-p)}\right) .$$

Thus

$$\cos^{-1}\sqrt{pS} + \sqrt{pS(1-pS)} = \frac{\pi}{2} - p^{3/2}\,\epsilon$$

i.e.

$$\sin^{-1}\sqrt{pS} - \sqrt{pS(1-pS)} = p^{3/2}\,\epsilon \quad .$$

Suppose that

$$S < < 1 .$$

Then

$$\sqrt{pS} + \frac{1}{6}(\sqrt{pS})^3 + \frac{3}{40}(\sqrt{pS})^5 - \sqrt{pS}\,[1 - \tfrac{1}{2}pS - \tfrac{1}{8}p^2S^2] = p^{3/2}\,\epsilon$$

i.e.

$$^2/_3 \; S^{3/2}(1 + \tfrac{1}{5}pS) = \epsilon$$

giving

$$S \approx \left(\frac{3}{2}\,\epsilon\right)^{2/3} \qquad (43)$$

neglecting terms of higher order in $\epsilon^{2/3}$.

However, putting (42) into the parabolic solution

$$S = (1 - \frac{3}{2}\,\alpha^{1/2}t)^{2/3}$$

we obtain equation (43), with an exact equality. The hyperbolic solution also gives (43) as an approximation.

For large

$$V/V_0$$

we must take S to be so small that

$$(qS) < < 1 .$$

Thus in any discussion of motion near the point singularity, we may, to all intents and purposes, consider the simplest, parabolic, solution only. This completes our consideration of the homogeneous solutions (16), (17), (18).

The inhomogeneous solutions satisfy equation (30). These are, as has been noted, just the homogeneous solutions, where now

$$\alpha, p, q, a, b$$

are functions of r.

There is now no one singular instant of time. Instead, we have a function

$$t_c(r) \qquad (33')$$

for which

$$\dot{S}(r, t_c)$$

first becomes infinite during the collapse. Note that

$$R(r, t_c)$$

vanishes, so we still get a singularity, though not really a "point" singularity. If we assume that the matter somehow ceases to exist when it reaches the point singularity, then the matter will steadily fall into the centre and disappear; if, however, matter emerges from the singularity, it will "pass through" matter still falling in. (The Relativistic treatment allows of other interpretations.) If

$$t_c(r)$$

is not monotonic increasing then it would seem that parts of the dust have overtaken other parts. We shall examine this question shortly.

It is unrealistic to assume that in extreme situations such as these, the dust fluid model, which ignores non-gravitational interactions between the particles, is physically acceptable.

We now consider this peculiar possibility of "overtaking". In this discussion we shall not take into account matter which has already reached the singularity and (possibly) come out again.

Let us study the function

$$\rho(r,t)$$

By equation (23)

$$\rho(r,t) = r^2 \rho_0(r)/R^2 R'$$

$$= \frac{\rho_0}{S^3}\left(1 + \frac{r}{S}\frac{\partial S}{\partial r}\right)^{-1}$$

Thus for given r, ρ becomes infinite either if

$$S(r,t) = 0$$

or if

$$1 + \frac{r}{S}\frac{\partial S}{\partial r} = 0 . \qquad (44)$$

If (44) is satisfied by some t^* such that

$$0 < t^* < t_0$$

then we might guess that overtaking has occurred. Indeed, if

$$1 + \frac{r}{S}\frac{\partial S}{\partial r} < 0$$

in any part of

$$t^* < t < t_0$$

"overtaking" must have occurred, for there is no other explanation of why $\bar{\rho}$, having become negative, should no longer represent the actual density of the fluid.

Firstly, let us suppose the body to be started from rest. Then, by (29)

$$S(r,t) = \xi(\alpha^{1/2} t) .$$

Hence

$$\frac{\partial S}{\partial r} = \frac{1}{2} \alpha^{1/2} \frac{d\alpha}{dr} t \dot{\xi} .$$

But

$$\alpha = \frac{8}{3} \pi G \bar{\rho}_0 (r)$$

giving

$$\frac{d\alpha}{dr} = \frac{8}{3} \pi G \frac{d\bar{\rho}_0}{dr} .$$

Now, according to (25),

$$\rho_0 - \bar{\rho}_0 = \frac{1}{3} r \frac{d\bar{\rho}_0}{dr}$$

so that

$$\frac{d\alpha}{dr} = \frac{8\pi G}{r} (\rho_0 - \bar{\rho}_0)$$

giving

$$\frac{\partial S}{\partial r} = \frac{4\pi G}{\alpha^{1/2} r} (\rho_0 - \bar{\rho}_0) t \dot{\xi}$$

$$= \frac{3}{2r} \left(\frac{\rho_0}{\bar{\rho}_0} - 1\right) \alpha^{1/2} t \dot{\xi} .$$

Hence

$$\frac{r}{R} \frac{\partial R}{\partial r} = 1 + \frac{3}{2} \left(\frac{\rho_0}{\bar{\rho}_0} - 1\right) \alpha^{1/2} t \, \dot{\xi}/\xi . \qquad (45)$$

Now, during contraction,

$$t > 0, \quad \dot{\xi} < 0$$

and therefore, if for all r,

$$\bar{\rho}_0 \geqslant \rho_0$$

R' will never vanish, and so the density never becomes infinite, except when

$$R = 0 .$$

Using (24),

$$\begin{aligned} \bar{\rho}_0 - \rho_0 &= \left(\frac{3}{r^3}\int_0^r r^2\rho_0\,dr\right) - \rho_0 \\ &= \frac{3}{r^3}\left(\int_0^r [r^2\rho_0 - \frac{d}{dr}(\tfrac{1}{3}r^3\rho_0)]dr\right) \\ &= -\frac{1}{r^3}\int_0^r r^3\frac{d\rho_0}{dr}\,dr \quad . \end{aligned} \qquad (46)$$

Thus a sufficient condition for non-singular density is that

$$\frac{d\rho_0}{dr} \leqslant 0 .$$

This is not, of course, a necessary condition.

Density singularities may arise when some region has an "excess of matter".

If in any region

$$\rho_0 > \bar{\rho}_0$$

then when $t = 0$,

$$1 + \frac{r}{S}\frac{\partial S}{\partial r} = 1.$$

Now

$$t/\xi \; \dot{\xi}$$

is a monotonically decreasing function of t for

$$0 \leqslant t < \frac{\pi}{2} \; \alpha^{-1/2}$$

and as

$$t \to \frac{\pi}{2} \alpha^{-1/2} \; ,$$

$$1 + \frac{r}{s} \frac{\partial S}{\partial r} \to -\infty$$

in this region.

Thus

$$\frac{\partial R}{\partial r}$$

vanishes for precisely one value of t during the collapse.

Hence any excess of matter will give rise to density singularities. Note further that in this case, if we suppose

$$R'(r_0, t^*(r_0)) = 0$$

then, since R' is a monotonic decreasing function of t for fixed r, we have

$$R'(r_0, t) < 0$$

for

$$t > t^* \; .$$

Fix

$$t = t_1$$

where

$$t^* < t_1 < t_c \; .$$

Then

$$R'(r_0,t_1) < 0.$$

But by (24) and (27), α is a continuous function of r even if ρ_0 is only piecewise continuous. Thus

$$R(r,t_1)$$

is a continuous function of r.

Now

$$R(0,t_1) = [r\xi(\alpha^{1/2}t)]_{r=0}$$

$$= 0$$

whereas

$$R(r_0,t_1) > 0$$

since

$$t_1 < t_c \quad .$$

It follows that there must exist a range of r, within

$$(0,r_0)$$

for which

$$R'(\tilde{r},t_1) > 0$$

whenever $\tilde{r}$ lies in this range. We deduce the existence of at least one $\tilde{r}_0$ in this range, for which

$$R(\tilde{r}_0,t_1) = R(r_0,t_1)$$

and so overtaking does indeed occur at this time.

Since here

$$\rho(r_0, t_1) < 0$$

this confirms our conjecture that "negative densities" really imply that overtaking is occurring, and has occurred. Furthermore, if ρ_0 is continuous, then by (45), so is R', and in this case there is a whole range of r for which overtaking occurs at this time.

If however

$$\bar{\rho}_0 \geqslant \rho_0$$

then by (45), R' is always positive for

$$t < t_c$$

and so R is a monotonic increasing function of r for fixed t; thus here overtaking never occurs.

There is one final query in this connection: is our criterion

$$R(r_1, t) = R(r_2, t) \tag{47}$$

sufficient for the phenomenon of "overtaking"? The only possible counterexample is when we have, in addition

$$\dot{R}(r_1, t) = \dot{R}(r_2, t) \tag{48}$$

$$[-\ddot{R}(r_1, t)] \geqslant [-\ddot{R}(r_2, t)] \tag{49}$$

taking

$$r_1 < r_2 ,$$

for then particles at r_1, r_2 would touch momentarily, and then recede relatively, without the r_2 particle overtaking the r_1 particle.

Write

$$R(r_1, t) \equiv R_1$$

etc.

Now

$$\ddot{R}_1 = -G\frac{M_{r_1}}{R_1^2}$$

$$= -\frac{1}{2}\frac{\alpha_1 r_1^3}{R_1^2}$$

using (24) and (27): similarly for $\ddot{R}_2$.

Thus

$$\ddot{R}_2 - \ddot{R}_1 = -\frac{1}{2}\cdot\frac{1}{R_1^2}(\alpha_2 r_2^3 - \alpha_1 r_1^3)$$

using (47). But

$$\alpha r^3 = 2GM_r$$

(see above), whence, since ρ_0 is non-negative

$$\alpha_2 r_2^3 \geqslant \alpha_1 r_1^3 .$$

Thus

$$\ddot{R}_2 - \ddot{R}_1 \leqslant 0$$

equality occurring only if initially there is no matter in the region between r_1, r_2.

Hence (49) cannot be satisfied, except if

$$\alpha_2 r_2^3 = \alpha_1 r_1^3 . \tag{50}$$

Even if this were so, and so (49) were satisfied with equality, we might still get "overtaking", depending upon the value of the next derivative. However, note

$$R_1 = r_1 \xi (\alpha_1^{1/2} t)$$

(see (2)), giving

$$\dot{R}_1 = -r_1 \, . \, \alpha_1^{1/2} \sqrt{\frac{1}{\xi} - 1}$$

$$= -\alpha_1^{1/2} r_1 \sqrt{\frac{r_1}{R_1} - 1} \qquad .$$

Thus, squaring both sides of (48),

$$\alpha_1 r_1^2 \left(\frac{r_1}{R_1} - 1 \right) = \alpha_2 r_2^2 \left(\frac{r_2}{R_2} - 1 \right)$$

and using (47):

$$\alpha_1 r_1^2 - \alpha_2 r_2^2 = \frac{1}{R_1} (\alpha_1 r_1^3 - \alpha_2 r_2^3)$$

so that, if (50) holds,

$$\alpha_1 r_1^2 - \alpha_2 r_2^2 = 0 \qquad (51)$$

and combining (50 and (51) gives

$$r_1 = r_2 ,$$

since r is not zero (otherwise R is). Thus (47) is indeed sufficient for overtaking to occur.

We could now proceed to a similar treatment of the more general elliptic solution. We would then obtain an equation similar to (45); unfortunately, however, the function R' is no longer monotonic in t, which makes the problem somewhat intractable analytically. A graphical treatment is possible, but the question does not seem to be of sufficient interest to justify the complications involved.

This chapter will be concluded by an examination of the "exterior solution". The results here have been known since the time of Newton – we are concerned merely with writing them in a manner which affords ready comparison with the corresponding Relativistic results.

What meaning can be given to the equation of motion when in some region

$$\rho_0(r) = 0 \; ?$$

Clearly it must refer to the motion of a test particle, of negligible mass, in the space within or exterior to the body in question. Since Newtonian Gravitation is a linear theory, we may confidently assert that a particle of small mass will not appreciably affect the gravitational field of the body itself. Such a particle, then, will follow a trajectory familiar from discussions of Newtonian Orbits. To give meaning to equations (16), (17), (18), we must assume that the test particle will share the symmetries of the general motion. This implies firstly that the initial, and therefore subsequent, motion will be radial.

In matching the interior and exterior "solutions" we would wish to impose an element of continuity in the behaviour of the particle vis-a-vis the dust fluid. If initially the fluid is at <u>rest</u>, we shall assume that the particle is likewise. If the fluid is homogeneous throughout the motion, then (cf.(11),(30))

$$V(r) = \frac{r}{r_b} V \qquad (r \leq r_b)$$

where V is constant. Here we need merely extend this condition to exterior test particles.

In general, if the initial velocity distribution is analytic (C^ω continuous) then the velocity-distribution function $V(r)$ can be uniquely continued analytically to the exterior. If, however, $V(r)$ is only C^K continuous at the boundary, for some K, then this extension is no longer unique: we could specify any function $V(r)$ in the exteriorregion which was C^K continuous at the boundary.

Suppose then that $V(r)$ is given throughout. When

$$r \geq r_b,$$

$$M_r = M_{r_b} = M$$

and (see (24)),

$$\rho_0(r) = \frac{3M}{4\pi r^3} \quad . \tag{52}$$

Then (27) becomes

$$\alpha(r) = \frac{2GM}{r^3} \tag{53}$$

or equally

$$= \left(\frac{r_b}{r}\right)^3 \cdot \alpha \tag{54}$$

where

$$\alpha = \bar{\rho}_0(r_b) \cdot \frac{8}{3}\pi G \quad .$$

In the homogeneous case, this last equation reduces to (7). Equation (50) applies generally, as do its solutions, using (16), together with (31) and (32), or their equivalents. If the initial velocity is zero throughout, the solution can be conveniently summarized as follows:

$$S = \xi\,(\alpha^{1/2}(r)t) \qquad (r \leq r_b)$$

$$= \xi\,((\frac{r_b}{r})^{3/2}\alpha^{1/2}t) \qquad (r \geq r_b)\ .$$

Note that S is an analytic function of r for $r > 0$, $t < t_c$ even though ρ may be discontinuous.

Generally, if the specified velocity distribution, $V(r)$ is C^k continuous, then the function p (or q) given by (31) is also C^k, and so the function S, for fixed t, is C^k continuous for

$$t < t_c(r).$$

Note that if the motion is from rest, (33) becomes, in the exterior

$$t_c = \frac{\pi}{2}\left(\frac{2GM}{r^3}\right)^{-1/2}$$

which is the standard result for the time taken for a test particle to fall onto particle of mass M.

Finally, we note that the gravitational potential is PIECEWISE C^1 continuous throughout, even if ρ_0 is only piecewise continuous. For

$$\Omega = -\frac{GM_r}{R} \qquad (35)$$

and

$$\frac{dM_r}{dr} = 4\pi r^2 \rho_0\ .$$

This result will prove useful in comparing the Classical and Relativistic approaches to the problem.

CHAPTER III

RELATIVISTIC EQUATIONS AND THEIR SOLUTIONS

The problem of the Relativistic dust-fluid sphere has been discussed in some detail since 1938; in particular, we may note the treatments of DATT (REF. 7), OPPENHEIMER & SNYDER (REF. 37), and more recently HOYLE & NARLIKAR (REF. 18) and NARIAI & TOMITA (REF. 30). Some of these authors consider only the case of uniform density; we shall discuss these articles in due course.

Throughout most of this chapter we shall be using the framework of "classical" General Relativity. This is essentially a local, rather than a global theory, in that we do not enquire too closely into the domain of validity of the particular system of coordinates employed, and assume implicitly an Euclidean topology of space-time (see Chapter V). In our later consideration of the problem of boundary conditions, and that of the "fate" of a collapsing dust-fluid, it will be seen that this framework is inadequate for a proper understanding. In order to tackle these problems effectively, we shall devote Chapter V to a development of General Relativity on the basis of modern differential geometry, though we shall exercise reasonable economy in the use of advanced concepts.

The first section of this chapter will be devoted to an exposition of the "tetrad" formalism in General Relativity, in order to obtain the standard expressions for the "stress-energy" tensor of the dust-fluid, and, more generally, of the perfect fluid. The present author would not wish to claim particular originality for this treatment. We next proceed to an examination of the form of the perfect fluid stress-energy tensor: the interpretation of the two scalars p, ρ is somewhat obscure, and does not appear to have been satisfactorily discussed in the literature. The question of whether ρ is the mass-density of classical physics is of some importance, especially if we wish to specify an equation of state, functionally relating p, ρ. Our method is to compare the Relativistic conservation equations with ones derived on a "semi-classical" basis; we conclude firstly that the pressure-field contributes a negative energy-density in the instantaneous rest-frame of the fluid; secondly, that ρ represents the <u>total</u> rest-frame energy-density, including the pressure field density and the density of inertial mass; lastly that the <u>number density</u> of fluid particles is proportional to a certain functional of p, ρ, and that it is this quantity which should be used in equations of state.

We then proceed to our principal task, which is the general solution of the Einstein field equations for a dust-fluid sphere. The approach appears to be original in that the analogy between the Newtonian and Relativistic solutions, well known in the homogeneous case (REF. 25) is extended to the general case. We shall also show why two types of coordinate system appear to be of fundamental importance.

The following conventions will be used: Greek indices will range in value from one to three and Roman indices from one to four. We shall interpret $\{x^\alpha\}$ as space coordinates, and x^4 as a time coordinate. The Einstein summation convention will be used throughout. We shall assume that the geometry of space-time is Riemannian, with normal hyperbolic metric - see LICHNEROWICZ (REF. 26). The signature will be taken to be $(+---)$. A vector u^a will be called TIMELIKE, NULL or SPACELIKE according as

$$g_{ab}\, u^a u^b \gtreqless 0$$

respectively. A hypersurface is said to be <u>null</u> if its normal is null; otherwise it is <u>timelike</u> if its normal is spacelike, and vice versa.

Physically, we shall assume the existence of a "stress-energy" tensor T_{ij}, which in some manner describes the distribution of matter throughout space-time, and satisfies the <u>conservation equations</u>

$$T^{ij}{}_{;j} = 0, \tag{1}$$

where the semicolon indicates covariant differentiation with respect to the Riemannian connexion (whereas a comma will indicate straightforward partial differentiation). We shall characterize the perfect fluid (with the dust-fluid as a special case) by reference to the eigenvalues of T_{ij}.

A tensor $F_{k\ell}$ is said to have an EIGENVALUE λ, with corresponding EIGENVECTOR p^i, if

$$F_{k\ell}\, p^\ell = \lambda p_k \quad . \tag{2}$$

It follows that λ must satisfy the EIGENVALUE EQUATION

$$\det[F_{k\ell} - \lambda g_{k\ell}] = 0 \quad . \tag{3}$$

If we multiply (2) by g^{ik}, we obtain

$$F^i{}_\ell\, p^\ell = \lambda p^i \quad ,$$

and instead of (3)

$$\det[F^i{}_l - \lambda \delta^i{}_l] = 0 .$$

This is the "mixed" form of the eigenvalue equation, and corresponds to familiar algebraic notions concerning endomorphisms. A contravariant form of these equations may be obtained similarly.

The tensor T_{ij} is by definition symmetric, and therefore possesses real eigenvalues. Note however that the mixed tensor $T^i{}_j$ is not in general symmetric, though of course it possesses the same eigenvalues as does T_{ij}.

The characterization of the Relativistic dust-fluid seems quite straightforward: it is completely specified by a scalar ρ, and a four-vector-field u^i, being the rest density and four-velocity of the fluid respectively. We shall therefore suppose that the tensor T_{ij} is degenerate, in that it has only one non-zero eigenvalue ρ, with corresponding time-like eigenvector u^i. At any point, all vectors in the (tangent) hyperplane perpendicular to u^i are then eigenvectors with corresponding eigenvalue zero. Such vectors must be spacelike; "perpendicular" here implies the vanishing of the metric scalar product - i.e. $v^i \perp u^i$ means

$$g_{ij} v^i u^j = 0.$$

The Relativistic perfect fluid, however, seems less straightforward. It is physically specified by two scalars, density and pressure, as well as the four-velocity vector field. Now, as we shall see, due to the Relativistic interaction (at the macroscopic level) between the matter, and the pressure field, there are various candidates for the title "density", and it may not turn out to be reasonable to expect the rest-density of inertial mass to be one of the eigenvalues of T_{ij}. However, we may define a perfect fluid as being specified by a stress-energy tensor which is degenerate in having only two eigenvalues; One will correspond to eigenvector u^i, and will be written ρ; all vectors perpendicular to u^i are eigenvectors corresponding to the other eigenvalue, which for reasons which will become apparent, we denote by $-p$.

Our next task is to obtain the algebraic form of T_{ij} in either of the cases discussed above (naturally, the dust-fluid is a degenerate case of the perfect fluid). In order to derive the form of T_{ij} in a straightforward fashion, it seems worthwhile to explain the use of the TETRAD formalism in General Relativity (cf. REF. 41).

In this section, we shall use a system of units in which the velocity of light is to be "understood", meaning that a unit of time is choasen, and lengths are to be measured in units of light-time. We shall also assume that u^i has been normalized, by the requirement that

$$g_{ij}u^i u^j = +1 . \qquad (4)$$

We may similarly normalise a spacelike eigenvector v^i by

$$g_{ij}v^i v^j = -1 . \qquad (5)$$

Generally, a vector-field u^i defines a congruence of "integral curves" – the world-lines of particles, such that u^i is everywhere tangent to the appropriate member of the congruence. When normalized, u^i is the unit tangent vector, $\frac{dx^i}{ds}$, to the world-line in question.

The cornerstone of the method is the use of four linearly independent vector-fields (vierbeins). We denote them by $\underset{a}{v}^i$ (a = 1 – 4), the index a being mere labelling, whilst i is a tensor index. For our purpose we choose $\underset{a}{v}^i$ to be orthonormal, and put

$$\underset{4}{v}^i = u^i, \tag{6}$$

the $\underset{\alpha}{v}^i$ being three mutually orthogonal unit spacelike vector-fields. In terms of $\underset{a}{v}^i$, we define the TETRAD METRIC $\underset{ab}{g}$ by

$$\underset{ab}{g} = \underset{a}{v}^i \underset{b}{v}_i = g_{ij} \underset{a}{v}^i \underset{b}{v}^j \quad . \tag{7}$$

Using (4), (5) and the orthogonality property, we see that in this case the $\underset{ab}{g}$ are the coefficients of the Minkowski metric diag (1,-1,-1,-1). This approach enables us to discuss algebraic properties of tensors in General Relativity in the language of Special Relativity.

The next step is to define $\overset{ab}{g}$ to be the components of the matrix inverse to $\underset{ab}{g}$; in our case the components of $\underset{ab}{g}$, $\overset{ab}{g}$ are identical. This enables us to define "upper" tetrad vectors $\overset{a}{v}^i$ (and similarly $\overset{a}{v}_i$) by

$$\overset{a}{v}^i = \overset{ab}{g} \underset{b}{v}^i .$$

We are now in a position to define upper and lower TETRAD COMPONENTS of a vector or tensor field:

$$\underset{a}{p} = p^i \underset{a}{v}_i = p_i \underset{a}{v}^i$$

$$\underset{ab}{F} = F_{ij} \underset{a}{v}^i \underset{b}{v}^j$$

$$\overset{abc}{K} = K_{ijk} \overset{a}{v}^i \overset{b}{v}^j \overset{c}{v}^k$$

and so on. Note that the tetrad components are scalars, whereas the tensor components are not.

A most important result is the following:

$$\overset{a}{v}{}^{i}\,\underset{a}{v}{}_{j} = \delta^{i}_{j} \qquad (8)$$

where the summation convention is understood for tetrad indices.

PROOF:

Using (7), we have

$$(\overset{a}{v}{}^{i}\,\underset{a}{v}{}_{j})\,\underset{b}{v}{}^{j} = \overset{a}{v}{}^{i}\,g_{ab}$$
$$= \underset{b}{v}{}^{i} \quad .$$

Thus $\overset{a}{v}{}^{i}\,\underset{a}{v}{}_{j}$ acts as the unit operator when operating on any tetrad vector. But the latter together form a basis, so that the expression acts as the unit operator when operating on <u>any</u> vector. Hence result.

A corollary is that any tensor may be "recovered" from its tetrad components:

$$K_{ijk\ldots} = \underset{abc\ldots}{K}\,\overset{a}{v}{}_{i}\,\overset{b}{v}{}_{j}\,\overset{c}{v}{}_{k}\;\ldots \qquad (9)$$

PROOF:

$$\underset{abc\ldots}{K}\,\overset{a}{v}{}_{i}\,\overset{b}{v}{}_{j}\,\overset{c}{v}{}_{k} = (K_{pqr\ldots}\,\underset{a}{v}{}^{p}\,\underset{b}{v}{}^{q}\,\underset{c}{v}{}^{r}\ldots)\,\overset{a}{v}{}_{i}\,\overset{b}{v}{}_{j}\,\overset{c}{v}{}_{k}\ldots$$
$$= K_{pqr\ldots}\,\delta^{p}_{i}\,\delta^{q}_{j}\,\delta^{r}_{k}\ldots \quad ,$$

using (8). Hence result.

A further consequence of (8) is that the expression $\overset{a}{v}{}_{i}\,\underset{a}{v}{}_{j}$ may reasonably be called the PROJECTION OPERATOR onto the hyperplane spanned by the $\underset{a}{v}_i$. For, if k^i is any vector, the expression

$$k^i \overset{a}{v}_i \underset{a}{v}_j$$

is a linear combination of the v_j, and hence is a vector in the hyperplane; whereas

$$k^i \overset{4}{v}_i \underset{4}{v}_j$$

is a vector in the direction of $\underset{4}{v}_j$. Furthermore,

$$\begin{aligned} k^i \overset{4}{v}_i \underset{4}{v}_j + k^i \overset{\alpha}{v}_i \underset{\alpha}{v}_j &= k^i \overset{a}{v}_i \underset{a}{v}_j \\ &= k^i g_{jk} \delta^k_i \\ &= k_j , \end{aligned}$$

so that k_j has been resolved into a component in the hyperplane, and one along $\underset{4}{v}_j$. All these results apply to a general tetrad basis; however, in our case we have in addition

$$\overset{44}{g} = 1 ,$$

so that

$$\overset{4}{v}_i = u_i \quad .$$

Hence

$$\begin{aligned} \overset{\alpha}{v}_i \underset{\alpha}{v}_j &= \overset{a}{v}_i \underset{a}{v}_j - \overset{4}{v}_i \underset{4}{v}_j \\ &= g_{ij} - u_i u_j \qquad (10) \end{aligned}$$

which is therefore the projection operator onto the plane perpendicular to u_i (cf. SYNGE, REF. 47).

Equation (8) therefore seems fundamental; in fact, the tetrad notation has served merely to disguise the fact that it is a standard result in elementary linear algebra. Regarding vector and tensor fields as being the prescription of a vector or tensor in each tangent space (see Chapter V), and using the usual notation for vectors, we may introduce the BASIS VECTOR FIELDS, $\underset{\sim}{e}_a$, associated with the coordinate system $\{x^\iota\}$, these fields having components

$$e_\alpha{}^\iota = \delta^\iota_a ,$$

where a is to be regarded as a tetrad, not a tensor, index. The metric induces a scalar product - strictly, a pseudoscalar product - in each tangent space, such that:

$$\underset{\sim}{e}_p \cdot \underset{\sim}{e}_q = g_{pq} ,$$

whence

$$\underset{\sim}{a} \cdot \underset{\sim}{b} = g_{\iota\jmath}\, a^\iota\, b^\jmath$$

Since the set of vectors $\underset{\sim}{v}_a$, being the tetrad with components $\underset{a}{v}^\iota$, form an orthonormal set in the case of interest to us, we would expect that

$$\underset{\sim}{e}_\iota = \sum_a (\underset{\sim}{e}_\iota \cdot \underset{\sim}{v}_a)\, \underset{\sim}{v}_a ,$$

$\{\underset{\sim}{v}_a\}$ being an orthonormal basis in each tangent space.

This equation is, however, incorrect, since we are dealing with a pseudo-scalar product; it must be modified to read:

$$\underset{\sim}{e}_\iota = (\underset{\sim}{e}_\iota \cdot \underset{\sim}{u})\underset{\sim}{u} - \Sigma(\underset{\sim}{e}_\iota \cdot \underset{\sim}{v}_\alpha)\underset{\sim}{v}_\alpha ,$$

as may readily be established by taking products with each of the $\underset{\sim}{v}_a$ in turn. This latter equation is equivalent to (8), with $\underset{\sim}{v}_a$ orthonormal, since

$$\underset{\sim}{e}_i \cdot \underset{\sim}{u} = g_{ik}\, u^k$$

$$= u_i \,,$$

so that, in terms of components with respect to the $\{\underset{\sim}{e}_j\}$,

$$\underset{i}{e}{}^j = u_i u^j - \sum_\alpha \underset{\alpha}{v}_i \, \underset{\alpha}{v}{}^j \quad .$$

But $e_i{}^j = \delta_i{}^j$; $\overset{4}{v}{}^i = u^i$, and $\overset{\alpha}{v}{}^i = -\underset{\alpha}{v}{}^i$, so that

$$\delta_i{}^j = \overset{a}{v}_i \, \underset{a}{v}{}^j$$

as required.

This has explained the significance of (8) in the case of orthonormal $\{\underset{\sim}{v}_a\}$; it remains to discuss the general case in the language of linear algebra. The tetrad metric $\underset{ab}{g}$ is given by

$$\underset{ab}{g} = \underset{\sim}{v}_a \cdot \underset{\sim}{v}_b$$

which are just metrical coefficients in the tangent space, derived from a basis $\{\underset{\sim}{v}_a\}$ instead of the basis $\{\underset{\sim}{e}_a\}$. We may define $\overset{ab}{g}$ as the inverse of $\underset{ab}{g}$, and $\{\underset{\sim}{v}{}^a\}$ may then be defined as the reciprocal basis to $\{\underset{\sim}{v}_a\}$, so that

$$\underset{\sim}{v}{}^a = \overset{ab}{g}\, \underset{\sim}{v}_b \;;$$

alternatively, we may use a far more powerful approach, and define $\{\underset{\sim}{v}{}^a\}$ to be the <u>dual</u> basis corresponding to $\{\underset{\sim}{v}_a\}$. In the latter approach,

$\underset{\sim}{v}^a$ is that linear functional for which

$$\langle \underset{\sim}{v}^a, \underset{\sim}{v}_b \rangle = \delta^a{}_b \quad .$$

Given the scalar product, it is easy to establish an isomorphism of the tangent space and its dual, under which the reciprocal basis and the dual basis correspond, so that in a sense both treatments are equivalent. We conclude that the expression

$$\underset{\sim}{v}^a \cdot \underset{\sim}{v}_b$$

may be interpreted either as a scalar product between vectors, or as a natural scalar product between a dual vectorand a vector (for further details, see Chapter V). Whichever interpretation we use, it is clear that we can write

$$\underset{\sim}{e}_i = \sum_a (\underset{\sim}{v}^a \cdot \underset{\sim}{e}_i)\, \underset{\sim}{v}_a \,, \tag{11}$$

since

$$\underset{\sim}{v}^a \cdot \underset{\sim}{v}_b = \delta^a{}_b .$$

However, equation (8) is obtained from equation (11) merely by taking j^{th} components. This completes our identification of tetrad-component results with standard vector ones.

Armed with this tetrad apparatus, it is easy to establish the algebraic structure of T_{ij}. Consider firstly, the dust-fluid, for which ρ is the sole non-zero eigenvalue. Here

$$T_{ik} \underset{a}{v}^k = \rho \underset{a4}{\delta} \underset{a}{v}_i \tag{12}$$

(no summation over a).

Using (8), (12),

$$T_{ik}\,\delta^k{}_j = \rho\,\delta_{a4}\,v_i\,\overset{a}{v}_j$$

$$= \rho u_i u_j ,$$

since $\overset{44}{g} = 1$ here. Thus

$$T_{ij} = \rho u_i u_j \quad . \tag{13}$$

For the perfect fluid, let ρ be the eigenvalue corresponding to u_i, and $(-p)$ be that corresponding to any $\underset{a}{v}_i$. Then

$$T_{ik}\,u^k = \rho\,u_i$$

and

$$T_{ik}\,\underset{\alpha}{v}^k = -p\,\underset{\alpha}{v}_i \quad .$$

Thus

$$T_{ij} = T_{ik}\,\overset{a}{v}_j\,\underset{a}{v}^k$$

(using (3)), whence

$$T_{ij} = \rho u_i u_j - p\,\underset{\alpha}{v}_i\,\overset{\alpha}{v}_j$$

$$= (\rho+p)u_i u_j - p\,g_{ij} \tag{14}$$

using (12). Of course, (13) is obtained from (14) by taking $p = 0$.

In order to interpret ρ and p in physical terms, we need to recall the significance of the electromagnetic stress-energy tensor in special relativity. Here we are dealing with Cartesian tensors in Minkowski space-time. Keeping to a system of units in which c is understood, the component T_{44} represents the density of electromagnetic field energy, namely,

$$\tfrac{1}{2}\,(\mathbf{D}\cdot\mathbf{E} + \mathbf{H}\cdot\mathbf{B});$$

$-T_{4\alpha}$ represents the 3-momentum (or energy flux); and $T_{\alpha\beta}$ is the electromagnetic stress - in any reference frame. By analogy, T_{44}, for any distribution of matter, is to be interpreted as an energy density, and $T_{\alpha\beta}$ as a three-dimensional stress tensor. (~~REF.~~)

All this may be carried over into General Relativity as follows: let $\underset{4}{w}^i$ be an arbitrary unit timelike vector-field, and complete an orthonormal tetrad with the addition of three mutually orthogonal vector-fields $\underset{\alpha}{w}^i$; these are determined up to a 3-rotation. Then the components of T_{ij} on this tetrad are to be interpreted as in Special Relativity, the "frame" referring to an observer with 4-velocity $\underset{4}{w}^i$. In the special case $\underset{4}{w}^i = u^i$, $\underset{44}{T}$ therefore represents the rest energy-density, $\underset{4\alpha}{T}$ should vanish, and $\underset{\alpha\beta}{T}$ represents the rest stress-tensor; for the perfect fluid this must have eigenvalues $(-p,-p,-p)$, where p is the pressure - this justifies the identification of the p in (14) with physical pressure.

The interpretation of ρ remains uncertain: the phrase "rest-energy" may imply mass density, or else a total energy arising from the interaction of matter with the pressure field. This uncertainty gives rise to some confusion in the literature (see REF. 49). To the best of the present author's knowledge, this question has not as yet been satisfactorily resolved: we devote the next section to an attempt to decide the identification of "inertial mass-density" and "particle number-density". Our results appear, in particular, to conflict with those of DOLAN (REF. 8), who uses variational techniques.

THE PERFECT FLUID

The equations of motion for the perfect fluid are a consequence of the conservation of T_{ij} (cf. REF. 49). According to (14)

$$T^{ij}{}_{;j} = [(\rho+p)u^i u^j]_{;j} - p^{,j}$$

$$= (\rho+p)_{,j} u^i u^j + (\rho+p)[u^i{}_{;j} u^j + u^i u^j{}_{;j}] - p^{,i} . \qquad (15)$$

Now write, as is customary (see e.g. REF. 20)

$$\frac{D}{ds} = \frac{\delta}{\delta x^i} \frac{dx^i}{ds}$$

$$= u^i \frac{\delta}{\delta x^i} , \qquad (16)$$

where $\frac{\delta}{\delta x^i}$ is the covariant derivative w.r.t. x^i. Then, since the covariant derivative of a scalar is just the partial derivative, (15) becomes

$$T^{ij}{}_{;j} = \frac{D(\rho+p)}{ds} u^i + (\rho+p)\left[\frac{Du^i}{ds} + (\mathrm{div}_4 \underset{\sim}{u})u^i\right] - p^{,i}, \qquad (17)$$

$\mathrm{div}_4 \underset{\sim}{u}$ being simply $u^j{}_{;j}$. Equating (17) to zero; multiplying by u_i and contracting then gives

$$0 = \frac{D(\rho+p)}{ds} u_i u^i + (\rho+p)\left[u_i \frac{Du^i}{ds} + (\mathrm{div}_4 \underset{\sim}{u})u_i u^i\right] - p^{,i} u_i$$

$$= \frac{D(\rho+p)}{ds} + (\rho+p)\left[u^i \frac{Du^i}{ds} + (\mathrm{div}_4 \underset{\sim}{u})\right] - \frac{Dp}{ds} . \qquad (18)$$

But
$$\frac{D}{ds}(u_i u^i) = u_i \frac{Du^i}{ds} + u^i \frac{Du_i}{ds}$$

$$= u_i \frac{Du^i}{ds} + u^i \frac{D}{ds}(g_{ik} u^k)$$

$$= 2u_i \frac{Du^i}{ds} ,$$

since the covariant derivatives of g_{ik} vanish. Hence

$$u_i \frac{Du^i}{ds} = 0 ;$$

substituting in (18),

$$\frac{Dp}{ds} + (\rho+p)\, \mathrm{div}_4 \underset{\sim}{u} = 0 \quad . \tag{19}$$

Equation (19) may be termed the "equation of continuity".

If (19) be now inserted into (17), equated to zero, one finds

$$\frac{D(\rho+p)}{ds} u^i + (\rho+p) \frac{Du^i}{ds} - \frac{Dp}{ds} u^i - p^{,i} = 0$$

whence

$$(\rho+p) \frac{Du^i}{ds} = p^{,i} - \frac{Dp}{ds} u^i \quad . \tag{20}$$

The equations (20) may be termed "equations of motion".

Thus far, we have merely obtained results which are standard in the literature. In order to proceed further, it is useful to revert to a system of coordinates in which the velocity of light appears explicitly. In this case, (19) and (20) should be modified to read:

$$\frac{Dp}{ds} + (\rho+p/c^2)\, \mathrm{div}_4 \underset{\sim}{u} = 0 \tag{21}$$

$$(\rho+p/c^2) \frac{Du^i}{ds} = \frac{1}{c^2} \left(p^{,i} - \frac{Dp}{ds} u^i\right) \tag{22}$$

where we interpret ds as proper time, for a timelike displacement.

As a first step in the interpretation of these equations we consider the non-relativistic, flat-space approximation, putting $u^4 = 1$, and assuming c to be negligible, and in particular:

$$\rho >> p/c^2 .$$

Here (21) becomes

$$\frac{Dp}{dt} = -\rho \operatorname{div} \underset{\sim}{q}$$

and (22) becomes, for i = 1,2,3

$$\rho \frac{D\underset{\sim}{q}}{dt} = -\underset{\sim}{\nabla}p, \tag{23}$$

since

$$p^{,\alpha} = g_M^{\alpha\alpha} \frac{\partial p}{\partial x^\alpha}$$

$$= -c^2 (\underset{\sim}{\nabla}p)_\alpha ,$$

the subscript "M" indicating Minkowskian metric components. These are just the classical equations of motion and continuity. However, the equation (22), with i = 4, requires more care. Put

$$\beta = (1 - q^2/c^2)^{-\frac{1}{2}} , \tag{24}$$

where of course

$$q = |\underset{\sim}{q}|,$$

and

$$\underset{\sim}{q} = \{\frac{dx^\alpha}{dt}\} .$$

Then

$$u^4 = \frac{dt}{ds} = \beta .$$

Thus the fourth equation of motion may be written:

$$(\rho + p/c^2)\frac{D\beta}{ds} = \frac{1}{c^2}\left(\frac{\partial p}{\partial t} - \beta \frac{Dp}{ds}\right) . \tag{25}$$

Now, according to (24)

$$\frac{D\beta}{ds} = \beta^3 q/c^2 \frac{Dq}{ds} , \tag{16}$$

which, inserted into (25), gives

$$\beta^3(\rho+p/c^2)\, q\frac{Dq}{ds} = \frac{\partial p}{\partial t} - \beta^2 \frac{\partial p}{\partial t} - \beta^2 \underset{\sim}{q} \cdot \underset{\sim}{\nabla} p , \tag{27}$$

since

$$\frac{D}{ds} = \beta \frac{D}{dt} = \beta \left(\frac{\partial}{\partial t} + \underset{\sim}{q} \cdot \nabla\right) .$$

We shall consider the full equation (17) shortly; for the present, we are interested in the non-relativistic approximation:

$$\rho q \frac{Dq}{dt} = - \underset{\sim}{q} \cdot \underset{\sim}{\nabla} p .$$

This equation states that the rate of change of kinetic energy/unit mass equals minus the rate of working of the pressure; it may, of course, be derived from (23) merely by scalar multiplication by $\underset{\sim}{q}$. Thus we recoup merely the three Eulerian equations of motion.

This completes ~~are~~ our examination of the "classical limit" $c \to \infty$. Clearly there should only be three independent relativistic equations of motion. Consider first the position in flat space-time. Equations (22) with i = 1,2,3, give

$$(\rho + p/c^2)\frac{D(\beta q)}{ds} = - \underset{\sim}{\nabla} p - \beta/c^2 \frac{Dp}{ds} \underset{\sim}{q} \tag{28}$$

and it may be conjectured that, without resorting to any approximations, (27) may be deduced from (23) simply by scalar multiplication by $\underset{\sim}{q}$.

To check this, we note that

$$\underset{\sim}{q} \cdot \frac{D(\beta \underset{\sim}{q})}{ds} = q^2 \frac{D\beta}{ds} + \beta q \frac{Dq}{ds}$$

$$= \beta q \frac{Dq}{ds} (\beta^2 q^2/c^2 + 1),$$

using (26). But

$$\beta^2 q^2/c^2 + 1 = \beta^2 (q^2/c^2 + 1 - q^2/c^2)$$

$$= \beta^2 .$$

Thus

$$\underset{\sim}{q} \cdot \frac{D(\beta \underset{\sim}{q})}{ds} = \beta^3 q \frac{Dq}{ds} \quad . \tag{29}$$

Now

$$\underset{\sim}{q} \cdot \left(\beta \frac{Dp}{ds} \underset{\sim}{q}\right) = \beta q^2 \left(\beta \frac{\partial p}{\partial t} + \beta \underset{\sim}{q} \cdot \underset{\sim}{\nabla} p\right)$$

whence

$$\underset{\sim}{q} \cdot \left(\underset{\sim}{\nabla} p + \beta/c^2 \frac{Dp}{ds} \underset{\sim}{q}\right) = \beta^2 q^2/c^2 \frac{\partial p}{\partial t} + (1 + \beta^2 q^2/c^2) \underset{\sim}{q} \cdot \underset{\sim}{\nabla} p$$

$$= (\beta^2 - 1)\frac{\partial p}{\partial t} + \beta^2 \underset{\sim}{q} \cdot \underset{\sim}{\nabla} p. \tag{30}$$

Hence, by (28), (29), 30),

$$(\rho + p/c^2)\beta^3 q \frac{Dq}{ds} = (1 - \beta^2) \frac{\partial p}{\partial t} - \beta^2 \underset{\sim}{q} \cdot \nabla p \ ,$$

which is just (27). This proves our conjecture.

A similar procedure can be adopted in the general case (curved space-time). According to (22),

$$(\rho + p/c^2)\frac{Du^\alpha}{ds} = \frac{1}{c^2} (p^{,\alpha} - \frac{Dp}{ds} u^\alpha) \ ,$$

whence

$$(\rho + p/c^2) u_\alpha \frac{Du^\alpha}{ds} = \frac{1}{c^2} (u_\alpha p^{,\alpha} - \frac{Dp}{ds}) \ .$$

But

$$u_k \frac{Du^k}{ds} = u_4 \frac{Du^4}{ds} + u_\alpha \frac{Du^\alpha}{ds},$$

and, as we have seen

$$u_k \frac{Du^k}{ds} = \frac{1}{2} \frac{D}{ds} (u_k u^k)$$

$$= 0.$$

Furthermore,

$$u_\alpha p^{,\alpha} = u_k p^{,k} - u_4 p^{,4}$$

$$= \frac{Dp}{ds} - u_4 p^{,4} ,$$

and so

$$u_\alpha (p^{,\alpha} - \frac{Dp}{ds} u^\alpha) = \frac{Dp}{ds} - u_4 p^{,4} - \frac{Dp}{ds}(1 - u_4 u^4)$$

$$= - u_4 (p^{,4} - \frac{Dp}{ds} u^4).$$

Provided $u_4 \neq 0$, then,

$$(\rho + p/c^2) \frac{Du^4}{ds} = \frac{1}{c^2} (p^{,4} - \frac{Dp}{ds} u^4)$$

which is just the final equation in (22).

THE INTERPRETATION OF ρ

The meaning of ρ in (21), (22) is not very clear even when we are dealing with flat space-time. As we have seen, ρ can be described as the "rest energy-density"; but does this mean the density of inertial mass, or does it not? It would seem that if we can solve the question on the basis of a flat space-time metric, our interpretation of ρ, p should be valid in general, since ρ, p are just tetrad components of T_{ik} on our orthonormal tetrad $\underset{a}{v}^i$ – for

$$T_{ik} \underset{4}{v}^i \underset{4}{v}^k = \rho \underset{4}{v}^i \underset{4}{v}_i$$

$$= \rho ;$$

$$T_{ik} \underset{\alpha}{v}^i \underset{\alpha}{v}^k = - p \underset{\alpha}{v}^i \underset{\alpha}{v}^i$$

$$= p,$$

all other components vanishing. We shall therefore put

$$\{u^i\} = (\beta, \beta \underset{\sim}{q})$$

as before.

Consider an element of fluid, proper volume τ. Let s be the proper time in the local instantaneous rest-frame of the element. Ignoring previous definitions, let $\rho c^2 \tau$ be the total rest-energy associated with the element τ, and let p be the pressure acting upon it. If, following the motion, the fluid proper-volume is changed from τ to $\tau + \delta\tau$, the pressure force will do work

$$\partial W = - p\, \delta\tau .$$

In a relativistic treatment, we must have

$$c^2 \delta(\rho\tau) = \delta W \tag{31}$$

so that

$$\delta\rho = -(\rho + p/c^2) \frac{\delta\tau}{\tau} . \tag{32}$$

It may be objected that (31) ought to be amended, to read

$$c^2\,\delta(\beta\rho\tau)/_{q=0} = \delta W$$

since $c^2(\beta\rho\tau)$ is the energy that would in fact be measured at any neighbouring event. However,

$$\delta\beta/_{q=0} = [\beta^3 q/c^2\,\delta q]_{q=0}$$
$$= 0$$

(cf.(26)); finally, $\beta/_{q=0}$ takes the value unity, and thus (31) and (32) can stand. We deduce that

$$\frac{D\rho}{ds} = -(\rho + p/c^2)\,\frac{1}{\tau}\frac{D\tau}{ds} \quad . \tag{33}$$

Now, classically, $\frac{1}{\tau}\frac{D\tau}{ds}$ is just div $\underset{\sim}{q}$, being specific rate of change of volume. Here, suppose that the point P is instantaneously at rest in the Lorentz frame in question, and let P' be a neighbouring point, position vector $\underset{\sim}{\eta}$ relative to P. Then

$$\frac{D\underset{\sim}{\eta}}{ds} = (\beta\underset{\sim}{q})_{P'}$$
$$= (\underset{\sim}{\eta}\cdot\nabla)(\beta\underset{\sim}{q})/_{P}$$

to the first order; since $q_P = 0$,

$$\frac{D\underset{\sim}{\eta}}{ds} = (\underset{\sim}{\eta}\cdot\nabla)\,\underset{\sim}{q}\ .$$

We can now proceed as in classical hydrodynamics (~~cf. ???.~~). Clearly, we have no need to insert the factor β at any stage; if η_α are the components of $\underset{\sim}{\eta}$ with respect to the principal ~~ones~~ axes of the quadric

$$(\underset{\sim}{\eta} \cdot \underset{\sim\sim}{\nabla q}) \cdot \underset{\sim}{\eta} = \text{const.},$$

and we put

$$\tau = \eta_1 \eta_2 \eta_3 ,$$

then

$$\frac{1}{\tau} \frac{D\tau}{ds} = \text{div } \underset{\sim}{q} \quad ,$$

as required. Note that in the above argument, the fact that p was at rest in the Lorentz frame is crucial; otherwise, in finding the relative velocity, we should have to take into account the relativistic addition law for velocities – an added complication. Clearly, a Lorentz transformation will convert div $\underset{\sim}{q}$ into a general $\text{div}_4 \underset{\sim}{u}$, which is therefore the correct expression for the <u>expansion</u> of particle world lines. Hence our "practical" equation (33) is identical with equation (21), thus confirming that p represents pressure, and ρc^2 represents <u>total</u> rest energy, arising from the interaction of matter with the pressure field.

We now seek to identify the density of <u>inertial</u> mass in the fluid. In the instantaneous rest-frame, the expression

$$- \underset{\sim}{\nabla} p$$

represents the surface force/unit volume and if m is the inertial mass of the volume τ, and in the absence of body forces,

$$\frac{D}{ds} (m \vartheta \underset{\sim}{q}) = -\tau \underset{\sim}{\nabla} p \; . \tag{34}$$

Now p has the dimensions of energy/density; on the macroscopic level, pressure represents a field of <u>negative</u> energy-density (this being related, on a microscopic level, to the kinetic energy of random motion). We may therefore expect that the total energy, ρ, comprises the rest-inertial-mass energy, plus the pressure field energy $(-\,p)$. If this is correct,

$$m = (\rho + p/c^2)\tau \ . \tag{35}$$

Thus on a priori grounds, we might expect an equation of motion of the form (34), with m given by (35).

However, this is precisely the equation of motion obtained from the conservation of the stress-energy tensor - i.e. (28). For

$$\frac{D}{ds}((\rho+p/c^2)\tau\beta \underset{\sim}{q}) = (\rho+p.c^2)\tau \frac{D}{ds}(\beta \underset{\sim}{q}) + \frac{D}{ds}[(\rho+p/c^2)\tau]\beta \underset{\sim}{q}$$

$$= \tau \left[(\rho+p/c^2)\frac{D}{ds}(\beta \underset{\sim}{q}) + \frac{1}{c^2}\frac{Dp}{ds}\beta \underset{\sim}{q}\right] ,$$

since by (32),

$$\frac{D}{ds}(\rho\tau) = -\,p/c^2 \frac{D\tau}{ds} \tag{36}$$

and

$$\frac{D}{ds}(p/c^2\tau) = p/c^2 \frac{D\tau}{ds} + \frac{\tau}{c^2}\frac{Dp}{ds}. \tag{37}$$

Clearly, for no other value of m can (28) and (34) agree. The conclusion seems inescapable, then, that

$$m = (\rho + p/c^2)\tau \ .$$

Note that neither $(\rho\tau)$ nor m are conserved, in general. In fact, using (36), (37)

$$\frac{Dm}{ds} = \frac{\tau}{c^2}\frac{Dp}{ds}$$

which vanishes only if p is constant on any particle path. Thus pressure gradients will increase or decrease inertial mass, locally.

If some equation of state in the form

$$p = p(\rho)$$

be given, then, according to (32)

$$\frac{d\tau}{\tau} = -\frac{dp}{\rho + p/c^2}$$

along any particle path. Hence

$$\tau \sim \exp\left(-\int \frac{dp}{\rho + p/c^2}\right) . \tag{38}$$

Furthermore, if N is the <u>number density</u> of particles in the fluid, then conservation of particle number implies that

$$\frac{D}{ds}(N\tau) = 0$$

and so, using (38),

$$N = N_0 \exp\left(\int \frac{dp}{\rho + p/c^2}\right) \tag{39}$$

N_0 being the number density at some event on the worldline in question. In general, N is the only conserved function of (ρ, p).

It would seem highly probable that most classical equations of state should be regarded as being functional relations between p and N, rather than p and ρ. Of course, in the classical limit,

$$N \propto \rho .$$

If p is given as a function of N - or equally as a function of τ, we have, using (32),

$$\frac{dp}{d\tau} + \frac{\rho}{\tau} + \frac{p(\tau)}{c^2 \tau} = 0$$

whence

$$\rho = \frac{M_0}{\tau} - \frac{1}{c^2\tau}\int p(\tau)d\tau, \tag{40}$$

where M_0 is constant on each worldline. For example, the equation of state of a polytrope is

$$p\tau^{1+1/n} = \lambda \tag{41}$$

where λ is constant; it is in this form that the equation of state may be taken over into relativistic physics (cf. THOMPSON & WHITROW, REF. 48). According to (40),

$$\rho = \frac{M_0}{\tau} + \frac{n\lambda}{c^2}\ \tau^{-(1+1/n)} \tag{42}$$

so that here the quantity $(\rho - n\ p/c^2)$ is conserved; this is naturally proportional to N.

An incompressible fluid is, strictly speaking, one for which

$$\frac{D\tau}{ds} = 0;$$

substituting this into (33), one sees that

$$\frac{D\rho}{ds} = 0 \tag{43}$$

as in classical physics. Generally, however, the term "incompressible" is supposed to imply that the density is in fact uniform, a stronger condition than (43). In relativity, as we have seen, the term "density" is ambiguous. Nevertheless, the fact that (43) holds makes it reasonable to suppose that the "strong" condition for incompressibility should be simply

$$\rho = \text{const.}$$

It must however be borne in mind that ρ is not the mass-density of classical physics.

Finally, we note that the trace, T, of the stress-energy tensor appears to have no special significance on the basis of the present discussion. Any attempt to equate T to a number-density (that of baryons, for example) seems lacking in foundation. Electromagnetic fields have a vanishing T, since

$$T^{ij} = F^{ik}F_k{}^j - \frac{1}{4} F_{k\ell}F^{k\ell} g^{ij} \;;$$

this does not mean (in theory, at any rate) that a vanishing T implies an electromagnetic field, and not a fluid comprised of non-zero rest mass particles. However, there are sound reasons (cf. REF. 4) for supposing that for physical fluids

$$\rho > 3p/c^2$$

which means that T must always be positive.

This concludes our investigation of the interpretation of the eigenvalues of T^{ij}. In all subsequent work, we shall revert to the system of units in which c is understood.

For the rest of this chapter we are concerned only with the dust fluid; our problem is to solve the Einstein Field Equations for a dust sphere (possibly inhomogeneous), with empty space exterior to it, and in the absence of rotation. At the present date, an empty space solution (the KERR metric, REF. 23) has been found, which appears to represent the field exterior to a rotating body, but little progress has been made in tackling, for example, the problem of a rotating dust cloud.

The next section will be devoted to a brief discussion concerning the type of coordinate system to be employed in solving the Einstein Equations, and the various assumptions which have to be made.

THE CHOICE OF COORDINATE SYSTEM

We consider first the analysis of <u>rotation</u> in General Relativity. The <u>rate</u> of rotation of the particle worldlines is given by the skew tensor ("bivector")

$$\omega_{ij} = u_{[i;j]} = u_{[i;j]} - \frac{Du_{[i}}{ds} u_{j]} \tag{44}$$

where the square brackets indicate the <u>skew</u> part of the appropriate tensors. Since u^i is a unit timelike vector-field, and by definition

$$\frac{Du_i}{ds} = u_{i;k} u^k ,$$

one finds that

$$\omega_{ij} u^j = 0, \tag{45}$$

(cf. SYNGE (REF. 47)).

At this stage it is useful to introduce the "permutation symbol" $\eta_{ijk\ell}$, where

$$\eta_{ijk\ell} = -(-g)^{\frac{1}{2}} \epsilon_{ijk\ell} . \tag{46}$$

We note the following results:

$$\eta^{ijk\ell} = (-g)^{-\frac{1}{2}} \epsilon_{ijk\ell} \tag{47}$$

$$\eta_{ijk\ell}\eta^{\ell abc} = \delta^{abc}_{ijk} \tag{48}$$

$$\eta_{ijk\ell}\eta^{k\ell ab} = -2\,\delta^{ab}_{ij} \tag{49}$$

where the δ symbols represent determinants formed from Kronecker deltas.

The dual bivector, ω^{ij*}, is defined by

$$\omega^{ij*} = \tfrac{1}{2}\,\eta^{ijk\ell}\omega_{k\ell} \quad . \tag{50}$$

According to (45), u^i is an eigenvector of ω_{ij}; however it is not an eigenvector of $\omega_{ij}{}^*$; we write

$$\begin{aligned}\omega^i &= \omega^{ij*}u_j \\ &= \tfrac{1}{2}\,\eta^{ijk\ell}u_j\omega_{k\ell} \ .\end{aligned} \tag{52}$$

If (44) be substituted into (52), we may omit the square brackets, and terms involving the affine connexion, on account of the alternating property of $\eta^{ijk\ell}$. We may also omit any term involving the vector u_i twice, whence (52) reduces to

$$\omega^i = \tfrac{1}{2}\,\eta^{ijk\ell}u_j u_{k,\ell} \quad . \tag{53}$$

We may call ω^i the "rate of rotation vector". From (49) and (50),

$$\omega^{ij**} = -\omega^{ij} \tag{54}$$

which is true for all bivectors.

Consider the bivector $2\,\omega^{[i}\,u^{j]}$. Examining (50), it might be conjectured that this is in fact the dual of ω^{ij}, since from (52) or (53),

$$\omega^j u_j = 0 \quad . \tag{55}$$

To test this conjecture, we note that

$$\begin{aligned}(2\,\omega^{[i}\,u^{j]})^* &= \eta^{ijk\ell}\omega_k\,u_\ell \\ &= \tfrac{1}{2}\,\eta^{ijk\ell}\eta_{kabc}u^a\omega^{bc}u_\ell \\ &= -\tfrac{1}{2}\,\delta^{ij\ell}_{abc}\,u^a\omega^{bc}u_\ell \ ,\end{aligned}$$

interchanging k, ℓ and using (48); expanding,

$$(2\,\omega^{[i}u^{j]})^{*} = -u_{\ell}\,(u^{i}\omega^{j\ell} - u^{j}\omega^{i\ell} + u^{\ell}\omega^{ij})$$

$$= -\omega^{ij},$$

using (45); reference to (54) shows that the bivector is indeed dual to ω^{ij}. This means that either ω^{i} or $\omega_{k\ell}$ may be used to describe the rate of rotation. Note that

$$\omega^{ij} = -\eta^{ijk\ell}\,\omega_{k}u_{\ell} \qquad (56)$$

whence

$$\omega^{ij}\omega_{j} = 0\ . \qquad (57)$$

According to (45), (57), ω_{ij} has eigenvectors u^{i}, ω^{i}, corresponding to eigenvalue zero. By (55), these vectors are orthogonal. It follows that, for all vectors σ^{j}, the vector $\omega_{ij}\sigma^{j}$ lies in the plane orthogonal to u^{i} and ω^{i}. Using (56),

$$\omega_{ij}\sigma^{j} = -\tfrac{1}{2}\,\eta_{ijk\ell}\sigma^{j}\,\omega^{k}u^{\ell}$$

$$= \tfrac{1}{2}\,\eta_{ijk\ell}\,\omega^{j}\,\sigma^{k}u^{\ell}\,,$$

which we may interpret as a rotation in the spacelike hyperplane orthogonal to u^{i}, the axis of rotation being ω^{i}.

Let a^{i}, b^{i} be mutually orthogonal unit spacelike vectors in the plane perpendicular to ω^{i}. Then, since ω_{ij} is skew,

$$\omega_{ij}\,a^{j} = \omega\,b_{i}$$

$$\omega_{ij}\,b^{j} = -\omega\,a_{i}$$

where ω is some scalar; without loss of generality we may take a^{i}, b^{i} in an order such that ω is positive. This shows that the complex vector

$$a^k + i\, b^k$$

and its conjugate, are eigenvectors of ω_{ij}, corresponding to eigenvalues $\pm\, i\omega$. If $\hat{\omega}^i$ be defined to be the unit vector in the direction of ω^i, so that

$$\hat{\omega}^i = (-\,\omega_k \omega^k)^{-\frac{1}{2}}\, \omega^i ,$$

then with respect to the orthonormal tetrad $\{u^i, \hat{\omega}^i, a^i, b^i\}$, the sole non-vanishing tetrad component of ω_{ij} is

$$\omega_{ij}\, a^i\, b^j = \omega \quad ;$$

thus

$$\omega_{ij} = 2\omega\, a_{[i} b_{j]} \,. \tag{58}$$

Since the vectors a^i, b^i are both spacelike, we may deduce that

$$\omega = (\tfrac{1}{2}\, \omega_{ij}\, \omega^{ij})^{\frac{1}{2}} . \tag{59}$$

However, using (56) and (49),

$$\omega_{ij}\omega^{ij} = -\,2\, \delta^{k\ell}_{ab}\, \omega_k u_\ell \omega^a u^b$$

$$= -\,2\, \omega_k \omega^k ;$$

thus

$$\omega = (-\,\omega_k \omega^k)^{\frac{1}{2}} . \tag{60}$$

Finally, we note that

$$\omega^{ij*} = 2\, \omega^{[i} u^{j]}$$

$$= 2\, \omega\, \hat{\omega}^{[i} u^{j]} , \tag{61}$$

and comparison with (58) exhibits the symmetrical relationship between the bivector and its dual. This completes our analysis of the rotation of particle worldlines.

The vector-field u^i is said to be HYPERSURFACE ORTHOGONAL if the equation

$$u_i \, dx^i = 0 \qquad (62)$$

is integrable; that is, if there exists a scalar field ξ such that

$$\xi u_i \, dx^i$$

is an exact differential. A necessary condition for this is that $\omega^i = 0$. For if $\xi u_i \, dx^i$ is an exact differential,

$$\left(\xi u_{[i}\right)_{,j]} = 0$$

whence

$$u_{[i,j]} = \xi^{-1} \xi_{,[i} u_{j]}$$

and so

$$\omega^i = \tfrac{1}{2} \eta^{ijk\ell} u_j u_{k,\ell}$$

$$= \tfrac{1}{2} \xi^{-1} \eta^{ijk\ell} u_j \xi_{,k} u_\ell$$

$$= 0 .$$

This is, in fact, a sufficient condition. This is particularly simple to demonstrate if u^i is geodesic - that is, if

$$\frac{Du_i}{ds} = 0 .$$

For, by (56), $\omega^i = 0$ implies $\omega_{ij} = 0$, whence if u^i is geodesic, $u_{[i,j]} = 0$. Thus in this case $u_i \, dx^i$ is itself an exact differential. The proof in the general case uses a somewhat different approach (see e.g. GOURSAT (REF. 11)). Suppose $u_p \neq 0$. Then (62) is equivalent to

$$dx^p = -\sum_{i \neq p} u_p^{-1} \; u_i \, dx^i \qquad (63)$$

so that (62) is integrable only if the right-hand-side of (63) is an exact differential. The condition for this is that

$$\frac{\partial x^p}{\partial x^i} = - u_p^{-1} u_i \;, \qquad (64)$$

where u_p, u_i are functionally dependant on x^p explicitly, as well as the $\{x^j\}$, and x^p is regarded as a function of the $\{x^j\}$. Equations (64) have a solution if and only if $\frac{\partial x^p}{\partial x^i}\frac{\partial}{\partial x^p}$

$$\left(\frac{\partial}{\partial x^j} + \frac{\partial x^p}{\partial x^j}\frac{\partial}{\partial x^p}\right)\left(u_p^{-1} u_i\right) = \left(\frac{\partial}{\partial x^i} + \frac{\partial x^p}{\partial x^p}\right)\left(u_p^{-1} u_j\right);$$

there are only three independent equations here, and, making further use of (64), these may be written in the form:

$$\sum_{i,j,k \neq p} \epsilon_{ijk} \left(\frac{\partial}{\partial x^k} - u_p^{-1} \; u_k \frac{\partial}{\partial x^p}\right)\left(u_p^{-1} \; u_j\right) = 0 \;. \qquad (65)$$

Evaluating the partial derivatives, (65) yields

$$\sum_{i,j,k \neq p}^{4} \epsilon_{ijk} \; u_p^{-2} \; [u_{j,k} u_p - u_{p,k} u_j - u_k u_{j,p} + u_p^{-1} u_{p,p} u_j u_k] = 0.$$

The last term of the left-hand-side vanishes since u^i appears twice, and after multiplication by $\pm u_p^2$, we find

$$\sum_{a,b,c=1}^{4} \epsilon_{iabc} \; u_a \; u_{b,c} = 0 \;, \qquad (66)$$

since we may put $p = a,b,c$ in turn. It might appear that i has to be restricted to values other than p; however, the above argument applies for <u>any</u> non-zero u_p, and so (66) applies for all i (unless there is only one non-zero u_p, in which case (66) is merely an identity and (62) is always integrable). We may write (66) in the tensor form

$$\eta^{ijk\ell} u_j u_{k,\ell} = 0$$

whence a sufficient condition for integrability is indeed $\omega^i = 0$. In Cosmology, it is usually assumed that the world-lines of "smoothed-out" matter are irrotational – the assumption is known as "Weyl's Postulate" (REF. 52).

Given that $\omega_{ij} = 0$, we may construct a system of coordinates known as "<u>comoving</u>", in the following manner: suppose that, for some ξ,

$$\xi \, u_i \, dx^i = d\phi \, . \tag{67}$$

Clearly, if we replace ϕ by some regular function of ϕ, the form of equation (67) is preserved, with a different ξ; we shall choose ϕ such that, within a particular region (our coordinate "patch") the scalar ξ is everywhere positive. The surfaces ϕ=const. are spacelike, and we may set $t = \phi(x^j)$, in order to define a time coordinate. Our choice of ϕ ensures that t is monotonic on the particle worldlines.

Next, we select one of the spacelike hypersurfaces, and introduce an arbitrary labelling of points in it, $\{x^\alpha\}$. Our coordinate system is to apply only within a region for which no two worldlines intersect; in such a region, the vector field generates a (1,1) correspondence between the hypersurfaces, corresponding points lying on the same worldline. We may thus label corresponding points on the hypersurfaces with the x^α coordinates of the initial hypersurface. The system (t, x^α) then provides coordinates for the region of space-time in question.

We call the hypersurfaces "space sections". On any such section, $dx^4 = 0$; thus (62) become

$$u_\alpha dx^\alpha = 0$$

where the dx^{α} are arbitrary. It follows that $u_{\alpha} = 0$. Furthermore, on each worldline, x^{α} = const., so that u^{α} must also vanish. But

$$u_{\alpha} = g_{4\alpha}u^{4} + g_{\beta\alpha}u^{\beta}$$

$$= g_{4\alpha}u^{4} ,$$

which shows that

$$g_{4\alpha} = g^{4\alpha} = 0 \quad . \tag{68}$$

Thus the line element takes the form

$$ds^{2} = g_{44}dt^{2} - d\sigma^{2} \tag{69}$$

where $d\sigma^{2}$ is positive definite (g_{44} must be positive, since u $u^{4}u_{4} = g_{44}(u^{4})^{2} = u^{\iota}u_{\iota}$). Note also that

$$u^{\iota} = \chi\delta^{\iota}{}_{4} \tag{70}$$

where χ is some scalar field.

The foregoing construction of comoving coordinates is valid for any fluid, subject only to some reasonable regularity conditions on u^{ι}. We are particularly concerned with the dust-fluid, for which we may effect §6 some considerable simplification, since u^{ι} must be geodesic, as is readily demonstrated: indeed, reference back to (20) shows that if p vanishes, then $\frac{Du_{\iota}}{ds} = 0$. Here, we may specialise our comoving coordinates in order to construct NORMAL GAUSSIAN COORDINATES (REF. 47). In terms of comoving coordinates,

$$0 = \frac{Du_{\alpha}}{ds}$$

$$= \Gamma^{4}_{\alpha 4}\, u_{4}\, u^{4} .$$

Now

$$\Gamma^4_{\alpha 4} = \tfrac{1}{2} g^{44}(g_{44,\alpha} + g_{4\alpha,4} - g_{\alpha 4,4}) \; ;$$

according to (68),(69)

$$g^{44} = g_{44}^{-1}$$

and so

$$\Gamma^4_{\alpha 4} = \tfrac{1}{2}(\log g_{44})_{,\alpha} \tag{71}$$

Thus g_{44} is a function of t alone. The remaining equation, $\frac{Du_4}{ds} = 0$, provides no further restriction. As we have noted, t is undetermined up to an arbitrary function of itself; if for geodesic motion, (69) reads

$$ds^2 = (A(t'))^2 dt'^2 - d\sigma^2$$

for some system of comoving coordinates (t', x^α), we shall set

$$t = \int A(t')dt' \; ,$$

whence t is proper time measured on a worldline. The coordinates (t, x^α) thus constructed are known as "Normal Gaussian"; we have

$$ds^2 = dt^2 - d\sigma^2 \tag{72}$$

and, since $u_4 u^4 = 1$,

$$u^i = \delta^i_4 \; . \tag{73}$$

It must be emphasised that (73) holds only for the dust-fluid.

For the most part, we shall be concerned with the case of spherical symmetry. The concept of symmetry has been discussed, among others, by SCHOUTEN (REF. 44). The only satisfactory treatment, from the point of view of Riemannian geometry, is that in terms of "groups of motion" - cf. EISENHART (REF. 9). For spherical symmetry one requires that the space-time should admit a 2-parameter group of motions, whose trajectories (surfaces of transmittivity) are spacelike 2-surfaces, locally isometric to the sphere. It follows that the line element of the space-time may be written in the form:

$$ds^2 = G(4,1) - R^2((dx^2)^2 + \sin^2 x^2 (dx^3)^2)$$

where $G(4,1)$ os a quadratic differential form, in the (x^4, x^1) coordinates, of signature (+ -) and R - the radius of curvature of a surface of transmittivity - is a function of x^4 and x^1 alone. Note that we must insist that the trajectories be isometric to the sphere: they must be spaces of constant curvature, but could otherwise be isometric to the Euclidean or the Lobatschewski plane. For a discussion of the special case of static vacuum fields, see JORDAN, EHLERS + KUNDT (REF. 22).

When the space-time admits a hypersurface - orthogonal vector field u^i, such that u^i "shares" in the symmetry (so that the LIE DERIVATIVE of u^i with respect to a killing vector (REF. 9) vanishes, then we may introduce a system of comoving coordinates $(t, r, \theta\ \phi)$ such that (69) and (70) become

$$ds^2 = e^{\nu} dt^2 - e^{\lambda} dr^2 - e^{\mu} d\Omega^2 \qquad (74)$$

$$u^i = e^{-\frac{1}{2}\nu} \delta^i_4 \qquad (75)$$

where ν, λ, μ are functions of (r, t) only, and

$$d\Omega^2 = \sin^2\theta\ d\phi^2 + d\theta^2.$$

The coordinates (θ, ϕ) are invariant, being constructs of the killing vectors; (r, t) are arbitrary labellings of the surfaces of transmittivity, and the space-sections, respectively. We may determine the coordinate r by selecting one space-section, and labelling it $t = 0$; we then require that r is the radius of curvature of surfaces of transmittivity in this hypersurface, i.e. that

$$r = R(r,0). \tag{76}$$

The requirement that the Lie Derivative of u^i should vanish implies that the fluid moves radially. In the case of <u>geodesic</u> motion, as for the dust fluid, (74) takes the simple form

$$ds^2 = dt^2 - e^{\lambda}dr^2 - e^{\mu}d\Omega^2. \tag{77}$$

Finally, it should be noted that, in (74),(77),

$$e^{\mu} = R^2. \tag{78}$$

In the remainder of this chapter, we shall find the functions λ, μ in (77) assuming the Einstein Field Equations. The Einstein tensor G^{ij}, where

$$G^{ij} = R^{ij} - \tfrac{1}{2} R g^{ij}, \tag{79}$$

satisfies

$$G^{ij}{}_{;j} = 0,$$

being the contracted Bianchi identities (~~ ~~). Note that the "R" in (79) is not to be confused with that in (76); in the former case, it is the trace of the Ricci tensor.

Following Einstein, we shall set

$$G^{ij} = -\kappa T^{ij} \tag{80}$$

where κ is a constant; in our system of units,

$$\kappa = 8\pi G \tag{81}$$

where G is the gravitational constant. Equation (80) may be read from right to left, or vice versa: in principle, one could determine the curvature of space-time induced by a particular distribution of matter, or equally deduce the physical matter corresponding to a mathematical curvature. In practice, however, we possess some information as to the structure of both G^{ij} and $T^{ij}_{\ \ ;}$ equations (80) may be regarded as partial differential equations for the metric components, and their solutions, giving functional forms for these components, provide further information as to the structure of the tensors.

THE DUST-FLUID SOLUTIONS

It is usually convenient to solve equations (80) in "mixed" tensor form. In terms of the metric (77), one obtains (REF. 44):

$$G^r_r = \frac{1}{4}\mu'^2 e^{-\lambda} - e^{-\mu} - (\ddot{\mu} + \frac{3}{4}\dot{\mu}^2) \tag{82}$$

$$G^\theta_\theta = e^{-\lambda}(\frac{1}{2}\mu'' + \frac{1}{4}\mu'^2 - \frac{1}{4}\lambda'\mu') - (\frac{1}{2}\ddot{\mu} + \frac{1}{4}\dot{\mu}^2 + \frac{1}{2}\ddot{\lambda} + \frac{1}{4}\dot{\lambda}^2 + \frac{1}{4}\dot{\lambda}\dot{\mu}) \tag{83}$$

$$G^\phi_\phi = G^\theta_\theta$$

$$G^t_t = e^{-\lambda}(\mu'' + \frac{3}{4}\mu'^2 - \frac{1}{2}\lambda'\mu') - e^{-\mu} - (\frac{1}{4}\dot{\mu}^2 + \frac{1}{2}\dot{\lambda}\dot{\mu}) \tag{84}$$

$$G^t_r = -e^{\lambda}G^r_t = \dot{\mu}' + \frac{1}{2}\dot{\mu}\mu' - \frac{1}{2}\dot{\lambda}\mu' \tag{85}$$

where dots signify partial differentiation with respect to t, and dashes with respect to r. For the dust-fluid,

$$\begin{aligned} T^i_{\ j} &= \rho\, u^i u_j \\ &= \rho\, \delta^i_4\, \delta^4_{\ j} \end{aligned}$$

in our comoving system of coordinates. Thus all components of $T^i{}_j$, except T^4_4, vanish. Applying the Einstein Equations, we find firstly that

$$(\mu')^{-1}\frac{\partial}{\partial t}(\mu') = \frac{1}{2}\frac{\partial}{\partial t}(\lambda - \mu)$$

whence

$$e^{\frac{1}{2}\lambda} = f(r)\frac{\partial}{\partial t}\left(e^{\frac{1}{2}\mu}\right) \qquad (86)$$

where $f(r)$ is arbitrary. In an early paper by OPPENHEIMER & SNYDER (REF. 31), the authors consider only the case $f(r) = 1$; they also suppose $\rho = \rho(t)$ alone - the homogeneous case.

In the rest of this chapter, R will be defined as in (78), and will not mean the trace of the Ricci tensor. According to (86)

$$e^{-\frac{1}{2}\lambda} = \frac{1}{f(r)}(R')^{-1}; \qquad (87)$$

also, from (78),

$$2\ddot{R} = (\ddot{\mu} + \tfrac{1}{2}\dot{\mu}^2)e^{\frac{1}{2}\mu} \quad . \qquad (88)$$

Equating (82) to zero, we find:

$$\frac{1}{4}\mu'^2 e^{\mu-\lambda} = 1 + (\ddot{\mu} + \frac{3}{4}\dot{\mu}^2)e^{\mu},$$

whence, using (88) and (87),

$$\frac{R'^2}{f^2(r)R'^2} = 1 + 2R\ddot{R} + \dot{R}^2 ,$$

so that, multiplying by $\dot{R}$,

$$\left(\frac{1}{f^2(r)} - 1\right).\dot{R} = \frac{\partial}{\partial t}(R\dot{R}^2) \qquad (89)$$

This equation may be integrated immediately, to give

$$\dot{R}^2 = \frac{1}{f^2(r)} - 1 + \frac{k(r)}{R} \quad , \tag{90}$$

where $k(r)$ is arbitrary. This equation may be further integrated, an operation which we postpone for the moment.

The next Einstein Equation we consider is

$$G_t^t = -\kappa\, T_t^t = -\kappa\rho \ .$$

From (84)

$$e^{-\lambda}(\mu'' + \frac{3}{4}\mu'^2 - \frac{1}{2}\lambda'\mu') - e^{-\mu} - (\frac{1}{4}\dot{\mu}^2 + \frac{1}{2}\dot{\lambda}\dot{\mu}) = -\kappa\rho \ .$$

However, (87) shows that

$$\lambda' = 2\frac{f'(r)}{f(r)} + 2\frac{\mu''}{\mu'} + \mu' \quad ,$$

so that the preceding equation becomes

$$\frac{4}{f^2(r)\mu'^2}\left[\frac{1}{4}\mu'^2 - \frac{f'(r)}{f(r)}\mu'\right] - 1 - e^{\mu}\left[\frac{3}{4}\dot{\mu}^2 + \frac{\dot{\mu}'\dot{\mu}}{\mu'}\right] = -\kappa\rho e^{\mu}.$$

If we now multiply both sides of this equation by μ', we find

$$\left(\frac{1}{f^2(r)} - 1\right)\mu' - 4\frac{f'(r)}{f^3(r)} - \frac{1}{2}e^{-\frac{1}{2}\mu}\frac{\partial}{\partial r}\left(\dot{\mu}^2 e^{3/2\,\mu}\right) = -\kappa\rho\mu' e^{\mu} \quad ,$$

which, in terms of R, reads

$$\left(\frac{1}{f^2(r)} - 1\right)\frac{R'}{R} - 2\frac{f'(r)}{f^3(r)} - \frac{1}{R}\frac{\partial}{\partial r}(R\dot{R}^2) = -\kappa\rho RR' \tag{91}$$

If (90) is now substituted in (91), most terms cancel, leaving

$$k'(r) = \kappa\rho\, R^2 R' \quad . \tag{92}$$

With the help of (81), this may be integrated in the form:

$$k(r) = k_0 + 8\pi G \int_0^r \rho R^2 R' \, dr \,, \tag{93}$$

where $k_0 = k(0)$. However, examination of (90) shows that we must normally set $k_0 = 0$, lest, for any t,

$$\dot{R}(r,t) \to \pm \infty$$

as $r \to o$!

Writing (cf. II, (21)):

$$\rho(r,0) = \rho_0(r)$$

which means that ρ is supposed to be given on the initial space-section, we see from (92) that

$$\frac{\rho_0(r)}{\rho} = \frac{R^2 R'}{R_0^2 R_0'} \quad . \tag{94}$$

If we choose the scaling for r as in (76) – which suggests an analogy with the Lagrangian description of motion used in Chapter II – this equation reads

$$\frac{\rho_0(r)}{\rho} = \frac{R^2 R'}{r^2} \quad . \tag{95}$$

We now write (cf. II, (26))

$$R = r\, S(r,t) \tag{96}$$

where

$$S(r,0) = 1 \quad .$$

Then (94) reads

$$\frac{\rho_0(r)}{\rho} = S^3 \left(1 + \frac{r}{S}\frac{\partial S}{\partial r}\right).$$

Substituting (95) into (93) gives (putting $k_0 = 0$) :

$$k(r) = 8\pi G \int_0^r \rho_0(r) r^2 dr$$

$$= 2 G M_r , \qquad (97)$$

where M_r is the gravitational mass, calculated on the initial hypersurface, within the "sphere", radius r. On the hypersurface, $\rho_0(r)$ and r are invariants, whence M_r is a scalar invariant, being a function of invariants. We may introduce the mean density $\bar{\rho}_0(r)$, as in II, (24), and also the function $\alpha(r)$, as in II, (27). In terms of the latter, (90) may be written, making use of (96):

$$\dot{S}^2 = \frac{\alpha(r)}{S} + F(r) \qquad (98)$$

where

$$F(r) = \frac{1}{r^2}\left(\frac{1}{f^2(r)} - 1\right).$$

If we now compare (98) with II, (30), it is clear that the form of the two equations is identical, $F(r)$ being related to the initial "velocity" distribution, $\dot{R}(r,0)$. All this suggests a powerful analogy between the Newtonian and the relativistic treatments of the dust-fluid; we shall postpone further investigation of this for the moment, and examine whether the remaining Einstein equation provides any further restriction.

Equating G_θ^θ to zero,

$$e^{-\lambda}\left(\frac{1}{2}\mu'' + \frac{1}{4}\mu'^2 - \frac{1}{4}\lambda'\mu'\right) + \left(\frac{1}{2}\ddot{\mu} + \frac{1}{4}\dot{\mu}^2 + \frac{1}{2}\ddot{\lambda} + \frac{1}{4}\dot{\lambda}^2 + \frac{1}{4}\dot{\lambda}\dot{\mu}\right) = 0.$$

Firstly, we shall substitute for λ' by means of (86); thus

$$\frac{1}{2}\, e^{-\lambda}\, \frac{f'(r)}{f(r)}\,\mu' + \left(\frac{1}{2}\ddot{\mu} + \frac{1}{4}\dot{\mu}^2 + \frac{1}{2}\ddot{\lambda} + \frac{1}{4}\dot{\lambda}^2 + \frac{1}{4}\dot{\lambda}\,\dot{\mu}\right) = 0. \qquad (99)$$

Next, we note that, in terms of R,

$$\frac{1}{2}\dot{\mu} = \dot{R}/R \quad ; \quad e^{-\lambda} = \frac{1}{f^2(r)R'^2} \quad ; \quad \frac{1}{2}\lambda = \log f(r) + \log R' \quad ;$$

giving

$$\frac{1}{2}\ddot{\lambda} = \frac{\ddot{R}'R' - \dot{R}'^2}{R'^2}$$

$$\frac{1}{4}\dot{\lambda}^2 = \frac{\dot{R}'^2}{R'^2}$$

and

$$\frac{1}{4}\dot{\lambda}\,\dot{\mu} = \frac{\dot{R}'\dot{R}}{RR'} \quad .$$

Thus (99) reduces to

$$\frac{\ddot{R}'}{R'} + \frac{\dot{R}'\dot{R}}{RR'} + \frac{\ddot{R}}{R} = -\frac{f'(r)}{f^3(r)}\,\frac{1}{RR'}$$

whence

$$\frac{\partial}{\partial r}\left(R\ddot{R} + \tfrac{1}{2}\dot{R}^2\right) = \frac{-f'(r)}{f^3(r)} \quad .$$

Integrating,

$$R\ddot{R} + \tfrac{1}{2}\dot{R}^2 = \frac{1}{2f^2(r)} + \tfrac{1}{2}a(t)$$

where a(t) is arbitrary; however, this equation may be written:

$$\frac{\partial}{\partial t}(R\dot{R}^2) = \left(\frac{1}{f^2(r)} + a(t)\right)\dot{R} \qquad (100)$$

and comparison of (100) with (89) shows that a(t) must equal minus one, except possibly when $\dot{R}$ vanishes. In any case, (100) provides no further restriction.

We conclude that the positions of the fluid particles, and also the metrical coefficients, are determined by equation (98). This is precisely the equation of motion in the Newtonian treatment, where our function R is analogous to the Lagrangian radial coordinate. Thus there are three types of solution, as in II, (16),(17),(18), together with II,(31),(32) and similar equations. Thus the non-linear equations of General Relativity lead to the <u>same</u> differential equation of motion as do the Newtonian equations. This is a remarkable result, but seems fairly plausible on the following grounds: in using "comoving" coordinates, we are relating rest-frames in the fluid one to another, and since the particle motion is geodesic the coordinate system could be regarded as "inertial", in the sense of Special Relativity. Furthermore, in the case of the dust-fluid, we need not take account of "coupling" between density and pressure - for the dust fluid, the equation of continuity is simply

$$\frac{D\rho}{ds} = -\rho\frac{D}{ds}(\log\tau)$$

where τ is an element of proper-volume; in comoving coordinates, this reduces to

$$\frac{\partial\rho}{\partial t} = -\rho\frac{\partial}{\partial t}(\log\tau)$$

as in Classical physics, except that here

$$\tau = R^2 e^{\frac{1}{2}\lambda} \delta r \delta \Omega$$

whereas, classically,

$$\tau = R^2 R' \, \delta r \delta \Omega \; .$$

Thus (86) alone, being a consequence of the simplest Einstein equation – (85) – shows that the Classical and Relativistic expressions for $\frac{\partial}{\partial t} \log \tau$ are identical. Thus we need not be too surprised at this strictly Newtonian behaviour, whereas we would doubt such behaviour possible for a perfect fluid, with equation of state <u>other</u> than $p = 0$.

It must however be emphasised that the space-time geometry is non-Euclidean, and, in general, not even conformally flat. We note that it has long been known that an <u>homogeneous</u> and isotropic dust-fluid will exhibit Newtonian behaviour: see e.g. ~~KERMACK~~ MILNE & McCREA (REF. 28). The mathematical conditions for homogeneity and isotropy are that the space-time should admit a 6-parameter group of motions, the trajectories being spacelike hypersurfaces; this leads to the conformally flat ROBERTSON-WALKER METRIC (cf. REF. 42):

$$ds^2 = dt^2 - \frac{S^2(t)}{(1 + \frac{1}{4} k \bar{r}^2)^2} \left(d\bar{r}^2 + \bar{r}^2 \, d\Omega^2 \right) \qquad (101)$$

where $k = 0, \pm 1$. In fact, this is a special case of our general solution, as we shall in due course demonstrate, save that in the form (101), the "isotropic" radial coordinate does not satisfy (76). However, it has not been generally realised that an arbitrary dust sphere exhibits Newtonian behaviour.

A question naturally arises: could this result be extended to an arbitrary accumulation of dust fluid? A reasonable conjecture is that a Newtonian analogy may be possible if the space-time admits a 1-parameter group of motions; but in the absence of this, the question may indeed be meaningless, for how then could one construct invariant coordinates with Newtonian analogues? The difficulties inherent in finding exact solutions in spaces without symmetries make it unlikely that this question can be satisfactorily answered. It seems particularly doubtful that any such analogy could hold if the fluid were rotating, for then we could not even use comoving coordinates.

We now determine the functional form of the metrical component e^{λ}. According to (86)

$$e^{\lambda} = f^2(r)R'^2 \quad .$$

Now, as we remarked previously, $f(r)$ is dependant upon the initial velocity distribution. If, for the moment, we restrict ourselves to the "elliptical" solution (II, case I):

$$S(r,t) = p^{-1}\xi(\alpha^{1/2}p^{3/2}t + a) \tag{102}$$

where

$$p = 1 - \frac{V^2(r)}{\alpha r^2} \tag{103}$$

and

$$\xi(a(r)) = p, \tag{104}$$

$V(r)$ being $\dot{R}(r,0)$, then (98) shows that

$$\alpha r^2 - p\alpha r^2 = \frac{1}{f^2(r)} - 1 + \alpha r^2$$

whence

$$f(r) = (1 - p\alpha r^2)^{-\frac{1}{2}} \quad . \tag{105}$$

Generally, the explicit expression for e^{λ}, substituting for R' by means of (102), is rather untidy. A concise expression is, however, obtained if the fluid is initially at rest, for, by (103),

$$p = 1 \;,$$

and use of II,(45) gives

$$e^{\lambda} = \frac{\xi^2}{1-\alpha r^2}\left(1 + \frac{3}{2}\left(\frac{\rho_0}{\bar{\rho}_0} - 1\right)\alpha^{\frac{1}{2}} t\, \dot{\xi}/\xi\right)^2 . \tag{106}$$

One consequence of (106) is that for any point on the initial hypersurface for which $\rho_0 = \bar{\rho}_0$, the worldline through that point is such that at any event upon it

$$R = r\xi\,(\alpha^{\frac{1}{2}} t)\;; \quad e^{\lambda} = \frac{\xi^2}{1-\alpha r^2}\;; \quad \text{and} \quad \rho = \rho_0\, \xi^{-3}\;;$$

these being identical to the expressions for the homogeneous, isotropic fluid, save that α is a function of r. If in fact the density is constant throughout, then the elliptical solution becomes

$$ds^2 = dt^2 - \xi^2(\alpha^{\frac{1}{2}} t)\left[\frac{dr^2}{1-\alpha r^2} + r^2 d\Omega^2\right]. \tag{107}$$

The substitution

$$r = \frac{\alpha^{-\frac{1}{2}}\,\bar{r}}{1 + \frac{1}{4}\bar{r}^2}$$

transforms (107) into (101), and similar reasoning applies to the hyperbolic case - the parabolic metric is automatically in Robertson-Walker form. This proves that the space-time of a homogeneous, isotropic dust sphere must ~~adyet~~ admit a 6-parameter group of motions. This result is obvious if the "sphere" were in fact the universe as a whole, where clearly there is no preferred centre; it might, however, have been supposed that the presence of the boundary would have had some effect upon the curvature of space-time, but this is not in fact the case.

In their discussion of the gravitational collapse of a dust-fluid, HOYLE & NARLIKAR (REF. 18) consider only the elliptical Robertson-Walker metric, on the grounds that this alone permits the initial condition $\dot{R}(r,0) = 0$.

However, it does not seem advisable to enquire too closely into the earlier history of a collapsing body whilst retaining the dust-fluid approximation: the circumstances giving rise to a highly collapsed state would, one might expect, vary considerably. Although elliptical solutions may be more realistic than hyperbolic ones, there seems to be no good reason for rejecting the latter outright; indeed, qualitative analysis of the collapse phase should be much the same whichever type of solution be considered. It should be noted that the simplest solution analytically is the parabolic one given by II, (17). We find

$$S(r,t) = (1 - 3/2\, \alpha^{\frac{1}{2}} t)^{2/3} \qquad (108)$$

(cf.II, (36), provided that

$$-\infty < t < 2/3\, \alpha^{-\frac{1}{2}} .$$

Here,

$$1 + \frac{r}{S}\frac{\partial S}{\partial r} = 1 - \frac{\frac{1}{2}(\alpha^{\frac{1}{2}})' r t}{1 - \frac{3}{2}\alpha^{\frac{1}{2}} t}$$

$$= 1 - \frac{\dfrac{(\alpha r^3)^{\frac{1}{2}\,\prime}\, t}{r^{\frac{1}{2}}}}{1 - \frac{3}{2}\alpha^{\frac{1}{2}} t}$$

$$= \frac{1 - \frac{3}{2}\dfrac{\rho_o}{\bar{\rho}_o}\, \alpha^{\frac{1}{2}} t}{1 - \frac{3}{2}\alpha^{\frac{1}{2}} t} ,$$

whence

$$ds^2 = dt^2 - (1 - \tfrac{3}{2} a^{\frac{1}{2}} t)^{4/3} \left[\frac{(1 - \frac{3}{2} \frac{\rho_0}{\bar{\rho}_0} \alpha^{\frac{1}{2}} t)^2}{(1 - \frac{3}{2} \alpha^{\frac{1}{2}} t)^2} \, dr^2 + r^2 d\Omega^2 \right]. \qquad (109)$$

(109) is in fact a generalisation of the solution found by OPPENHEIMER & SNYDER (REF. 37) - their solution is just the uniform density one - in fact, therefore, a Robertson- Walker metric. It will be seen that the comoving coordinate system employed is admissable. in the sense of LICHNEROWICZ (REF. 26), in the space-time region given by

$$-\infty < t < \tfrac{2}{3} \alpha^{-\frac{1}{2}}; \quad 0 < r < r_b; \quad 0 < \theta < \pi; \quad 0 < \phi < 2\pi$$

(where r_b is the boundary of the dust-sphere), provided that throughout, ρ_0 is C' continuous, and $\rho_0 < \bar{\rho}_0$. The latter condition refers to the possibility of "overtaking", discussed in chapter II: in fact, II,(46) shows that a sufficient condition for overtaking not to occur is that ρ_0 be strictly decreasing.

Clearly, the comoving coordinate system must break down at events for which "overtaking" is occurring. After such an event (physically absurd), it is clear, as in the Newtonian treatment, that $M_R \neq M_r$, so that matching of solutions before and after such a "singularity" seems complicated.

It should be said that DATT (REF. 7), quoted also in LANDAU & LIFSCHITZ (REF. 24), obtained the general solution of the Field Equations for the dust-fluid, but not in a particularly helpful form. BONDI (REF. 3) also obtained a general solution, but again not in a form for which the Newtonian analogy was apparent. The present solution, which in elliptical form, is given by (102) – (105) was obtained wholly independently by the present author.

The full consequences of our solutions will be explored in the next chapter. The text topic to be considered is that of the <u>exterior</u> solution.

EXTERIOR SOLUTIONS & SCHWARZSCHILD COORDINATES

Comoving exterior solutions, as in II,(54), are an immediate consequence of our general solution. If we write

$$\alpha(r) = \frac{2GM}{r^3} = \alpha\left(\frac{r_b}{r}\right)^{3/2} \qquad (110)$$

where $\alpha = \bar{\rho}_0(r_b)$, we obtain all the possible exterior solutions to a dust sphere, radius r_b. In each of them, t is proper time along an empty-space geodesic. If the motion is from rest, then we may match each interior solution with a unique "comoving" exterior solution such that:

$$\lim_{r \uparrow r_b} R(r,t) = \lim_{r \downarrow r_b} R(r,t)$$

for fixed t, and such that $\dot{R}(r,0) = 0$ for $r \geq r_b$ (as well as for $r < r_b$). Again, if $V(r)$ is analytic, we can continue it in the region $r \geq r_b$, and therefore obtain a unique exterior solution to "match" the interior one.

BIRKHOFF'S THEOREM (REF. 2) states that any (reasonable) metric of a spherically symmetric, <u>empty</u> space-time must be <u>static</u> (see also

BONNOR, REF.). The criterion for a static space-time is that it admits a 1-parameter group of motions, whose trajectories are hypersurface orthogonal and timelike; as a consequence, a coordinate system $\{T, x^\alpha\}$ can be found such that the space-sections have equations $T = \text{const.}$, and the metric coefficients are independent of T. Thus constructed, the coordinate T is an invariant (excluding the trivial case of flat space-time). In terms of such a coordinate system, the exterior metric must take the form of the SCHWARZSCHILD EXTERIOR SOLUTION (REF. 45):

$$ds^2 = \left(1 - \frac{2GM}{R}\right) dT^2 - \left(1 - \frac{2GM}{R}\right)^{-1} dR^2 - R^2 d\Omega^2 \quad . \qquad (111)$$

It follows that all our solutions with $\alpha(r)$ as in (110) are local isometries (coordinate transformations) of one another, and of (111).

One particular merit of (111) is that the invariant R (radius of curvature of the 2-surfaces of transmittivity) is now adopted as the radial coordinate. The question naturally arises, is it possible that the interior solutions are locally isometric to metrics analogous to (111)? Clearly, these are non-static, so that at first sight, the coordinate T employed would not be an invariant. We therefore attempt to find a transformation of coordinates which enables us to write (77) locally in the form:

$$ds^2 = e^{\Lambda} dT^2 - e^{K} dR^2 - R^2 d\Omega^2 , \qquad (112)$$

where (Λ, K) are functions of (R, T). We require

$$g^{RT} = 0 ; \quad g^{RR} = -e^{-K} ;$$

since g^{ij} is a contravariant second-rank tensor field, if the transformation we seek exists, we must have

$$\left(\frac{\partial R}{\partial t}\right)^2 g^{tt} + \left(\frac{\partial R}{\partial r}\right)^2 g^{rr} = -e^{-K},$$

for R is a function of (r,t) alone, and $g^{rt} = 0$. But the value of g^{tt} is simply unity, and $g^{rr} = -e^{-\lambda}$, so that, by (87) this equation reads:

$$\dot{R}^2 - \frac{1}{f^2(r)} = -e^{-K};$$

using (96) and (98) this reduces to

$$e^{-K} = 1 - \frac{\alpha r^3}{R} \qquad (113)$$

where r is now regarded as a function of (R,T). An alternative form of (113) is

$$e^{-K} = 1 - \frac{2GM_r}{R} \qquad (114)$$

as is seen by comparing (90), (97) and (98). In an empty region of space-time, M_r is simply the gravitational mass M, as one would expect from Birkhoff's theorem.

Clearly, we seek a function $T(r,t)$ in order to effect the transformation. It is determined by the equation $g^{RT} = 0$, which imposes the condition

$$\frac{\partial R}{\partial t}\frac{\partial T}{\partial t} g^{tt} + \frac{\partial R}{\partial r}\frac{\partial T}{\partial r} g^{rr} = 0,$$

i.e.

$$\frac{\partial T}{\partial t} = \frac{1}{f^2(r)R'\dot{R}}\frac{\partial T}{\partial r} \qquad (115)$$

Since R is a known function of (r,t), (115) is in principle soluble, and so such a transformation exists. We shall shortly give explicit solutions in two special cases. Finally, the coefficient e^{Λ} is obtained from the equation for $g^{TT}{}_L$

$$\left(\frac{\partial T}{\partial t}\right)^2 - \frac{1}{f^2(r)R'^2}\left(\frac{\partial T}{\partial r}\right)^2 = e^{-\Lambda} \;;$$

substituting for $\frac{\partial T}{\partial r}$ from (115),

$$e^{-\Lambda} = -f^2(r)\left(\dot{R}^2 - \frac{1}{f^2(r)}\right)\left(\frac{\partial T}{\partial t}\right)^2$$

$$= f^2(r)\left(1 - \frac{\alpha r^3}{R}\right)\left(\frac{\partial T}{\partial t}\right)^2 \quad . \tag{116}$$

Because (115) is a first-order linear equation, the function T is determined up to an arbitrary function of itself. As our first example of an explicit solution of (115) we consider the case of an empty region of space-time, with spherical symmetry. Here, Birkhoff's theorem applies, and so it must be possible to choose the function T such that

$$e^{\Lambda} = 1 - \frac{2GM}{R} \quad ,$$

as in (111). According to (116) then,

$$\frac{\partial T}{\partial t} = \frac{1}{f(r)}\left(1 - \frac{2GM}{R}\right)^{-1} \tag{117}$$

which may be solved for T to give

$$T = \frac{1}{f(r)}\int \frac{dR}{\left(\frac{1}{f^2(r)} - 1 + \frac{2GM}{R}\right)^{1/2}\left(1 - \frac{2GM}{R}\right)} + A(r) \tag{118}$$

where r is kept constant in the integral, and $A(r)$ is to be determined from (115). However, it ought to be checked that (118) really does provide a solution of (115) - i.e. that (117) and (115) are consistent.

We require that the functions P, Q given by

$$P(r,t) = \frac{1}{f(r)}\left(1 - \frac{2GM}{R}\right)^{-1} \tag{119}$$

$$Q(r,t) = f^2(r)\, R'\dot{R}\, P(r,t) \tag{120}$$

should satisfy

$$\frac{\partial P}{\partial r} = \frac{\partial Q}{\partial t} \tag{121}$$

(so that $P\,dt + Q\,dr$ is to be an exact differential). Now, differentiating (120) with respect to t, and making use of (119),

$$\frac{\partial Q}{\partial t} = f(r)\left[\frac{\partial}{\partial t}(R'\dot{R})\left(1 - \frac{2GM}{R}\right)^{-1} - \frac{2GM}{R^2}R'\dot{R}^2\left(1 - \frac{2GM}{R}\right)^{-2}\right].$$

But

$$\frac{\partial}{\partial t}(R'\dot{R}) = \tfrac{1}{2}\frac{\partial}{\partial r}(\dot{R}^2) + R'\ddot{R}$$

$$= \frac{-f'(r)}{f^3(r)} - \frac{2GM}{R^2}R'$$

using (90) together with (97), and $M_r = M$. Hence

$$\frac{\partial Q}{\partial t} = f(r)\left(1 - \frac{2GM}{R}\right)^{-2}\left[\left(\frac{-f'(r)}{f^3(r)} - \frac{2GM}{R^2}R'\right)\left(1 - \frac{2GM}{R}\right) - \frac{2GM}{R^2}R'\left(\frac{1}{f^2(r)} - 1 + \frac{2GM}{R}\right)\right]$$

$$= f(r)\left(1 - \frac{2GM}{R}\right)^{-2}\left[\frac{-f'(r)}{f^3(r)}\left(1 - \frac{2GM}{R}\right) - \frac{1}{f^2(r)}R'\frac{2GM}{R^2}\right]. \tag{122}$$

By inspection of (119), the right-hand side of (122) is just $\frac{\partial P}{\partial r}$. Hence result.

It may be noted that the foregoing argument provides the basis of a novel proof of Birkhoff's theorem, provided that it be accepted that any empty spherically symmetric space-time must admit locally a system of normal Gaussian coordinates. If so, then as we have seen, the metric must be of the dust-fluid type with $\alpha(r)$ given by (110). The transformation of it to the Schwarzschild type (112) is governed by the equation (115), and we have proved this to have the solution (118). Finally, if $\bar{T}(r,t)$ is this particular solution, then the general solution of the linear equation (115) must be

$$T = \Phi\ (\bar{T})$$

where Φ' exists, and is non-zero in the domain of applicability of the metric (112): thus Φ^{-1} exists, and so the general solution may be converted into (111) by transformation of T only.

Returning to our general solution, we note that the "gravitational mass out to radius r" – the invariant M_r given by (97) is a fundamental parameter of the motion. It may be compared with the <u>inertial</u> mass, $M_r^{(i)}$, out to radius r:

$$M_r^{(i)} = \int_{\substack{t=0 \\ r' \leqslant r}} \rho_0 \sqrt{-g}\ d\tau'$$

$$= 4\pi \int_0^r 4\pi\rho_0 f(r) r^2 dr \ . \qquad (123)$$

The discrepancy between the expressions (123),(97) may be accounted for in terms of "gravitational binding energy;" see e.g. HOYLE (REF. 16). Note that $M_r^{(i)} = M_r = 0$ when $r = 0$, and (taking $f(r) = +\sqrt{f^2(r)}$),

$$\frac{d}{dr}\left(M_r^{(1)} - M_r\right) = 4\pi r^2 \rho_0 (f-1).$$

Now

$$f \gtreqless 1$$

accordingly as to whether the solution is elliptic, parabolic or hyperbolic, respectively; hence

$$M_r^{(1)} \gtreqless M_r$$

in each of these cases respectively. The quantity

$$\Omega = -\frac{GM_r}{R}$$

may be called the "gravitational potential" as in classical theory. As we remarked at the end of II, Ω is continuous even if ρ_0 is not – thus, if we could conceive of the instantaneous annihilation of matter in the region $r > r^*$ (or, perhaps, annihilation spreading outwards with the speed of light), then

$$\Omega = -\frac{GM_{r^*}}{R}$$

in the new exterior region, M_{r^*} being constant. This quantity could be measured by observing geodesic deviation of test particles. In fact, an empty spherical shell inside the matter would, in principle, be sufficient for such a measurement.

To close this chapter, we give a further example of a solution of (115). The special case of the homogeneous isotropic fluid is particularly straightforward: since here

$$R = rS(t)$$

(115) becomes

$$\frac{\partial T}{\partial t} = (rf^2(r)S\dot{S})^{-1} \frac{\partial T}{\partial r}$$

so that

$$\frac{\partial T}{\partial \int \frac{dt}{S\dot{S}}} = \frac{\partial T}{\partial \int f^2(r) r dr}$$

whence

$$T = \Phi\left(\int \frac{dt}{S\dot{S}} + \int f^2(r)\, r\, dr\right). \tag{124}$$

We may distinguish between the three types of solution: in the elliptic case,

$$\dot{S}^2 = \alpha\left(\frac{1}{S} - p\right)$$

and so

$$T = \Phi\left[\int \frac{\alpha^{-1} dS}{1 - pS} + \int \frac{r dr}{1 - p\alpha r^2}\right]$$

$$= \Psi\left[(1 - pS)\sqrt{1 - p\alpha r^2}\right], \tag{125}$$

where Ψ is an arbitrary function. Similarly, in the hyperbolic case,

$$T = \Psi\left[(1 + qS)\sqrt{1 + q\alpha r^2}\right] \tag{126}$$

whereas in the parabolic, Oppenheimer-Snyder case,

$$T = \Phi\left[\left(1 - \frac{3}{2}\alpha^{\frac{1}{2}} t\right)^{2/3} + \frac{1}{2}\alpha r^2\right] \tag{127}$$

where, of course, α is constant in these equations. We call the (R, T, θ, ϕ) system of coordinates "Schwarzschild coordinates".

In the next chapter, the mathematical and physical consequences of our solutions will be examined.

CHAPTER IV

RELATIVISTIC COLLAPSE - VARIOUS PROBLEMS

Our principal tasks in this chapter are, firstly, the full description of the motion of a dust fluid during the collapse phase and, secondly, to investigate the question of the boundary conditions linking an interior to an exterior solution. There is, however, a preliminary point that seems worthy of a brief discussion: as we have seen, the exterior solutions are all locally transformations of the Schwarzschild one, and are therefore transformations of one another. Can we therefore be certain that our class of interior solutions is such that no particular solution is isometric to some other? On physical grounds, this seems highly unlikely: each solution appears to correspond to a distinct ρ or u^i, whereas in the exterior no such field is defined.

If, in fact,

$$R = R_1(r,t); \quad R = R_2(r',t')$$

define two solutions of III, (90), where (r',t') are functions of (r,t), then if g_{ij}, $g_{i'j'}$ are isometric, their surfaces of transmittivity must have equal curvatures, whence the values of R_1, R_2 must be equal for corresponding (r,t), (r',t'). Now each metric is locally isometric to one of the form III, (112), where

$$(e^{-K})_1 = 1 - \frac{2GM_r}{R_1} \tag{1}$$

$$(e^{-K})_2 = 1 - \frac{2GM_{r'}}{R_2} \tag{2}$$

each of those being unique. If the two solutions are isometric, then (1) and (2) must be <u>identical</u>, and since $R_1 = R_2$,

$$M_r = M_{r'} \quad . \tag{3}$$

Excluding any range of r for which M_r is constant, this shows that r' is a function of r alone. By II, (55),

$$\frac{dM_r}{dr} \sim \rho_0(r)$$

so that

$$M_r = \text{const}, \quad r_a \leqslant r \leqslant r_a$$

implies

$$\rho_0(r) = 0 \quad \text{whenever} \quad r_a < r < r_b \ .$$

Thus (3) provides no restriction for empty regions only.

Otherwise, we have

$$r = h(r'), \qquad t = k(r',t'), \quad \text{say.}$$

It seems that this is inconsistent with

$$g_{rt} = g_{r't'} = 0 \ ;$$

in fact,

$$g_{r't'} = \frac{\partial t}{\partial r'}\frac{\partial t}{\partial t'}\, g_{tt} + \frac{\partial r}{\partial r'}\frac{\partial r}{\partial t'}\, g_{rr}$$

and since $\frac{\partial r}{\partial t'}$ vanishes, and $g_{tt} \neq 0$, t is independent of either r' or t'.

Furthermore,

$$g_{t't'} = \left(\frac{\partial t}{\partial t'}\right)^2 g_{tt} + \left(\frac{\partial r}{\partial t'}\right)^2 g_{rr}$$

whence

$$\frac{\partial t}{\partial t'} = \pm 1$$

and since t must therefore be a function of t' alone, we find

$$t = \pm t'.$$

Finally, then, t and t' vanish together, whence

$$r = R_1(r,0) = R_2(r',0) = r'$$

from our scaling of the radial coordinate. Thus the two solutions can differ only trivially (that is, according to our orientation for time).

We now consider the detailed description of the motion, treating the collapse phase only (an expansion phase is obtained merely by a time-reversal - which means that there is little of special interest in the consideration of "anticollapse" ~~(ref.~~). One of the chief difficulties in General Relativity is that coordinate systems employed in the description of a particular system may not be related in any obvious way to "observable" quantities. Now, invariants constructed from our various vector and tensor fields represent intrinsic mathematical properties of our geometrical model - one would therefore expect that they are, at least in principle, physically measurable. In particular, the tetrad components of any tensor field (or even, generally, an arbitrary "Geometrical Quantity" - Schouten (ref. 44)), are measurable; one would relate these quantities to observations made by an individual whose 4-velocity was u^i. But great care must be exercised in interpreting results obtained by means of the use of a particular set of coordinates, for there is no reason to suppose that there is some form of preferred "grid" in space-time.

The significance of comoving coordinates is quite straightforward: the θ,ϕ coordinates are invariants, labelling points on a surface of transmittivity in the same manner as the usual labelling on the unit sphere. The t coordinate represents proper time on the preferred u^i - geodesic passing through the event in question. The r coordinate is an invariant on the initial hypersurface - the radius of curvature of the transmittivity surfaces - and generally represents a unique labelling of the u^i-geodesics, with the given initial hypersurface property. We may therefore expect that results expressed in terms of comoving coordinates will be of physical significance, except the r coordinate is not necessarily related to a physical measure of radial distance.

Now the function $R(r,t)$ is, as we have seen, analogous to the Lagrangian coordinate in the Newtonian treatment, and will vanish at time

$$t = t_0(r) ,$$

(cf.II,(33')). Furthermore, an observer comoving with the fluid on the geodesic $r = r_0$ could in principle measure $\rho(r_0,t)$; according to our solution, if he survived to tell the tale, he would have measured

$\rho(r_0, t_c)$ to be infinite. Thus the collapse of a dust fluid, as viewed by a comoving observer, leads to a Newtonian-type catastrophe.

There seems no doubt that the set of events $\{t = t_c(r)\}$ represents a set of Space-Time singularities: if we wish to avoid such a property in our geometrical model, we must either regard the Dust Fluid Model as being unphysical, despite the considerations of Chapter I, or else seek to modify the Field Equations of General Relativity (cf. Hoyle & Narlikar – ref.(18)).

It is instructive to consider the behaviour of other invariants close to a singular event: the most important in this connexion arises from the relative motion of the fluid. The rate of expansion of a fluid element is given by

$$3\theta = u^i{}_{;i}$$

and since according to III(17),

$$(\rho u^i)_{;i} = 0$$

– conservation of mass for the dust fluid – we see that

$$\frac{1}{\rho}\frac{D\rho}{ds} = -u^i{}_{;i} = -3\theta .$$

Adopting our comoving coordinate system, this reduces to

$$\theta = -\frac{1}{3}\,\frac{1}{\rho}\,\frac{\partial \rho}{\partial t}$$

$$= \frac{1}{3}\,\frac{\partial}{\partial t}\,\log(R^2 R'), \qquad (5)$$

substituting for ρ by means of II(23). This result may also be verified by direct calculation:

$$u^{i}{}_{;i} = u^{i}{}_{,i} + \left(\frac{\partial}{\partial x^{i}} \log\sqrt{-g}\right)u^{i}$$

$$= \frac{\partial}{\partial t} \log\sqrt{-g}$$

in comoving coordinates.

The homogeneous and isotropic case is straightforward: here $R = rS(t)$, whence

$$\theta = \frac{\dot{S}}{S} \tag{6}$$

which clearly tends to minus infinity as $t \to t_c$. In other words, for the homogeneous, isotropic dust-fluid, the specific rate of volume shrinkage becomes infinite at the "space-time singularity". Can this result be extended to the inhomogeneous case?

Note that the general solution can always be written in the form

$$R = \phi Z\,(\psi t + a) \tag{7}$$

where ϕ, ψ, a are positive valued functions of r, and

$$\ddot{Z} = -\tfrac{1}{2}\,Z^{-2}\ .$$

We deduce that

$$R' = \phi' Z + \phi(\psi' t + a')\dot{Z}\ ,$$

and

$$\dot{R}' = (\phi\psi)'\dot{Z} + \phi\psi(\psi' t + a')\ddot{Z}\ .$$

Thus

$$\frac{\dot{R}'}{R'} = \frac{(\phi\psi)' - \frac{1}{2}\phi\,\psi(\psi' t + a')(Z^{2}\dot{Z})^{-1}}{\phi' Z\dot{Z}^{-1} + \phi(\psi' t + a')}$$

In order to investigate the limiting behaviour of this expression, we shall assume that

$$\phi' t_c(r) + a' \neq 0 \;;$$

the vanishing of $(\phi' t + a')$ near a singular event is clearly highly exceptional. This special case could be investigated by considering the detailed solutions of the differential equation for Z, but we may deem it unlikely that it would exhibit radically differing behaviour near a singularity.

Since the first integral of the equation for Z takes the form

$$\dot{Z}^2 = \frac{1}{Z} + \text{const},$$

and $\quad Z\xi(\phi t_c + a) = 0,\;$ we see that

$$Z^2\dot{Z} \to 0$$

and

$$Z\dot{Z}^{-1} \to 0$$

as $t \to t_c$. It follows that

$$\frac{\dot{R}'}{R'} = Z^{-2}\dot{Z}^{-1}\left(-\tfrac{1}{2}\phi + o(1)\right)$$

as $t \to t_c$, provided that $\phi' t_c + a' \neq 0$.

Now

$$\theta = \frac{1}{3}\left(\frac{2\dot{R}}{R} + \frac{\dot{R}'}{R'}\right)$$

$$= \frac{1}{3}\left(2\phi Z^{-1}\dot{Z} + \left(-\tfrac{1}{2}\phi + o(1)\right)Z^{-2}\dot{Z}^{-1}\right)$$

as $t \to t_c$,

$$= \frac{1}{3}Z^{-1}\dot{Z}\left(2\phi + \left(-\tfrac{1}{2}\phi + o(1)\right)Z^{-1}\dot{Z}^{-2}\right).$$

But
$$Z\dot{Z}^2 = 1 + kZ$$

where k is a constant. Hence
$$Z^{-1}\dot{Z}^{-2} = 1 + o(1)$$

and so
$$\theta = \frac{1}{3} Z^{-1}\dot{Z}\left(\frac{3}{2}\psi + o(1)\right)$$

$$= \frac{1}{2} Z^{-1}\dot{Z}(\psi + o(1)). \qquad (8)$$

Finally, $\psi > 0$, and (in the collapse phase), $\dot{Z} < 0$.

Thus
$$\theta \to -\infty \quad .$$

Note, however, that for nearly homogeneous solutions, the asymptotic behaviour is quantitively different from that of the homogeneous solution, for if we let $\psi \to 1$, $a \to 0$ in (8), we find
$$\theta \sim \frac{1}{2}\frac{\dot{S}}{S} \qquad (9)$$

in the usual notation. Comparison with (5) shows that the effective collapse is at half the rate for homogeneous isotropic collapse. This curious result throws doubt on the validity of the homogeneous solution near a singularity - the homogeneous configuration must be unstable under these conditions. It must be emphasised that we have been dealing only with the case of spherical symmetry, however. This result might have applications in the study of evolutionary cosmologies close to a "space-time origin". Although we have considered an artificial form of inhomogeneity, some doubt must be thrown on the validity of the Robertson-Walker assumptions in these extreme circumstances. We shall not consider this point any further, but note there is a problem here which may require investigation.

THE OBSERVATIONAL SITUATION

The quasi-Newtonian behaviour of the Relativistic Dust-Fluid is, strictly speaking, the behaviour that would be noted by an observer freely falling with the fluid and in its proximity. This has little relevance to astronomical observation.

In considering observations made by a distant observer, we should strictly speaking consider a succession of light-experiments. As we shall see, there is little difficulty in determining the red shift of the light emitted by a collapsing star, but an analysis of non-radial light radiation (null geodesics) - resulting in some tiresome calculations - would be required if we were to consider theoretically the appearance of an extended collapsing source of radiation. However, since the <u>exterior</u> metric of the empty region exterior to a spherically symmetric body is just the Schwarzschild metric (assuming that one can ignore the disturbance engendered by the radiation), we can analyse collapse in terms of the preferred coordinates (R,T).

Following the notation of Jordan, Ehlers and ~~Sachs~~ Kundt (Ref. 22) we let k^i be the tangent vector to a null congruence which emerges radially from a point. Then the "parallax distance" ρ_P is defined to be the inverse divergence of the rays:

$$\rho_P = \left(\tfrac{1}{2} k^{i}{}_{;i}\right)^{-1} \quad . \tag{10}$$

Strictly, the stress energy tensor in the exterior region, which is empty save for the radiation, takes the form

$$T^{ij} = \mu k^i k^j$$

where μ is the energy-density, but we neglect μ as a first approximation.

Inspecting the usual form of the Schwarzschild metric, it will be seen that the radial null rays must satisfy the equation

$$\left(1 - \frac{n^2}{R}\right)^{\frac{1}{2}} dT = \pm \left(1 - \frac{n^2}{R}\right)^{-\frac{1}{2}} dR \tag{11}$$

and, indeed, the solutions of this are just the radial null geodesics. For, since T is the static time coordinate, it is ignorable, and one of the geodesic equations must reduce to

$$\left(1 - \frac{n^2}{R}\right) \frac{dT}{du} = h = \text{const.}, \tag{12}$$

where u is an affine parameter. The radial null geodesics are therefore given by the solutions of (12), together with

$$\left(1 - \frac{n^2}{R}\right)^{\frac{1}{2}} \frac{dT}{du} = \pm \left(1 - \frac{n^2}{R}\right)^{-\frac{1}{2}} \frac{dR}{du}, \tag{13}$$

if solutions exist. Substituting (12) into (13) gives

$$\frac{dR}{du} = \pm h$$

which determines u up to an affine transformation; the remaining equation simply reduces to (11). For convenience, we may set $h = 1$.

Now, by definition,

$$k^i = \frac{dx^i}{du}$$

where u is the affine parameter; hence, for radial light-rays,

$$k^i = \left(\left(1 - \frac{n^2}{R}\right)^{-\frac{1}{2}}, 1, 0, 0\right)$$

where the first component is the T-component. Thus

$$k^i{}_{;i} = \Gamma^i{}_{ij}\, k^j$$

$$= \frac{\partial}{\partial x^i} (\log\sqrt{-g}\; k^j$$

$$= 2/R \;.$$

Hence, according to (10),

$$\rho_P = R,$$

and so R is indeed the parallax distance. For any spherically symmetric metric, this coincides with the luminosity distance (~~Sachs, Ref. ———~~).

Thus R is, in the astronomical sense, an observable, and it is meaningful, in deriving the Schwarzschild metric, to impose the familiar condition

$$g_{ij} \to \eta_{ij} \quad \text{as} \quad R \to \infty \;,$$

where $\{\eta_{ij}\}$ are the components of the Minkowski metric.

For an observer exterior to the body, but "resisting" its gravitational field so that his affine distance from the centre remains constant, the coordinate T is simply a multiple of his proper-time:

$$ds = \left(1 - \frac{n^2}{R_s}\right)^{\frac{1}{2}} dT \;, \qquad (14)$$

where his worldline satisfies

$$R = R_s = \text{const.}$$

Clearly this breaks down if $R_S \leqslant n^2$ (the "Schwarzschild Limit"), and so we can deduce that we cannot have a timelike worldline $R = R_S$ if $R_S \leqslant n^2$. It used to be thought that this implied that "nothing can pass through the Schwarzschild singularity", but this is an oversimplification.

Clearly, we may regard a "frame of reference" based on worldlines $R =$ const. as a "static frame" (whether or not the metric is static). More precisely a static frame is determined by an orthogonal viertein containing the tangent vector to the curve congruence $R =$ const., $\theta =$ const., $\phi =$ const. Since R is well-defined for all spherically symmetric space-times, we may regard this definition as meaningful even in the interior region of the body; the sole proviso being that the concept "static frame" is to be restricted so as to apply only to regions throughout which the congruence is timelike. Associated with such frames are the Schwarzschild coordinates (T, R, θ, ϕ), where T is determined up to an arbitrary function of itself (see Chapter III, (112) and (115)).

In the exterior region, as we have noted, there is a preferred coordinate T as determined by the asymptotic condition for the Schwarzschild solution: T is simply the proper time on $R =$ const. "at infinity". Equation (14) suggests a procedure for determining the form of T in the interior. Consider two neighbouring curves $R = R_1$, $R = R_2$ (with $\theta -$ const., $\phi -$ const.). Suppose that in the neighbourhood of a certain space-time event $R = R_1$ lies in the interior region, and $R = R_2$ lies in the exterior. Clearly the proper-time intervals on neighbouring segments of the two curves must be comparable, if the coordinate T is used as a parameter, provided that both curves are timelike. Since throughout

$$ds = g_{TT}^{\frac{1}{2}} dT,$$

it follows that for the T coordinates in the interior and the exterior regions to be compatible, g_{TT} must be continuous across the boundary. Note that this argument is based on the geometrical construction for Schwarzschild coordinates, and not upon some general "boundary conditions".

Since our coordinate systems are orthogonal, continuity of g_{TT} implies continuity of g^{TT}. We shall now explore the consequences for the special case of the homogeneous, isotropic dust-fluid.

Now in terms of the transformation given by III (115),

$$g^{TT} = \left(\frac{\partial T}{\partial t}\right)^2 - \frac{1}{f^2(r)R'^2_4}\left(\frac{\partial T}{\partial r}\right)^2$$

$$= \Phi'^2_a \left(\frac{1}{S^2\dot{S}^2} - \frac{1}{f^2(r)S^2} \cdot r^2f^4(r)\right)$$

using III, (124), which applies to the case of a uniform body. Rearranging,

$$g^{TT} = \frac{r^2f^2(r)}{S^2\dot{S}^2} \quad \Phi'^2_a \left(\frac{1}{r^2f^2(r)} - \dot{S}^2\right)$$

$$= \frac{r^2f^2(r)}{S^2\dot{S}^2} \quad \Phi'^2_a \left(\frac{1}{r^2} - \frac{\alpha}{S}\right)$$

$$= \frac{f^2(r)}{S^2\dot{S}^2} \quad \Phi'^2_a \left(1 - \frac{\alpha r^2}{S}\right)$$

using III, (98).

Hence $$\lim_{r \uparrow r_b} g^{TT} = \frac{f^2(r_b)}{S^2\dot{S}^2} \quad \Phi_b'^2 \left(1 - \frac{\alpha r_b^2}{S}\right) . \tag{15}$$

Comparison with the exterior Schwarzschild Exterior Solution then shows that

$$\frac{f^2(r_b)}{S^2\dot{S}^2} \quad \Phi_b'^2 \quad \left(1 - \frac{\alpha r_b^2}{S}\right) = 1 \tag{16}$$

since

$$\frac{2GM}{\mathcal{R}_b} = \frac{\alpha r_b^3}{r_b S}$$

$$= \frac{\alpha r_b^2}{S} \quad .$$

Equation (16) can in principle be solved in order to determine the functional form of Φ. In order to obtain explicit solutions we must distinguish between the three cases determined by III (125), (126),(127). In the "elliptic" case, (16) gives

$$\Phi_b' \left(1 - \frac{\alpha r_b^2}{S}\right) = \pm (1 - p\alpha r_b^2)^{\frac{1}{2}} S\dot{S}$$

and writing

$$\Phi\left(\frac{1}{p\alpha} \log \frac{1}{u}\right) = \Psi (u)$$

as in II (125),

$$v\Psi'(v) \cdot -p\alpha \cdot \left(1 - \frac{\alpha r_b^2}{S}\right) = \pm (1 - p\alpha r_b^2)^{\frac{1}{2}} S\dot{S}$$

where

$$v = (1 - pS)(1 - p\alpha r_b^2)^{\frac{1}{2}} \quad .$$

In order to derive a differential equation for Ψ as a function of v, we note that

$$\dot{S} = -\alpha^{\frac{1}{2}}\left(\frac{1}{S} - p\right)^{\frac{1}{2}}$$

$$= -\frac{\alpha^{\frac{1}{2}}}{S^{\frac{1}{2}}}(1 - pS)^{\frac{1}{2}} \quad .$$

Thus

$$\Psi'(v).p\alpha = \pm \frac{\alpha^{\frac{1}{2}}\, S^{\frac{1}{2}}}{(1-pS)^{\frac{1}{2}}\left(1-\frac{\alpha r_b^2}{S}\right)}$$

$$= \pm \frac{\alpha^{\frac{1}{2}}\left(1-\frac{v}{\lambda}\right)^{3/2} p^{-3/2}}{\left(\frac{v}{\lambda}\right)^{\frac{1}{2}}\left(\frac{1-v/\lambda}{p} - \alpha r_b^2\right)}$$

$$= \pm\, p\alpha^{\frac{1}{2}}.p^{-3/2}\frac{\left(1-\frac{v}{\lambda}\right)^{3/2}}{\left(\frac{v}{\lambda}\right)^{\frac{1}{2}}\left(\lambda^2 - \frac{v}{\lambda}\right)}$$

where $\lambda^2 = 1 - p\alpha r_b^2$. (17)

Simplifying,

$$\Psi' = \pm\, \alpha^{-\frac{1}{2}} p^{-3/2}\frac{(\lambda - v)^{3/2}}{v^{\frac{1}{2}}(\lambda^3 - v)} \quad . \qquad (18)$$

In general, if

$$u = (1 - pS)(1 - p\alpha r^2)^{\frac{1}{2}}, \qquad (19)$$

our transformation of coordinates is determined by

$$p^{3/2}\alpha^{\frac{1}{2}}T = \pm \int^{u} \frac{(\lambda - v)^{3/2}}{v^{\frac{1}{2}}(\lambda^3 - v)}\,dv \qquad (20)$$

The sign ambiguity may be removed by insisting that T-time and t-time should "run the same way". A necessary condition for this is that, throughout the region of applicability of the T-coordinate,

$$\frac{dT}{dt} > 0$$

on each worldline given by

$$r = \text{const.}, \quad \theta = \text{const.}, \quad \phi = \text{const.}$$

On such a geodesic,

$$\frac{dT}{dt} = \frac{\partial T}{\partial t}$$
$$= \pm \frac{(\lambda - u)^{3/2}}{u^{\frac{1}{2}}(\lambda^2 - u)} \frac{\partial u}{\partial t}$$

using (19). But

$$\frac{\partial u}{\partial t} = -p\dot{S}\,(1 - p\alpha r^2)^{\frac{1}{2}}$$

by differentiation of (18). Since during the collapse phase $\dot{S} < 0$, it is clear that we must take the positive sign in (19).

The question naturally arises - are $\dot{T}$ and t aligned for all observers? The answer is in the positive, provided that we remain in the space-time "coordinate patch" in which T is defined (as a time coordinate) - we must not intersect a singularity of the transformation

$$T = T(r,t)\ .$$

For

$$\frac{dT}{dt} = \frac{\partial T}{\partial t} + \frac{\partial T}{\partial r}\frac{dr}{dt}$$
$$= \frac{\partial T}{\partial t}\left(1 + \frac{rS\dot{S}}{1-p\alpha r^2}\ \frac{dr}{dt}\right) \qquad (21)$$

since

$$0 = g^{RT} = r\dot{S}\,\frac{\partial T}{\partial r} + S\,.\,\frac{(1-p\alpha r^2)}{S^2}\,\frac{\partial T}{\partial t}\,.$$

Furthermore, $\dot{S} < 0$, and so $\frac{dT}{dt}, \frac{\partial T}{\partial t}$ can be of opposite sign only if $\frac{dr}{dt} > 0$, and

$$\frac{r^2 S^2 \dot{S}^2}{(1-p\alpha r^2)^2}\left(\frac{dr}{dt}\right)^2 > 1\,. \qquad (22)$$

Now, tangent to any timelike worldline,

$$\left(\frac{ds}{dt}\right)^2 = 1 - S^2\left({}^1/1-p\alpha r^2\left(\frac{dr}{dt}\right)^2 + r^2\,\frac{(d\Omega)^2}{dt^2}\right)$$

whence

$$\frac{S^2}{1-p\alpha r^2}\left(\frac{dr}{dt}\right)^2 < 1\,. \qquad (23)$$

Thus

$$\frac{r^2 S^2 \dot{S}^2}{(1-p\alpha r^2)^2}\left(\frac{dr}{dt}\right)^2 < \frac{r^2 \dot{S}^2}{1-p\alpha r^2}$$

$$= \frac{\alpha r^2}{S}\,\frac{(1-pS)}{1-p\alpha r^2}$$

$$< 1$$

in contradiction of (21), provided that $S > \alpha r^2$.

But,

$$S \leqslant \alpha r^2 \Rightarrow S \leqslant \alpha r_b^2$$

giving

$$R_b \leqslant \alpha r_b^2 = 2GM,$$

so that the collapse has "passed through" the Schwarzschild "limit".

Clearly, the T-transformation is invalid here, since if $S \leqslant \alpha r^2$,

$$u \geqslant (1 - p\alpha r^2)^{3/2} \geqslant (1 - p\alpha r_b^2)^{3/2} = \lambda^3,$$

using the notation of (18). The integral in (19) is therefore improper, if the constant of integration has been chosen so as to permit T to be defined outside the Schwarzschild limit.

Hence

$$\frac{dT}{dt} \Big/ \frac{\partial T}{\partial t} > 0$$

for all observers who do not pass through the "Schwarzschild Singularity", and since T and t are aligned for "comoving" observers, they are aligned for all "qualifying" observers.

It should be noted that the foregoing argument can be adapted to the general case: we have restricted ourselves to the homogeneous case in order to find an explicit functional form for T, which cannot be achieved in general. In general, (21) may be written:

$$f^4(r)R'^2\dot{R}^2\left(\frac{dr}{dt}\right)^2 > 1 ;$$

(22) becomes

$$f^2(r)R'^2\left(\frac{dr}{dt}\right)^2 < 1$$

whence

$$f^4(r)R'^2\dot{R}^2\left(\frac{dr}{dt}\right)^2 < 1 - f^2(r)(1 - \frac{\alpha r^3}{R}) < 1$$

outside the "Schwarzschild singularity", as before (cf.III, (98)).

Particular cases of (19) were obtained by Oppenheimer & Snyder (Ref. 37), Hoyle & Narlikar (Ref. 18) and Nariai & Tomita (Ref. 30).

Returning to the homogeneous case, we shall proceed to an evaluation of the integral (19) in terms of elementary functions. Substituting $v = \lambda \sin^2\delta$, one finds that

$$p^{3/2}\alpha^{\frac{1}{2}}T = \lambda \int^{u} \frac{2\sin\delta \cos^4\delta d\delta}{\sin\delta(\lambda^2-\sin^2\delta)}$$

(choosing the plus sign in (19)); after some reduction, one obtains

$$p^{3/2}\alpha^{\frac{1}{2}}T = 2\lambda \left[(2-\lambda^2)\delta\Big|_u + \tfrac{1}{2}(\sin\delta\cos\delta - \delta)\Big|_u + (1-\lambda^2)^2 \int^{u} \frac{d\delta}{\lambda^2-\sin^2\delta}\right] \qquad (24)$$

$$= \lambda(3-2\lambda^2)\psi + \lambda\sin\psi\cos\psi + (1-\lambda^2)^{3/2}\log(\cos(\psi-\epsilon)/\cos(\psi+\epsilon)) + \text{const.} \qquad (25)$$

where

$$\psi = \sin^{-1}\sqrt{(u/\lambda)} \qquad (26)$$

and

$$\epsilon = \cos^{-1}\lambda\,. \qquad (27)$$

Since $u=0$ for $\dot{S}=0$, if we take the principal value of the function $\sin^{-1}$ in (25), and set $T=0$ when $\dot{S}=0$, we can dispose of the arbitrary constant in (24).

Note that as $u \uparrow \lambda^2$, $\cos(\psi+\epsilon)\downarrow 0$, and $T \to +\infty$.

This shows quite clearly the singular behaviour of the coordinate system (R,T). Hoyle and Fowler (Ref. 17) consider the possibility of invisible massive objects which have already collapsed beyond the Schwarzschild limit, yet are detectable from the astronomical point of view. As we have seen, although T can be defined in a coordinate patch within the Schwarzschild limit (by applying (19), with appropriate change of sign, provided that $u < \lambda$), it cannot be defined in any patch which simultaneously includes regions within and without the Schwarzschild limit, whence collapse through this "limit" could not be observed from an outside observer in a static frame, (cf. Oppenheimer & Snyder (Ref. 37)). Hoyle subsequently appears to have discarded this notion - see, for example, Narlikar's article in Discovery (Ref. 32).

THE SCHWARZSCHILD LIMIT

We have used the phrases "Schwarzschild Limit" and "Schwarzschild Singularity" somewhat freely in this chapter, and the concept requires clarification. The usual static form of the exterior Schwarzschild metric breaks down on the hypersurface R = 2M, and it used to be considered that this surface represented some form of geometrical singularity, or limiting configuration for a massive, dense body. This viewpoint is not generally accepted nowadays. This is not to say that the hypersurface does not have geometrical significance - indeed, its properties indicate that General Relativity has curious consequences in certain extreme situations.

It is convenient to examine the exterior Schwarzschild metric in "comoving" – i.e. "free-falling" coordinates. There are a whole class of such coordinate systems, each system corresponding to a choice of initial velocity for the free-falling frame. In the notation of Chapter II, each corresponds to a choice of p (or q). Purely for convenience, we shall consider first the system in which the metric takes the form:

$$ds^2 = dt^2 - \frac{1}{1 - \frac{2M}{r}}\left(\frac{\partial R}{\partial r}\right)^2 dr^2 - R^2 d\Omega^2 \qquad (28)$$

where

$$R = r\,\xi\left(\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t\right) \qquad (29)$$

and the function ξ is defined in II, (8). This metrical form was obtained also by Hoyle and Narlikar (Ref. 18), who appear to regard it as a unique "comoving" solution. It is, of course, merely a special case of the class of "elliptical" solutions: Oppenheimer & Snyder (Ref. 37) obtained the parabolic solution, which corresponds to a frame which free-falls from rest at infinity.

(28) has the inherent disadvantage that the coordinate patch defined by (r, t, θ, ϕ) with $r > 0$, $|t| < \frac{\pi}{2}\left(\frac{r^3}{2M}\right)$, does not map $(1,1)$ into the whole Schwarzschild Space-Time. The coordinate system has a singularity at $r = 2M$ (not to be confused with the null surface $R = 2M$ – the "Schwarzschild Limit"), but is quite acceptable as a form of the exterior metric to a spherical dust-body which at time

$t = 0$ is outside the hypersurface $R = 2M$. This point will be explained further in the next chapter.

The parabolic form is ideal in that it spans the whole of space-time (other than the curve $r = 0$), excluding the essential space-time singularity $R = 0$. The Schwarzschild metric takes the form

$$ds^2 = dt^2 - \left(\frac{\partial R}{\partial r}\right)^2 dr^2 - R^2 d\Omega^2 \tag{30}$$

where

$$R = r\left(1 - \frac{3}{2}\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t\right)^{2/3} \tag{31}$$

and

$$r > 0, \quad -\infty < t < \frac{2}{3}\left(\frac{r^3}{2M}\right)^{\frac{1}{2}}. \tag{32}$$

Clearly, $\frac{\partial R}{\partial r}$ cannot vanish without the phenomenon of "overtaking" discussed in Chapter II occurring — and this is quite impossible for a case of a free falling frame. As a check, we can evaluate $\frac{\partial R}{\partial r}$:

$$\frac{\partial R}{\partial r} = \left(1 - \frac{3}{2}\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t\right)^{2/3} + \frac{3}{2}\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t \left(1 - \frac{3}{2}\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t\right)^{-1/3}$$

$$= \left(1 - \frac{3}{2}\left(\frac{2M}{r^3}\right)^{\frac{1}{2}} t\right)^{-1/3} \tag{33}$$

which cannot vanish for finite t, and becomes infinite only at $R = 0$. The singularity at $r = 0$ is no coincidence, since there must be a mass at the initial origin in order to give rise to a Schwarzschild solution in the first place.

Inspection of (29) and (30) indicates that there is no objection in principle to passing through the "Schwarzschild Limit". An observer whose worldline satisfied $r = r_0$, $\theta = \theta_0$, $\phi = \phi_0$, where $r_0 > 2GM$, would in finite proper time pass through the surface $R = 2GM$. The worldline would throughout be a timelike geodesic. However, it should be noted that such an observer will always pass into the hypersurface, never out of it. Indeed, any observer once inside the hypersurface can not escape without firstly passing through the essential singularity $R = 0$. For let k^a be any timelike unit vector, so that $k^a k_a = 1$. Suppose further that k^a be future orientated, so that $k^t = \frac{dt}{ds} > 0$. Then

$$\frac{dR}{ds} = k^t \dot{R} + k^r R' \ , \tag{34}$$

since R is independent of (θ, ϕ). Substituting $k^a k_a = 1$ into (29) shows that

$$R'^2 (k^r)^2 \leqslant (k^t)^2 - 1. \tag{35}$$

Now according to (32), $R' > 0$: thus certainly

$$|k^r| R' < |k^t| \ . \tag{36}$$

But $k^t > 0$, $\dot{R} < 0$, whence

$$\frac{dR}{ds} < k^t (1 + \dot{R})$$

whatever the sign of k^r. From (30) one finds $\dot{R} = -\left(\frac{2GM}{R}\right)^{\frac{1}{2}}$, and so if $R \leqslant 2GM$, $\frac{dR}{ds} < 0$. Hence any observer once inside $R = 2GM$ will remain inside, and fall into $R = 0$ whatever his worldline.

Even a light ray will be subject to this behaviour, for according to General Relativity this is represented by a null geodesic. In fact, if k^a is any future-orientated null-vector, we can replace (34) by

$$R'^2 (k^r)^2 \leq (k^t)^2$$

from which we deduce (35), and proceed as before.

The null hypersurface $R = 2GM$ provides an example of what Hawking terms a "closed trapped surface" (Ref. 12). For an observer inside it, the hypersurface represents a space-time horizon. In fact, if we consider the situation with time reversed, we have a form of violent explosion, with (29) representing the exterior to a finite evolutionary universe – the "horizon" phenomenon assuming significance in the early stages.

EXTERNAL OBSERVATION

As we have noted, the Schwarzschild coordinate T becomes infinite as the hypersurface $R = 2GM$ is approached. It seems natural to suppose that an observer comoving with the boundary of the collapsing body will observe events on a body fixed in a static frame as if they were increasingly "speeded up" and conversely for an observer in the static frame. However, the situations of the observers are not conjugate, as may be seen from an analysis of the red-shift photons emerging from the collapsing body as being increasingly red-shifted since the "Doppler" red shift due to the receding boundary will combine with the "gravitational" red shift. However, the "collapsing" observer will receive light from the static body which may be red- or violet-shifted, depending on the circumstances - the gravitational violet shift will oppose the Doppler red-shift.

Analysis of the red-shift of light emitted by a collapsing sphere is complicated by the question of the non-radial light paths. In classical language, the frequency observed will be subject to a Doppler and transverse Doppler effect which will vary with the angle of emission. This will give rise to a form of chromatic ~~observation~~ aberration. Precise analysis of non-radial null geodesics proves somewhat tiresome, and it is clear that the bulk of the radiation received will have been emitted in a direction almost aligned with the radial. Accordingly, in the following discussion we shall consider radial emission only.

The exterior metric is most conveniently handled in the form of equation (29). Using (30),

$$\frac{\partial R}{\partial t} = -\frac{n}{R^{\frac{1}{2}}} \qquad (37)$$

where

$$n^2 = 2GM. \qquad (38)$$

Also, inspection of (32) yields

$$\frac{\partial R}{\partial r} = \left(\frac{r}{R}\right)^{\frac{1}{2}} \qquad (39)$$

and so

$$ds^2 = dt^2 - \frac{r}{R}\,dr^2 - R^2 d\Omega^2. \qquad (40)$$

The geodesic equations are always satisfied by

$$R^2 \sin^2\theta \, \dot{\phi}^2 = R^2 \dot{\theta}^2 = 0 \, ,$$

and in particular the congruence of radial and null geodesics is given by

$$dt^2 - r/R \, dr^2 = 0 \, . \tag{41}$$

Integration of (40) is straightforward: we have

$$dR = \frac{\partial R}{\partial r} dr + \frac{\partial R}{\partial t} dt$$

$$= (r/R)^{\frac{1}{2}} dr - n/R^{\frac{1}{2}} dt$$

and so the equation of a null geodesic becomes

$$dt = \pm \, (dR + n/R^{\frac{1}{2}} dt)$$

giving either

$$(1 - n/R^{\frac{1}{2}}) dt = dR$$

or

$$(1 + n/R^{\frac{1}{2}}) dt = -dR \, .$$

Note that for $R > n^2$ there is one outward pointing and one inward pointing null geodesic through any point (in the sense that $\frac{dR}{dt} \gtrless 0$) whilst for $R < n^2$ both geodesics (and the associated null tangent vectors) point inwards.

From the point of view of the external observer, the outward pointing geodesic is the significant one. If a photon is emitted radially at $R = R_0$ at time t_0, and received at the event (R, t), then

$$f(R) - f(R_0) = t - t_0 \tag{42}$$

where

$$f'(R) = (1 - n/R^{\frac{1}{2}})^{-1}.$$

Suppose now that (R_0, t_0) is situated on the boundary of the collapsing body, and that a neighbouring photon is emitted from the same point on the boundary at time $t_0 + \Delta t_0$. Reception takes place on the fixed surface $R = \text{const.}$ To first order,

$$\Delta R = 0 \;;$$

$$\Delta R_0 = \frac{\partial R_0}{\partial t_0}\, \Delta t_0$$

$$= - n/R_0^{\frac{1}{2}} \Delta t_0 \tag{43}$$

using (36). Thus

$$f'(R)\Delta R - f'(R_0)\Delta R_0 = \Delta t - \Delta t_0$$

i.e. $$-(1 - n/R_0^{\frac{1}{2}})^{-1} \cdot -n/R_0^{\frac{1}{2}} \Delta t_0 = \Delta t - \Delta t_0 \,.$$

Hence $$\Delta t = \left(1 + \frac{n/R_0^{\frac{1}{2}}}{1 - n/R_0^{\frac{1}{2}}}\right)\Delta t_0$$

$$= (1 - n/R_0^{\frac{1}{2}})^{-1} \Delta t_0 \,. \tag{44}$$

Now the red-shift, z, of light observed on the surface $R = \text{const.}$ is given by

$$1 + z = \frac{\Delta s}{\Delta s_0}$$

where Δs_0, Δs are the proper time intervals between neighbouring emissions and their receptions (cf. Ref. 27)). On the collapsing body, $\Delta r = 0$, and so (39) gives

$$\Delta s_0 = \Delta t_0 .$$

However, on the receiving surface $\Delta R = 0$, giving

$$(r/R)^{\frac{1}{2}}\Delta r = n/R^{\frac{1}{2}}\Delta t, \text{ to first order,}$$

Substituting into (39), setting $\Delta\theta = \Delta\phi = 0$

$$(\Delta s)^2 = (1 - n^2/R)(\Delta t)^2$$

and assuming that the t coordinate has been aligned with the proper-time s,

$$\Delta s = (1 - n^2/R)^{\frac{1}{2}}\Delta t .$$

Substitution into (43) gives

$$1 + z = \frac{\Delta s}{\Delta s_0} = \frac{(1 - n^2/R)^{\frac{1}{2}}}{1 - n/R_0^{\frac{1}{2}}} . \qquad (45)$$

The red-shift determined by (45) is simply a combination of a Doppler shift – between an observer comoving with the boundary of the body and an observer, instantaneously coincident, but with fixed R – and a gravitational red-shift. This will become clearer if we write (44) in the following form:

$$1 + z = \left(\frac{1 - n^2/R}{1 - n^2/R_0}\right)^{\frac{1}{2}} \cdot \left(\frac{1 + n/R_0^{\frac{1}{2}}}{1 - n/R_0^{\frac{1}{2}}}\right)^{\frac{1}{2}}$$

$$= \left(\frac{1 - n^2/R}{1 - n^2 R_0}\right)^{\frac{1}{2}} \cdot \left(\frac{1 + v/c}{1 - v/c}\right)^{\frac{1}{2}}$$

where $v/c = \left|\frac{\partial R_0}{\partial t_0}\right|$. Note that this interpretation illustrates the connection between the classical Lagrangian variables (r,t) and their

Relativistic counterparts.

Examination of (45) confirms that $z \to \infty$ as $R_0 \to 2GM$ (from above). Thus a collapsing body viewed from a distance will become progressively fainter: insofar as the collapse itself could be observed, this would appear to slow down eventually. If we assume that observation takes place at a considerable distance from the body, then the term n^2/R can be ignored, and we have:

$$n/R_0^{\frac{1}{2}} = \frac{z}{1+z} ;$$

from (43) and (44) may be deduced:

$$\frac{dR_0}{ds} = -\frac{z}{(1+z)^2} \qquad (46)$$

Thus $\left|\frac{dR_0}{ds}\right|$ increases with z until $z = 1$, when it decreases. At $z = 1$, in normal units, $\left|\frac{dR_0}{ds}\right| = \frac{1}{4} c$.

The red-shift question is somewhat complex, however. It is instructive to consider the rate of change of z with time: we have

$$\frac{dz}{ds} = -\frac{1}{2} \frac{n}{R_0^{3/2}(1-n/R_0^{\frac{1}{2}})^2} \frac{dR_0}{ds}$$

$$= \frac{1}{2n^2} \frac{z^4}{(1+z)^3} \qquad (47)$$

using (45) and (46).

For large z, $\frac{dz}{ds} \sim z$. Accordingly there would be no hope of identifying spectral lines, and the red-shift would rapidly become so great that the body would effectively disappear. It must be emphasised, however, that it would not then have passed through its gravitational radius: (45) shows that it will <u>never</u> do so <u>in the time of the distant observer</u>. This analysis should settle the hoary question of "collapsed bodies" (Ref. 17), which has been revived recently (cf. Ref. 39). There certainly could exist "Black Holes", being matter which had collapsed to near the Schwarzschild limit and was therefore invisible, but such matter would not have passed through this limit.

Furthermore, for $z \sim 1$, $\frac{dz}{ds}$ may be very large indeed, depending on the mass of the body. In standard units, the typical time interval for changes in z will be given by

$$T_z \sim 16\, GM/c^3 .$$

Clearly, only for very large masses will it be possible to observe a red-shift of the order of unity - if M is small, the body will rapidly disappear, as discussed in the preceding paragraph. For a reasonably stable observational situation, the gravitational radius should be of order of magnitude about 1/100th light year (or more). This would require a mass of, say $10^8 M_{\odot}$. A red-shift of one-half would be stable observationally if the mass was about $10^7 M_{\odot}$ or more.

We conclude that if a body undergoing <u>catastrophic</u> collapse exhibits a stable spectrum, with a substantial red-shift arising largely from the collapse itself (rather than the normal cosmological expansion), then the mass of the body must be of a magnitude approaching that of a typical galaxy. This conclusion may be of significance in theoretical discussions about quasars (see the introduction). In fact no variation in red shift has as yet been observed in a quasar.

It should be noted that we have restricted this analysis to the case of the parabolic solution for pressure-free collapse: as has been noted, the other solutions normally result in similar behaviour. The equivalent analysis is best performed by the use of alternative methods (using the usual form of the exterior Schwarzschild metric): it suffices for our purposes to note that (5) is replaced in the elliptic case by

$$1 + z = \frac{(1 - n^2/R)^{\frac{1}{2}}}{\lambda - n\left(\frac{1}{R_0} - \frac{p}{r_b}\right)} \tag{48}$$

where λ is given by (17). Here

$$\frac{dR_0}{ds} = - \frac{n\left(\frac{1}{R_0} - \frac{p}{r}\right)^{\frac{1}{2}}}{(1 + z)} \tag{49}$$

and

$$\frac{dz}{ds} = \frac{z^2}{2n^2(1+z)^4}\left[z^2 + 2(1-\lambda)(1+z)\right]^2 \tag{50}$$

(for $R >> 1$).

Normally $(1-\lambda)$ is very small (initial radius >> gravitational radius), and so the difference between (50) and (47) is negligible. If, however, the collapse became "catastrophic" only when the gravitational radius had almost been reached (in comoving time), then for $z \sim 1$,

$$\frac{dz}{ds} \sim \frac{25}{32n^2} ,$$

and so the red-shift is less stable than the normal case as given by (47).

We now turn to the situation as observed by an observer on the surface of the collapsing body.

Suppose that a photon is emitted radially at (R, t) and reception takes place on the body at (R_0, t_0). Then (42) gives the relation between the events, except that now

$$f'(R) = -(1 + n/R^{\frac{1}{2}})^{-1} .$$

Considering two neighbouring emissions, we find as before

$$f'(R)\Delta R - f'(R_0)\Delta R_0 = \Delta t - \Delta t_0 ,$$

and using (43) this shows that

$$\Delta t_0 - \Delta t = n/R_0^{\frac{1}{2}} (1 + n/R_0^{\frac{1}{2}})^{-1} \Delta t_0$$

and so $$\Delta t_0 = (1 + n/R_0^{\frac{1}{2}})\Delta t \qquad (51)$$

which may be compared with (44) above.

Here

$$1 + z = \frac{\Delta s_0}{\Delta s}$$

and proceeding as before we find

$$1 + z = \frac{1 + n/R_0^{\frac{1}{2}}}{(1 - n^2/R)^{\frac{1}{2}}} . \qquad (52)$$

Note that the light is always red-shifted, rather than violet-shifted. (52) may be regarded as a combination of a gravitational violet shift and a (stronger) Doppler red-shift

$$1 + z = \left(\frac{1 - n^2 R_0}{1 - n^2/R}\right)^{\frac{1}{2}} \left(\frac{1 + v/c}{1 - v/c}\right)^{\frac{1}{2}}$$

where $v/c = \left|\frac{\partial R_0}{\partial t_0}\right|$.

If the light is emitted from a considerable distance from the surface of the body, then effectively

$$z = n/R_0^{\frac{1}{2}} .$$

The rate of change of z is here given by

$$\frac{dz}{ds_0} = +\frac{1}{2}\frac{n^2}{R_0^2}$$

$$= \frac{z^4}{2n^2} \qquad (53)$$

which is to be compared with (47). Here z becomes of order unity at the gravitational radius, when as explained before it would normally not be possible to observe a stable red-shift.

The two observers are not truly "conjugate" therefore: both will observe a red-shift, but light emitted from the body will become infinitely red-shifted as the gravitational radius is approached, whilst light received at the surface of the body will be finitely red-shifted when the gravitational radius is reached, the spectral shift becoming infinite only as the "point singularity" $R = 0$ $(r > 0)$ is reached.

It should be mentioned here that the type of collapse - elliptic etc. - determines the nature of the spectral shift in the early stages of collapse. For example, in the elliptic case

$$1 + z = \frac{(1 - p\alpha r_b^2)^{\frac{1}{2}} + n/R_0^{\frac{1}{2}} (1 - \frac{pR_0}{r_b})^{\frac{1}{2}}}{(1 - n^2/R)^{\frac{1}{2}}} \qquad (54)$$

so that in the early stages the gravitational shift may predominate, resulting in a violet shift.

It is clear from the preceding discussion that there is no theoretical obstacle to the collapse of a dust-fluid ~~containing~~ continuing beyond the extent of the "Schwarzschild radius" - but only from the viewpoint of a neighbouring observer free-falling with the body. A static observer will note that the body appears increasingly faint, and spectral lines will become impossible to detect; however, insofar as this observer can measure the radius of body, he will observe the collapse to be slowing down (cf. equation (46)). The radius will appear to asymptotically approach the gravitational radius but in a short space of time the emitted light will be so intensely red-shifted so as to make the body effectively invisible.

As has been noted, only a comoving or free-falling observer could observe the ultimate "Newtonian" catastrophe - though in practice any living creature would be crushed by the gravitational potential at an early stage. Although catastrophic collapse to a point singularity could never in fact be observed, the fact that our equations predict this behaviour (in the comoving frame of reference) may seem unfortunate. The obvious difficulties with any concept of collapse to a point singularity are concerned with the implication of an infinite density - and the unlimited possibilities for the subsequent behaviour of the body (if any!). The conceptual difficulties will be discussed further in the concluding Chapter, after the mathematical exposition of Chapter V.

We shall remark here that Hoyle & Narlikar have proposed a modification of the Einstein Field Equations - the "c" field of their steady-state cosmology - and have noted (Ref. 18) that their equations admit a solution which does not give rise to "continuous creation", and which appears relevant to the collapse question. They discuss the homogeneous, isotropic case, which gives rise to a Robertson-Walker metric with a modified differential equation for S(t). The authors seem to believe that the "elliptic" case is of particular significance, but as usual the "parabolic" case is easier to handle. Their methods would give rise to a "parabolic" equation of the form

$$\dot{S}^2 = \alpha/S - \lambda/S^4 ;$$

Hoyle & Narlikar argue that the body would therefore oscillate between extremes determined by the roots of this equation. It is assumed that the λ term is negligible except for very small S. However, this would imply that the body fell within its gravitational radius and later emerged, with all the associated conceptual difficulties. A further point is that it would appear that a matching exterior solution would be non-static. In any case, the approach has been attacked by

Hawking (Ref. 14) on the grounds that it would give rise to infinite fields.

We now turn to the other topic of our investigation: boundary conditions across the surface of discontinuity (i.e. $r = r_b$)

BOUNDARY CONDITIONS

The question of Boundary Conditions in General Relativity has been considered by a number of authors, notably O'Brien & Synge, (Ref. 35), Lichnerowicz (Ref. 26) and Nariai and Tomita (Ref. 30, 31)

In electromagnetic field theory, the differential equations admit a family of solutions: a particular solution is found by applying the derived boundary conditions (e.g. those applicable at a surface of discontinuity). The partial differential equations of General Relativity admit families of solutions, and it is tempting to search for analogous conditions. Difficulties arise in that we are dealing with a family of coordinate systems, and so any set of conditions which are expressed in terms of a particular system of coordinates rather than in terms of invariants is necessarily suspect.

There is considerable misunderstanding in the literature about the significance of boundary conditions. For example, the early paper on the Problem of Gravitational Collapse by Nariai & Tomita (Ref. 30) seems based on highly dubious premises. They point out that Hoyle & Narlikar's special solutions in comoving coordinates result in a discontinuity in the metric tensor across the boundary surface, in these particular coordinates. They state, inter alia, that "continuous fitting" of interior and exterior metrics in a particular coordinate system is a "physical fact", which betrays a lack of understanding of the rôle of coordinate systems in a physical theory based on differential geometry.

The significance of coordinate systems will be discussed in the next chapter. Ikariai & Tomita's approach is interesting in that they express the Schwarzschild exterior solution in the form

$$ds^2 = e^{\nu}dt^2 - e^{\lambda}dr^2 - R^2 d\Omega^2$$

where $R = rR(t)$ and ν, λ are functions of r, t; they then demonstrate that there exists a particular solution for which the SYNGE & O'BRIEN CONDITIONS (Ref. 35) are satisfied at $r = r_b$. It should be noted that the coordinates (r, t) are NOT comoving coordinates, and generally the curve $r = r_b$, θ=const., ϕ=const. is not a geodesic. In fact, their "fitting" of an interior and an exterior metric would have been illusory were it not that for their particular solution, $r = r_b$, θ=const., ϕ=const. turns out to be a geodesic (a point which they do not check!). Curiously enough, the requisite condition is equivalent to the critical Synge & O'Brien condition – continuity of the function $\frac{\partial}{\partial r} e^{\nu}$ – i.e.

$$e^{\nu} = 1, \ \frac{\partial \nu}{\partial r} = 0 \quad \text{at} \ r = r_b.$$

The authors of the paper show that their version of the exterior solution is smooth at the surface $R = 2M$, and suggest that their version is therefore superior to others in the literature. However, their solution is extremely involved analytically: our comoving form (43) of the exterior solution is far superior, and mention must also be made of the Finkelstein metric (Ref. 10).

It should be noted that the Mariot - Tomita exterior metric is such that $\frac{\partial}{\partial r} e^{\lambda}$ is not continuous at the boundary. It is indeed difficult to find a form of the exterior metric to a uniform dust-fluid which satisfies the full Lichnerowicz conditions, which require continuity of the first derivatives of all metric components.

How can one choose between the Synge - O'Brien approach to boundary conditions and that of Lichnerowicz? The former approach considers a succession of smooth surfaces with the discontinuity surface as a sort of limit. The approach to the limit is treated geometrically, by considering paths in a 5-dimensional Space-Time, but is open to the objection that a number of arbitrary, but critical, assumptions are made. The Lichnerowicz approach is to insist that differential geometry as such requires that there should be at least one coordinate system for which, in a region containing any portion of the discontinuity surface, the metric tensor and <u>all</u> its first derivatives are continuous - the stress-energy discontinuity appearing as a discontinuity in certain of the second derivatives. This approach is also open to the objection of seeming arbitrary, and appears almost to "fit the physics to the mathematics". ~~However, these conditions now appear in a new light, thanks to the researches~~

~~of~~

In the next chapter, we shall show that the foundations of differential geometry do not necessarily imply a continuously differentiable metric.

It seems advisable to clarify the function of boundary conditions in General Relativity: there is surely no reason to require conditions which enable one to select a particular system of coordinates; the question of "matching" only arises in choosing from disjunct families of solutions - a pair of solutions in the same family being obtained from another by a coordinate transformation. For example, the essential parameter 2GM in the exterior Schwarzschild solution is clearly to be equated to the interior parameter ar_b^3: this could be demonstrated by the use of integral theorems.

In fact, any further selection of solutions is purely a matter of convenience. For example, the comoving interior solution for a uniform dust fluid may, as we have seen, be transformed into the general Schwarzschild type of metric (cf. Schwarzschild's own interior solution) - though there is a family of such solutions, just as there is a family of "Schwarzschild'type" exterior solutions. There is, however, a preferred exterior solution - the one for which g_{π} is independent of T, and we can choose an interior metric form so that this T coordinate is compatible on both sides of the boundary. It must be emphasised that this 'coordinate matching' is wholly for ease of description, and does not imply any sort of physical limitation.

Our final remarks on the "boundary conditions" question will follow the exposition of the next chapter.

CHAPTER V

MATHEMATICAL BASIS OF GENERAL RELATIVITY

1. HISTORICAL

The central purpose of General Relativity is the construction of a geometrical model of Space, Time and Gravitation. A geometry consists of a set of elements called points, together with some prescribed relationships between subsets of this set of points. The Theory of Relativity utilises Riemannian Geometry, the defining relationships of which have until recently remained somewhat obscure. This Geometry arose out of the study of surfaces embedded in Euclidean 3-space, and more generally of hypersurfaces in Euclidean n-space. The only difficulties here were the precise definition of such surfaces or hypersurfaces, which requires some care. However, Differential Geometers became increasingly interested in the intrinsic properties of such surfaces, that is, those properties which are independent of any particular embedding. For example, a Euclidean Plane and the surface of a right cylinder in Euclidean 3-space are of identical intrinsic character, differing only in extrinsic properties.

Thus Riemannian Geometry in general came to be regarded as being in some way analogous to surface geometry, but using a metric which is arbitrarily prescribed (apart from some regularity conditions), rather than derived from a Euclidean embedding. However, although the tensor calculus employed in the description of surface geometry may readily be extended so as to use arbitrary metric tensors, the new geometry requires an abstract definition before the tensor formulae can be interpreted in geometrical terms.

Before the appearance of the 1932 Cambridge Tract by Veblen & Whitehead, "The Foundations of Differential Geometry" (Ref. 51), most workers in the field seemed to regard a definition on the following lines as adequate:

(1) A Manifold is a Set of Points which may be put into (1,1) correspondence with the set of all n-tuples of real numbers, where n is a given positive integer.

(2) Co- and contra-variant vector and tensor fields are defined on the manifold.

(3) A (1,1) correspondence between co- and contra-variant tensor fields is given by means of second rank covariant tensor field.

By (2) was meant that sets of n^p $(p \geq 1)$ functions of the n-tuples were given, together with the appropriate transformation laws between these and functions of a new set of n-tuples, which are obtained from differentiable functions of the original n-tuples.

Instead of (3), it was often stated that a second order differential form was postulated; however, this latter concept has only recently been made precise.

2. EUCLIDEAN SPACE

This approach can be criticised on a number of grounds. To define vectors and tensors by means of "components" (with respect to what?) and some arbitrary-looking transformation laws between these and other "components" goes against the spirit of modern algebra. The situation is considerably clearer when our "manifold" is in fact

Euclidean Space. This may be regarded as a Vector Space; the space of position vectors, $\underline{P}$. Since the space is of finite dimension n, we may choose a set of n basis vectors (frame) ; $\underline{e}_i$ and a set of n variables x^i such that

$$\underline{P} = x^i \underline{e}_i \quad . \tag{1}$$

The x^i form a set of n coordinates for points $\underline{P}$, relative to the basis $\underline{e}_i$. It follows that $\underline{P}$ may be regarded as a differentiable function of the x^k, and

$$\underline{e}_i = \frac{\partial \underline{P}}{\partial x^i} \quad . \tag{2}$$

We may also show that given any (allowable) coordinate system (x^k) there exists a corresponding basis, as given by (2) above. That is, the n vectors

$$\frac{\partial \underline{P}}{\partial x^i}$$

are linearly independent. We defer the proof for the present.

We now define a contravariant vector field on this Euclidean Space to be a vector-valued function of position – i.e. a mapping of the space into itself. Let $\underline{e}_i$, $\underline{e}'_i$ be two bases of the space, and (x^k), (x'^k) the corresponding coordinate systems. Let $\underline{Q}(\underline{P})$ be a vector field.

Then
$$\begin{aligned}\underline{Q}(\underline{P}) &= Q^i(\underline{P})\, \underline{e}_i \\ &= Q^i(\underline{P}) \frac{\partial \underline{P}}{\partial x^i} \\ &= Q^i(\underline{P}) \frac{\partial x'^k}{\partial x^i} \frac{\partial \underline{P}}{\partial x'^k} \\ &= \frac{\partial x'^k}{\partial x^i} Q^i(\underline{P})\, \underline{e}'_k \quad .\end{aligned}$$

But we may write $$\underline{q}(\underline{p}) = q'^{i}(\underline{p})\, \underline{e}'_i$$

whence $$q'^{k}(\underline{p}) = \frac{\partial x'^{k}}{\partial x^{i}}\, q^{i}(\underline{p})$$

or, writing $$q^{i}(\underline{p}) = q^{i}(x^{j})$$

$$q'^{i}(\underline{p}) = q'^{i}(x'^{j})$$

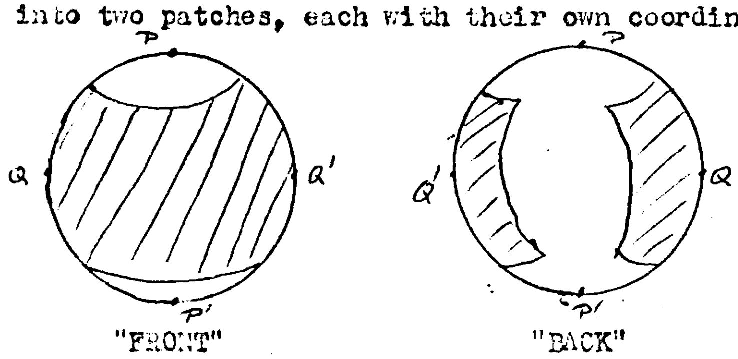
we have $$q'^{k}(x'^{j}) = \frac{\partial x'^{k}}{\partial x^{i}}\, q^{i}(x^{j}). \qquad (3)$$

We have recovered the usual transformation law for a contravariant vector-field. However, there is no need to work with components except in examples, since we have a perfectly straightforward definition of such a field. We could further define a covariant vector field as a mapping of the space into the dual vector-space, and tensor fields as mappings of the Euclidean space into tensor-product spaces. "Dual Space" and "Tensor Product" are purely algebraic concepts, which we shall discuss presently.

3. GEOMETRICAL IDEAS

Unfortunately, Riemannian Space is not in general a vector space, and the association of vectors with points on such a space requires some care; nevertheless, this can be achieved in a particularly elegant fashion (as shown by Chevalley, "The Theory of Lie Groups" (Ref. 6). We could of course define a vector field as a mapping of points into vectors of an arbitrary vector space, but this would be irrelevant in our attempt to construct a geometry, since this vector space would bear no relation (except for dimension) with our point space. To regard vector and tensor fields as geometrical objects they must be derived from the geometrical structure itself.

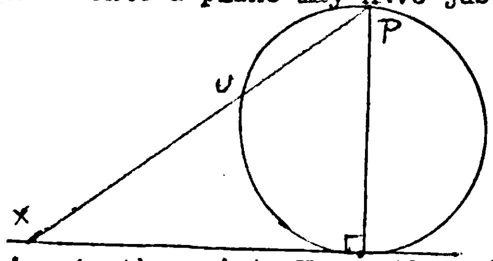
The other class of objections relate to definition (1), and are more fundamental. Firstly, it has to be modified if it is to include some typical examples of surfaces embedded in Euclidean 3-space. The sphere, for example, cannot be totally described by one such 'coordinate system'. Spherical polar coordinates on a unit sphere fail to provide a (1,1) correspondence between pairs of real numbers and points on the sphere at $\theta = 0$ and $\theta = \pi$. The best that can be done is to divide the sphere into two patches, each with their own coordinate system



The above diagram indicates how this can be done using two systems of spherical polar coordinates. In the shaded patch we choose coordinates such that P, P' are $\theta = 0$, $\theta = \pi$, and (θ,ϕ) take values in the appropriate subset of the normal ranges; in the other patch we use coordinates (θ',ϕ') with Q,Q' corresponding to $\theta' = 0,\pi$. Since P,P' do not lie within the shaded patch, nor Q,Q' within the unshaded one, the correspondence is (1,1) throughout each patch separately.

1. In fact, it can be provided that no coordinate system can be found which covers the sphere completely. Whereas spherical polars excluded

two points on the sphere, coordinates obtained by sterographic projection of the sphere onto a plane may have just one excluded point



If we assign to the point U on the sphere the cartesian coordinates of the point X, where X is the point of intersection of $\mathcal{P}$U and the plane ($\mathcal{P}$ fixed), then the correspondence is (1,1) for all points U except U = $\mathcal{P}$. Note that the projection of $\mathcal{P}$ itself may be regarded as the intersection of the plane with the tangent plane of $\mathcal{P}$, which explains why the Euclidean plane is said to possess a <u>line</u> at infinity. Note also the line element in this system of coordinates:

$$ds^2 = \frac{1}{(1 + (1/4)r^2)^2}(dx^2 + dy^2) \quad , \quad (r^2 = x^2 + y^2).$$

Thus (1) should be amended to read:

" A manifold is a set of points which is covered by a finite family of subsets, for each of which there exists a (1,1) correspondence between the points and a set of real n-tuples."

However, this still is not good enough. We wish to regard a "manifold" as being in some sense a <u>continuous</u> structure; to require that a portion of it can be "labelled" by continuous variables does not ensure that the mapping itself is performed in a continuous fashion. Simply because two points have coordinates (x^k) and $(x^k + \delta x^k)$ does not imply that the two points are in some "physical" sense "neighbouring". All it stipulates is that the set of points has the Cardinal Number of the continuum. Since n continuous variables have the same cardinal number as one such variable, we have not succeeded in determining the dimension of our space. It might equally be described by a set of m variables, where $m \neq n$. This was demonstrated by Georg Cantor, who found explicitly a (1,1) map of E_m (Euclidean m-space) into E_n — i.e. E_n can be described by m variables. Dimension of Euclidean spaces is invariant only under maps which are continuous, as well as (1,1) (strictly, "(1,1) onto"). More generally, our abstract space requires a neighbourhood concept already associated with it, to ensure that it is locally related to n-tuples in a continuous and (1,1) onto fashion.

Both Euclidean and Minkowski Spaces have a "built in" neighbourhood structure, though in the latter case this is in no way related to the "metric".

4. SCALAR PRODUCTS

We shall define a space to be Euclidean if it is a Vector Space of finite dimension, over the reals, with a given positive definite scalar product.

This means that given any two vectors $\underline{u},\underline{v}$ of our space, they determine a unique real number ϕ satisfying

E.1. (i) $\phi(\lambda\underline{u} + \mu\underline{u}',\underline{v}) = \lambda\phi(\underline{u},\underline{v}) + \mu\phi(\underline{u}',\underline{v})$

(ii) $\phi(\underline{u},\lambda\underline{v} + \mu\underline{v}') = \lambda\phi(\underline{u},\underline{v}) + \mu\phi(\underline{u},\underline{v}')$

E.2 $\phi(\underline{u},\underline{v}) = \phi(\underline{v},\underline{u})$

E.3 $\phi(\underline{u},\underline{u}) \geqslant 0$

E.4 $\phi(\underline{u},\underline{u}) = 0 \iff \underline{u} = 0$

The vector space of n-tuples of real numbers is called CARTESIAN n-SPACE (R^n). If we specify a scalar product, Cartesian space becomes a special case of Euclidean space. The NATURAL Scalar Product is defined as follows:

$$\phi(\{x^i\},\{y^i\}) = \sum_i x^i y^i \quad . \tag{4}$$

Now any two n-dimensional vector spaces are isomorphic — i.e. there exists a linear mapping between them which is (1,1) onto. Thus E^n must be isomorphic to R^n. In fact if $\underline{e}_i$ is any basis of E^n, then if $\underline{u} \in E^n$

$$\underline{u} = u^i \underline{e}_i$$

and the mapping $\underline{u} \longleftrightarrow \{u^i\}$ is an isomorphism. Note that we can choose $\underline{e}_i$ so as to form an orthonormal basis — i.e.

$$\phi(\underline{e}_i,\underline{e}_j) = \delta_{ij} \quad .$$

Then
$$\begin{aligned} \phi(\underline{u},\underline{v}) &= u^i v^j \phi(\underline{e}_i,\underline{e}_j) \\ &= \sum_i u^i v^i \quad . \end{aligned} \tag{5}$$

So with this choice of basis, and choosing the natural scalar product, the scalar product is preserved by the isomorphism.

Cartesian n-space being just the set of n-tuples, we have a (1,1) and dimension preserving map between E^n and n-tuples, which are therefore "coordinates". Although this is quite familiar, we state it for purposes of comparison. Isomorphic mappings of vector spaces do indeed map "neighbourhoods" of points in one space onto neighbourhoods in the other.

Note that we used the linear relationship

$$\underline{u} = u^i \underline{e}_i$$

in order to establish the isomorphism

$$\underline{u} \longleftrightarrow \{u^i\} \ .$$

This mapping exists for all finite dimensional vector spaces; in particular, for Minkowski Space; but here the isomorphism does not preserve the (pseudo) scalar product — it is not, therefore, an "isometry". A general scalar product satisfies axioms E.1, E.2 above, but in place of E.3,4 we have E'.3:

$$\text{For all } \underline{x}, \ \ \phi(\underline{u},\underline{x}) = 0 \Longleftrightarrow \underline{u} = 0.$$

Riemannian space cannot be regarded as a vector space (unless it is "flat"). We have to postulate a neighbourhood structure a priori, in order to assign coordinates in a "continuous", i.e. neighbourhood preserving fashion. This could be accomplished by postulating a distance function or "topological metric"; this would not, however, enable us to construct the hyperbolic geometry we require for our relativistic world-model.

5. [illegible]

Essentially we wish to "weaken" the vector-space structure of Euclidean or Minkowski space, and then "rebuild" it by adding a Riemannian structure. Unfortunately, we have to drop all the vector space axioms. From our definition of a "geometry" at the beginning of this chapter it would seem that we wish to characterise a collection of subsets of the space as "neighbourhoods" by means of relationships between them. In Cartesian space, there are any number of suitable definitions of "neighbourhood". However, we cannot find any characteristic property of such neighbourhoods which does not depend on the vector-space structure in some way. However, a more general notion, that of the "Open Set", fulfils our requirements. An open set in Cartesian space can be thought of as being "built up" from neighbourhoods by set "addition" and "subtraction" (Union & Intersection): a set in this space is defined to be open if every point of it has a neighbourhood around it lying wholly within the set. This definition is independent of which definition of neighbourhood we use, provided it is reasonable (e.g. "spherical" or "cubical" neighbourhood). A particular case of an open set in R^3 would be the interior of any "respectable" closed surface. Such open sets have the following properties:

1. The Union of a finite, or countably infinite number of Open Sets is an Open Set.

2. The Intersection of a finite number of Open Sets is an Open Set.

We may paraphrase these as follows: "The Collection of Open Sets is "closed" under the operations of set union & intersection." If these are now used as general definitions of open sets in any space, we are ready to commence the "topological" treatment of differential geometry, which we shall shortly sketch.

Of course, the familiar coordinate formalism is eventually extracted. The topological approach is, as we have seen, essential to the logical development of the subject. It provides clear insight into the geometrical significance of much of the tensor formalism, and is indispensable to understanding of such concepts as "extension" and "completion" of Riemannian spaces, and, indeed, to tackling any "global" question. Finally, as we shall see, it clarifies questions we have already discussed concerning coordinate and metric conditions at boundaries, and the introduction of the concept of "time orientation".

[illegible]

A set is said to be a TOPOLOGICAL SPACE if we select a collection of subsets such that the collection is closed under the operations of set union (repeated any finite or countably infinite number of times) and of set intersection (repeated any finite number of times). Having selected such a collection, the constituent members will be described as <u>the</u> Open Sets. When one selects such a collection, one is said to assign a TOPOLOGY to the set.

A HAUSDORFF SPACE is a topological space which satisfies the HAUSDORFF SEPARATION AXIOMS:

A. If P is any point of the space, there exists at least one open set containing P.

B. If P, Q are any two points, there exists an open set containing P, and one containing Q, with no points in common.

(A) is obviously essential if all points are to have neighbourhoods. (B) is necessary if any limit process is to be defined. It might be thought that B is a consequence of the coordinate postulates (to be discussed shortly), but this is not the case.

A (1,1) correspondence between two topological spaces is said to be a HOMEOMORPHISM if each point of any open set in the one space is mapped onto a point in an open set of the other space. If two spaces possess such a (1,1) relation, they are said to be HOMEOMORPHIC or TOPOLOGICALLY EQUIVALENT.

The notion of homeomorphism provides precisely the neighbourhood preserving (1,1) mapping that we require in order to assign coordinates. Vector space isomorphism is a special case of homeomorphism provided we assign the "usual" topology to the vector spaces — i.e. that the open sets be defined in terms of neighbourhoods, in the usual way.

A topology is the most basic piece of structure we can give to a space. Nothing has been said so far about affine connections or Riemannian metrics; two spaces which are topologically equivalent need have no metrical similarity. Surfaces in E^3 are homeomorphic if they can be continuously deformed into each other: this can be visualised as involving operations of bending and twisting, though without permanent breakage. Thus a sphere and an ellipsoid are topologically equivalent, but neither is equivalent to an anchor ring (torus).

We shall now ensure that our space does not consist of a number of wholly separate portions (some accounts postpone this to a later stage). To make this precise, we postulate that it is impossible to find two or more non-intersecting (disjoint) open sets whose union forms the whole space. If this (negative) condition is satisfied, the space is said to be CONNECTED. The definition of a "path" logically belongs to a later stage of our discussion, but it is intuitively clear that in such a space, any two points may be joined together by a path consisting only of points of the space. Thus we would exclude such surfaces in R^3 as a hyperboloid of two sheets.

We now define Cartesian n-Space R^n (cf. above) as the set of all real n-tuples $\{x^k\}$, together with the straightforward addition and scalar multiplication axioms:

C.1. $\{x^k\} + \{y^k\} = \{x^k + y^k\}$

C.2. $\lambda\{x^k\} = \{\lambda x^k\}$ (λ real)

so as to form a vector space. As we have seen, given a scalar product, such as the natural product

$$\phi(\{x^k\},\{y^k\}) = \sum_k x^k y^k$$

this "coordinate" space is a special case of our abstractly defined "Euclidean space".

It may be proved (see Newman (ref. 33)) that R^n is not homeomorphic to R^m $(m \neq n)$. Thus dimension in this sense is a topological invariant.

If there exists a collection of open sets in a topological space, whose union is the whole space, the collection is called a BASE of the space. If the number of sets in the collection is finite or countably infinite, the base is said to be COUNTABLE.

We now give the following definition:

An n-DIMENSIONAL MANIFOLD is a connected, HAUSDORFF Topological Space with a countable base such that each open set in this base is homeomorphic to an open set in R^n.

This means that the space can be divided into (overlapping) patches, each of which is topologically equivalent to an open portion of R^n (or C^n). In each patch, to every point P corresponds one element of R^n, i.e. a 'coordinate label'. Each open set of the base, U_α, together with its mapping onto the set of n-tuples, ϕ_α is said to form a coordinate CHART on the manifold.

Clearly the minimum number of charts needed to form a base is a topological invariant of the manifold – for the Euclidean Sphere, as we have seen, this number is two.

7. DIFFERENTIABLE STRUCTURE

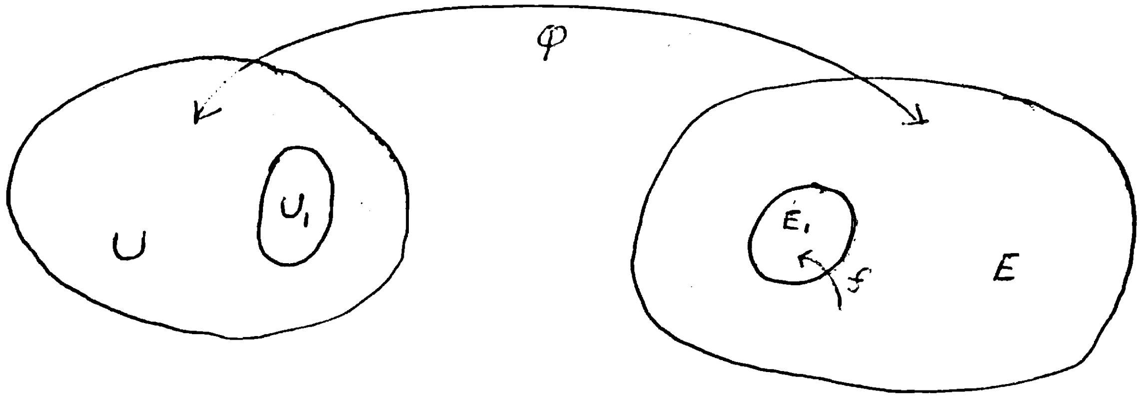
We now come to discuss differentiable systems of coordinates. Firstly, we consider one chart (U,ϕ). Let

$$\phi(U) = E$$

i.e. E is the open set on R^n corresponding to U. Let E_1 be an open subset of E. Then there exists an open subset of U : U_1, say, with

$$\phi(U_1) = E_1 .$$

Suppose we have a continuous function f which maps E onto E_1 .



This function can be represented by a set of n equations of the form

$$y^\alpha = f^\alpha(x^1, \ldots x^n) \qquad (\alpha = 1, \ldots n)$$

where $\{x^k\}$ belongs to E, $\{y^k\}$ to E_1, and the f^α are continuous functions. If there corresponds to the function f an inverse function f^{-1} which maps E_1 onto E, then f is a homeomorphism of E onto E_1, provided that f^{-1} is also continuous (see Newman, Ref(33)).

A sufficient (though not necessary) condition for f to be a homeomorphism is that the f^α be differentiable, with non-vanishing Jacobian over all n-tuples of E, since in that case the equations possess an inverse, and the functions

$$x^\alpha = f^{-1\,\alpha}(y^1 \ldots y^n)$$

are differentiable, and therefore continuous.

We note here that homeomorphism is an EQUIVALENCE RELATION, that is, if $\longleftrightarrow$ indicates that two spaces are topologically equivalent, H.1 $\quad U \longleftrightarrow U$

H.2. $\quad U \longleftrightarrow V$ implies $V \longleftrightarrow U$

H.3 $\quad U \longleftrightarrow V$ and $V \longleftrightarrow W$ implies $U \longleftrightarrow W$.

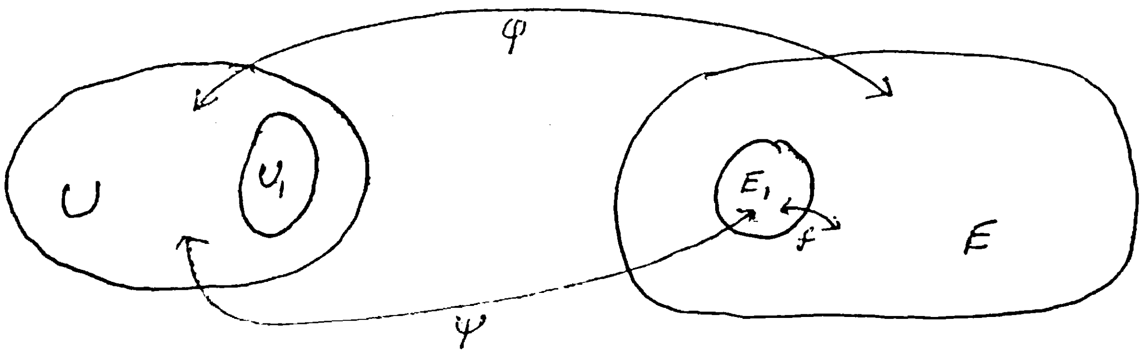
If f is differentiable, then by property (3), U is homeomorphic to E_1, the correspondence being denoted by

$$\psi = f \circ \phi \tag{6}$$

which signifies that

$$\phi(U) = E \text{ and } f(E) = E_1 \tag{7}$$

Thus U has two homeomorphisms onto open sets of R^n, ϕ and ψ.



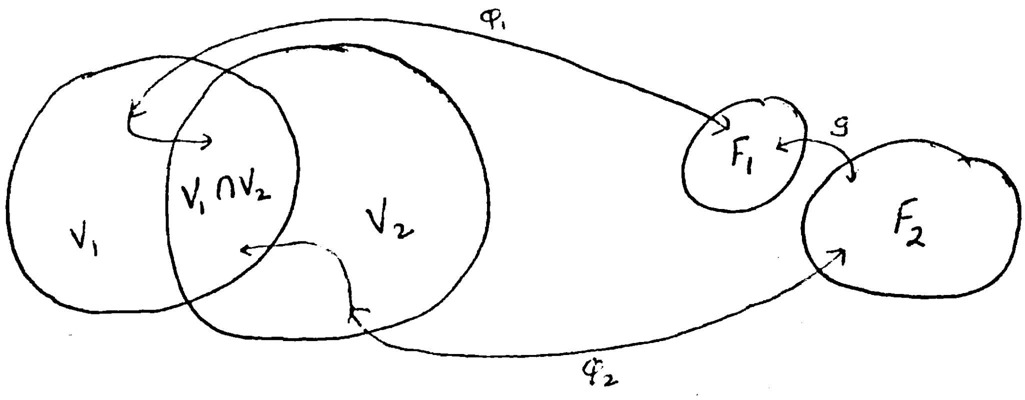
Since there are any number of differentiable functions f, and corresponding open sets E_1, there are any number of coordinate systems on U, related to one another by differentiable functions. As we have stated, coordinate systems do not have to be mutually differentiable; however, since we shall be concerned with differential properties of curves and surfaces, we restrict ourselves to these differentiable coordinates. Except where stated otherwise, we shall now assume that "differentiable" implies "differentiable any number of times". This, of course, restricts the class of "allowable" coordinate systems.

Now consider the question of coordinate systems on the manifold as a whole, not just in one patch. Consider two overlapping charts (V_1, ϕ_1); (V_2, ϕ_2). In the overlap region $V_1 \cap V_2$ we have two mappings onto open Cartesian sets — i.e. two coordinate systems. Let

$$\phi_1(V_1 \cap V_2) = F_1 \qquad \phi_2(V_1 \cap V_2) = F_2 \tag{8}$$

By property H.1. above, $V_1 \cap V_2$ is homeomorphic to itself. By property H.3., F_1 must be homeomorphic to F_2. The situation is the reverse of the previous one, where we constructed a second homeomorphism of U onto a Cartesian set by means of an arbitrary differentiable function. Here we already have two coordinate systems on $V_1 \cap V_2$, and have deduced the existence of a (1,1) and bicontinuous function g, with

$$g(F_1) = F_2 \tag{9}$$



(Note: whether or not F_1, F_2 intersect in no way affects this discussion.)

Now it may happen that g is not differentiable. In that case we are unable to produce mutually differentiable coordinates over the manifold as a whole. However, it may be possible to find a new collection of charts

$$(V_\alpha, \phi_\alpha)$$

which cover the manifold, such that in each overlap region, the corresponding function g is differentiable. If the manifold admits such a collection of charts, it is said to be a DIFFERENTIABLE MANIFOLD. In symbols, the condition is that for all α, β the function

$$\phi_\beta \circ \phi_\alpha^{-1} \tag{10}$$

which maps $\phi_\alpha(V_\alpha \cap V_\beta)$ onto $\phi_\beta(V_\alpha \cap V_\beta)$ should be differentiable.

Given the charts (V_α, ϕ_α) we can extend them to a complete set of differentiable coordinate systems on the manifold. This collection consists of all charts (W, ψ) such that for all α the function

$$\phi_\alpha \circ \psi^{-1} \tag{11}$$

which maps $\psi(W \cap V_\alpha)$ onto $\phi_\alpha(W \cap V_\alpha)$ should be differentiable.

We clearly have to prove that if (W, ψ) and $(\bar{W}, \bar{\psi})$ are two such charts, then the function

$$\bar{\psi} \circ \psi^{-1} \tag{12}$$

which maps $\psi(W \cap \bar{W})$ onto $\bar{\psi}(W \cap \bar{W})$ is differentiable. If this is the case, then the collection contains all differentiable coordinate systems. Since none of the standard treatments of the subject discuss this point, it seems worth while to give a proof here. It appears that a couple of preliminary results are needed first of all. We note firstly that the operation of set intersection is distributive over that of set union (e.g. Newman, (Ref. 33)), i.e. if A,B,C are sets,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad . \qquad (13)$$

Also, both operations are associative.

Lemma I

If Z is any set

$$Z = \bigcup_{\alpha}(Z \cap V_{\alpha}) \qquad (14)$$

For $(Z \cap V_{\alpha}) \cup (Z \cap V_{\beta}) = Z \cap (V_{\alpha} \cup V_{\beta})$ and by induction $\bigcup_{\alpha} V_{\alpha} = M$, the whole manifold.

Finally $Z \cap M = Z$, which establishes our result.

Lemma II

$$\phi(Z) = \bigcup_{\alpha} \phi(Z \cap V_{\alpha}) \qquad (15)$$

where ϕ is a homeomorphism.

Now if A,B are sets,

$$\phi(A \cup B) \subset \phi(A) \cup \phi(B) \qquad (16)$$

since if $x \in A \cup B$, either $x \in A$ or $x \in B$. Thus either $\phi(x) \in \phi(A)$ or $\phi(x) \in \phi(B)$. Since ϕ is a homeomorphism, it possesses an inverse, ϕ^{-1}. Consider the (Cartesian) sets L,M, where

$$L = \phi(A), \qquad M = \phi(B) \ .$$

Then

$$\phi^{-1}(L \cup M) \subset \phi^{-1}(L) \cup \phi^{-1}(M)$$

by the above argument, replacing ϕ by ϕ^{-1}.

Thus

$$\phi \circ \phi^{-1}(L \cup M) \subset \phi \circ [\phi^{-1}(L) \cup \phi^{-1}(M)]$$

That is,

$$\phi(A) \cup \phi(B) \subset \phi(A \cup B) \tag{17}$$

Combining (16) and (17),

$$\phi(A \cup B) = \phi(A) \cup \phi(B).$$

By induction, if A_α is a collection of sets

$$\phi(\bigcup_\alpha A_\alpha) = \bigcup_\alpha \phi(A_\alpha)$$

and so, using Lemma I,

$$\phi(Z) = \bigcup_\alpha \phi(Z \cap V_\alpha)$$

Hence result.

We now proceed to our theorem.

By Lemma II,

$$\phi(W \cap \bar{W}) = \bigcup_\alpha \phi\ (W \cap \bar{W} \cap V_\alpha) \tag{18}$$

that is, we can build up the Cartesian image of

$$W \cap \bar{W}$$

from that of the portions lying in each coordinate patch.

In each V_α the mapping ϕ_α is defined

In W the mapping ϕ is defined

In $\bar{W}$ the mapping $\bar{\phi}$ is defined

Thus $\phi_\alpha, \phi, \bar{\phi}$ are simultaneously defined on

$$W \cap \bar{W} \cap V_\alpha .$$

Let $P \in W \cap \overline{W} \cap V_\alpha$.

Then $\phi_\alpha(P) = \phi_\alpha \circ \psi^{-1} \circ \psi(P)$

$$= \phi_\alpha \circ \overline{\psi}^{-1} \circ \overline{\psi}(P)$$

whence

$$\phi_\alpha \circ \psi^{-1} \circ \psi(P) = \phi_\alpha \circ \overline{\psi}^{-1} \circ \overline{\psi}(P)$$

Thus

$$\overline{\psi}(P) = (\phi_\alpha \circ \overline{\psi}^{-1})^{-1} \circ (\phi_\alpha \circ \psi^{-1}) \circ \psi(P)$$

But $\overline{\psi}(P) = \overline{\psi} \circ \psi^{-1} \circ \psi(P)$

and so

$$\overline{\psi} \circ \psi^{-1} = (\phi_\alpha \circ \overline{\psi}^{-1})^{-1} \circ (\phi_\alpha \circ \psi^{-1}) \qquad (19)$$

where each side of the equation operates on elements of $\psi(W \cap \overline{W} \cap V_\alpha)$.

Note that given

$$f(Q) = g(Q)$$

for all $Q \in \psi(W \cap \overline{W} \cap V_\alpha)$, we can deduce

$$f = g .$$

However, if Z is any set, and

$$f(Z) = g(Z)$$

we cannot infer that $f = g$, for

$$h(Z) = Z$$

does not imply that h is the identity function — it is simply a homeomorphism of the set onto itself. Thus it is essential in establishing our identity to consider operations on <u>any</u> element of $W \cap \bar{W} \cap V_\alpha$, rather than on the set itself.

Now $\phi_\alpha \circ \bar{\psi}^{-1}$ maps

$$\bar{\psi}(W \cap \bar{W} \cap V_\alpha) \quad \text{onto} \quad \phi_\alpha(W \cap \bar{W} \cap V_\alpha)$$

and is, by definition, differentiable, with a differentiable inverse. Similarly, $\phi_\alpha \circ \phi^{-1}$ is differentiable.

Thus

$$\bar{\psi}\, \phi^{-1}$$

is differentiable when operating on the Cartesian set

$$\phi(W \cap \bar{W} \cap V_\alpha) \quad .$$

If a function is differentiable on each of a collection of sets A_α, it is clearly differentiable on their union.

Thus

$$\bar{\psi}\, \phi^{-1}$$

is differentiable when operating on

$$\bigcup_\alpha \phi(W \cap \bar{W} \cap V_\alpha)$$

and so, by Lemma II, is differentiable when operating on

$$\phi(W \cap \bar{W})$$

which is the required result.

We have obtained a maximal class of relatively differentiable coordinates from a collection of charts with differentiable coordinates. It is not clear whether or not we will obtain the same complete class if we start with another collection of charts. Such a class is called a DIFFERENTIABLE STRUCTURE; for surfaces in Euclidean Spaces, it was found by Milnor (Ref. 29)) that whereas the "sphere" (and all surfaces homeomorphic to it) in E^n possessed only one differentiable structure for $n \leq 6$, it possesses several distinct ones for $n = 7$. However, for our purposes it is sufficient that the manifold be differentiable, and one differentiable structure be assigned to it.

We note in recapitulation that in developing differential geometry we have been imposing two types of condition upon our original set concept: firstly, restrictions on the type of space, such as the requirements that it be Hausdorff, connected, have a countable basis, be a manifold, and be differentiable. Then we have the structure conditions, such as those which assign a topology and a differentiable structure to the space: each of these signifies that we have selected one of several alternatives and the geometry is regarded as consisting of the space together with the chosen relationship. For example, a vector space has any number of scalar products (bilinear functions): Euclidean space is formed by choosing one such function. In our future discussion we shall be more concerned with further structure conditions.

8. ORIENTATION

To conclude this section on coordinates, we introduce the concept of orientation. As we have seen, on any manifold each coordinate patch admits any number of mutually differentiable coordinates, where "differentiable" implies a non-vanishing Jacobian: i.e. one which is either positive or negative throughout the patch. We can select out those systems of coordinates for which the Jacobian of any pair is always positive. These are called the ORIENTATED systems of coordinates. There are of course two such collections of coordinates — in Euclidean Space these are the "left handed" and the "right handed" systems. Note that the positive Jacobian property is an equivalence relation, as we have defined it; since the members of such a collection are mutually related in this way, each collection is called an EQUIVALENCE CLASS. Thus we can separate the differentiable coordinates into two disjoint equivalence classes.

When we come to consider the manifold as a whole, however, just as we could not guarantee the existence of differentiable coordinates throughout, we cannot expect that in general, orientated coordinate systems exist. If they do (a "restriction") the manifold is called ORIENTABLE, and to choose one equivalence class of orientated coordinates is called "giving an ORIENTATION" to the manifold — a "structure" condition. These considerations form the starting point for our investigation of "time orientation" in our final chapter.

We now seek to introduce the concepts of vector and tensor fields on manifolds. A preliminary sketch of the relevant algebraic notions seems desirable.

9. [illegible]

Linear algebra starts with the vector space concept; "vector" simply means some element of a given vector space. We can construct vector spaces which are intrinsically related to a given space: the method is to characterise a set of real-valued functions of vector variables as a new vector space. We shall use a similar technique when deriving the 'tangent space' at a point of a manifold. Occasionally we shall use vector-valued functions of a vector — mappings of the space into itself, or into other spaces.

The LINEAR FUNCTIONS on a vector space V are the real valued functions

$$h(\underline{v})$$

satisfying

$$h(a\underline{v} + b\underline{w}) = ah(\underline{v}) + bh(\underline{w}) \qquad (20)$$

where $\underline{v},\underline{w} \in V$ and a,b are real numbers.

If $\underline{e}_i$ $(i = 1,\ldots,n)$ is a basis of the space, then

$$\begin{aligned} h(\underline{v}) &= h(v^i\underline{e}_i) \\ &= v^i h(\underline{e}_i) \ . \end{aligned}$$

Each $h(\underline{e}_i)$ is a real number, h_i, say.

Thus

$$h(\underline{v}) = v^i h_i \ . \qquad (21)$$

This strongly suggests that the function, h, (not the number $h(\underline{v})$) should be regarded as a vector: we would wish to interpret $h(\underline{v})$ as $\underline{h}.\underline{v}$. The set of all functions h forms a vector space if we define addition, and scalar multiplication in the obvious way:

L.1. $\quad (h + g)(\underline{v}) = h(\underline{v}) + g(\underline{v})$

L.2. $\quad (ah)(\underline{v}) = ah(\underline{v})$

where $(h+g)$ and (ah) are now linear functions, as is easily verified.

The linear function "0" is defined by

L.3 $$0(\underline{v}) = 0$$

The vector space axioms are then satisfied, as is verified using L.1, L.2, L.3.

The set of all linear functions, together with definitions L.1, L.2, L.3., is called the DUAL SPACE of V.

Clearly a linear function h is fully determined if we specify the n values $h(\underline{e}_i)$ for any basis $\underline{e}_i$. Let f^1 be the function for which

$$f^1(\underline{e}_1) = 1, \quad f^1(\underline{e}_2) = 0, \; \ldots\ldots \; f^1(\underline{e}_n) = 0$$

with similar definitions of $f^2, \ldots f^n$: i.e.

$$f^i(\underline{e}_j) = \delta^i{}_j \quad .$$

Then the n functions f^i are linearly independent — for if not there exist n real numbers a_i, not all zero, such that

$$a_i \underline{f}^i = 0$$

the zero function. But

$$a_i f^i(\underline{e}_j) = a_i \delta^i{}_j = a_j \qquad (j = 1 - n)$$

and since at least one a_j is not zero, $a_i f^i$ cannot be the zero function. Since there are n functions f^i, and the dual space is also of dimension n (proved in a straightforward manner), the f^i form a basis of the dual space V^*.

This basis is called the one dual to the $\underline{e}_i$, and is usually written $\underline{e}^i$. Note $\underline{e}^i$ is not a basis of V — it is a different sort of vector.

We note that if h is any linear function, we wrote

$$h(\underline{e}_i) = h_i$$

But there exist n numbers $\bar{h}_i$, say, for which

$$h = \bar{h}_j \underline{e}^j$$

so that

$$h(\underline{e}_i) = \bar{h}_j \underline{e}^j(\underline{e}_i)$$

$$= \bar{h}_j \delta^j{}_i = \bar{h}_i$$

whence

$$h_i = \bar{h}_i$$

Thus the h_i are indeed the components of a vector of V^*.

We shall use the following notation:

If A, B are sets we denote by

$$A \times B$$

the set of all pairs

$$(a, b)$$

where $a \in A$, $b \in B$. It is called the CARTESIAN PRODUCT of A and B. The pairs are to be regarded as ordered, but without any algebraic properties.

10. [illegible]

We wished to regard

$$h(\underline{v})$$

as a scalar product of $\underline{h}$ and $\underline{v}$. We earlier defined a scalar product on $V \times V$ as a bilinear function

$$\phi : V \times V \to R$$

where ϕ was symmetric, and such that if

$$\phi(\underline{u},\underline{x}) = 0$$

for all x, then $\underline{u} = \underline{0}$. (Non-singular property).

Clearly the linear functions on V determine a unique bilinear function on $V^* \times V$:

$$\phi(\underline{h},\underline{v}) = h(\underline{v}) \quad . \tag{22}$$

The non-singular property holds, since if

$$h(\underline{v}) = \underline{0}$$

for all v, then for $i = 1 - n$,

$$h(\underline{e}_i) = 0 \ .$$

But as we have seen, these are the components of $\underline{h}$ w.r.t. the dual basis. Thus $\underline{h}$ is the zero function.

However, the symmetric property poses difficulties. Namely

$$\phi(\underline{v},\underline{h})$$

is meaningless, since $\underline{v}$ is not a function of h. If, however, we could find some reason for identifying

$$V^{**} \text{ and } V$$

then we could give a meaning to

$$v(h)$$

for $\underline{v}$ could equally be regarded as an element of V or as an element of V^{**}.

Now all n-dimension vector-spaces are isomorphic: if $\underline{e}_i$ is a basis of V and $\underline{f}_i$ is a basis of W, then if $\underline{x} \in V$

$$\underline{x} = x^i \underline{e}_i$$

where $\{x^i\}$ are n real numbers. Consider the element $\underline{X}$ (say) of W with the same "components"

$$\underline{X} = x^i \underline{f}_i \ .$$

Given the x^i this element $\underline{X}$ is unique — we have established a (1,1) correspondence between V and W, and it is clear that

$$(\lambda \underline{x}_1 + \mu \underline{x}_2) \longleftrightarrow (\lambda \underline{Y}_1 + \mu \underline{Y}_2) \qquad (24)$$

where

$$\underline{x}_1 \longleftrightarrow \underline{Y}_1 \quad \text{and} \quad \underline{x}_2 \longleftrightarrow \underline{Y}_2$$

The correspondence is therefore an isomorphism. This relationship is highly artificial, since if we choose a new basis of V, $\underline{e}_i'$, then $\underline{x}$ no longer corresponds to $\underline{X}$ but to some new element $\underline{X}^1$, where

$$\underline{X}^1 = x^{i'} \underline{f}_i$$

i.e. we have a new isomorphism of V and W. The relationship is determined by a choice of basis for V and W, and there are as many such relationships as there are bases for V. Generally, there is no way of singling out one such isomorphism. If, however, V and W are so related that we can establish an isomorphism in a way which makes no reference to an arbitrary choice of basis, the V and W are said to be NATURALLY ISOMORPHIC. As an example consider Cartesian 3-space. The set of all triples

$$(a,b,o)$$

with the same addition and scalar multiplication laws forms a vector sub-space of this space; it is naturally isomorphic to Cartesian two-space, as are the sets

$$\{(a,o,b)\} \quad \text{and} \quad \{(o,a,b)\}$$

We shall also describe two spaces as being naturally isomorphic if the relation has been established by means of a basis, but a change of basis leaves us with the same correspondence. We identify naturally isomorphic spaces.

Let $\underline{v} \in V$.

Corresponding to the basis $\underline{e}^i$ of V^* we have the dual basis

$$\underline{\tilde{e}}_i \quad \text{(say)}$$

of V^{**}, and for all elements $\underline{h}$ of V^*

$$\underline{\tilde{e}}_i(h) = h_i$$

where $\underline{h} = h_i \underline{e}^i$.

If v^i are the components of the given vector $\underline{v}$ w.r.t. the basis $\underline{e}_i$,

$$\begin{aligned} v^i \underline{\tilde{e}}_i(h) &= v^i h_i \\ &= h(\underline{v}) \end{aligned}$$

write $$\underline{\tilde{v}} = v^i \underline{\tilde{e}}_i \quad .$$

Thus there exists a $\underline{\tilde{v}} \in V^{**}$ such that

$$\underline{\tilde{v}}(h) = \underline{h}(v)$$

for all $\underline{h} \in V^*$.

Suppose there exists a $\underline{\tilde{\tilde{v}}} \in V^{**}$ such that this relation holds.

Then $$(\underline{\tilde{\tilde{v}}} - \underline{\tilde{v}})(h) = 0$$

and so $$\underline{\tilde{\tilde{v}}} = \underline{\tilde{v}} \quad .$$

Also if

$$\underline{\tilde{v}}(\underline{h}) = \underline{h}(\underline{w})$$

then $$\underline{h}(\underline{v}) = \underline{h}(\underline{w}) .$$

This holds for all $\underline{h}$; in particular for each $\underline{e}^i$ in turn.
But

$$\underline{e}^i(\underline{v}) = v^i \quad \text{and} \quad \underline{e}^i(\underline{w}) = w^i .$$

Thus

$$\underline{v} = \underline{w} .$$

We have therefore established a (1,1) correspondence between elements $\underline{v}$ and elements $\underline{\tilde{v}}$, where

$$\underline{\tilde{v}}(\underline{h}) = \underline{h}(\underline{v}) \qquad (25)$$

Clearly this correspondence is linear, and since we establish it from the above equation, with no reference to a basis, V and V^{**} are naturally isomorphic.

This concept of "natural isomorphism" is not particularly satisfactory, even though it appears in most elementary textbooks on Linear Algebra. It is clear that V and V^{**} are related in a stronger manner than are V and V^*, say. However, we normally characterise morphic relationships between sets (with structures) by requiring them to preserve some structure (e.g. that an isomorphism be an isometry). To inquire into the manner in which the relationship is established seems very odd, and the requirement that it be "independent of basis" is difficult to make precise. An excellent treatment of this and related questions is found in the appendix to Phillip Tondeur's "Introduction to Lie Groups", Springer-Verlag (1965), (Ref 50) where the concepts of " Categories" and "Functions" are discussed.

We shall sometimes denote by

$$\langle \underline{h}, \underline{v} \rangle$$

the scalar product $\phi(\underline{h},\underline{v})$ determined by the duality of V^* relative to V. We can regard

$$\phi : V^* \times V \to R$$

as a genuine, symmetric, scalar product; we shall write

$$v(\underline{h})$$

instead of

$$\tilde{v}(\underline{h}) \, .$$

11. TENSOR PRODUCTS

Just as the set of all linear functions on V (together with the appropriate definitions for addition and scalar multiplication) forms a vector space, so does the set of all bilinear functions (real valued functions of two vector variables, linear in each). Given a basis $\underline{e}_i$ of V, just as the functions $\underline{e}^i$ such that

$$\underline{e}^i(\underline{e}_j) = \delta^i{}_j$$

form a basis of V^* the bilinear functions $\underline{e}^{ij}$, where

$$\underline{e}^{ij}(\underline{e}_k, \underline{e}_l) = \delta^i{}_k \delta^j{}_l \qquad (26)$$

form a basis of this space.

If T is any bilinear function

$$T(\underline{v},\underline{w}) = v^i w^j T(\underline{e}_i, \underline{e}_j)$$
$$= v^i w^j T_{ij}, \quad \text{say,}$$

From the Classical approach to tensors, it is clear that we wish to regard the T_{ij} as components of a second rank tensor.

We may introduce a straightforward addition and scalar multiplication of bilinear functions. Note that if

$$F(\underline{e}_k,\underline{e}_l) = G(\underline{e}_k,\underline{e}_l) \qquad (\forall\, k,l)$$

then

$$F(\underline{v},\underline{w}) = G(\underline{v},\underline{w})$$

i.e.

$$F = G .$$

Now

$$T_{ij}\,\underline{e}^{ij})(\underline{e}_k,\underline{e}_l) = T_{ij}\,\delta^i{}_k\delta^j{}_l$$

$$= T_{kl}$$

$$= T(\underline{e}_k,\underline{e}_l) .$$

Hence

$$T = T_{ij}\,\underline{e}^{ij} \tag{27}$$

and so T_{ij} are the components of T w.r.t. the basis $\underline{e}^{ij}$.

Given any two linear functions (elements of V^*)

$$f, g$$

we may define a bilinear function F by

$$F(\underline{v},\underline{w}) = f(\underline{v})g(\underline{w}).$$

In this case we shall write

$$F = f \otimes g \tag{28}$$

(read : "f tensor g").

F is called the TENSOR PRODUCT of f and g. It is a very natural definition of multiplication, since

$$(f \times g)(\underline{v},\underline{w}) = f(\underline{v})g(\underline{w}) \quad . \tag{29}$$

Clearly tensor multiplication is not commutative. We postpone consideration of an associative law. It is obvious that the multiplication is distributive over addition of linear and bilinear functions. The class of all tensor products forms a subset of the set of all bilinear functions; if a bilinear function is of this form it is said to be DECOMPOSABLE.

In terms of components w.r.t. the bases $\underline{e}_i\underline{e}^{ij}$ we would then have

$$F_{ij} = f_i g_i \quad . \tag{30}$$

In this case the matrix F_{ij} is of rank one, and there are (n-1) independent vectors $\underline{v}$ of V such that

$$F(\underline{e}_i,\underline{v}) = 0 \quad .$$

Conversely, if F_{ij} is of rank one, then F is decomposable. Any matrix can be written as a sum of matrices of rank one, so any bilinear function can be written as a sum of tensor products.

Consider the bilinear function

$$\underline{e}^i \times \underline{e}^j \quad .$$

We have

$$\begin{aligned}(\underline{e}^i \times \underline{e}^j)(\underline{e}_k,\underline{e}_j) &= e^i(\underline{e}_k).e^j(\underline{e}_j) \\ &= \delta^i{}_k\delta^j{}_j \quad .\end{aligned}$$

Thus

$$\underline{e}^i \otimes \underline{e}^j = \underline{e}^{ij} \tag{31}$$

and we can write, generally,

$$G = G_{ij}\, \underline{e}^i \otimes \underline{e}^j \tag{32}$$

G is said to be SECOND RANK COVARIANT TENSOR, and the vector space of bilinear functions on $V \times V$ is said to be the TENSOR PRODUCT SPACE

$$V^* \otimes V^*$$

(a more sensible notation than $V \times V$, since we have written the basis elements as $\underline{e}^i \otimes \underline{e}^j$, not $\underline{e}_i \otimes \underline{e}_j$).

More generally, given two vector spaces V, W, the set of bilinear functions on $V \times W$, turned into a vector-space in the usual way, forms the space

$$V^* \otimes W^* \ .$$

We may also consider functions on $V^* \times V^*$. Since we have identified V^{**} with V, a basis of the vector space of bilinear functions on $V^* \times V^*$ will be

$$\underline{e}_i \otimes \underline{e}_j$$

and so we denote this space by

$$V \otimes V \ .$$

It remains to consider the associative law for tensor products. Here

$$(U \times V) \times W = U \times (V \times W) \tag{33}$$

is true in the sense that the two spaces are naturally isomorphic. For simplicity, we consider only tensor product spaces of V with itself.

Given two bases

$$\underline{e}_{i}, \quad \underline{e}_{i'}$$

of V, there must exist a relationship of the form

$$\underline{e}_{i'} = p^{i}_{i'}\, \underline{e}_{i}$$

where

$p^{i}_{i'}$ is a non-singular matrix.

Clearly

$$\underline{e}_{i} = (p^{i}_{i'})^{-1}\, \underline{e}_{i'} \quad .$$

We shall write

$p^{i'}_{i}$ instead of $(p^{i}_{i'})^{-1}$.

Thus

$$p^{i'}_{i}\, p^{i}_{j'} = \delta^{i'}_{j'} \quad .$$

If $\underline{v} \in V$,

$$\underline{v} = v^{i}\, \underline{e}_{i} = v^{i'}\, \underline{e}_{i'}$$

$$= v^{i} p^{i'}_{i}\, \underline{e}_{i'}$$

and if $\underline{f} \in V^{*}$, we have

$$e^{i'}(\underline{e}_{j'}) = \delta^{i'}_{j'}$$

whence

$$e^{i'}(\underline{e}_{j}) = p^{i'}_{j}$$

giving

$$\underline{e}^{i'} = p^{i'}_{i}\underline{e}^{i} \text{ and } \underline{e}^{i} = p^{i}_{i'}\underline{e}^{i'} \quad .$$

Thus

$$f = f_\iota \underline{e}^\iota = f_{\iota'} \underline{e}^{\iota'}$$

$$= f_\iota p^\iota_{\iota'} \underline{e}^{\iota'} \quad .$$

Hence

$$v^{\iota'} = p^{\iota'}_\iota v^\iota \qquad (34)$$

$$f_{\iota'} = p^\iota_{\iota'} f_\iota \qquad (35)$$

which are the usual component transformation laws.

These considerations provide another way of demonstrating the natural isomorphism between V and V^{**}. The duality relation gives a mapping of a basis $\underline{e}_\iota \in V$ onto a basis $\underline{e}^\iota \in V^*$, and provides an isomorphism of V and V^* by

$$\underline{v} \longleftrightarrow \underline{\xi}$$

where

$$\underline{v} = v^\iota \underline{e}_\iota, \qquad \underline{\xi} = \xi_\iota \underline{e}^\iota$$

and

$$v^\iota = \xi_\iota \text{ numerically.}$$

With respect to a new basis

$$\underline{e}_{\iota'} \text{ of } V$$

$$\underline{v} = p^{\iota'}_\iota v^\iota \underline{e}_{\iota'} , \quad \underline{\xi} = p^\iota_{\iota'} \xi_\iota \underline{e}^{\iota'}$$

and since

$$p^{\iota'}_\iota \neq p^\iota_{\iota'} \text{ unless } \underline{e}_\iota = e_{\iota'} ,$$

the mapping

$$\underline{e}_{\iota'} \rightarrow \underline{e}^{\iota'} \quad \text{provides an isomorphism of } V \text{ and } V^*$$

under which $\underline{v}$ does not correspond to $\underline{\xi}$. Thus we get an isomorphism corresponding to each choice of basis, and so obtain no natural isomorphism.

However, if $\tilde{\underline{e}}_\iota$ is the basis of V^{**} dual to $\underline{e}^\iota$, then it is easy to show that if $\tilde{\underline{e}}_{\iota'}$ is dual to $\underline{e}^{\iota'}$, then

$$\tilde{\underline{e}}_{\iota'} = p^{\iota}_{\iota'} \, \underline{e}_\iota \, . \tag{36}$$

The bases

$$\underline{e}_\iota, \tilde{\underline{e}}_\iota \quad \text{provide an isomorphism of } V, V^{**}$$

by

$$\underline{v} \leftrightarrow \tilde{\underline{v}}$$

where

$$\underline{v} = v^\iota \, \underline{e}_\iota \, , \quad \tilde{\underline{v}} = v^\iota \, \tilde{\underline{e}}_\iota$$

Choice of a new basis $\underline{e}_{\iota'}$ gives

$$\underline{v} = p^{\iota'}_{\iota} \, v^\iota \, \underline{e}_{\iota'} \, , \quad \tilde{\underline{v}} = p^{\iota'}_{\iota} \, v^\iota \, \tilde{\underline{e}}_{\iota'}$$

whence the mapping $\underline{e}_{\iota'} \rightarrow \tilde{\underline{e}}_{\iota'}$ provides the same isomorphism as does $\underline{e}_\iota \rightarrow \tilde{\underline{e}}_\iota$. Here we have an isomorphism independent of choice of basis.

12. CONTRAVARIANT AND COVARIANT VECTORS

A similar technique can be used to show that

$$V \otimes V \quad \text{and} \quad (V^* \otimes V^*)^*$$

are naturally isomorphic.

Use the notation $\underline{e}^{ij}$ for the tensor-product basis of $V^* \times V^*$, and the notation $\underline{e}_i \times \underline{e}_j$ for the basis of $V \times V$, each constructed from $\underline{e}_i$.

Now $V^* \times V^*$ is a vector space of dimension n^2 ($\dim V = n$). The dual space has a dual basis consisting of the n^2 elements

$$\underline{e}_{ij}$$

where

$$e_{ij}(\underline{e}^{k\ell}) = \delta_i^k \, \delta_j^\ell \tag{37}$$

Since

$$e^{k\ell}(\underline{e}_m, \underline{e}_n) = \delta_m^k \, \delta_n^\ell$$

we have

$$\underline{e}^{k'\ell'} = p_k^{k'} p_\ell^{\ell'} \, \underline{e}^{k\ell}$$

We deduce

$$\underline{e}_{i'j'} = p_{i'}^{i} \, p_{j'}^{j} \, \underline{e}_{ij} \tag{38}$$

Similarly

$$\underline{e}_{i'} \otimes \underline{e}_{j'} = p_{i'}^{i} \, p_{j'}^{j} \, \underline{e}_i \times \underline{e}_j \tag{39}$$

Thus the mapping

$$\underline{e}_{ij} \rightarrow \underline{e}_i \times \underline{e}_j$$

provides a natural isomorphism of the two spaces.

Consider now

$$(V \times V) \times V \quad \text{and} \quad V \times (V \times V) \quad .$$

Denote bases of each by

$$(\underline{e}_i \times \underline{e}_j) \times \underline{e}_k \,, \quad \underline{e}_i \times (\underline{e}_l \times \underline{e}_k)$$

respectively.

Then it is easy to show that

$$(\underline{e}_{i'} \times \underline{e}_{j'}) \times \underline{e}_{k'} = p^{i}_{i'} p^{j}_{j'} p^{k}_{k'} (\underline{e}_i \times \underline{e}_j) \times \underline{e}_k \qquad (40)$$

$$\underline{e}_{i'} \times (\underline{e}_{j'} \times \underline{e}_{k'}) = p^{i}_{i'} p^{j}_{j'} p^{k}_{k'} \underline{e}_i \times (\underline{e}_j \times \underline{e}_k) \qquad (41)$$

whence the mapping

$$(\underline{e}_i \times \underline{e}_j) \times \underline{e}_k \rightarrow \underline{e}_i \times (\underline{e}_j \times \underline{e}_k)$$

provides a natural isomorphism of the two spaces.

Clearly the set of trilinear functions over V can be turned into a vector space, which we denote by

$$V \times V \times V \,.$$

It is easy to show that this space is naturally isomorphic to $(V \times V) \times V$ — hence in future we drop the brackets.

Generally we shall define a CONTRAVARIANT TENSOR of ORDER (RANK) p, over V, to be a multilinear function of p variables, each an element of V^*. A COVARIANT TENSOR is defined similarly, except that the variables are elements of V. MIXED TENSORS are functions of vector variables, some of which are elements of V and some of V^*. The vector space of tensors of contravariant and covariant ranks (p,q) respectively will be denoted

$$\times^{p} V \times {}^{q} V^{*} \qquad (42)$$

In term of the tensor-product bases, such a tensor T can be written in the form:

$$T = T^{ab\ldots f}_{ij\ldots r}\, \underline{e}_a \times \underline{e}_b \times \ldots \underline{e}_f \times \underline{e}^i \times \ldots \underline{e}^r \qquad (43)$$

We can deduce the transformation law for the components of such a tensor:

$$T^{a'b'\ldots f'}_{i'j'\ldots r'} = p^{a'}_{a}p^{b'}_{b}\ldots p^{f'}_{f}\,p^{i}_{i'}\ldots p^{r}_{r'}\;T^{ab\ldots f}_{ij\ldots r} \tag{44}$$

"Tensors" must not be confused with "tensor fields" — yet to be defined. The latter is some form of assignment of tensors to points of a space.

We may establish other interesting natural isomorphisms: e.g. tensors of type (1,1) are equivalent to linear mappings

$$V \to V$$

and generally tensors of type (p,q) to multilinear mappings

$$\underbrace{V \times V \times \ldots V}_{p \text{ times}} \to \otimes^{q} V$$

or equally

$$\underbrace{V^{*} \times \ldots \times V^{*}}_{q \text{ times}} \to \otimes^{p} V^{*}$$

Thus we can derive new vector-spaces from a given one, by considering sets of real-valued functions on it.

<u>13. [illegible] OF FUNCTIONS</u>

We now examine the set of <u>all</u> real-valued <u>differentiable</u> functions on a differentiable manifold.

Let $f(p)$ be a real-valued function. Then in any local chart (v,ϕ) it is represented by its "coordinate image" $\tilde{f}(x^{1},\ldots x^{n})$, where

$$\tilde{f} = f \circ \phi^{-1} . \tag{45}$$

This is because

$$\phi(p) = \{x^k\}$$

implies

$$p = \phi^{-1}(x^1, \ldots x^n).$$

If the function $\tilde{f}$ is a differentiable function of the (differentiable) coordinates $\{x^k\}$, the function f is said to be DIFFERENTIABLE. We shall assume, except where otherwise stated, that this means "differentiable any number of times". The set of all such functions f is denoted by $C^\infty(M)$. This set can be turned into a vector space by means of three definitions, as for functions on a vector space:

F.1. $(f + g)(P) = f(P) + g(P)$

F.2. $(kf)(P) = kf(P)$

F.3. $0(P) = 0$

for any point P.

Since there are now no linearity assumptions on such functions, this vector space is not of finite dimension. We can extend F.3. as follows:

To every real number f_0 there corresponds a function $f_0(P)$ where

F.3a. $f_0(P) = f_0$

for all P. Such a function is called a CONSTANT FUNCTION.

Finally we can add a multiplication definition

F.4. $(f.g)(P) = f(P)g(P)$

It can be verified that the associative and distributive axioms of addition and multiplication are satisfied. Such a set is called a LINEAR ALGEBRA over the real numbers. This is just a vector space with a multiplication. If we ignore definition F.2. for the moment we may also regard the set as a RING. This is a commutative group under an addition operation, and a semi-group under multiplication, with the distributive law satisfied. If furthermore the full group requirements hold for multiplication the set is called a FIELD. Examples: the integers (positive and negative) form a ring, the rationals form a field, as do the real numbers, and also the complex numbers.

A vector space over a field K has the usual axioms, except that "scalar now means "element of K" instead of "real number". All the normal theorems still hold, for example we can define the dual space to be the set of all K-valued linear functions.

If however we weaken the assumptions, supposing that the normal vector axioms hold but the scalars are drawn from a ring R, the set is said to form a MODULE over the ring. A module satisfies some of the vector theorems; in particular every element is linearly expressible in terms of a set of basis elements. However, the following vector space result does not hold for modules, in general: if the space is of dimension n, and we are given r(< n) linearly independent vectors, then we can find another n-r vector such that the n vectors are independent, and so form a basis. The proof of this uses division of scalars, which is not always possible in a ring.

Definition F.3a. ensures that our ring has a unit element. It is not a field, because not all elements possess a multiplicative inverse: for a field, to any element f there corresponds an element f^{-1} (say), with

$$f^{-1}.f = 1$$

i.e. $$(f^{-1}.f)(P) = 1 \quad \text{for all } P$$

which implies $$(f^{-1})(P)f(P) = 1$$

and so $$(f^{-1})(P) = \frac{1}{f(P)} .$$

This must hold except if f is the zero function. However $f(P)$ can have value zero even though f is not the zero function; since the above relation must hold for all P, (f^{-1}) does not exist whenever f has a zero.

The vector space of the C^{∞} functions, being infinite dimensional has no obvious geometrical significance. We earlier defined a vector field on Euclidean space to be a mapping of the space into itself; a similar definition will not do for manifolds, because this would no longer imply that vectors were being "attached" to each point of space. However, there is an equivalent, though more subtle definition of vector fields which can be generalised to provide a definition of vector fields on a manifold.

14. DERIVATIONS

Given an algebra A, a function D which maps A into itself will be called a DERIVATION if the following requirements are satisfied.

D.1. $\quad D(\lambda v + \mu w) = \lambda D(v) + \mu D(w)$

D.2. $\quad D(v \circ w) = w \circ Dv + v \circ Dw$ (Leibniz rule)

for all $v, w \in A$. Here the dot product is the multiplication operation – not a scalar product.

Now consider the set of all such derivations of the algebra C^∞. We can define addition and scalar multiplication of derivations as follows:

D.3. $\quad (X + Y)(f) = X(f) + Y(f)$

D.4. $\quad (hX)(f) = h\,X(f)$

where X, Y are derivations; f, h are C^∞ functions. Here the scalars are elements of C^∞ – i.e. differentiable functions.

Add

D.5. $\quad 0(f) = 0$

where the right-hand side is the zero function. As with all such definitions, the remaining vector-space axioms are satisfied; except that the scalars are now only a <u>ring</u>, and not a <u>field</u>, as we have seen. Thus the derivations form a module over the C^∞ functions.

As an example, take for our manifold simply the real line. The C^∞ functions are the differentiable functions of one variable x. Now clearly a <u>special case</u> of a derivation is simply the ordinary derivative

$$\frac{d}{dx} .$$

This maps any function $f(x)$ into the derived function. We may also define an operator

$$\left.\frac{d}{dx}\right]_a$$

which given a point a maps the function f into the number $f'(a)$. More generally, given as derivation D we may define an associated operator

$$D_a$$

which maps the function f into the number

$$[Df]_{x=a}$$

The operators D_a satisfy D.1., D.2. above; also D.3, D.4, D.5., except that in D.4. h is to mean any real number, and not any C^∞ function.

We now prove the following interesting result: any derivation on the real line is simply a multiple of the derivative.

Let a be any point. Then the mean value theorem states that

$$f(x) = f(a) + (x-a)f'(a + \theta(x-a)) \qquad (46)$$

where $0 < \theta < 1$.

Let D be any derivation. Any constant C counts as a C^∞ function. Now

$$D(C.1) = D(C).1 + C.D(1)$$

by the Leibniz property. Also

$$D(C.1) = CD(1)$$

by the linear property.

Hence

$$D(C).1 = 0$$

i.e.

$$D(C) = 0$$

just as for the derivative.

Now

$$Df(x) = Df(a) + (x-a)Df'(a + v(x-a)$$

$$+ f'(a + \theta(x-a))D(x - a)$$

But $f(a)$ is a constant — thus $Df(a)$ vanishes.

Now put $x = a$. Then

$$D_a f(x) = 0.Df'(a + \theta(x-a)) + f'(a)D_a(n)$$

$$= f'(a)\, D_a(x) \qquad (47)$$

Thus

$$D_a = D_a(x) \left.\frac{d}{dx}\right]_a$$

$$= \text{const.} \left.\frac{d}{dx}\right]_a .$$

This is true for every point a.

Therefore

$$D = D(x) \frac{d}{dx} .$$

$D(x)$ is a fixed function for given D.

15. DERIVATIONS ON A MANIFOLD

For general manifolds we may prove that any derivation X is merely a linear sum of partial derivative operators. The proof is similar to the above.

If C is any constant function on the manifold, the foregoing proof that

$$D(C) = 0 \tag{48}$$

still applies.

We could use the n-variable mean-value theorem, but an analogous proof suffices.

Let P be a fixed point in a coordinate chart (V,ϕ) with coordinates $\{a^i\}$. Let Q be any point of V such that all the "intervening" points, with coordinates

$$\{a^i + t(x^i - a^i)\}, \qquad 0 \leqslant t \leqslant 1$$

(where Q has coordinates $\{x^i\}$) lie within V.

Then as we have seen

$$f(Q) = \tilde{f}(x^i)$$

where

$$\tilde{f} = f \circ \phi^{-1} .$$

Now

$$\tilde{f}(x^i) = \tilde{f}(a^i) + \int_0^1 \frac{d}{dt} \tilde{f}(a^i + t(x^i - a^i))dt \tag{49}$$

Write

$$\xi^i = a^i + t(x^i - a^i) .$$

Then

$$\frac{d\tilde{f}}{dt} = \sum_1^n \frac{\partial \tilde{f}}{\partial \xi^k} \frac{d\xi^k}{dt}$$

$$= \sum_1^n (x^k - a^k) \frac{\partial \tilde{f}}{\partial \xi^k} \ .$$

Thus $$\tilde{f}(x^i) = \tilde{f}(a^i) + \sum_1^n (x^k - a^k) \int_0^1 \frac{\partial}{\partial \xi^k} \tilde{f}(\xi^i)\, dt \ .$$

Write $$E_k(x^i, a^i) = \int_0^1 \frac{\partial}{\partial \xi^k} \tilde{f}(\xi^i) dt \ . \qquad (50)$$

Then $$E_k(a^i, a^i) = \int_0^1 \left.\frac{\partial \tilde{f}(x^i)}{\partial x^k}\right]_{a^i} dt$$

$$= \frac{\partial \tilde{f}}{\partial a^k} \ .$$

The function $\frac{\partial}{\partial x^k}(\tilde{f})$ on Cartesian space corresponds to a certain function on the manifold, which we denote by

$$\frac{\partial}{\partial x^k}(f) \ .$$

By definition

$$\frac{\partial}{\partial x^k} f = \left[\frac{\partial}{\partial x^k}(f \circ \phi^{-1})\right] \circ \phi \ . \qquad (51)$$

This operator $\frac{\partial}{\partial x^k}$ is clearly a derivation; its coordinate image is the real-variable partial derivative.

Now

$$\tilde{f}(x^i) = \tilde{f}(a^i) + \sum_1^n (x^k - a^k) E_k(x^i, a^i) \qquad (52)$$

which is equivalent to

$$f(Q) = f(P) + \sum_1^n (x^k - a^k) E_k(x^i, a^i) \qquad (53)$$

Let X be any derivation of the C^{∞} functions.

Then

$$X(f(Q)) = X(f(P)) + \sum_{1}^{n}(x^{k} - a^{k})\, X\,(\Xi_{k}(x^{l},a^{l}))$$
$$+ \sum_{1}^{n} \Xi_{k}(x^{l},a^{l})\, X\,(x^{k} - a^{k}) .$$

But $f(P)$ and a^{k} are constant functions.

Thus

$$X(f(Q)) = \sum_{1}^{n} \left\{ (x^{k} - a^{k}) X\,(\Xi_{k}(x^{l},a^{l})) + \Xi_{k}(x^{l},a^{l}) \right. \quad (54)$$

Now consider the one value $Q = P$.

Here $x^{k} = a^{k}$.

Thus $$X_{P}(f(Q)) = \sum_{1}^{n} \Xi_{k}(a^{l},a^{l})\, X_{P}(x^{k})$$

where X_P is the derivation "evaluated" at P. Just as for derivations on Cartesian space, X_P associated with a point P and any function f a real number.

Hence $$X_{P}(f) = \sum_{1}^{n} \frac{\partial}{\partial x^{k}}\Big]_{P}(f) \cdot X_{P}(x^{k}) \; .. \qquad (55)$$

Since this is true for all P, we could equally write

$$X(f(P)) = \sum_{1}^{n} \frac{\partial}{\partial x^{k}}(f(P))\; X\,(x^{k}(P)) \qquad (56)$$

(55) implies that the vector space of operators X_P has as a basis the set of operators

$$\frac{\partial}{\partial x^{k}}\Big]_{P}$$

(56) implies that the module of derivations X has as a basis the set of derivations

$$\frac{\partial}{\partial x^k} .$$

16. TANGENT SPACES

We denote the module of derivations on a manifold M by

$$\mathscr{D}^1(M)$$

and the vector space of operators X_P by

$$\mathscr{D}^1(P)$$

Any element of $\mathscr{D}^1(M)$ will be called a CONTRAVARIANT VECTOR FIELD on M. That is, a vector field is a derivation of M.

Any element of $\mathscr{D}^1(P)$ will be called a TANGENT VECTOR at P. $\mathscr{D}^1(P)$ itself is called the TANGENT SPACE to M at P. A vector field clearly enables us to select one tangent vector at each point — indeed providing a mapping of tangent vectors onto the points of M.

We now check the component transformation law for contravariant vector fields. By (56) above

$$X = X(x^k) \frac{\partial}{\partial x^k} \tag{57}$$

(using the summation convention).

Suppose our coordinate patch admits the coordinates $(x^{k'})$ also. Then

$$X = X(x^{k'}) \frac{\partial}{\partial x^{k'}} \tag{58}$$

But each $x^{i'}(P)$ is a function on the manifold, so we have

$$X(x^{i'}(P)) = X(x^k)\frac{\partial}{\partial x^k}(x^{i'}(P)) \qquad (59)$$

using (57).

However, $\frac{\partial}{\partial x^k}(x^{i'}(P))$ is in fact just $\frac{\partial}{\partial x^k}(x^{i'}(x^j))$ where the former is in our symbolic notation, and the latter is the straightforward partial derivative.

Hence (59) is simply

$$X(x^{i'}) = X(x^k)\frac{\partial x^{i'}}{\partial x^k} \qquad (60)$$

Writing $X(x^i) = u^i$ and $X(x^{i'}) = u^{i'}$, then by (57),(58) $\{u^i\}$, $\{u^{i'}\}$ are each components of the vector field X w.r.t. our coordinate-derived bases.

Using (60)

$$u^{i'} = \frac{\partial x^{i'}}{\partial x^k}u^k \qquad (61)$$

which is the familiar component result for contravariant vector fields.

17. [illegible]

Consider any derivation X.

We have

$$X = X(x^i)\frac{\partial}{\partial x^i}$$

Consider derivations of functions taking values along some curve.

$$x^i = h^i(t) \quad .$$

Write $X^i(x^k)$ for the operation $X(x^i)$

The X^i are each C^∞ functions.

Now the system of equations

$$\frac{dh^i}{dt} = X^i(h^k) \qquad (62)$$

has at least one solution: thus these equations determine at least one curve.

For all points lying on this curve we may write

$$X(x^i(t)) = \frac{dx^i}{dt} = \dot{h}^i(t) .$$

If P is any point on this curve corresponding to $t = t_0$, say, then

$$X_P(x^i) = \dot{h}^i(t_0) .$$

But $X_P(x^i)$ are just the components of X_P w.r.t. the natural basis. Thus given any X, we can find a curve for which the components of X_P are the "coordinate components" of the tangent vector to the curve at P, for all points P lying on the curve.

Now we can always find a solution of the equations (62) passing through any point for which the X^i are not all zero (from the theory of such a system of equations). Thus for all points P with $X_P \neq 0$ we can find a curve for which X is the tangent vector (in the coordinate component sense) for points on the curve at and around P.

Note that along this curve, if F is any C^∞ function

$$X(F(t)) = \frac{dx^i}{dt}\frac{\partial F}{\partial x^i}$$

$$= \frac{d}{dt}(F(t))$$

whence the operator X is to be thought of as the tangential derivative along this curve.

In particular, of course, putting $F = x^i$ for each coordinate in turn,

$$X(x^i(t)) = \frac{dx^i}{dt} \qquad (63)$$

which is consistent with (1). However, this follows directly from

$$X = X(x^i) \frac{\partial}{\partial x^i}$$

whence

$$X(x^j) = X(x^i) \delta_i{}^j \qquad (64)$$

1[illegible]. [illegible] FORMS

The dual module of $\mathscr{D}^1(M)$ is denoted by

$$\mathscr{D}_1(M) .$$

Any element of this module is called a COVARIANT VECTOR FIELD (the transformation law is obtained in a straightforward manner). If $\mathscr{D}_1(P)$ denotes the vector space obtained from $\mathscr{D}_1(M)$ by replacing all scalar functions by their values at P, then an obvious result (Helgason (ref. 15)) is that $\mathscr{D}_1(P)$ is the dual vector-space of $\mathscr{D}^1(P)$.

Corresponding to our coordinate basis of $\mathscr{D}^1(M)$ there is the dual basis of $\mathscr{D}_1(M)$. This will be denoted by

$$\{dx^i\} .$$

Each dx^i is a linear operator on tangent vectors, and is not, therefore, any sort of coordinate increment. However it is in many ways analogous to the increment concept.

Since dx^i is the dual basis to $\frac{\partial}{\partial x^i}$, we have

$$(dx^i)\left(\frac{\partial}{\partial x^j}\right) = \delta^i{}_j \quad (65)$$

Corresponding to each C^∞ function F, we define an operator (dF) by the equation

$$(dF)(X) = X(F) \quad (66)$$

where X is a derivation. It follows that dF is a linear operator, and therefore an element of $\mathcal{D}_1(M)$. Note that (66) is consistent with (65), since (66) gives

$$(dx^i)\left(\frac{\partial}{\partial x^j}\right) = \frac{\partial}{\partial x^j}(x^i) = \delta^i_j . \quad (67)$$

More generally, (66) gives $(dF)\left(\frac{\partial}{\partial x^j}\right) = \frac{\partial F}{\partial x^j}$

whence our notation allows us to "multiply" derivatives and differentials in an obvious manner.

Now $$X = X(x^i)\frac{\partial}{\partial x^i} .$$

Thus

$$(dF)(X) = X(x^i)\frac{\partial F}{\partial x^i} \quad (68)$$

But using (65)

$$\begin{aligned}(dx^i)(X) &= (dx^i)\left(X(x^j)\frac{\partial}{\partial x^j}\right)\\ &= X(x^j)\,\delta^i{}_j\\ &= X(x^i) . \end{aligned} \quad (69)$$

This follows immediately from (66), providing a full check of consistency of (65) and (66).

Using (68) and (69)

$$(dF)(X) = \frac{\partial F}{\partial x^i}\, dx^i\,(X)$$

whence $\left\{\frac{\partial F}{\partial x^i}\right\}$ are the components of dF with respect to the dual basis. The earlier approach to vectors described $\left\{\frac{\partial F}{\partial k^i}\right\}$ itself as a covariant vector — we regard them as the components of the covariant vector field dF w.r.t. the basis of differentials.

The relation

$$dF = \frac{\partial F}{\partial x^i}\, dx^i \tag{70}$$

provides the justification for our notation.

19. TENSOR FIELDS

Co- and Contravariant Tensor Fields may now be defined in a straightforward manner: the <u>tensor product module</u> $\mathcal{D}^r_s(M)$ is defined to be

$$\otimes^r \mathcal{D}^1(M) \otimes^s \mathcal{D}_1(M) \tag{71}$$

and consists of the appropriate set of multilinear functions on $\mathcal{D}^1(M)$. The module $\mathcal{D}^r_s(M)$ defines a vector space $\mathcal{D}^r_s(P)$ and it may be proved (Helgason, (Ref. 15)) that $\mathcal{D}^r_s(P)$ is the appropriate tensor-product space of $\mathcal{D}^1(P)$ and $\mathcal{D}_1(P)$; also that $\mathcal{D}^r_s(M)$ is the dual of $\mathcal{D}^s_r(M)$.

For the vector-space $\mathcal{D}^1(P)$, if $\underline{e}_i$, $\underline{e}_{i'}$ are any two bases, then if $\xi^i, \xi^{i'}$ are the components of a vector X_P

$$\xi^{i'} = A^{i'}_i\, \xi^i$$

where

$$\underline{e}_i = A^{i'}_{i}\underline{e}_{i'} \quad .$$

For the module $\mathscr{D}^1(M)$, given any two coordinate bases $\frac{\partial}{\partial x^i}$, $\frac{\partial}{\partial x^{i'}}$, then the component transformation law, as we have seen is the same as for $\mathscr{D}^1(P)$ except that the matrix $A^i_{i'}$ is no longer a set of numbers, but the set of functions

$$\left\{ \frac{\partial x^{i'}}{\partial x^i} \right\} \quad .$$

Similarly, the component transformation law for $\mathscr{D}^r_s(P)$ involves, as we have seen, the product of the matrix A r times, and the product of A^{-1} s times. For $\mathscr{D}^r_s(M)$ we simply replace A by the above Jacobian Matrix. Thus "tensor fields" are indeed "tensors" in the earlier sense. The operation of contraction can be defined in an intrinsic manner (Helgason, (Ref. 15)).

20. AFFINE CONNECTIONS

We next discuss the meaning of an Affine Connection. To assign one is clearly a "structure", condition. Covariant Differentiation is clearly a directional operation — in the early treatment of tensor calculus the directional character was expressed in terms of coordinate increments. We shall express the directional nature of such differentiation in terms of tangent vectors.

An AFFINE CONNECTION is a collection of operators ∇_X, each associated with an arbitrary vector field (of $\mathscr{D}^1(M)$), X, which maps $\mathscr{D}^1(M)$ into itself, such that the following requirements are satisfied.

A.1. $\nabla_X(aY + bZ) = a\nabla_X(Y) + b\nabla_X(Z)$

where a,b are constant functions, Y,Z are contravariant vector-fields.

A.2. $\nabla_X(F.Y) = F\nabla_X(Y) + X(F)Y$

where F is any C^∞ function.

A.3. $\nabla_{FX+GY}(Z) = F\nabla_X(Z) + G\nabla_Y(Z)$

Of course we need not introduce constants a,b into A.1., since

$$\nabla_X(aY) = a\nabla_X(Y)$$

may be deduced from A.2.

∇_X is called the COVARIANT DERIVATIVE w.r.t. X. In any coordinate patch, $\mathcal{D}^1(M)$ has the basis

$$\left\{ \frac{\partial}{\partial x^i} \right\}$$

Write $$\nabla_i = \nabla_{\partial/\partial x^i} \qquad (72)$$

Then $$\nabla_i \left(\frac{\partial}{\partial x^j}\right)$$

is a vector of $\mathcal{D}^1(M)$; we may write

$$\nabla_i \left(\frac{\partial}{\partial x^j} \right) = \Gamma_{ij}{}^k \frac{\partial}{\partial x^k} \qquad (73)$$

where the Γ's are C^∞ functions.

If $x^{k'}$ are a new set of coordinates

$$\nabla_{i'} \left(\frac{\partial}{\partial x^{j'}} \right) = \Gamma_{i'j'}{}^{k'} \frac{\partial}{\partial x^{k'}} \qquad (74)$$

But $$\frac{\partial}{\partial x^{j}} = \frac{\partial x^{j}}{\partial x^{j'}} \frac{\partial}{\partial x^{j}}$$

and so, using A.1. and A.3.,

$$\nabla_{i'}\left(\frac{\partial}{\partial x^{j'}}\right) = \frac{\partial x^{j}}{\partial x^{j'}} \nabla_{i'}\left(\frac{\partial}{\partial x^{j}}\right) + \frac{\partial}{\partial x^{i'}}\left(\frac{\partial x^{j}}{\partial x^{j'}}\right)\frac{\partial}{\partial x^{j}}$$

whence, using A.2.

$$= \frac{\partial x^{i}}{\partial x^{i'}} \frac{\partial x^{j}}{\partial x^{j'}} \nabla_{i}\left(\frac{\partial}{\partial x^{j}}\right) + \frac{\partial^{2} x^{j}}{\partial x^{i'}\partial x^{j'}} \frac{\partial}{\partial x^{j}}$$

i.e. $$\Gamma_{i'j'}{}^{k'} \frac{\partial}{\partial x^{k'}} = \Gamma_{ij}{}^{k} \frac{\partial x^{i}}{\partial x^{i'}} \frac{\partial x^{j}}{\partial x^{j'}} \frac{\partial}{\partial x^{k}} + \frac{\partial^{2} x^{j}}{\partial x^{i'}\partial x^{j'}} \frac{\partial}{\partial x^{j}} \qquad (75)$$

Let both sides operate on the function $(x^{\ell'})$.

Then

$$\Gamma_{i'j'}{}^{\ell'} = \Gamma_{ij}{}^{\ell} \frac{\partial x^{i}}{\partial x^{i'}} \frac{\partial x^{j}}{\partial x^{j'}} \frac{\partial x^{\ell'}}{\partial x^{\ell}} + \frac{\partial^{2} x^{j}}{\partial x^{i'}\partial x^{j'}} \frac{\partial x^{\ell'}}{\partial x^{j}} \qquad (76)$$

(where we have written ℓ instead of k). This is, of course, the transformation law for connection components in the older sense. Note how straightforward our concept of covariant derivative is: A.2. is really the Leibniz Rule, if we allow ∇_X to operate on SCALARS — i.e. C^{∞} functions — also:

$$\nabla_X(F) = X(F) \qquad (77)$$

that is, ∇_X is the derivative in the 'X' direction.

The notion of <u>parallelism</u> requires some care: we shall give the general lines of the argument (for a fuller discussion see Helgason (Ref. (15))

21. PARALLELISM

Consider a curve $H(t)$; then at each point we can choose a tangent vector which is tangent to this curve in the coordinate-component sense. We may also choose a vector field X such that for each point of the curve X_p is this tangent vector. Suppose we have another collection of tangent vectors Y_p for each point of the curve, which is such that $Y_{H(t)}$ as t is taken round the curve is to be regarded as a differentiable function of t. Then we may choose a vector field Y such that the associated Y_p are these tangent vectors.

The collection of vectors $Y(t)$ is said to be PARALLEL along the curve if

$$\nabla_X(Y) = 0 \tag{73}$$

along the curve. In other words, Y is COVARIANTLY CONSTANT in the tangential direction to the curve. Starting with some vector $Y(t_0)$ at $t = t_0$, the solution of the above equation will define a new vector $Y(t)$ at each point of the curve. The vector $Y(t_0)$ is then said to be PARALLELY PROPAGATED along the curve.

In a coordinate patch, write

$$X^i = X(x^i) \ ; \qquad Y^i = Y(y^i)$$

X^i, Y^i are then the components of their respective vector-fields.

Then

$$X = X^i \frac{\partial}{\partial x^i} \qquad\qquad Y = Y^j \frac{\partial}{\partial x^j}$$

whence

$$\begin{aligned}\nabla_X(Y) &= Y^j \nabla_X\left(\frac{\partial}{\partial x^j}\right) + X(Y^j)\frac{\partial}{\partial x^j}\\ &= X^i Y^j \nabla_i\left(\frac{\partial}{\partial x^j}\right) + X^i \frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j}\\ &= X^i Y^j \Gamma_{ij}{}^k \frac{\partial}{\partial x^k} + X^i \frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j}\end{aligned}$$

Along the curve, $X^i = \frac{dx^i}{dt}$. Since Y is parallel along the curve,

$$0 = \left(\frac{dx^i}{dt} Y^j \Gamma_{ij}{}^k \frac{\partial}{\partial x^k} + \frac{dY^j}{dt}\frac{\partial}{\partial x^j}\right)(x^\ell)$$

for each x^ℓ; thus

$$\frac{dY^\ell}{dt} + \Gamma_{ij}{}^\ell \frac{dx^i}{dt} Y^j = 0 \qquad (79)$$

These equations could enable us to determine the parallel propagation of a vector $Y(t_0)$ along the whole length of the curve, for suppose the curve lies in a collection of charts (V_α,ϕ_α) where α ranges from 1 to r and t_0 lies within V_1. Then we could solve the equations in the coordinates ϕ_1 valid throughout V_1. In particular this gives us a solution at some one point of $V_1 \cap V_2$, so we can solve the equations in the system ϕ_2, using this one value to eliminate the arbitrary constants, and so on through the charts

Note that the above equation involves the vector fields X,Y only as regards their values on the curve. Thus in the equation

$$\nabla_X(Y) = 0$$

it is of no consequence which fields X,Y we choose, provided that X_P is the tangent vector to the curve at each point P.

If the curve is such that the tangent vector itself is parellely propagated, the curve is said to be a GEODESIC. The condition for this is that

$$\nabla_X(X) = 0 \qquad (80)$$

or in a coordinate patch,

$$\frac{d^2x^\ell}{dt^2} + \Gamma_{ij}{}^{\ell}\ \frac{dx^i}{dt}\frac{dx^j}{dt} = 0 \qquad (81)$$

A complete solution of equation (80), X(t), provides a MAXIMAL geodesic.

Let P, Q be two points, and H(t) a curve connecting P, Q. Then parellelism w.r.t. a given affine connection and the curve H provides an isomorphism between $\mathfrak{D}^1(P)$ and $\mathfrak{D}^1(Q)$, that is, a linear rule for selecting a tangent vector at Q given one at P, and vice versa. Since

$$\nabla_X(Y) = 0$$

possesses a unique solution, given $Y(t_0)$ (where $P = H(t_0)$), there is a (1,1) relationship between a tangent vector at P and the parallel vector at Q. Since the equivalent coordinate equations (A) are linear the solution is a linear function of the initial parameter, which are just the coordinate components $Y^i(t_0)$. Thus if P, Q lie in the same patch, the relationship between Y_P and Y_Q is linear, and the argument can readily be extended to the general case.

One important theorem (see Helgason (Ref. 15)) is that given an affine connection, a point P, and a non-zero tangent vector at

P, then there exists a unique maximal geodesic through P with the given vector as tangent.

2[illegible]. COVARIANT DERIVATIVES

We next show that covariant differentiation is a most natural definition of differentiation.

Suppose we are given vector fields X, Y. Let P be a point for which $X_P \neq 0$. Then, as we have seen, we can find a curve $H(t)$ in the neighbourhood of P with $X(t)$ as its tangent vector (in the coordinate component sense). Given an affine connection, then to every vector R of $\mathcal{D}^1(P)$ corresponds the parallel tangent vector at any point t of the curve, which we denote by

$$\tau_t(R) .$$

Each τ_t, as we have seen, is an isomorphism. Conversely, to each $Y(t)$ corresponds a vector

$$\tau_t^{-1}(Y(t))$$

of the tangent space of P.

One cannot define a derivative of Y as

$$\lim_{t \to 0} \frac{1}{t} (Y(t) - Y(0))$$

since $Y(t)$ and $Y(0)$ belong to two different vector spaces, and no meaning can be attached to adding or subtracting them. However, if $P = H(0)$

$$Y(0) \text{ and } \tau_t^{-1}(Y(t))$$

belong to the same vector space – a natural definition of a derivative would be

$$\lim_{t \to 0} \frac{1}{t}\left(\tau_t^{-1}(Y(t)) - Y(0)\right) .$$

We shall prove that this is in fact the covariant derivative of Y at the point P. The vector X comes in because τ_t is determined by a curve, which in turn is determined by X.

It is sufficient to consider the portion of the curve lying within the patch containing P.

Let T be some point on the curve. Take the vector $\tau_T^{-1}(Y(T))$ at P and propagate it parallely along the curve for all points up to T. Denote these vectors by

$$Z(t)$$

where $$Z(0) = \tau_T^{-1}(Y(T)) .$$

Then $$\frac{dZ^\ell}{dt} + \Gamma_{ij}{}^{\ell} \frac{dx^i}{dt} Z^j = 0 \tag{32}$$

Also, the parallel propagation of $Z(0)$ along the curve to T is just

$$\tau_T(Z(0))$$

i.e.

$$Y(T)$$

whence

$$Z(T) = Y(T) .$$

By the mean value theorem

$$Z^\ell(T) = Z^\ell(0) + T \frac{dZ^\ell}{dt}(u) \tag{33}$$

where

$$0 < u < T .$$

Now $$\frac{1}{T}\left(Z^{\ell}(T) - Z^{\ell}(0)\right) = \frac{dZ^{\ell}(u)}{dt}$$

$$= -\Gamma^{\ell}_{ij}\ \frac{dx^{i}(u)}{dt}\ Z^{j}(u)$$

We have

$$\tau_T^{-1}(Y(T)) = Z(0), \quad \text{whence}$$

$$\frac{1}{T}\left(\tau_T^{-1}(Y(T)) - Y(0)\right) = \frac{1}{T}\left(Z(0) - Y(0)\right).$$

Taking ℓ^{th} components,

$$= \frac{1}{T}\left(Z^{\ell}(0) - Z^{\ell}(T)\right) + \frac{1}{T}\left(Z^{\ell}(T) - Y^{\ell}(0)\right)$$

$$= \Gamma_{ij}{}^{\ell}\ \frac{dx^{i}(u)}{dt}\ Z^{j}(u) + \frac{1}{T}\left(Y^{\ell}(T) - Y^{\ell}(0)\right)$$

$$= \Gamma_{ij}{}^{\ell}\ \frac{dx^{i}(T)}{dt}\ Z^{j}(T) + \frac{dY^{\ell}(0)}{dt} + 0(T)$$

whence in the limit

$$\lim_{t\to 0}\frac{1}{t}\left[\left(\tau_t^{-1}(Y(t)) - Y(0)\right)\right]^{\ell} = \left.\frac{dY^{\ell}}{dt} + \Gamma_{ij}{}^{\ell}\ \frac{dx^{i}}{dt}\ Y^{j}\right]_P ,$$

since $Y^{j}(T) = Z^{j}(T)$.

Thus $$\lim_{t\to 0}\frac{1}{t}\left(\tau_t^{-1}(Y(t)) - Y(0)\right) = \left[\nabla_X(Y)\right]_P \qquad (84)$$

We may now extend the definition of ∇_X to operate on all vector and tensor fields.

If P, Q are two points, and $H(t)$ is a curve joining them, let τ denote parallel translation of a tangent vector at P. If ω denotes a member of the dual space $\mathcal{D}_1(P)$ we shall <u>define</u> the element of $\mathcal{D}_1(Q)$: $\tau(\omega)$ by

$$(\tau(\omega))\ (X_Q) = (\omega)(\tau^{-1}(X_Q)) \tag{85}$$

for all tangent vectors X_Q at Q.

This means that a parallely translated dual vector, operating on a tangent vector, is equivalent to the dual vector, operating on the parallely translated tangent vector.

Generally, a tensor field of type $\binom{r}{s}$ is a multilinear function of r contravariant vector-fields and s covariant vector-fields. If T is such a tensor field, then the parallel translation of T_P will be defined by

$$(\tau(T_P))\left(X_Q^{(1)},\dots, X_Q^{(r)}, \omega_Q^{(1)},\dots, \omega_Q^{(s)}\right)$$
$$= T_P\left(\tau^{-1}(X_Q^{(1)}),\dots, \tau^{-1}(X_Q^{(r)}, \tau^{-1}(\omega_Q^{(1)}),\dots \tau^{-1}(\omega_Q^{(s)})\right) \tag{86}$$

Then, by analogy with the foregoing, we <u>define</u>

$$\Big[\nabla_X(T)\Big]_P = \lim_{t\to 0} \frac{1}{t}\left(\tau_t^{-1}(T(t)) - T(0)\right) \tag{87}$$

with the notation as before.

In particular, if F is a scalar (C^∞ function) then the "parallel translation" of $F(0)$ can only mean $F(t)$, so we write

$$\Big[\nabla_X(F)\Big]_P = \lim_{t\to 0}\frac{1}{t}\Big(F(t) - F(0)\Big)$$
$$= \frac{dF}{dt}\Big]_P \quad .$$

But $X_P = \frac{d}{dt}\Big]_P$, since the curve is tangent to the tangent vector.

Thus

$$\nabla_X(F) = X(F) \tag{88}$$

a definition suggested previously.

We now show that these definitions give the normal covariant derivative component formulae.

If Y is any contravariant vector-field, we have shown that

$$\left[\nabla_X(Y)\right]_P = \lim_{t\to o} \frac{1}{t}\left(\tau_t^{-1}(Y(t)) - Y(0)\right).$$

Thus (if L.H.S. is finite)

$$\tau_t^{-1}(Y(t)) = Y(0) + O(t)$$

and so, equally

$$Y(t) = \tau_t(Y(0)) + O(t).$$

This holds for any such vector field: in particular for $\nabla_X(Y)$.

Now

$$\tau_t(Y(0)) = Y(t) - t\,\tau_t\left[\nabla_X(Y)\right]_P + O(t^2)$$

$$= Y(t) - t\left[\nabla_X(Y)\right]_t + O(t^2)$$

which shows that

$$\left[\nabla_X(Y)\right]_P = \lim_{t\to o} \frac{1}{t}\left(Y(t) - \tau_t Y(0)\right) \qquad (89)$$

The argument above breaks down if $\Gamma_{ij}{}^k$ are not continuous along the curve; however, we could still establish the result from first principles.

Let ω be a covariant tensor-field.

Then

$$\left[\nabla_X(\omega)\right]_P(Y)_P = \lim_{t\to o} \frac{1}{t}\left(\tau_t^{-1}(\omega(Y(t)) - \omega_P(Y_P)\right)$$

$$= \lim_{t\to o} \frac{1}{t}\left(\omega(\tau_t(Y_P)) - \omega_P(Y_P)\right) \qquad (90)$$

by the definition of parallel translation of covariant vectors.

Now

$$\tau_t(Y_P) = Y(t) - t\left[\nabla_X(Y)\right]_P + O(t^2)$$

as we have seen:

$$= \left[Y^k(t) - t(\dot{Y}^k(0) + \Gamma_{ij}{}^k \dot{x}^i Y^j(0)) + O(t^2)\right]\left(\frac{\partial}{\partial x^k}\right)_t$$

since ω is linear, and $\omega\left(\frac{\partial}{\partial x^k}\right) = \omega_k$, where $\omega_k \, dx^k = \omega$, we have

$$\omega(\tau_t(Y_P)) = Y^k\omega_k - t\,\dot{Y}_P^k\,\omega_k - t\omega_k\,\Gamma_{ij}{}^k \dot{x}_P^i\, Y_P^j + O(t^2)$$

Thus

$$\omega(\tau_t(Y_P)) - \omega_P(Y_P) = (Y^k\omega_k)_t - (Y^k\omega_k)_P - t\,\dot{Y}_P^k\,\omega_k(t) -$$

$$- t\,\omega_k(t)(\Gamma_{ij}{}^k \dot{x}^i Y^j)_P$$

proceeding to the limit

$$\left[\nabla_X(\omega)\right]_P(Y)_P = \frac{d}{dt}(Y^k\omega_k) - \frac{dY^k}{dt}\omega_k - \Gamma_{ij}{}^k \frac{dx^i}{dt} Y^j \omega_k$$

$$= Y^k \frac{d\omega_k}{dt} - \Gamma_{ij}{}^k \frac{dx^i}{dt} Y^j \omega_k$$

Thus

$$\left(\left[\nabla_X(\omega)\right]_P\right)_j = \frac{d\omega_j}{dt} - \Gamma_{ij}{}^k \frac{dx^i}{dt}\omega_k \qquad (91)$$

A similar argument using multilinear instead of linear functions proves the usual result for tensor components.

This treatment in terms of parallel displacements is due to Chevalley (Ref. 6)). An alternative treatment of the concept of affine connection is that using the concept of "frame bundles" (see e.g. Nomizu (Ref. 34)): the approach is certainly elegant, but requires a great deal of preparatory development.

23. THE RIEMANNIAN MANIFOLD

Thus far we have developed the concept of the CONNECTED Differentiable Manifold: the next step represents a further specialisation.

A Riemannian structure is a scalar valued symmetric second rank tensor which is nowhere degenerate (i.e. for all points P, $\{\forall Y, g_P(X,Y) = 0\} \Rightarrow X_P = 0$). A further requirement normally is that the scalar field g is C^∞ continuous.

The Riemannian Connection is the unique connection which preserves g on parallel displacement, and for which the TORSION TENSOR is zero. In Levi-Civita notation, we require

$$\Gamma_{ij}{}^{k} - \Gamma_{ji}{}^{k} = 0 \; ; \qquad (92)$$

$$g_{ij;k} = 0. \qquad (93)$$

How do we establish these requirements in the tensor-field notation? Firstly, it might be thought that (92) could be written

$$\nabla_X Y - \nabla_Y X = 0 \; . \qquad (94)$$

However, the Liebniz property necessitates a complication: writing $X = \xi^i \frac{\partial}{\partial x_i}$, $Y = \eta^j \frac{\partial}{\partial x_j}$ we have

$$\nabla_X Y - \nabla_Y X = \xi^i \eta^j \nabla_i \frac{\partial}{\partial x^j} - \eta^j \xi^i \nabla_j \frac{\partial}{\partial x^i} + \xi^i \frac{\partial \eta^j}{\partial x^i} \frac{\partial}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^j} \frac{\partial}{\partial x^i}$$

$$= \xi^i \eta^j \left(\nabla_i \frac{\partial}{\partial x^j} - \nabla_j \frac{\partial}{\partial x^i}\right) + XY - YX$$

We are therefore led to introduce the Torsion Tensor $T(X,Y)$, defined by:

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] \tag{95}$$

where

$$[X,Y] = XY - YX \quad . \tag{96}$$

The alternating operator product $[X,Y]$ satisfies the Liebniz property Y : we have, for all scalar fields F,

$$\begin{aligned} [X,FY] &= X(F)Y + FXY - FYX \\ &= F[X,Y] + X(F)Y \quad . \end{aligned} \tag{97}$$

The product itself defines an operator known as the LIE DERIVATIVE:

$$\begin{aligned} \mathscr{L}(X)Y &= [X,Y] \\ &= -\mathscr{L}(Y)X \quad . \end{aligned} \tag{98}$$

θ is a map $\mathscr{D}^1 \to \mathscr{D}^1$ and might appear to satisfy the axioms for an affine connection. However, (X) is also Liebniz in X, and not linear as required for an affine connection.

The tensor character of the mapping T is established by demonstrating its linearity:

$$\begin{aligned} T(X,FY) &= F\nabla_X Y + X(F)Y - [X,FY] - F\nabla_Y X \\ &= F(\nabla_X Y - \nabla_Y X) - F[X,Y] \\ &= F\,T(X,Y) \quad . \end{aligned} \tag{99}$$

We have

$$\begin{aligned} T\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) &= \nabla_i \frac{\partial}{\partial x^j} - \nabla_j \frac{\partial}{\partial x^i} + \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] \\ &= \left(\Gamma_{ij}{}^k - \Gamma^k{}_{ji}\right)\frac{\partial}{\partial x^k} \end{aligned} \tag{100}$$

which identifies T with the classical torsion tensor.

Having assigned a Riemannian structure g, there is, as we have noted, a unique affine connection which preserves g on parallel displacement, and for which the torsion tensor vanishes. In component notation, the familiar result is that $\Gamma^k{}_{ij}$ is given by

$$\Gamma^k{}_{ij} = \tfrac{1}{2} g^{k\ell}(g_{\ell j,i} + g_{i\ell,j} - g_{ij,\ell}) \ . \qquad (101)$$

We shall endeavour to rewrite this equation in the notation of covariant differentiation. We have

$$2g_{mk}\Gamma^k{}_{ij} = \frac{\partial}{\partial x^i} g_{mj} + \frac{\partial}{\partial x^j} g_{mi} - \frac{\partial}{\partial x^m} g_{ij}$$

and the left-hand side may be written: $2g\left(\frac{\partial}{\partial x^m}, \nabla_i \frac{\partial}{\partial x^j} \right)$.

We are led to try for a relationship of the form

$$2g(X,\nabla_Z Y) = Zg(X,Y) + Yg(X,Z) - Xg(Y,Z) \qquad (102)$$

but again the Liebniz property creates complications. The tensor version of (101) is in fact

$$\begin{aligned} 2g(X,\nabla_Z Y) = {} & Zg(X,Y) + g(X,[Z,Y]) \\ & + Yg(X,Z) + g(Y,[X,Z]) \\ & - Xg(Y,Z) - g(X[Y,Z]). \end{aligned} \qquad (103)$$

Interchanging Y,Z one finds

$$\begin{aligned} 2g(X,\nabla_Z Y - \nabla_Y Z) &= -X(g(Y,Z) - g(Z,Y)) - 2g(X,[Y,Z]) \\ &= 2g(X,[Z,Y]), \end{aligned} \qquad (104)$$

since g is symmetric. Since this relationship holds for all X,

$$\nabla_Y Z - \nabla_Z Y = [Y,Z] \qquad (105)$$

and hence the torsion tensor vanishes.

We also have to establish that $\nabla_Z g = 0$. Now the covariant d derivative of a tensor may be defined by equation (87), and by analogy with the elementary theorem on the product of two functions, the generalised Leibniz property may be established:

$$\nabla_Z\{T(X,Y,\ldots)\} = (\nabla_Z T)(X,Y,\ldots) + T(\nabla_Z X,Y,\ldots) + T(X,\nabla_Z Y,\ldots) + \ldots \qquad (106)$$

Alternatively, this equation could be used to <u>define</u> $\nabla_Z T$. A particular case is that when T is a 1-form ω :

$$\nabla_Z\{\omega(Y)\} = (\nabla_Z \omega)(Y) + \omega(\nabla_Z Y)$$

so that

$$(\nabla_Z \omega)(Y) = Z\omega(Y) - \omega(\nabla_Z Y). \qquad (107)$$

One deduces that in particular

$$(\nabla_k dx^i)\left(\frac{\partial}{\partial x^j}\right) = -dx^i\left(\nabla_k \frac{\partial}{\partial x^j}\right)$$

$$= -\Gamma^i{}_{kj}$$

and so

$$\nabla_k dx^i = -\Gamma^i{}_{kj}\, dx^j , \qquad (108)$$

which leads to the familiar result on the derivative of a covariant vector field in Levi-Civita notation.

Returning to the question of the Riemannian Affine Connection, we have

$$\nabla_Z\{g(X,Y)\} = (\nabla_Z g)(X,Y) + g(\nabla_Z X,Y) + g(X,\nabla_Z Y). \qquad (109)$$

Now, by (103),

$$2g(X,\nabla_Z Y) + 2g(Y,\nabla_Z X) = 2Zg(X,Y)$$

all other terms cancelling. But

$$\nabla_Z\{g(X,Y)\} = Zg(X,Y)$$

since $g(X,Y)$ is a scalar field. Thus $\nabla_Z g = 0$.

The argument may also be inverted, which proves that there is one and only one affine connection satisfying the two properties simultaneously.

Since $g\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right) = g_{ij}$, we have

$$g = g_{ij}dx^i \otimes dx^j. \tag{110}$$

If $\delta\underline{s}$ is a tangent vector given by

$$\delta\underline{s} = \delta x^i \frac{\partial}{\partial x^i}$$

where $\delta x^i = x^{i\prime} - x^i$, and $x^{i\prime}, x^i$ are coordinates of neighbouring points on a curve, we have

$$g(\delta\underline{s},\delta\underline{s}) = g_{ij}\delta x^i\delta x^j \tag{111}$$

in accordance with the Levi-Civita approach.

Since g is a symmetric tensor field, at any point $\underline{P}$ we may choose a basis $\underline{e}_i$ of the tangent space at P such that the matrix g_{ij} takes the diagonal form with ± 1 along the leading diagonal. In General Relativity we are concerned with 'Normal Hyperbolic' Riemannian Space, where

$$g_P(\underline{e}_i,\underline{e}_j) = \eta_{ij} \,, \tag{112}$$

where η_{ij} is a diagonal matrix with terms (1,-1,-1,-1).

Normally it is not possible to choose a coordinate system such that

$$g_P\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \eta_{ij}$$

for all P. If it is possible, we have achieved a further specialisation — Minkowski (flat) space. In terms of this "preferred" coordinate system, we find

$$\nabla_i \frac{\partial}{\partial x^j} = 0 \qquad (113)$$

and so the tangent vector $\frac{\partial}{\partial x^j}$ at a point P is parallely transported to $\frac{\partial}{\partial x^j}$, at Q. Thus for Minkowski Space, the preferred coordinate system generates a (1,1) mapping of all the tangent spaces onto one another. It is simple to show that the relationship (112) is invariant under the affine group of coordinate transformations (generating "quasi-Cartesian coordinates"), and that the mappings of the tangent spaces are preserved under this group. The manifold is thus a "copy" of its unique tangent space, and can be regarded as a vector space with scalar product.

A similar argument holds when g is positive definite: the manifold is then merely a copy of Euclidean Space.

24. THE RIEMANN TENSOR

Just as the third rank Torsion Tensor was defined by

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$

we can similarly define a fourth rank tensor — the Riemann Tensor — as follows:

$$R(X,Y) = \nabla_X\nabla_Y - \nabla_Y\nabla_X - \nabla_{[X,Y]}. \qquad (114)$$

Linearity is easily established. It should be noted that R maps tangent vectors into tangent vectors, the mapping being determined by pairs of tangent vectors (X,Y). The standard components of R are given by

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^\ell} = R^k_{\ell ij}\frac{\partial}{\partial x^k}. \qquad (115)$$

Since $\nabla_{\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right]} = 0$ trivially,

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \nabla_i\nabla_j - \nabla_j\nabla_i \qquad (116)$$

which is another way of writing the familiar component result:

$$\xi_{\ell;[i,j]} = R^k_{\ell ij}\xi_k \qquad (117)$$

where the semicolon means covariant differentiation. The standard formula for $R^k_{\ell ij}$ in terms of derivatives of the Γ^k_{ij} follows immediately.

In Minkowski Space, we can find a coordinate system for which $\nabla_i\, {}^{\partial}/\partial x^j = 0$. It follows that

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} = \nabla_i\nabla_j\frac{\partial}{\partial x^k} - \nabla_j\nabla_i\frac{\partial}{\partial x^k}$$

$$= 0$$

since $\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0$.

Hence $R(X,Y) = 0$. Conversely, it may be shown that if the Riemann Tensor vanishes, then there exists a coordinate system for which (118) holds.

25. [illegible]

The modern theory of Differential Geometry has now been developed sufficiently for our purposes — all the normal "component" results of General Relativity may now be interpreted in terms of the above framework.

However, the assumption that our vector and scalar fields are all C^∞ is too restrictive for our purposes, and we must check to what extent this assumption can be weakened.

Basically, we wish to permit surfaces of discontinuity, in order to give meaning to jump conditions. We may suppose that our functions are C^∞ except on the points of a submanifold (hypersurface) given by $S(P) = 0$, where S is a C^∞ function (without discontinuity) and in the chart (V,ϕ) the Jacobian of S is non-vanishing.

Now if a point P lies outside the submanifold we may deduce by means of the Mean Value Theorem that there exists a point Q such that the points between P and Q all lie outside the sub-manifold. The proof that the derivations of the functions form a module spanned by the operators $\partial/\partial x^i$ remains unchanged, provided that the functions are defined on all points not inside the submanifold.

If furthermore the submanifold divides the manifold into 2 disjunct regions, which we shall label (+) and (−), and if the partial derivatives of a particular function f tend to finite limits on each side of the submanifold, then we may define

$$X_p^+ f = \left(\frac{\partial f}{\partial x^i}\right)_p^+ X_p(x^i) \; ; \quad X_p^- f = \left(\frac{\partial f}{\partial x^i}\right)_p^- X_p(x^i) \qquad (113)$$

where P lies in the submanifold. We have assumed that X is a derivation of the subalgebra of the precisely C^∞ functions, and so $X_p(x^i)$ is defined. If our class of functions is restricted to those which are continuous everywhere except on this submanifold, then we will have "left" and "right" tangent spaces on the submanifold.

Whilst this approach does not seem to have been pursued in the literature, it would appear that our treatment of Differential Geometry would support such a weakening of the basic assumptions. A tensor field would be defined on the "left" or "right" tangent space at points on the submanifold, and the component functions would have "left" and "right" values.

Thus it would seem that there is no mathematical objection to the Synge-O'Brien formulation of the boundary conditions problem. It must however be assumed that the functions g_{ij} in the particular coordinate patch used by the authors are such that all derivatives tend to a limit on each side of the interface.

Indeed, the more restrictive Lichnerowicz conditions appear to be no more soundly based. Further discussion of this point is reserved for our concluding Chapter, when we will also consider the question of the ultimate "fate" of a collapsing body in the light of this mathematical exposition.

CHAPTER VI

CONCLUSIONS

In this final chapter, we proceed to conclusions regarding our two main problems and develop the concept of "time orientation"; finally we review the principal results which have been obtained.

1. JUMP CONDITIONS

Many problems in Field Theory involve regions with notional sharp boundaries. Various Field Equations are used, admitting families of solutions: the use of junction conditions enables the appropriate solution to be selected. There have been various attempts to derive meaningful junction conditions in General Relativity, but these are beset with difficulties, mainly because of requirements of coordinate invariance.

A notable attempt at obtaining junction conditions is that of Synge & O'Brien. Their method involves the use of a coordinate system for which the (local) equation of the boundary takes the form $x^1 = 0$. Unfortunately, the method involves a number of assumptions which may not be satisfied in the particular coordinate patch selected. For example, although the use of comoving coordinates seems logical, the dust fluid solutions in comoving form cannot satisfy the conditions. In fact, g_{rr} itself must be discontinuous. A continuous metric tensor may be obtained by the use of Schwarzschild coordinates, or else in the shape of the non-diagonal form obtained by use of the invariant coordinates (R,t) — but there are no suitable conditions on the first derivatives.

The Lichnerowicz conditions are even more restrictive, and are clearly of no value in themselves in selecting the appropriate solution. Furthermore, as demonstrated in the previous Chapter, they cannot be said to be an absolute requirement mathematically. On the other hand, it may well be possible in certain cases to demonstrate the existence of a coordinate system under which these conditions are satisfied.

In a problem involving spherical symmetry with a spherical shell as a boundary, it is clear that on the boundary surface the curvature of the symmetry surfaces must be the same whether the boundary surface is treated as being imbedded in the interior or the exterior region. The comoving time coordinate may be defined as proper time along the orthogonal trajectories of the symmetry surfaces (and this definition will hold generally, and not merely for our dust fluid model). It follows that the coordinates (t, θ, ϕ) specify a point on the boundary surface independent of which imbedding is under consideration, provided that the origin for t is chosen suitably.

Accordingly, we may equate the functional form — $R(t)$ — of the space-section curvature on both sides of the boundary. This critical condition contains both an invariant condition and a coordinate restriction. The coordinate restriction arises from the choice of origin for the t parameter.

In the case of the dust-fluid, the boundary surface is of free-fall elliptic, parabolic or hyperbolic type (see Chapter II). In order that the t-parameters should match, the appropriate type of function $R(t)$ must be used in the exterior, and the values of the functions $p(r)$

(used for the elliptic case) and therefore $a(r)$ must equate on either side of the boundary. The condition $R(t)\big|_+ = R(t)\big|_-$ then provides the physical condition $M = M_{T_V}$ which selects the correct exterior solution.

In such problems involving spherical symmetry, this is the sole boundary condition of any value. Can this approach be adapted to other problems? The difficulty is to find the "preferred" coordinate curves that correspond to the (t, θ, ϕ) curves in the spherically symmetric cases.

The solution is straightforward. These curves are in fact the "lines of curvature" of the hypersurface (see for example Willmore (Ref.53)) These are generated by the eigenvectors of the Second Fundamental Form (sometimes known as the Third Fundamental Form)

$$c = c_{\alpha\beta} du^{\alpha} \otimes du^{b}$$

where

$$c_{\alpha\beta} = g_{ij} N^{i}{}_{;\alpha} N^{j}{}_{;\beta} \qquad (1)$$

and $\{u^{\alpha}\}$ $(\alpha = 1,2,3)$ is a system of coordinates in some coordinate patch on the hypersurface, $\{N^{i}\}$ being the unit normal to the hypersurface. The eigenvalues of c are the "principal curvatures" of the hypersurface.

There is clearly a physical condition here. The principal curvatures determine the curvature of curves which lie within the hypersurface and so must be independent of the two imbeddings under consideration (exterior and interior regions). Hence the Second Fundamental Form must be independent of imbedding. Another requirement is that the induced metric, or First Fundamental Form,

$$a = a_{\alpha\beta} du^{\alpha} \otimes du^{\beta}$$

where

$$a_{\alpha\beta} = g_{ij} \frac{\partial x^i}{\partial u^{\alpha}} \frac{\partial x^j}{\partial u^{\beta}} \qquad (2)$$

should be independent of the imbedding.

These conditions are not straightforward in their application. It is necessary to identify points on the hypersurface under the two imbeddings: the equations of the surface may take the forms

$$x^i = f^i(u^{\alpha}); \quad x^i = f'^i(u'^{\alpha})$$

and so attention must be paid to the functional forms f^i, f'^i. In the collapse problem under consideration we may set

$$u^1 = t, \quad u^2 = \theta, \quad u^3 = \phi$$

but we must check that the coordinate t has an identical meaning under the two imbeddings. Thus the initial conditions are paramount, and the parameter p must equate on the two sides (elliptic case).

Consider our dust-fluid model and the empty exterior region. The First Fundamental Form of the boundary surface is

$$a = dt^2 - R^2(t)(d\theta^2 + \sin^2\theta d\phi^2) \qquad (3)$$

where, as is the convention, dt^2 is a shorthand for $dt \otimes dt$, and similarly $d\theta^2, d\phi^2$. Inspection of (3) provides the condition $R(t)\big|_+ = R(t)\big|_-$ previously discussed.

The requirement

$$c^+_{\alpha\beta} = c^-_{\alpha\beta}$$

should now be satisfied identically. To check this, we calculate $c_{\alpha\beta}$

from first principles.

In terms of our comoving coordinates (t, r, θ, ϕ) in that order

$$N^i = N(0,1,0,0) \qquad (4)$$

since the metric is diagonal, and the condition on the hypersurface

$$N_i dx^i = 0$$

reduces simply to

$$N_t dt + N_\theta d\theta + N_\phi d\phi = 0$$

where dt, $d\theta$, $d\phi$ are arbitrary. Since N^i is the unit normal, we have

$$g_{ij} N^i N^j = -1$$

and so

$$N = |g_{rr}|^{-\frac{1}{2}} \quad . \qquad (5)$$

Writing

$$N^i = N\delta^i_r$$

where r denotes the particular coordinate, we have

$$N^i{}_{;\alpha} N^j{}_{;\beta} = (N_{,\alpha}\delta^i_r + \Gamma^i_{\alpha\lambda} N\delta^\lambda_r)(N_{,\beta}\delta^j_r + \Gamma^j_{\beta\mu} N\delta^\mu_r)$$

$$= (N_{,\alpha}\delta^i_r + N\Gamma^i_{\alpha r})(N_{,\beta}\delta^j_r + N\Gamma^j_{\beta r}) \ .$$

Thus

$$g_{ij} N^i{}_{;\alpha} N^j{}_{;b} = g_{rr} N_{,\alpha} N_{,\beta} + NN_{,\alpha} g_{rj}\Gamma^j_{\beta r} + NN_{,\beta} g_{ri}\Gamma^i_{\alpha r} + N^2 g_{ij}\Gamma^i_{\alpha r}\Gamma^j_{\beta r}$$

$$= g_{rr}(N_{,\alpha} N_{,\beta} + NN_{,\alpha}\Gamma^r_{\beta r} + NN_{,\beta}\Gamma^r_{\alpha r} + N^2\Gamma^r_{\alpha r}\Gamma^r_{\beta r})$$

$$+ N^2(g_{tt}\Gamma^t_{\alpha r}\Gamma^t_{\beta r} + g_{\theta\theta}\Gamma^\theta_{\alpha r}\Gamma^\theta_{\beta r} + g_{\phi\phi}\Gamma^\phi_{\alpha r}\Gamma^\phi_{\beta r})$$

Now

$$\Gamma^{r}_{\alpha r} = [\log|g_{rr}|^{\frac{1}{2}}]_{,\alpha}$$

$$= \frac{-N_{,\alpha}}{N}$$

using (5)

Thus

$$N_{,\alpha}N_{,\beta} + NN_{,\alpha}\Gamma^{r}_{\beta r} + NN_{,\beta}\Gamma^{r}_{\alpha r} + N^2\Gamma^{r}_{\alpha r}\Gamma^{r}_{\beta r} = 0$$

Furthermore,

$$\Gamma^{t}_{tr} = 0$$

since g_{tt} is constant.

Hence

$$c_{\alpha\beta} = N^2(g_{\theta\theta}\Gamma^{\theta}_{\alpha r}\Gamma^{\theta}_{\beta r} + g_{\phi\phi}\Gamma^{\phi}_{\alpha r}\Gamma^{\phi}_{\beta r})$$

and so

$$c_{\theta\theta} = N^2 g_{\theta\theta}(\Gamma^{\theta}_{\theta r})^2; \quad c_{\phi\phi} = N^2 g_{\phi\phi}(\Gamma^{\phi}_{\phi r})^2;$$

all other components vanishing. Substituting for $g_{\theta\theta}, g_{\phi\phi}$,

$$c_{\theta\theta} = N^2 R_r^2 \; ; \quad c_{\phi\phi} = N^2 \sin^2\theta R_r^2 \quad . \tag{6}$$

Now, according to (5),

$$N^2 = g_{rr}^{-1}$$

$$= k/R_r^2$$

where k is the value of $(1 - p\alpha r^2)$ on the boundary in the elliptic case (or the value of the corresponding expression in the hyperbolic case, or unity for the parabolic case).

Thus

$$c_{\theta\theta} = k \; ; \quad c_{\phi\phi} = \sin^{-2}\theta c_{\theta\theta}$$

all other components of $c_{\alpha\beta}$ vanishing.

The principal curvatures of the boundary hypersurface are given by

$$\det|c_{\alpha\beta} - \kappa a_{\alpha\beta}| = 0 \tag{8}$$

and are therefore 0, k/R and k/R. Clearly these eigenvalues are independent of the imbedding, provided that R and k equate on either side of the boundary.

The conditions $a^{+} = a^{-}$; $c^{+} = c^{-}$ thus enable the correct matching to be obtained at the interface between two dust fluid regions of differing ρ (in other words, involving a jump discontinuity in the Riemann tensor). In particular the correct matching is obtained when one region consists of empty space.

We may conclude that the geometrical conditions on the matching of the first and second fundamental forms (arising from the two imbeddings) represent a physical restriction: in the dust fluid model under consideration this is equivalent to the equating of gravitational mass on both sides of the boundary.

Although it is not clear that the restriction has the same interpretation in the more general case of the perfect fluid, this seems highly probable.

The conditions on the first and second fundamental forms seem of general application, and are quite independent of any coordinate system employed. We may adopt these conditions as a fundamental postulate of General Relativity, along with the Field Equations.

If this approach should be generally accepted, the coordinate conditions of Synge O'Brien and Lichnerowicz will be seen to be unnecessary. Furthermore, the particular researches of Nariai and Tomita into pressure-free collapse will be seen to have started from false premises, and their conclusions will seem of doubtful significance.

2. CATASTROPHIC COLLAPSE

In Chapter IV we discussed the "fate" of a collapsing body, and noted that the description of the motion was highly sensitive to the reference frame. In a frame associated with distant astronomical observation the body will appear to collapse asymptotically to the gravitational radius, although it will effectively disappear due to rapidly increasing red-shift long before this is reached.

In a frame which is comoving with the body, or is effectively trapped within its gravitational field, a quasi-Newtonian behaviour will in principle be observed, in that the collapse will accelerate and proceed towards a point-event singularity which will act as a kind of vortex (speaking somewhat inaccurately). Ultimately everything inside the light-horizon of this frame will be sucked inexorably into the singularity.

The question naturally arises, what next? It may be objected that the singular event is itself incredible, and the concept simply indicates that we have no acceptable physical theory regarding the behaviour of matter at extreme densities. However, an examination of the exterior region, which is apparently trapped, may be more realistic. If collapse does not proceed to the singularity it may perhaps be reversed at some stage.

Discussion of these matters must be concerned with possible topological structures of the manifold under consideration; in particular, with possible connectivities of Space Time. A full treatment of this subject is outside the scope of this Thesis, and we shall confine ourselves to a few aspects. The examination of the foundations of Differential Geometry conducted in the previous Chapter was a necessary preliminary.

Mention at this stage must be made of the work of Wheeler ~~(ref.~~).
It is however only marginally relevant to the problem under discussion.

Most treatments of the collapse problem assume by implication that the manifold under consideration is of Euclidean topology. That is, the manifold is globally homeomorphic to 4-dimensional Euclidean space. Although this is perhaps the simplest assumption as to the neighbourhood structure of the space, there is no bound to the number of alternative possibilities.

For example, the manifold might be topologically equivalent to a subspace of a higher dimensional Euclidean Space. Alternatively, independent of any such imbedding, it may be equivalent to a region of Euclidean 4-space for which certain sets of points have been identified with each other. For example, if a rectangular region in Euclidean 2-space has 2 parallel edges identified, the resultant 2-space is equivalent to a cylinder.

One such identification is the "Antipodal Map". In Euclidean 3-space, consider the region exterior to a certain sphere, and suppose that points on opposite sides of a diameter are identified. The resultant space clearly does not possess the Euclidean topology, since any sphere concentric with the "missing" sphere cannot be shrunk down to a point.

This concept may be of value in General Relativity, but its application in 4-dimensional space requires some care.

It is necessary to map two sets of <u>events</u>, not merely spatial points. In the case of "elliptic" collapse it may seem reasonable to map points on the null surface $R = 2GM$ so as to identify events (t, θ, ϕ) and $(t + \lambda, \theta + \pi, \phi)$, where λ is the "rebound time", equal on the boundary to π/α. The effect of this is that an inward falling observer, on passing through the "Schwarzschild Radius", will then be flung "outwards", though he will detect no sudden change in velocity. He would however observe a reversal in his acceleration and that neighbouring geodesics were now diverging instead of converging.

Such sharp changes are outside our normal experience. However, collapse to a point singularity poses even more serious difficulties. It seems reasonable to seek methods by which the presence of "catastrophe" type singularities could be avoided in our model and the antipodal map is one such method.

There is reason to suppose that the space-time region in which the observer would emerge under the antipodal map hypothesis is essentially disjunct from the region in which collapse took place. In the normal course, if an observer falls towards the Schwarzschild hypersurface and returns without having intersected the "singularity", his proper time will have been dilated as compared to that of an observer with constant R. The effect will increase with the closeness of approach to the critical hypersurface. This is demonstrated by the analysis of the apparent red shift of any light emitted or reflected by the first observer and received by the second one (considered in Chapter IV).

However, these considerations do not conclusively establish that the regions of space-time would be disjunct: although it seems probable that the contrary hypothesis would imply a violation of causality (in Chapter IV we demonstrated that an observer at constant R could never observe collapse to the Schwarzschild hypersurface in finite proper time), the concept of the antipodal map seems to require a thorough investigation — an investigation which is outside the scope of the present Thesis.

3. TIME ORIENTATION

Mention of the causality question brings us to the final topic to be considered.

Birkhoff's Theorem (cf. Chapter III) states that an empty spherically symmetric space-time must be static. Certainly it can be proved that such a manifold must be locally isometric to the Schwarzschild manifold, but this in fact is static only within the region labelled $R > 2GM$. The region $R \leqslant 2GM$ is not static — one cannot find a time coordinate T and space coordinates $\{x^\alpha\}$ such that the metric coefficients in this system of coordinates are functionally independent of T. Using the more sophisticated language (cf. Chapter III), one cannot find a 1-parameter group of motions for which the trajectories are hypersurface-orthogonal and timelike. The differential equations will still admit a solution, but in the region $R < 2GM$ the trajectories are spacelike.

In the physical sense, the interior region, being non-static, is time-directed. Either catastrophic collapse is taking place or else catastrophic explosion.

The "collapse" space-time, and the "explosion" or "anticollapse" space-time may be regarded as distinct time-orientated manifolds.

As we have seen in Chapter V, an orientated manifold is one which admits two disjunct equivalence classes of coordinate systems, the equivalence relation being a positive Jacobian. A time-orientated manifold also admits two disjunct equivalence classes of coordinate systems, the equivalence relation being that for all time-like curves,

$$\frac{d\eta/ds}{|d\eta/ds|} = \frac{d\bar{\eta}/ds}{|d\bar{\eta}/ds|} \qquad (10)$$

where $\eta, \bar{\eta}$ are corresponding time-coordinates (η is a time coordinate if and only if $g(d\eta, d\eta) > 0$ — i.e. $g^{\eta\eta} > 0$). This equivalence relation implies that the future-pointing sense of timelike curves is preserved.

The following Proposition is central to the present discussion: the two metrics representing collapse and anticollapse in the portion $R < 2GM$ of the exterior Schwarzschild manifold cannot be transformed from one to another by means of a time-orientation preserving coordinate transformation.

Consider a spherically symmetric metric of the form

$$ds^2 = A(\xi, \eta)d\eta^2 + 2B(\xi, \eta)d\eta d\xi + C(\xi, \eta)d\xi^2 + F(\xi, \eta)d\Omega^2 \qquad (11)$$

where $B^2 > AC$ and $C < 0$, $F < 0$. Then $g^{\eta\eta} > 0$, and so η is a time coordinate.

Let $\frac{d}{ds}$ be the tangent vector to a time-like radial curve. In terms of its components with respect to the coordinate system (ξ, η), we have

$$A\left(\frac{d\eta}{ds}\right)^2 + 2B\frac{d\eta}{ds}\frac{d\xi}{ds} + C\left(\frac{d\xi}{ds}\right)^2 = 1 \qquad (12)$$

Let $\alpha, \beta, (\alpha > \beta)$ be the roots of the equation

$$Cx^2 + 2Bx + A = 0.$$

Then $$C\left(\frac{d\xi}{ds} - \alpha\frac{d\eta}{ds}\right)\left(\frac{d\xi}{ds} - \beta\frac{d\eta}{ds}\right) = 1, \text{ and therefore } > 0. \qquad (13)$$

We deduce that if $\frac{d\eta}{ds} > 0$,

$$\beta\frac{d\eta}{ds} < \frac{d\xi}{ds} < \alpha\frac{d\eta}{ds} \qquad (14a)$$

whilst if $\frac{d\eta}{ds} < 0$

$$\alpha\frac{d\eta}{ds} < \frac{d\xi}{ds} < \beta\frac{d\eta}{ds}. \qquad (14b)$$

Let $(\xi, \eta) \to (\bar{\xi}, \bar{\eta})$ be a coordinate transformation in some coordinate patch, and suppose that $\bar{\eta}$ is a time coordinate. Then

$$\frac{d\bar{\eta}}{ds} = \frac{\partial\bar{\eta}}{\partial\eta}\frac{d\eta}{ds} + \frac{\partial\bar{\eta}}{\partial\xi}\frac{d\xi}{ds}.$$

If $\frac{\partial\eta}{\partial\xi} > 0$ then if

$$\left(\frac{\partial\bar{\eta}}{\partial\eta} + \alpha\,\frac{\partial\bar{\eta}}{\partial\xi}\right)\frac{d\eta}{ds} > \frac{d\bar{\eta}}{ds} > \left(\frac{\partial\bar{\eta}}{\partial\eta} + \beta\,\frac{\partial\bar{\eta}}{\partial\xi}\right)\frac{d\eta}{ds} \tag{15a}$$

whilst if $\frac{d\eta}{ds} < 0$ we have

$$\left(\frac{\partial\bar{\eta}}{\partial\eta} + \alpha\,\frac{\partial\bar{\eta}}{\partial\xi}\right)\frac{d\eta}{ds} < \frac{d\bar{\eta}}{ds} < \left(\frac{\partial\bar{\eta}}{\partial\eta} + \beta\,\frac{\partial\bar{\eta}}{\partial\xi}\right)\frac{d\eta}{ds}\;. \tag{15b}$$

Thus if $\frac{\partial\bar{\eta}}{\partial\xi} > 0$, a sufficient condition for the transformation to be time-orientation preserving is that

$$\frac{\partial\bar{\eta}}{\partial\eta} + \beta\,\frac{\partial\bar{\eta}}{\partial\xi} > 0 \tag{16a}$$

whilst a sufficient condition for it to be time-orientation reversing is that

$$\frac{\partial\bar{\eta}}{\partial\eta} + \alpha\,\frac{\partial\bar{\eta}}{\partial\xi} < 0. \tag{16c}$$

Similarly, if $\frac{\partial\bar{\eta}}{\partial\xi} < 0$, the sufficient condition for time-orientation preservation is

$$\frac{\partial\bar{\eta}}{\partial\eta} + \alpha\,\frac{\partial\bar{\eta}}{\partial\xi} > 0 \tag{16b}$$

and the sufficient condition for time-orientation reversal is

$$\frac{\partial\bar{\eta}}{\partial\eta} + \beta\,\frac{\partial\bar{\eta}}{\partial\xi} < 0. \tag{16d}$$

Consider now the pair of metrics

$$ds^2 = \left(1 - \frac{n^2}{R}\right)dt^2 - \frac{2n}{R^{\frac{1}{2}}}\,dt\,dR - dR^2 - R^2 d\Omega^2 \tag{17}$$

$$ds^2 = \left(1 - \frac{n^2}{R}\right)d\bar{t}^2 + \frac{2n}{\bar{R}^{\frac{1}{2}}}\,d\bar{t}\,d\bar{R} - d\bar{R}^2 - \bar{R}^2 d\Omega^2 \tag{18}$$

As has been noted, the two manifolds with the respective metrics are isometric. By inspection, a mutual transformation must take the form

$$\bar{R} = R, \quad \bar{t} = \bar{t}(t,R) .$$

Accordingly,

$$\left(1 - \frac{n^2}{R}\right)\left(d\bar{t}^2 - dt^2\right) + \frac{2n}{R^{\frac{1}{2}}}\,(d\bar{t} + dt)dR = 0 . \qquad (19)$$

Thus either

$$\text{(a)} \qquad d\bar{t} = -dt \qquad (20)$$

$$\text{and so } \frac{\partial \bar{t}}{\partial t} = -1, \quad \frac{\partial \bar{t}}{\partial r} = 0$$

or

$$\text{(b)} \qquad \left(1 - \frac{n^2}{R}\right)(d\bar{t} - dt) + \frac{2n}{R^{\frac{1}{2}}}\,dR = 0$$

whence

$$\frac{\partial \bar{t}}{\partial t} = 1, \left(1 - \frac{n^2}{R}\right)\frac{\partial \bar{t}}{\partial R} = -\frac{2n}{R^{\frac{1}{2}}} \quad . \qquad (21)$$

Referring back to equation (16c), we see that if (20) applies, the transformation is always time-orientation reversing (as may be obtained directly). If instead equations (21) apply, then the transformation is time-orientation preserving if

$$1 - \frac{2n\alpha/R^{\frac{1}{2}}}{1 - n^2/R} > 0$$

provided that $R > n^2$: otherwise the same relation holds with β substituted for α.

Now in fact

$$\alpha = 1 - n/R^{\frac{1}{2}}, \quad \beta = -(1 + n/R^{\frac{1}{2}}) \qquad (22)$$

and so for $R > n^2$ the transformation is time-orientation-preserving if

$$\frac{1 - n/R^{\frac{1}{2}}}{1 + n/R^{\frac{1}{2}}} > 0 \tag{23a}$$

and for $R < n^2$ the condition becomes

$$\frac{1 + n/R^{\frac{1}{2}}}{1 - n/R^{\frac{1}{2}}} > 0. \tag{23b}$$

Thus if $R > n^2$, the transformation is either of the form $\bar{t} = -t + \text{const.}$, or else is time-orientation preserving.

According to (16c), if $R < n^2$, and

$$1 - \frac{2n\alpha/R^{\frac{1}{2}}}{1-n^2/R} < 0$$

then the transformation is time-orientation reversing.

It follows that if $R < n^2$, the transformation is always time-orientation reversing.

Thus in the region $R < 2GM$ of the exterior Schwarzschild manifold, the collapse and anticollapse metrics can only be transformed into one another by means of a time-reversing isometry. The proposition is therefore established.

We may conclude that, within the Schwarzschild "singularity", the collapse and anticollapse metrics represent disjunct regions of space-time (unless time-reversing transformations are accepted, which would appear to violate causality).

This analysis seems to throw doubt on some recent work concerning the possible conversion of the collapse phase into the anti-collapse phase after passing through the essential singularity $R = 0$. Further work on this problem is indicated, but there appear to be grounds for believing that the "multiple connectedness" hypothesis discussed in the previous section may pose less serious difficulties than are posed by more straightforward hypotheses.

4. SUMMARY OF RESULTS

This thesis has throughout been concerned with a highly idealised model of the collapse situation. Nevertheless, the only serious element omitted from the analysis would appear to be the factor of rotation. A full mathematical treatment of a rotating dust fluid is still awaited, but if the Newtonian analogy is reliable it would seem that the collapse would be halted within a plane perpendicular to the axis of rotation. Nevertheless, qualtiatively similar results may seem probable, with collapse to a disc substituting for collapse to a point.

In Chapter II, it was shown that interesting results may still be obtained from a Newtonian treatment of the collapse problem. Even a straightforward sounding Proposition - that the collapse time decreases as the initial velocity increases - requires considerable care in its proof.

However, the most interesting Newtonian results concern the question of "overtaking". We consider only the case where the body is initially at rest and show that this situation cannot occur if the initial distribution of matter is such that the density is a monotonically increasing function of r. If, on the other hand, the initial density at some radius exceeds the mean density of the sphere within this radius then infinite densities are predicted during the collapse. Furthermore, it is shown that at least one shell of matter will pass through other shells of matter.

Chapter III considers the Relativistic treatment of the collapse problem. Although this Thesis is concerned almost exclusively with the dust fluid, the discussion naturally leads to an examination of the stress-energy tensor of the "Perfect Fluid". Classical analogues of the General Relativistic concept are developed, using the local instantaneous rest-frame. It is shown that the eigenvalue ρ cannot reasonably be identified with the rest-inertial mass, and instead the quantity $\rho + p/c^2$ (using the conventional system of units) appears to meet this description.

In this chapter, the general solution of the Field Equations for the spherically symmetric dust-fluid is obtained, independently of earlier derivations. It is shown that in comoving coordinates the solutions can be expressed in terms of an ordinary differential equation identical to the corresponding Newtonian equation. A close analogy is drawn between comoving coordinates (carefully defined) and the Lagrangian coordinates of the classical treatment. Schwarzschild coordinates are introduced, and transformation formulae established. A novel proof of Birkhoff's Theorem is obtained.

Various consequences of the general solution are explored in Chapter IV. Firstly, it is proved that interior solutions cannot be isometric to one another. Next, it is shown that from the viewpoint of the comoving frame, each solution implies a Newtonian-type catastrophe.

Next, we come to one of the most surprising results of our investigations. The asymptotic behaviour of the inhomogeneous dust-fluid is investigated, and it is shown that (to the first order) an imhomogeneous sphere will collapse at one-half the rate of the homogeneous sphere. It would appear that there are two distinct modes of collapse, with the homogeneous mode being unstable close to the point singularity.

The use of Schwarzschild coordinates is considered with particular reference to the homogeneous case, and results obtained by earlier workers in the field are generalised.

The question of the so-called "Schwarzschild singularity" is examined in detail, and it is shown that from the point of view of the rest frame of a collapsing body the Schwarzschild null-hypersurface is indeed penetrated, but once only.

The next topic is concerned with the observational situation. A formula for the red-shift of radially emitted radiation from a collapsing dust sphere is obtained, and it is shown that the red-shift and its rate of change with respect to (observational) time become infinite as the gravitational radius is approached. It is demonstrated that if a collapsing body exhibits a stable spectrum without noticeable variation in spectral shift, and if the red-shift is of order unity, then the mass of the body must be of order of magnitude greater than 10^8 times the mass of the Sun. A smaller body would rapidly vanish as the collapse proceeded, although in theory it would never reach the gravitational radius in finite observational time.

Finally in this Chapter, the problem of Boundary Conditions in its relation to the problem under investigation is formulated.

The greater part of Chapter V is concerned with an exposition of the mathematical foundations of General Relativity. Many of the concepts developed enable the outstanding problems to be clarified. The standard treatment is varied by the introduction of "left" and "right" tangent spaces on a "piecewise C^{∞} manifold". We deduce that the Lichnerowicz boundary conditions are not a necessary consequence of the mathematical framework, and indeed there is no basic objection to the alternative Synge & O'Brien conditions, which are less restrictive.

In Chapter VI, we consider the boundary condition problem further. Criteria are developed for the special case of spherical symmetry under consideration, and these are then generalised to give the following condition: under the two imbeddings representing the two sides of the boundary, the First and Second Fundamental Forms of the boundary hypersurface must equate. It is suggested that this condition can effectively replace all others adopted to date.

The question of the "fate" of a collapsing body is then considered in detail, and a possible approach is suggested by the use of multiply connected manifolds. In particular, attention is drawn to the relevance of the Antipodal Map, particularly in relation to the null hypersurface at the gravitational radius. Further lines of enquiry are suggested in this connection.

Finally, the concept of "Time Orientation" is developed. If time reversals are prohibited on causality grounds, it is shown that catastrophic collapse and catastrophic expansion are essentially disjunct. The two submanifolds of the Schwarzschild exterior manifold contained within the gravitational radius cannot be related by means of a time-translation or indeed any transformation preserving causality. Thus collapse within the Schwarzschild radius can under no circumstances be followed by expansion.

R.H.C.
LIBRARY

References

1 BARNES, A., J.Phys.A (G.B.), 3 (6), 653 (1970).

2 BIRKHOFF, G.D., "Relativity and Modern Physics", Harvard, Cambridge (Mass), 1923.

3 BONDI, H., Mon.Not.Roy.Astron.Soc. 107, 410 (1947).

4 BONDI, H., Proc.Royal Soc., A281, 39 (1965).

5 BONNOR, W.B., Leopold Infeld "Festschrift", 167, Warsaw (1962).

6 CHEVALLEY, C., "Theory of Lie Groups", Princeton (1946).

7 DATT, B., Zeit.Phys., 108, 314 (1938).

8 DOLAN, P., "Problems in Relativistic and Steady State Cosmology", University of London (1963).

9 EISENHART, L.P., "Riemannian Geometry", Princeton (1926).

10 FINKELSTEIN, R.J., Phys.Rev.110, 965 (1958).

11 GOURSAT, E., "Cours d'Analyse Mathématique", Vol II, Pt II (Hedrick & Dunkel trans), Ginn, New York (1917); original (1902).

12 HAWKING, S.W., Proc.Royal Soc., A294, 511 (1966).

13 HAWKING, S.W., Proc.Royal Soc., A295, 490 (1966).

14 HAWKING, S.W., Proc.Royal Soc., A300, 187 (1967).

15 HELGASON, S., "Differential Geometry and Symmetric Space", Academic Press, New York (1962).

16 HOYLE, F., and FOWLER, W., Mon.Not.Roy.Astron.Soc. 125, 169 (1963a).

17 HOYLE, F., and FOWLER, W., Nature, 197, 533 (1963b).

18 HOYLE, F., and NARLIKAR, J.V., Proc.Roy.Soc. A273, 1 (1963).

19 HOYLE, F., FOWLER, W., BURBIDGE, G. and BURBIDGE, E., Astrophys.J. 139, 909 (1964).

20 INFELD, L., and PLEBANSKI, J., "Motion and Relativity", Pergamon, Warsaw (1960),

21 ISRAEL, W., Phys.Rev.164, 1776 (1967).

22 JORDAN, P., EHLERS, J. and KUNDT, W., Akad.Wiss.Lit.Mainz, Abhandl., mat.-nat. K1(2), 3 (1960).

23 KERR, R.P., Phys.Rev.Lett., 11, 522 (1963).

24 LANDAU, L., and LIFSHITZ, E., "The Classical Theory of Fields", Addison-Wesley, Cambridge (Mass)., (1951).

25 LICHNEROWICZ, A., C.R.Acad.Sci.Paris, 212, 421 (1941).

26 LICHNEROWICZ, A., "Theories reletivistes de la gravitation et de l'électromagnétisme", Masson, Paris (1955).

27 McVITTIE, G.C., "General Relativity and Cosmology", Wiley, New York (1956).

28 MILNE, E.A., and McCREA, W.H., Quat.Journ.Math.(Oxford), 5, 73 (1934).

29 MILNOR, J., On Manifolds Homeomorphic to the 7-sphere. Ann. of Math. (2), 44 (1956).

30 NARIAI, H. and TOMITA, K., Prog.Theor.Phys., 34 (1), 155 (1965a).

31 NARIAI, H. and TOMITA, K., Prog.Theor.Phys., 34 (6) (1965).

32 NARLIKAR, J.V., Discovery (1965).

33 NEWMAN, M.H.A., "Elements of the Topology of Plane Sets of Points", Cambridge (1961).

34 NOMIZU, K., Amer.J.Math., 76, 33 (1954).

35 O'BRIEN, S., and SYNGE, J.L., Comm.Dublin Inst. A, 9 (1952).

36 OPPENHEIMER, J.R., and VOLKOFF, G.M., Phys.Rev.55, 374 (1939).

37 OPPENHEIMER, J.R. and SNYDER, H., Phys.Rev.56, 455 (1939).

38 PENROSE, R., Phys.Rev.Lett., 14. 57 (1965).

39 PENROSE, R., Nuovo Cimento, serie 1, 1, 252 (1969).

40 PENROSE, R., and HAWKING, S.W., Proc.Royal Soc. A314, 529 (1970).

41 PIRANI, F.A.E., Bull.Acad.Polon.Sci., 5, 143 (1957).

42 ROBERTSON, H.P., Rev.Mod.Phys. 5, 62 (1933).

43 ROBINSON, I., SCHILD, A., and SCHUCKING, E., "Quasi-Stellar Sources and Gravitational Collapse", Univ.Chicago Press (1965)

44 SCHOUTEN, J.A., "Ricci-Calculus", Springer-Verlag, Berlin (1954),

45 SCHWARZSCHILD, K., Sitz.Preuss.Akad.Wiss, 424 (1916).

46 SEXL, R.V., "General Relativity and Gravitational Collapse", Springer-Verlag, Berlin (1970).

47 SYNGE, J.L., "Relativity: The General Theory", North-Holland (1960).

48 THOMPSON, I.H., and WHITROW, G.J., Mon.Not.Roy.Astron.Soc.

49 TOLMAN, R.C., "Relativity, thermodynamics and cosmology", Clarendon Press, Oxford (1934).

50 TONDEUR, P., 'Introduction to Lie Groups", Springer-Verlag (1965).

51 VEBLEN, O., and WHITEHEAD, J.H.C., "The Foundations of Differential Geometry", Cambridge (1931).

52 WEYL, H., "Raum-Zeit-Materie", Springer-Verlag, Berlin (1918).

53 WILLMORE, T.J., "An Introduction to Differential Geometry", Oxford Univ. (1959).